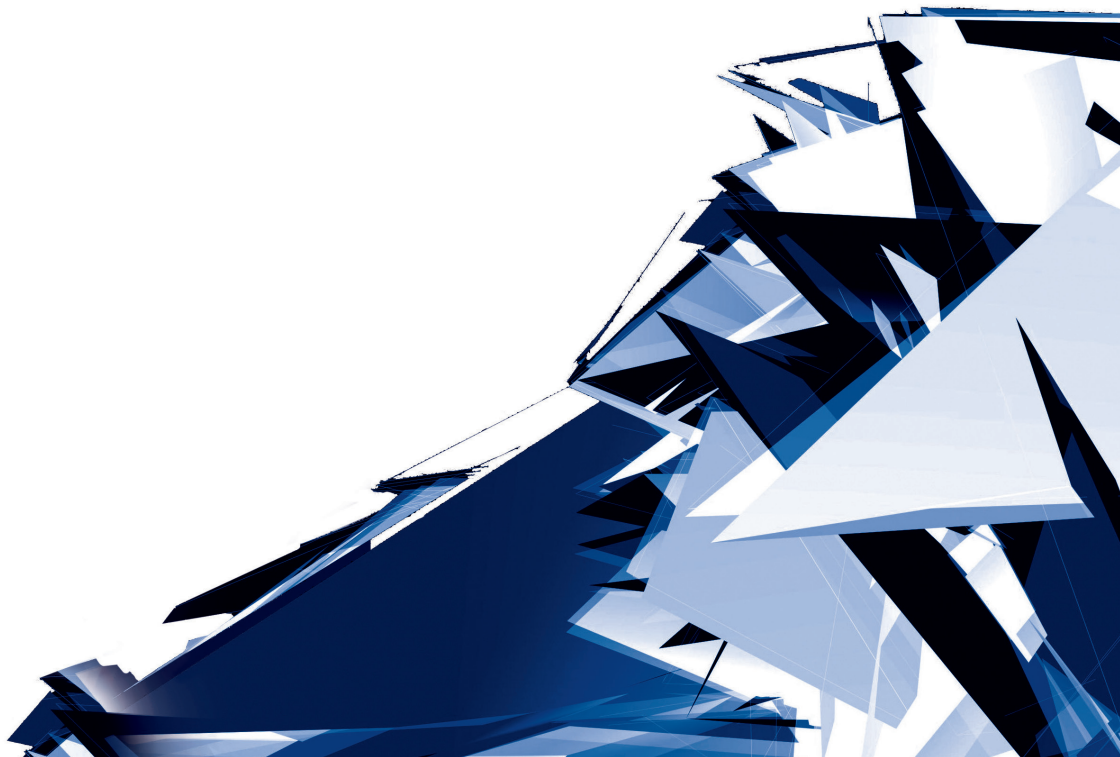
Technical Transactions

Czasopismo Techniczne

Volume 11

Year 2017 (114)



www.ejournals.eu/Czasopismo-Techniczne
www.technicaltransactions.com
www.czasopismotechniczne.pl

ISSN 0011-4561

Basic version of each Technical Transactions magazine is its online version
Pierwotną wersją każdego zeszytu Czasopisma Technicznego jest jego wersja online

Contents

MECHANICS

PHYSICS

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.17.185.7414

Artur Jasiński
Department of Architecture and Fine Arts, The Andrzej Frycz Modrzewski Krakow University College

Colour manipulation as a tool for creating the city image on examples of Miami Beach and Tel Aviv

Manipulacje kolorem jako narzędzie kształtowania wizerunku miasta na przykładach Miami Beach i Tel Awiwu

Abstract

The two famous, coastal cities – Miami Beach in the United States and Tel Aviv in Israel – have much in common: they have a similar history, topography and architecture. Their contemporary image was shaped relatively recently, in the second half of the 20th century. Thanks to the changes to the originally grey colour scheme, the sordid and unkempt cities were transformed into world-class tourist attractions and trendy style icons. That is when their current reputations were established: Miami Beach as the pastel-shaded hub of American Art Deco and white Tel Aviv as the symbol of modernist, Bauhaus-like architecture.

Keywords: colours of the city, creation of image, urban identity

Streszczenie

Dwa sławne nadmorskie miasta: amerykańskie Miami Beach i izraelski Tel Awiw, łączy wiele wspólnego: posiadają podobną historię, topografię i architekturę. Ich współczesny wizerunek został ukształtowany stosunkowo niedawno – w drugiej połowie XX wieku. Dzięki manipulacji dokonanej w ich pierwotnej kolorystyce szare, brudne i zaniedbane miasta zostały przekształcone w globalne atrakcje turystyczne i modne ikony stylu. Ukształtowane wtedy zostały towarzyszące im obecnie legendy: mieniącego się pastelowymi kolorami Miami Beach, stolicy amerykańskiego Art Deco, i białego Tel Awiwu, symbolu modernistycznej architektury rodem z Bauhausu.

Słowa kluczowe: kolory miasta, kształtowanie wizerunku, tożsamość miasta

1. The origins of Miami Beach

The City of Miami Beach was built on a sandy oceanfront shielding the port city of Miami from the East and was located at the southern tip of Florida. The first settler to move onto the land belonging to Miami Beach was John Collins, a farmer from New Jersey who came there to establish a vast mango and avocado plantation. It was the Lummus brothers, investors and developers, who operated there on a bigger scale. In 1912, they founded the Ocean Beach Realty Company and bought 500 acres of land. They had the mangrove forests cleared, had the land drained and parcelled out to create luxurious housing estates. Another leading investor was a landowner tycoon, Carl Fisher. He had quite a different view on development: he built small summer-house estates designed for middle-class American families and brought sand to create the famous beaches. In 1915, the City of Miami Beach was officially incorporated and expanded in 1924, when its administrative limits were moved north and reached 87th Street. This is when the town grew to its current size [17, p. 219].

Fig. 1. Miami Beach: housing development along Ocean Drive, as in 2016 (photo by the author)

During the interwar period, the southern and central part of Miami Beach, up to 44th Street, became built up. Most buildings were erected during the so-called Great Florida Land Boom in the 1920s and the so-called Depression Era Land Boom in the late 1930s. Local architecture was dominated by two main stylistic movements: Mediterranean Revival and Art Deco. The buildings of the 1920s were put up predominantly in a Mediterranean style. They were characterized by arcades supported by columns, arched windows, balconies with forged,

cast iron railings, white elevations, stone details and roofs covered with red tiles. Examples of such architecture can be found in the Spanish Village, along Española Way.

The second half of the 1920s saw the introduction of Art Deco in the United States, which owes its popularity to the influence the *Exposition Internationale des Arts Décoratifs et Industriels Modernes* of 1925 in Paris had on American architects. Buildings in the Art Deco style can be found in many American cities. Probably the most famous of them is the Chrysler Building built in Manhattan in 1930 according to William van Alen's design. Art Deco reached Miami Beach with the second wave of rapid development of the city in the late 1930s. The late type of the American Art Deco style, known as Streamline Moderne, was more austere in terms of style and made use of the achievements of industrial design, promoted during the *America World Fair* in 1930. Its local variation, known as Tropical Deco, also used motifs inspired by the tropical flora and fauna, as well as details from ship architecture, thus strengthening the coastal and vacation atmosphere of Miami Beach (http://www.mdpl.org/about-us/about-miami-beach-design-styles/what-is-art-deco/, access: 13.08.2017).

After World War II, Miami Beach developed yet another local style, which was a variation on modernism called MiMo (Miami Modernism). The idea behind it was to adapt the international style, traditionally taught in American design schools of the times, to the local conditions. It was characterized by asymmetric structures, round window openings, metal blinds and brise-soleil, often made out of aluminium or copper sheets, and walls covered in ceramic mosaics, open balconies and catwalks. A local typology of apartment buildings was created (the so-called Garden Style) in which the apartments, accessible through open galleries, were built around a centrally-positioned garden.

After the post-war surge of investments fuelled by cheap loans in the 1940s and 1950s, there came a long period of stagnation and degradation. Only in the second half of the 1980s did Miami Beach start to grow rapidly once again and its original skyline formed by short buildings was deformed by numerous high housing estates and hotels. The most colossal ones were put up on the southern tip of the city (South Point Tower, Y.H. Lee Associates 1986; Portofino Tower Group, Sieger Architectural Partnership and Skidmore, Owings and Merrill 1994-2008). Further major projects are being carried out and planned, mainly along the coastline and the port area in the western part of the city (e.g. Waverly at South Beach, Architectonica 2001). In 2000, the city of 7.1 square miles had a total population of 88 thousand inhabitants.

2. The origins of Tel Aviv

It is usually said that the origins of Tel Aviv date back between 1906-1909 when a group of local notables supported by the National Jewish Fund established the Ahuzat Bayit neighbourhood just a few kilometres north of Jaffa. It was a well-thought-out and a professional urban planning project, modern in character and inspired by the idea of a garden city. The author of the project, which served as the basis for the parcelling out of the first 60 plots of land, was engineer Avraham Goldman, while the guidelines were laid down by the famous

Zionist activist Arthur Ruppin. The axis of the neighbourhood was Rothschild Boulevard and the main side street was named after Herzl. The plots were large, much bigger than in Jaffa, amounting to 500 metres square each and the density of buildings was limited to 33 percent. Development lines were created and moved away from the frontage, so that each detached house had enough access to air and light [7, p. 17]. In 1909, another housing estate, Nahalat Binyamin, was built by the sea upon the nearby dunes. The one-storey houses had simple shapes, gable or hip roofs, and entrance porches with upper balconies. Additionally, the elevations were adorned by stylised ornaments. From the front side, there were forged fences supported by stone pillars. The only public building erected in the new housing estate was the Herzliya Hebrew Gymnasium, a monumental building in eclectic style with numerous references to the art of Islam, designed by the architect Joseph Barski [8, p. 122].

When the new neighbourhoods started merging, they were given the common name of Tel Aviv (in Hebrew: spring mound), which derived from the title of the Hebrew translation of Theodor Herzl's *Altneuland* [15, p. 4]. In 1920, in order to ensure the harmonious growth of Tel Aviv, the architect Richard Kauffmann was asked to come up with a spatial management plan for its northern part. Kauffmann introduced a block-based regulation system, which exists until today and consists of a network of main roads built on the coastline along the north-south axis, intersected by a network of perpendicular streets going from the hills towards the seashore. After the violent riots between the Jews and the Arabs which broke out on May 1, 1921, many Jewish families based in Jaffa moved to Tel Aviv in fear of persecution. In the same year, Tel Aviv received city status and in 1922, six neighbouring Jewish districts were annexed into Tel Aviv, including the eclectic Neve Tsedek.

A strong impulse for Tel Aviv's development came with the influx of Jewish immigrants from Palestine who started settling under the *haawara* agreement reached between Zionist organisations and the Nazi German government in 1933 [16, pp. 23-26]. These were mainly wealthy middle-class citizens, rich bourgeoisie and professionals. Many of them settled in Tel Aviv where they posed a challenge to the traditional lifestyle and collectivised economy. It is thanks to those immigrants that private enterprises started developing, new shops and malls were being opened, as was the stock market. This is when German-language newspapers started coming out, a number of cafés and a philharmonic hall were opened and the city of Tel Aviv, which had been a provincial town until then, finally started resembling a cosmopolitan metropolis [16, pp. 51-52]. Among the architects who arrived in Palestine in the 1920s and 1930s were young kibbutzniks who had studied architecture in Europe. It is the second generation of the Yishuv – Yoseph Neufeld, Zeev Rechter and Arieh Sharon – who, upon their return to Tel Aviv, established the influential architects' club called "Chung" (in Hebrew: circle, ring), which was actively promoting modernist style in architecture, associating it and linking it with the ideas of Zionism [14, p. 23].

For young architects educated in Europe, the rapidly growing Tel Aviv was not only a perfect place to start their professional careers, but also an ideological battlefield, which allowed them to introduce the rules of Zionism into the architectural practice: to design a new form for a new nation. The modernist idea of breaking all ties with history also provided them with the best possible opportunity to cut themselves off from the local building tradition,

"contaminated" by centuries-long Arabic influences, as well as from colonial architecture and the European diaspora, which Zionist activists considered bourgeois and rotten. For the first time in centuries, Jews had a chance to break free from the influences of foreign cultures and to build – literally and metaphorically speaking – their own, Jewish home. The International Style architecture, drawing on Mediterranean traditions, became a convenient and ideologically-potent platform, which made it possible to combine the ideals of the modernist style in architecture with the left-wing, pragmatic and often rough in style Labour Zionism of *Erec Israel*, represented by the two most influential Zionist organisations at the time: the centre-left labour party Mapai and the trade union movement Histadrut. The rules of the new architecture proposed by Le Corbusier received an original touch in Jewish Palestine and, in the late 1930s, Tel Aviv was heavy with hundreds of modern buildings designed in the local, simplified version of the International Style.

Fig. 2. Tel Aviv. Gottgold House on the corner of Allneby Street and Sheinkin Street, designed by Yehuda Magidovitch, 1935. As in 2014 (photo by the author)

Contrary to the label they later received, Tel Aviv's modernist buildings were never white. The buildings were covered with plaster and their elevations usually painted in bright colours: from shades of beige to shades of grey and sandy brown, and sometimes even rich colours were used, examples of which are the currently non-existent buildings of the "Blue Villa" at Bialik Street designed by Genia Averbuch and the "Red House" at Hayarkon Street. Some buildings were decorated in fancy ways; particularly interesting were the architectural details of arcades, entrances and stairways, which were often finished off with decorative terrazzo.

The spacious Tel Aviv was supposed to be the opposite of the Arabic tightness and untidiness. The aim of the Zionists was to create a place *full of energy and vigour, beautiful and built in accordance with Hebrew law. The new Jewish city was supposed to naturally dominate over the neighbouring Jaffa and become the hub of business, industry and trade, as well as leisure, health, sports, tourism and administration* [10, p. 195].

Adam LeBor [9] claims that the entangled histories of Jaffa and Tel Aviv can be seen as a metaphor for the contemporary history of the two nations: in the late 19th century, Jaffa, which at the time was the hub of Palestine's economy and Arabic culture, was inhabited by very few Jews, while today, the old Jaffa is merely a tiny enclave in Tel Aviv, a tourist attraction of sorts, with the city's population reaching almost half a million inhabitants, less than 5% of whom are Arabs. Due to the high prices of real estate in the snobbish Jaffa, the Arabic population inhabits the neighbouring district of Ajami, but even there, they are gradually being pushed out. The ratio changed not only in terms of space and demography; the importance and symbolism of both cities, merged into one urban organism in 1950, changed as well: the old Arabic Jaffa became the *picturesque, artistic neighbourhood where you can escape the hustle and bustle of the jam-packed Tel Aviv*[1], while Tel Aviv, which was meant to be a clean and hygienic, European-style antidote to the mess of the oriental Jaffa, turned into a noisy, partly shabby and dirty cosmopolitan metropolis.

3. The use of colour to shape the city image

In the 1970s, South Beach was a run-down residential district inhabited by pensioners of mainly Jewish origin. The wealthy ones moved north, while the poor ones stayed there without any hope for a better life. American historian M. Baron Stofik once called this place "God's waiting room" [12, p. 128]. The old buildings were either falling into ruin or were gradually demolished to make room for building investments planned along the seacoast. At the same time, drugs, crime and poverty were running rampant in the streets.

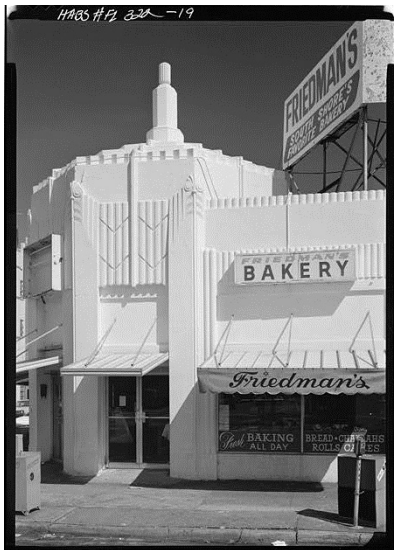


Fig. 3. Friedman's Bakery (on the corner of 7th St. and Washington Ave.) circa 1980, before being painted with Leonard Horowitz's design. Source: Library of Congress, HABS FLA, 13-MIAM, 5-19

This is the city that Barbara Baer Capitman, art historian, journalist and urban activist from Chicago, saw when she arrived in 1973. When her husband, a professor of economics at the local Florida International University, died two years later, she

[1] Quoted by: www.izrael.badacz.org/turystyka/telaviv_jaffa.html (access: 15.08.2017).

focused all her attention on the protection of the architectural heritage of Miami Beach. She received help from Leonard Horowitz, a student of architecture from New York who was fascinated by Art Deco architecture. Together, they founded the Miami Design Preservation League in 1976. She created an inventory list of period buildings [2] and made sure that in 1979 the Miami Beach Architectural Historical District – now popularly known as "Art Deco District" – was listed in the National Register of Historic Places. It was the first urban complex built in the USA in the 20th century to be officially recognised as a historical monument.

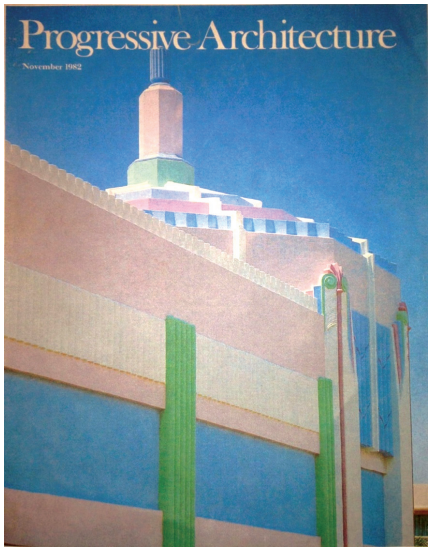


Fig. 4. The cover of Progressive Architecture magazine from 1982 showing Friedman's Bakery after it was painted according to Leonard Horowitz's design. Source: http://wlrn.org/post/meet-man-behind-all-those-south-beach-pastels

Capitman was a fighter who did not make any compromises. She led demonstrations and protests and she organised sit-in protests to prevent the demolition of buildings she viewed as valuable. Her distinctive, screechy voice and high-handed manner did not gain her many friends, nevertheless, she was consistent and persistent. In 1980, she invited a famous avant-garde artist, Andy Warhol, to Miami Beach and gave him a tour of the historic buildings. This was the first event to attract the wider public's attention to the architecture of Miami Beach, but the demolitions kept on; already a month after Warhol left Miami Beach, Boulevard Hotel was taken down, as was The New York Hotel a year after. It was only the premiere of the cult TV series *Miami Vice* in 1984 that brought a halt to this destructive trend. Miami was then the centre of drug trafficking and the cocaine trade, which was the central issue of many of the show's plots. The series was shot on the streets of Miami, against the backdrop of Art Deco buildings on South Beach, and the show's intro was inspired by the neon signs of the local hotels. The dynamic film shot with a hand-held camera brought fame to the actors and the city of Miami Beach alike. Jumping on the show's bandwagon, Capitman and Horowitz travelled all across the United States to promote Art Deco architecture and convince the Americans see its value. In 1988, Capitman published a book *Deco Delights: Preserving The Beauty and Joy of Miami Beach Architecture* and in 1994, a few years after her death, another important publication was put into print – *Rediscovering Art Deco: A Nationwide Tour of Architectural Delights* – which she had written with Michael D. Kinerk and Dennis W. Wilhelm. In 1996, the central part of Tenth Street in South Miami Beach, between Washington Avenue and Ocean Drive, was named after her [5].

The key tool used in the renovation of the historic buildings of Miami Beach was the colour palette created by Leonard Horowitz in 1980 and applied for the first time during the renovation of the Friedman's Bakery. Horowitz drew inspiration from the colours of the sand, the sea and the sky at sunset. His ideas were supported by the city council, which allowed him to carry out his

first project. It involved repainting the white building with Friedman's Bakery at 686 Washington Avenue using a pastel colour palette. The initial reactions were rather critical: a woman passing by the newly renovated building commented: *This is not us. This is not what I remember for as long as I've lived here. Deco schmeco*. The most pointed remark said: *I hate that building. It looks like a whorehouse* [6]. However, when in 1982 a picture of the renovated elevation of Friedman's Bakery made the cover of an influential magazine *Progressive Architecture*, the new style of Miami Beach gained enormous popularity and spread across the streets of the city. The pastel-coloured, geometric Art Deco buildings with distinctive architectural details reflected the postmodern trend in American culture and architecture and attracted a number of celebrities, fashion photographers, as well as film crews to Miami Beach. Horowitz died of AIDS at the age of 43, a year after the release of the documentary film *Pastel Paradise* depicting his accomplishments and the history of South Beach renovation. A person who used to be considered an eccentric shaped the contemporary landscape of Miami Beach, the international style symbol, where over 150 buildings have already been repainted on the basis of his pastel colour palette. The once declining city which had been home to both pensioners and mobsters transformed into a fashionable, bustling resort swarming with tourists from all over the world.

Fig. 5. The Horowitz's Pastel Palette created in 1980 by Leonard Horowitz. Source: http://wlrn.org/post/meet-man-behind-all-those-south-beach-pastels

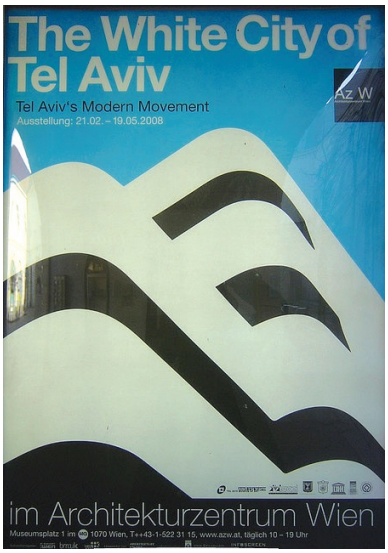
Israel's modernist architecture was discovered in the second half of the 20th century due to an exhibition held in Tel Aviv in 1984 entitled *White City. International Style Architecture in Israel*. The exhibition was accompanied by a catalogue written by its curator, an Israeli

architecture historian Michael Lewin, in which he describes the development of the agglomeration of modernist buildings in Tel Aviv as the symbolic birth of Israeli architecture. It cannot be stressed enough how important the exhibition and Lewin's short publication were to the great surge of fascination with modernist architecture in Israel and the beginning of the world-wide success of the term 'White City'.

As a result of the exhibition, the topic of International Style architecture in Israel was raised by popular magazines. Esther Zandberg, the only architecture critic at that time, presented the modernist heritage of Tel Aviv in a series of articles under a shared title *White Box*. At that time, the buildings erected in the 1930s were gradually falling into ruin, being rebuild, altered or demolished to make room for new projects. Zanberg's writings had a huge impact on the way architecture and city planning were perceived by regular citizens, contributed to raising historical awareness and inspired protests against the demolition of the old buildings for the purposes of redevelopment, which started appearing in Tel Aviv in the 1960s and 1970s.

Another person who substantially contributed to the preservation of modernist architecture was Nitza Metzger-Szmuk, a conservator-restorer educated in Italy. Soon upon her return to Israel in 1990, she was offered the position of the city conservator of historic buildings and was put in charge of cataloguing and preserving buildings put up in Tel Aviv during the 1930s. She was an unwearied promoter of modernism who gave lectures and organised exhibitions; she also authored a monumental, bilingual book *Dwelling on the Dunes: Tel Aviv – Modern Movements and Bauhaus Ideals* published in 2004.

Fig. 6. Renovated buildings on Rothschild Boulevard in Tel Aviv, 2014 (photo by the author)



Thus, International Style architecture became the official element of Israel's national heritage and a must-see stop in all tourist programmes, reinforcing the myth of Tel Aviv as the "white city built on sand" and concealing – conveniently for the Zionists – the centuries-long, colourful history of the neighbouring Arabic Jaffa [15, pp. 11-12].

Fig. 7. The poster for "The White City of Tel Aviv" exhibition, held in Vienna in 2008. Source: https://pl.pinterest.com/pin/214061788513159861/

4. Conclusion

The new narratives developed in the mid-20th century based on sophisticated colour schemes had ground-breaking implications for the contemporary history of Miami Beach and Tel Aviv. Not only did they contribute to shaping an appealing media image of the cities and to develop the interest of public opinion, but also led to increasing tourist traffic, reviving the economy, boosting real estate prices and, finally, to transforming the landscape of both cities. This is because the schemes were actually put into practice: the centre of Tel Aviv turned in part into "the white city", as many modernist houses were repainted white, and the majority of historic Art Deco buildings in Miami Beach were painted in keeping with the pastel colour palette designed by Leonard Horowitz. In that way, the cities started actually resembling, to a certain extent, the legends that had been built around them.

Creating the pastel-coloured myth of Miami Beach was aimed at bringing a halt to the surge of demolitions and inspire a revitalisation of decaying buildings, but it also brought about the city's economic revival and sparked the public opinion's interest in the short, yet prolific and successful style in American architecture, namely the local variation on Art Deco. During the wave of popularity of the city's architecture in the 1980s, Miami Beach became a style icon, the summer capital of Florida and a major tourist destination.

The contemporary myth of Tel Aviv as the "White City", apart from marketing and economic significance, has ideological and political meaning, is supposed to contrast with the dirty and "black" Arabic Jaffa. Sharon Rotbard, an Israeli architect and critic, writes: "to change the city, we must first change its history" [15, p. 4]. The Israelis are therefore rewriting Tel Aviv's history anew, referring predominantly to its 20th-century, modernist roots, erasing its colonial past, diminishing the importance of the centuries-long history of Jaffa and avoiding any mention of its previous Arabic inhabitants. The old Jaffa, stripped of its original character,

turned into a themed imitation for tourists and a culinary centre of the cosmopolitan Tel Aviv, which is currently advertised as the world capital of entertainment and nightlife.

The changes that occurred in both Miami Beach and Tel Aviv shared a similar, distinctive rhythm and dynamic: after the stagnation of the 1960s and 1970s, they both saw a period of revitalisation and growth, which later led to massive commercialisation and gentrification. This applies particularly to the historic districts and the coastlines: South Beach along Ocean Drive and Collins Avenue in Miami Beach and the central neighbourhood in Tel Aviv along Rothschild Boulevard and the old Jaffa's coastline. During the revitalisation of the cities, the major impulse for change was brought by manipulation on its original colour schemes. As a result, the cities not only reshaped their images, but also reconstructed their original identities.

References

[1] Azaryahu M., *Tel Aviv: Mythography of a City*, Syracuse University Press, New York 2007.

[2] Capitman B. (ed.), *Portfolio of Art Deco Historic District*, Miami Design Preservation Leaque, Miami Beach 1979.

[3] Capitman B., Brooke S., *Deco Delights: Preserving The Beauty and Joy of Miami Beach Architecture*, E.P. Dutton, New York 1988.

[4] Capitman B., Kinerk M.D., Wilhelm D.W., *Rediscovering Art Deco: A Nationwide Tour of Architectural Delights*, Viking Studio Books, New York 1994.

[5] Dembling S., *Barbara Baer Capitman: South Beach's Art Deco Hero*, National Trust for Historical Preservation, September 3, 2014, https://savingplaces.org/stories/barbara-baer-capitman-south-beach-art-deco-hero#.WZA5e_m7SVp (access: 4.09.2017).

[6] Duba M., *Meet The Man Behind All Those South Beach Pastels*, WLRN Miami/South Florida, Nov. 16, 2013, http://wlrn.org/post/meet-man-behind-all-those-south-beach-pastels (access: 4.09.2017).

[7] Eliakim T., *The Urban Planning of Tel Aviv up to 1948*, [in:] Metzger-Szmuk N., *Tel Aviv. Dwelling on the Dunes. Modern Movement and Bauhaus Ideas*, Editions de l'éclat, Paris–Tel Aviv 2004, 17-26.

[8] Harpaz N., *Zionist Architecture and Town Planning. The Building of Tel Aviv (1919-1929)*, Purdue University Press, West Lafayette 2013.

[9] LeBor A., *City of Oranges. An Intimate History of Arabs and Jews in Jaffa*, W.W. Norton Company, New York 2007.

[10] LeVine M., *Overthrowing Geography. Jaffa, Tel Aviv, and the Struggle for Palestine 1880-1948*, University of California Press, Berkeley–Los Angeles–London 2005.

[11] Lewin M., *White City: The International Style Architecture in Israel. A Portrait of an Era*, Tel Aviv Museum of Art, Tel Aviv 1984.

[12] Lin J., *The Power of Urban Ethnic Places: Cultural Heritage and Community Life*, Routledge, New York 2011.

[13] Metzger-Szmuk N., *Tel Aviv. Dwelling on the Dunes. Modern Movement and Bauhaus Ideas*, Editions de l'éclat, Paris–Tel Aviv 2004.

[14] Nitzan-Shiftan A., *Contested Zionism – Alternative Modernism: Erich Mendelsohn and the Tel Aviv Chug In Mandate Palestine,* [in:] Yacobi H. (ed.), *Constructing a Sense of Place: Architecture and Zionist Discourse,* Ashgate Publishing Ltd., London 2004, 17-51.
[15] Rotbard S., *White City, Black City. Architecture and War in Tel Aviv and Jaffa,* The MIT Press, Cambridge Massachusetts 2015.
[16] Segev T., *Siódmy milion. Izrael – piętno zagłady,* Wydawnictwo Naukowe PWN, Warszawa 2012.
[17] Shulman A.T., Robinson R.C., Donnely J.F., *Miami Architecture: An AIA guide Featuring Downtown, the Beaches, and Coconut Grove,* University Press of Florida, Gainesville 2010.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.17.186.7415

Przemysław Markiewicz (p.markiewicz@pk.edu.pl)
Institute of Building Design, Faculty of Architecture, Cracow University of Technology

BIM STANDARD AND THE NEW SPECIALISATIONS THAT HAVE EMERGED BECAUSE OF IT AROUND THE ARCHITECTURAL PROFESSION

STANDARD BIM I POWSTAŁE DZIĘKI NIEMU NOWE SPECJALIZACJE W OTOCZENIU ZAWODU ARCHITEKTA

Abstract

Civilisational development brings new challenges in the discipline of architecture. In order to meet them in architectural and construction design, it is becoming inevitable to introduce the broadly understood BIM standard as widely as possible. New, previously unknown professional specialisations have emerged around the architectural profession.

- More and more often, the redevelopment and modernisation of existing buildings is performed in the BIM standard, which is why it is necessary to convert existing 2D documentation into a virtual 3D model of a building.
- Laser rangefinders coupled with appropriate computer software allow the performing of intelligent measurement, which is based on the simultaneous measuring and modelling of a virtual building's elements.
- The technologically latest, most accurate, quickest and non-invasive method of gathering data on a building is performed using measurements that employ a laser scanner, which creates a point cloud within a 3D space.
- The creation of object libraries which represent specific commercial products for the most popular computer aided design programs has become one of the most dynamically developing specialisations surrounding the architectural profession.
- The possibility of presenting a design using an interactive multimedia presentation broadens the capabilities of using it in ways that are different from the traditional manner.

The new specialisations that are emerging in association with technological progress around the architectural profession broaden the traditionally understood market for architectural services. The subject of design in the BIM standard and the new specialisations should be introduced into the curriculum of architecture students as quickly and as broadly as possible.

Keywords: architect's professional toolkit, 3D design documentation conversion, laser scanning, object libraries, Building Information Model eXplorer

Streszczenie

Rozwój cywilizacyjny przynosi w dziedzinie architektury nowe wyzwania. Aby im sprostać, w projektowaniu architektoniczno-budowlanym nieuchronne staje się powszechne wprowadzenie szeroko rozumianego standardu BIM. Wokół zawodu architekta powstają nieznane wcześniej nowe specjalizacje zawodowe.

- Coraz częściej projekty przebudowy i modernizacji budynków istniejących wykonywane są w standardzie BIM i z tego względu konieczne jest na wstępie przekształcenie istniejącej dokumentacji 2D w wirtualny model budynku 3D.
- Dalmierze laserowe, sprzężone z odpowiednim oprogramowaniem komputerowym, pozwalają na inteligentne inwentaryzacje, polegające na jednoczesnym pomiarze i modelowaniu elementów wirtualnego budynku.
- Najnowsza technologicznie, najdokładniejsza, najszybsza i bezinwazyjna metoda zbierania danych na temat budynku to inwentaryzacja za pomocą skanera laserowego, który tworzy chmurę punktów w przestrzeni 3D.
- Tworzenie obiektów bibliotecznych do najpopularniejszych programów wspomagających projektowanie, które dotyczą konkretnych produktów handlowych, stało się jedną z najbardziej dynamicznie rozwijających się specjalizacji na obrzeżach zawodu architekta.
- Możliwość przedstawienia projektu w interaktywny i multimedialny sposób poszerza możliwości wykorzystania go poza tradycyjne zastosowania.

Nowe specjalizacje które powstają w związku z postępem technologicznym wokół zawodu architekta, poszerzają tradycyjnie rozumiany rynek usług architektonicznych. Problematyka projektowania w standardzie BIM i nowe specjalizacje powinny być jak najszybciej i w możliwie najszerszym zakresie wprowadzone do programu nauczania studentów architektury.

Słowa kluczowe: warsztat architekta, konwersja dokumentacji projektowej do 3D, skanowanie laserowe, obiekty biblioteczne, Building Information Model eXplorer

1. Introduction

Civilisational development brings new challenges in numerous disciplines. In order to face these challenges, each profession changes and evolves, with new, previously unknown professional specialisations emerging around them.

The situation is similar in the discipline of architecture and construction. Over the course of recent years, a series of new, previously unknown specialisations have emerged around the architectural profession. The cause of these ongoing changes is, chiefly, an extraordinary development in the field of information technology, which has revolutionised the professional toolkit that supports architectural and construction design. These changes, to be as brief as possible, are based on replacing traditional 2D CAD systems with BIM standard software that allows the creation of virtual models of buildings linked with an enormous database, which precisely parameterises all the elements of a design.

New programming tools dedicated for architects are being supported by more and more modern peripheral devices, such as laser scanners and drones. All of this – when put together – stimulates the emergence of new professional specialisations. Among the new specialisations that have emerged around the architectural profession, we should mention, for instance, spatial building measurement techniques, both on the urban scale, as well as on that of an individual structure, the development of object libraries for specific commercial products, as well as the use of virtual design presentation techniques in association with databases, used for various new applications.

The new information technologies are implemented most quickly and easily by the youngest generation of architects. This is the basis of the immensely important role of universities that are responsible for the quality of the education that the graduates who are entering the employment market receive. Engineer-level studies (level I) and Master's-level ones (II level) in architecture and urban design, which culminate in the awarding of the title of *magister inżynier architekt* (equiv. to MS Eng. Arch. – transl. note), should provide graduates with full professional knowledge, which is not only adapted to the reality of today, understood as a state-of-the-art professional toolkit, but is also one that takes into account the direction of changes that are going to take place in the following decade at the minimum. Proper education can provide current students of architecture with immense professional opportunities and a generational competitive advantage. On the other hand, an improper configuration of courses during studies and the omission of state-of-the-art practical, professional knowledge regarding the profession's tools can waste these opportunities.

2. Configuration of tools that aid in design

Along with the currently ongoing changes, it is becoming necessary to configure professional tools in the form of computer-aided architectural and construction design software in a manner that can meet current needs.

The vast majority of professionally active architects are very reluctant to change the computer-aided design software that they work on. Furthermore, they often use them in the wrong configuration.

An example of such a wrong configuration of computer-aided design software can be as follows:

- The spatial conceptual design is developed in the freely available Sketchup program by Google. It is a program that enables relatively easy and quick three-dimensional modelling; however, it does not give the ability to provide a detailed parameterisation of each of the elements and the material from which they are made.
- All the more detailed architectural and construction solutions are developed as 2D drawings – independent floor plans and cross-sections – in programs following the 2D AutoCAD standard.
- Sketchup is once again used to develop photo-realistic visualisations; however, the virtual model is built completely separately and independently from the 2D architectural and construction documentation, developed in AutoCAD.
- 3DMax Studio, expanded in order to obtain better effects with such programs like V-ray, etc., is used to create photorealistic visualisations.
- The final production of sheets is performed in a bitmap and digital photography editing software, e.g. Photoshop.

The verification of the appropriateness of the configuration of the tools that are being used and that support design occurs during attempts to introduce modifications and designs changes. In the configuration of an architect's digital professional toolkit described above – let's call it the "typical" configuration – actions like moving, changing the size and shape of windows and/or slight modifications to a building's massing entail a very large number of independent and tedious activities and, furthermore, increase the risk of making numerous mistakes in the form of the mutual inconsistency between each of the elements that comprise the design documentation (floor plans/cross-sections/facades/3D model).

The alternative, which eliminates the problems and inconveniences listed above, is the use of a cohesive, BIM-standard digital platform that supports architectural and construction design. In Poland, this is most often a choice between Autodesk's Revit and Graphisoft's Archicad. The performance of analogous modifications and changes in a design is simple and quick. Furthermore, the modifications are performed on a virtual building model, ruling out mistakes in the form of inconsistencies between drawings – when a 3D model is the basis for the generation of 2D floor plans, cross-sections and facades, then mistakes simply cannot be made!

3. Design using the bim standard

BIM (abbreviation of *Building Information Modelling*)[1] is based on the creation, with the use of appropriate software, of an immense database, precisely defining each part of a building (its structure, materials and their properties, fittings, etc.) and ordered in a three-dimensional space in the form of a virtual 3D model. A BIM-standard design is developed using three-

[1] *BIM Curriculum*, Graphisoft, 2013, http://www.graphisoft.com/learning/training-materials/bim-curriculum/ (access: 21.03.2016).

dimensional objects, such as walls, floor slabs, roofs, ceiling surfaces, windows, doors, etc., which, apart from their geometric dimensions, are being assigned with appropriate parameters (physical properties, technical properties, etc.). Integrating information within a single database makes it possible to automatically identify modifications and detect possible collisions. A building's model can be appended and verified using various types of schedules, timetables, cost estimates, etc.

The distinctive element of designing in BIM-standard computer-aided design programs is the shift of the main workload on the early phases of the design process (the conceptual design stage), which provides the greatest capability of influencing effectiveness, at the lowest cost and at the lowest difficulties that are associated with it[2].

Designing a building as a virtual spatial model, in which all the parameters of a real structure have been introduced, such as the layers of partitions with defined sets of parameters and their physical properties, technical parameters, prices, etc. allows us to carry out various types of analyses[3] and simulations already during the early stages of design, something which is completely unavailable in the case of traditional designs, developed in 2D CAD software (*Computer Aided Design*)[4], in the form of 2D floor plans, cross-sections and facades. One example of such an analysis is the daylighting and shade analysis of a building. In a design developed using BIM-standard software, the virtual model has a precisely defined geographical location, along with its height above sea level and orientation in regard to the cardinal directions. Thanks to this, we have a defined angle of incidence of solar rays which is characteristic of the design, and which can be determined for any given hour and day within a year. Daylighting analysis is particularly crucial and key in the design of residential buildings, which have a legally regulated amount of hours of direct access to sunlight. At the same time, the prescribed time of access to sunlight is often difficult to achieve, particularly in the case of apartments in a quarter-based, intense urban built environment.

The revolutionary change that has occurred in the professional toolkit of architects is best illustrated in the form of the typical parameters that are contained in a BIM-standard design[5]:

- a precisely parameterised structure of the external elements of a building (layer structures for all partitions):
  - specific thicknesses for each layer of a partition,
  - a selection of construction materials with their associated physical parameters (heat transfer coefficient for the purposes of calculations, density per cubic metre, etc.),
- openings with specified parameters regarding windows and doors:
  - translucent elements are introduced along with their U coefficient and thermal gains from sunlight,

---

2 Nordby A., *Przewodnik MaTrID – Zintegrowane projektowanie*, European Commission Executive Agency for Competitiveness and Innovation, 2013, p. 3.

3 *BIM Curriculum*, Graphisoft, 2013, *op. cit.*

4 Sydor M., *Wprowadzenie do CAD. Podstawy komputerowo wspomaganego projektowania*, Wydawnictwo Naukowe PWN, Warszawa 2009, p. 47.

5 *BIM Curriculum*, Graphisoft, 2013, http://www.graphisoft.com/learning/training-materials/bim-curriculum/ (access: 21.03.2016).

- opaque elements are presented along with their U coefficient, Psi values (linear thermal energy transfer coefficient [W/mK], which is used to calculate the effect of thermal bridges that occur at the connection between the window frame and the wall around the opening) as well as linear infiltration characteristics,
- the more important external elements, which form a mass that accumulates heat, divided into:
  - heavy (concrete structures), > 400 [kg/1 m^2] of useable area,
  - medium (masonry structures), 250-400 [kg/1 m^2] of useable area,
  - light (post and beam structures), < 250 [kg/1 m^2] of useable area,
- the level of air infiltration through a partition over an hourly energy balance, as well as the total air change rate per hour (ACH), with the following division:
  - 0.6 [l/s,m^2] is a low value (for a passive building),
  - 1.0 [l/s,m^2] is a medium value (recommended building),
  - 1.5 [l/s,m^2] is a high infiltration value (similar to a building fitted with mechanical ventilation),
- the properties of the designed material on the external surface of a layered structure, which affect the absorption characteristics of a given structure – the capacity to absorb solar energy by a given partition [%].

4. New specialisations within the architectural profession

New programming tools, which support BIM-standard design and the peripheral devices that cooperative with it, have caused immense progress in recent years, along with the emergence of numerous specialisations that surround the architectural profession. The demand for architects educated in the new specialisations grows each year not only in Poland, but also, and perhaps the most, in the most developed countries of the world, which are the quickest to adopt new technologies.

The most distinct of these new specialisations in the architectural profession have been described below.

4.1. Converting 2D documentation to 3D

All manners of building redevelopment and modernisation projects require the development of design documentation. Insofar as in the case of historical buildings, the basis for the development of such a documentation is a record of its extant state, in the case of buildings that have been built in recent decades, the basis for the development of design documentation is usually such a building's existing documentation, developed using digital support in accordance with AutoCAD's 2D standards. Due to the fact that building redevelopment and modernisation designs are being developed in accordance with BIM standards, it is necessary to perform an initial conversion of the existing 2D documentation into a virtual 3D model of the building. This poses numerous problems. Practical experience

shows that 2D documentation typically contains enough mistakes that it requires thorough and precise verification over the course of the development of a virtual model, as well as in confrontation with the verification measurements of the existing structure.

Upgrading existing documentation to 3D, as the building of an existing building's virtual model is colloquially called, is tedious, time-consuming and requires a lot of work. The lack of designers in this specialisation is widely felt. However, the capabilities of designing the redevelopment and modernisation of a building on the basis of a BIM spatial model are absolutely incomparable and will become a widely established standard in the future.

A virtual BIM-standard model of a building is, at the same time, a database that allows the precise parameterisation of, among other things, the following elements:

- The numerical model of the terrain
- The construction materials that form the layered partitions of a building along with their physical characteristics
- The parameters of windows and doors
- The technical infrastructure and systems of a building
- The operating temperatures of each room as well as operation schedules and timetables
- The climate parameters associated with the geographical location and immediate surroundings of a building
- etc.

4.2. 3D extant state documentation development

3D extant state documentations of buildings can be divided depending on the scale of the documentation, such as extant state documentation on the urban scale, in which the virtual spatial model includes a large area with numerous structures that are presented with limited precision in accordance with the scale, as well as extant state documentation of individual buildings, which is developed in a fairly detailed manner, taking into account technical infrastructure, etc.

4.2.1. 3D extant state documentation on the urban scale

Modern extant state documentation development on the urban scale should be performed in the following order:

- **Aerial photography** – The Central Surveying and Cartographic Facility possesses an archive of aerial photographs from various years, which cover most of Poland's surface area. The photogrammetric measurement precision of such photographs (x, y and h) is estimated at around 0.5 m (RMS[6]). Areas for which documentation is being developed and for which archival materials can be obtained, can also have a proper photogrammetric air run plotted and performed.

[6] RMS – (root-mean square) – interpreted as an area of the root mean – for instance, RMS = 0.25 means that 65% of the observations are within a root-mean square with a 25 mm radius.

- **Aerial triangulation** – in order to perform precise measurements based on aerial photography, it is necessary to perform their georeferencing[7] with points measured at ground level. A surveyor is usually commissioned to perform this.
- **Measurement of buildings** – the measurement of buildings can be performed with various degrees of precision. There are three basic levels, which have been presented in Figure 1.

Fig. 1. The three degrees of precision of 3D extant state documentation on the urban scale

- The most appropriate one seems to be the intermediate variant (for the performance of analyses as well). The measurement of chimneys or of the detailed elements of roofs yields poor results with a photograph that has a pixel size equivalent of 0.24 m. Should it be necessary to reduce the outlines of the buildings to their ground floors as seen in the Building and Land Registry (BaLR – EGiB in Polish – transl. note) – the complicated and work-intensive nature of such documentation can only be exacerbated. It should also be pointed out that aerial photographs cannot be used to ascertain whether a visible roof has walls, or is it purely a freestanding cover – just as it cannot be determined whether it is a registered building or not. It can, however, be determined (by comparing with the BaLR) whether a building visible on the BaLR exists, or whether the outline of the ground floor does not extend beyond the outline of the roof (an error in the BaLR or an outdated outline of the ground floor), locate buildings, structures or freestanding roofs.
- **NTM measurements (Numerical Terrain Model)** can be performed in the form of documentation on a stereo model (the highest degree of precision) or can be generated on the basis of aerial photography either automatically or semi-automatically – depending on the expected precision. Each of the methods has its advantages and disadvantages. For instance, preparing data for the performance of a shade analysis in the case when a model of a group of buildings is being built – the NTM data that we are able to obtain should ideally be capable of being simplified into a network of points spaced 2-5 metres apart.
- **Orthophotomap**[8] – provided that we have photographs and an NTM, the preparation of an orthophotomap from photographs with a pixel size equivalent of 0,24 m is not as work-demanding. An orthophotomap constitutes excellent reference material.
- **Streets** – in the case of having an orthophotomap, the street grid can be presented using it as a background, in the form of vectors of street axes. To this end, we can consider

[7] Georeferencing is based on providing a raster or vector file with a predetermined coordinate system.

[8] The map whose content is being presented is an aerial photographic image (ordinary aerial or satellite images of the Earth's surface).

importing data from Open-Street-Map that can be easily downloaded in vector format. Any blank spots or inconsistencies can be identified and re-measured from aerial photographs or from stereo measurements. The streets should then be superimposed on the NTM so that they can be assigned their respective height values.

Fig. 2. The superimposition of a vector road grid on an orthophotomap – the Market in Słomniki

- **Spatial urban model** – a virtual model – mock-up – of the entire area of a commune can be developed in various graphical styles (depending on particular needs). Vector street axes, along with the names of streets, can then be superimposed on the virtual mock-up.

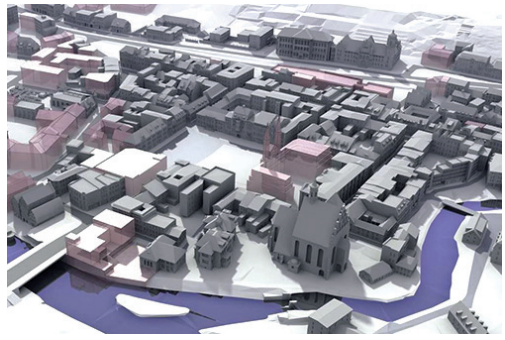


Fig. 3. Various graphical styles of an urban scale virtual mock-up

4.2.2. Developing extant state documentation using measurements performed by a Flexijet laser rangefinder

Modern equipment used to develop extant state documentation has revolutionised traditional methods of measuring and recording the extant state of buildings. One example of such a device is Flexijet[9]. The system is composed of a laser rangefinding device which cooperates with ArchiCAD software installed on a laptop computer that is connected to said device[10]. This enables us to perform an intelligent measurement, which is based on the simultaneous measurement and modeling of the elements of a virtual building. It is the first solution that allows the creation of intelligent BIM models directly during field measurements performed using a laser rangefinder.

[9] http://bimm-gmbh.de/en/ (access: 7.04.2017).

[10] http://www.archicad.pl/sprzet/flexijet/76-1/286-inteligentne-inwentaryzacje (access: 7.04.2017).

The Flexijet device performs laser-based measurements. During measuring, the Flexijet device plays the part of a computer mouse – each measurement is a "click", which defines the location of a point within a three-dimensional space. The measurements have a precision range of 2 mm. They can be used to construct the elements of a BIM model, such as walls, beams, pillars, etc. with the help of ArchiCAD tools. The model that we obtain is complete. It can also contain complex, parametric and intelligent BIM objects, such as windows, doors, stairs, railings, etc. This makes the data that is obtained easy to modify later on, serving as a basis for further design work.

Over half of the construction projects that are being carried out in urban areas are redevelopment projects. These often encompass the modernisation of complicated structures, including historical buildings. Flexijet guarantees the high precision of an extant state documentation regardless of the complexity of the measured building. Measurement precision is particularly important in the case of detailed construction documentation.

Fig. 4. Extant state documentation of the frontal facade of the Palace in Łobzów, which houses the Faculty of Architecture of the Cracow University of Technology, developed using the Flexijet device in 2014 at the Descriptive Geometry, Technical Drawing and Engineering Graphics Laboratory (A-43). Source: http://arch.pk.edu.pl/blog/2014/10/13/kurs-obslugi-systemu-pomiarowego-flexijet/ (access: 7.04.2017)

Flexijet also allows the performance of terrain surveying. This makes determining the shape and the location of key elements of a site quick and easy. Considerable benefits offered by the use of innovative BIM measurement technologies are also gained in the form of measuring complicated structural systems. The BIM model that is thus obtained can be used to easily extract a range of additional information, for instance, a schedule of the elements of a timber roof structure.

4.2.3. Laser scanning

The latest in terms of technology, the most precise, quickest and non-invasive method of gathering data about a building is measurement using a laser scanner. The effect of scanning various places (which takes only a couple of minutes) is measurement data in the form of millions of measurement points, which create a so-called point cloud within a three-

dimensional space. After initial refining, which takes the form of "superimposing" the results of measurements from mutually supplementing measurement stations, we can very quickly obtain a visualisation of the scanned structure. Scanning using this method can be used in the case of many types of structures, from telegraph lines and circulation routes, through industrial and engineering structures and their infrastructure, to historical and modern buildings of various types and sizes. The end product of the scanning process can be both the point cloud itself as the final product, as well as, after further refining – plans, maps, drawings, terrain models, visualisations and many other things.

In the case of laser scans, measurement is performed with a laser beam (which usually operates in the near infrared spectrum), which pulses at a very high frequency, thanks to a prism that rotates in the vertical plane. The maximum speed of scanning is dependent on the model of the scanner, usually amounting to a couple of thousands of points per second. The distance to the measured point is determined on the basis of measuring the time it takes for the beam to travel to the object and back. The measuring toolkit includes a laptop computer with specialist software. The data obtained using laser scanning is stored on a data storage device, which is most commonly a hard drive. This is caused by the sheer size of this information. However, along with the progress regarding flash cards, more and more scanners provide this option in terms of data storage. Afterwards, we are given the ability to export the data in the form of a point cloud to various formats, along with characteristic qualities. The products of laser scanning can be: a point cloud within a three-dimensional space, three-dimensional models, drawings and plans, orthoscans and virtual models.

Laser scanning is a relatively new method of obtaining a large amount of data using laser beams. It can be divided into:

- **Terrestrial Laser Scanning (TLS)**, which is performed using ground-based 3D laser scanners,
- **Mobile Laser Scanning (MLS)**, which is performed using an integrated scanner that travels on such platforms like boats, trains or cars,
- **Airborne Laser Scanning (ALS)**, which is a method obtaining 3D data for very large areas in terms of surface,
- **Satellite Laser Scanning (SLS)**[11], which is performed from Earth's orbit[12].

Laser scanning technology is continuously improving, while equipment is becoming more accessible and the software provides more and more capabilities.

One example of BIM-standard design support software that makes it possible to import a point cloud is ArchiCAD by Graphisoft. This program makes it possible to quickly create a BIM model on the basis of an imported point cloud. Especially concerning large and complicated buildings, this method of developing an extant state documentation greatly shortens the time that is necessary to perform work and eliminates mistakes made during the traditional measurement process.

[11] Kurczyński Z., *Lotnicze i satelitarne obrazowanie Ziemi*, Oficyna Wydawnictwa Politechniki Warszawskiej, Warszawa 2006.

[12] So far the only dinstrument using this technique was the Geoscience Laser Altimeter System (GLAS) which is fitted onto the ICESat satellite. The beam footprint of the laser sent from a 600 km high orbit on the surface of the earth is around 70 m, and their centres are placed around 172 m apart from each other.

Fig. 5. Example of a scan of a building in the form of a point cloud. Source: https://blog.graphisoftus.com/graphisoft/archicad-19-new-features-point-cloud-introduction-to-point-clouds (access: 7.04.2017)

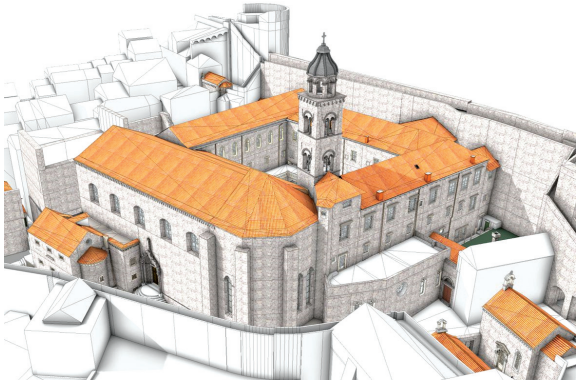


Fig. 6. Extant state measurement of the Dominican Monastery in Dubrovnik, with a surface area of 5500 m^2, performed by BimPoint from Krakow, which required carrying out around 1100 laser scans. Source: http://bimpoint.pl/ (access: 7.04.2017)

The most common applications of 3D scanning using point clouds are:

- the modelling of existing buildings before renovating or redeveloping them,
- site modelling before designing new structures,
- the modelling of existing structures of buildings in order to identify inconsistencies in existing documentation.

We can obtain the following information from a BIM model developed on the basis of a point cloud:

- a database of information on a structure required for further renovation, conservation or design work,
- a cohesive and precise technical documentation,
- all types of technical drawings (floor plans of any level, cross-sections through any location, external and internal facades),
- quantitative schedules: of rooms, doors and windows, as well as other elements of a building,

- analyses: daylighting, energy balance, structural, etc.,
- a bill of quantities and a cost estimate of construction work,
- all types of perspective or axonometric views, virtual walkthroughs, visualisations and animations,
- etc.

A typical scope of a building's extant state documentation includes:

- a BIM model, which is a virtual building (.ifc, .pla, formats, etc.),
- technical documentation in a scale of 1:50 or 1:100 (.dwg, .pdf, .jpg format) – floor plans of characteristic levels, facades, cross-sections through any location, etc.,
- schedules: of rooms, volume, windows and doors,
- perspective and axonometric views of a structure and additional 3D cross-sections,
- a 3D model of a structure useable on mobile devices, enabling virtual walkthroughs,
- additional data packages depending on specific needs.

The market's demand for designers who can effectively work within the scope described above is becoming higher each year.

4.2.4. Creating library objects

The creation of object libraries which are meant to represent specific commercial products for the most popular computer-aided design software has become one of the most dynamically developing specialisations around the architectural and interior design professions.

- From the manufacturers' point of view, professional libraries with systematised objects that represent specific commercial products that can be placed in a design developed by an architect have become one of the most effective means of reaching clients and increasing the sales of offered products.
- From the point of view of the architect, the capability of inserting specific commercial products that retain their true dimensions and technical descriptions into a design greatly eases and quickens design work. This applies to both elements like all sorts of construction systems, windows, doors and gates, as well as all types of infrastructural and fitting elements of a building, etc.

The basic idea behind online services that offer object libraries that represent specific commercial products can be described in the following manner[13]:

- Manufacturers supply a service with their products in digital form – CAD drawings, technical data, descriptions, photographs, calculations. The online service provides special tools that enable this information to be entered.
- Manufacturers reach out to designers with their offering – they send emails to the inboxes of designers who are registered with a service, order their own CD-ROMs through the services, "opening their windows" to the designers with their offerings on their own websites, etc.

[13] https://www.archispace.pl/node/31231_Oferta (access: 7.04.2017).

- Designers search for and select specific products through an online service's search engine, the manufacturers' websites, or by reading messages sent by manufacturers.
- Designers download products from the online services of manufacturers – in the form of CAD drawings, drawings showing the details of construction solutions, layout patterns, etc.
- Designers include the manufacturers' products in their design documentation in the form of premade product drawings – 3D models, floor plans, cross-sections, views, drawings of construction details and their descriptions.
- Manufacturers deliver their products to a construction site.

A complete offering of online services that supply object libraries should include file formats for all the most popular CAD programs, starting with 2D CAD files, we should expect to find among search results files for AutoCAD, BIM-standard software should include files for Revit and ArchiCAD, while in the case of programs which are used to make photorealistic visualisations, files for 3ds MAX and other CAD systems.

Object libraries, depending on the product that they represent, should be developed in different ways:

- **2D object libraries** – for designs developed in AutoCAD, as flat drawings of top-bottom and side views, as well as cross-sections.
- **3D object libraries** – for designs developed in BIM programs (for instance ArchiCAD and Revit), as well as for Sketchup and 3ds MAX.
- **Parametric object libraries** – for designs developed in BIM programs (for instance ArchiCAD and Revit). Parametric objects make it possible to change certain parameters within the objects, such as width and height of doorframes or the type of door. They require adapting the objects to the library module of the program for which they are designed.

An example of the use of object libraries in the design of a toilet for disabled persons has been presented in Figure 7.

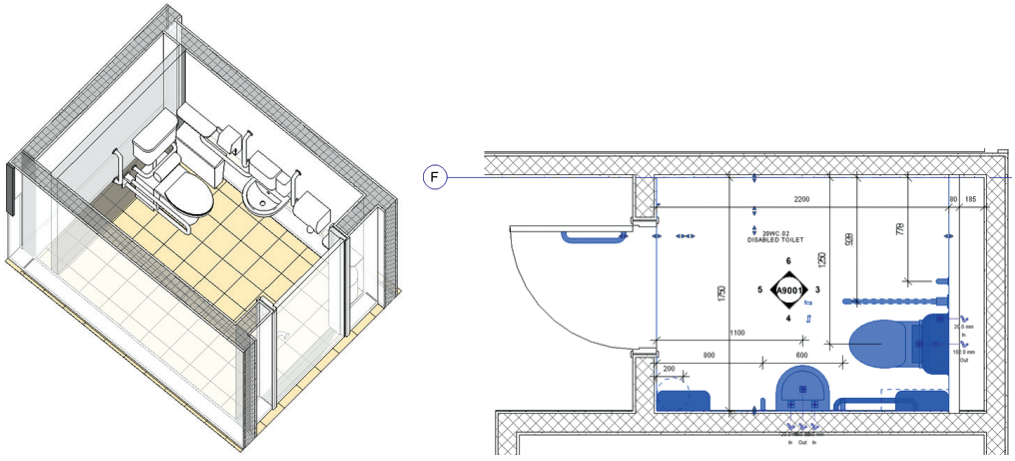


Fig. 7. Working with a BIM Library – highlighted in blue is the BIM Library plumbing fixture accessible toilet family with shared nested plumbing fixtures; the accessories are available as individual families. Source: https://www.thenbs.com/knowledge/working-with-nbs-national-bim-library-content (access: 21.04.2017)

Using libraries featuring premade elements is becoming widespread among architects, even already at the stage of their architecture studies. We can download all sorts of elements from the numerous object library sharing forums, so that we can use them in the designs that we work on. However, despite a very large and constantly increasing offering of online services with object libraries, a very large amount of products are still not available. Thus, the ability to create our own object libraries is greatly needed and should be introduced into the educational curriculum. Furthermore, due to the large market demand for designers who can create such objects, it is an interesting possibility of the commercial application of one's skills and cooperation with professional services.

4.2.5. New presentation methods – BIMx

According to architecture students, the printing of course design theses is associated with a heavy financial cost. The cost of preparing a complete set of full-colour sheets glued to a stiff sheet of foam that presents a typical design work along with visualisations amounts to around 200 PLN. Over the course of a semester, students typically produce around 5 of such designs. When the costs are multiplied by the amount of students, this produces truly sizeable amounts of money...

As it turns out, it is not necessary to pay these costs, as new technologies in the presentation and archiving of design work allow us to make multimedia presentations of a design in a manner which is qualitatively incomparable to traditional paper prints.

The ability to present a design in the form of an interactive multimedia presentation broadens the possibilities of using it so that they exceed traditional applications.

One example of such a technology is BIMx software (Building Information Model eXplorer)[14] – it cooperates with ArchiCAD and allows the creation of a virtual presentation of a designed building. The program converts the model of a building into an .exe file, which can be sent to a client. After opening the file we can take a walk around the building like in a computer game, moving between rooms, using doors. By using appropriate virtual reality goggles like Oculus, we can obtain an illusion of depth. Such a virtual visit to a building which does not exist yet is an extraordinary experience for a client, while enabling architects to more effectively sell their designs. The GRAPHISOFT BIMx application has led to a situation in which BIM has also reached parties which are interested in, but that are not directly participating in the process of a building's design, for instance to contractors, manufacturers, clients and owners of buildings or real estate agencies.

The program is available in Polish.

BIMx models can be viewed on mobile devices such as iPads/iPhones. Thanks to a dedicated free application, BIMx files can be opened on any device using iOS or Android, as well as on Mac or Windows computers. The application is linked with the Facebook user community which shares these types of files.

[14] GRAPHISOFT BIMx Desktop Viewer User's Manual, GRAPHISOFT, 2016.

Fig. 8. BIMx Application, used to present a BIM-standard design on mobile devices like iPads/iPhones. Source: https://itunes.apple.com/us/app/bimx-building-information-model-explorer (access: 7.04.2017)

New uses of programs of this type, like the aforementioned BIMx, are still growing. Thanks to the ability to "plug in" any database and additional information to the presentation of a virtual design saved in a BIMx program, we can use such a presentation in numerous, entirely new ways:

- **The presentation of real estate developer designs** – a virtual design of a building can be visited using Oculus-type goggles, selecting furniture and fittings.
- **Designs of exhibitions** – exhibits that we can "approach" in a virtual model of an exhibition can have any and all descriptive information, multimedia presentations and videos "plugged in".
- **Interactive urban mock-ups** – the ability to "attach" various types of interactive information into any selected points of a virtual urban mock-up can help both in the creation of designs used to tour an urban space across various education routes, as well as commercial projects – that promote certain structures or services.
- **Instruction manuals regarding the equipment and infrastructure of a building** – a virtual design of a building can be used during the operation phase, for instance by "attaching" various types of instruction manuals to the technical elements of a building, as well as infrastructure drivers.
- etc.

Architects who can create these types of innovative virtual presentations for iPad-type mobile devices have a wide field from which to obtain new commissions and broaden the market for their services.

5. Summary and conclusions

Even a cursory review of new technologies that support architectural and construction design as well as the new specialisations that have emerged around the architectural profession indicate that currently ongoing changes are revolutionary in character – to a much wider

degree than the changes that occurred a couple of decades ago when manual drafting had been replaced by computers with the first 2D CAD systems.

The ongoing changes are invariably leading to alterations in the current design standards, based on 2D CAD systems, in the direction towards the widely understood BIM standard.

The new specialisations which emerge in association with the technological progress around the architectural profession broaden the traditionally understood market for architectural services.

The subject of designing using the BIM standard and the new specialisations should be, in the author's opinion, introduced into the curriculum of architecture students as fast and as broadly as possible.

References

[1] ds, *BIM Curriculum*, Graphisoft, 2013, http://www.graphisoft.com/learning/training-materials/bim-curriculum/ (access: 21.03.2016).

[2] Nordby A., *Przewodnik MaTrID – Zintegrowane projektowanie*, European Commission Executive Agency for Competitiveness and Innovation, 2013, 3.

[3] Sydor M., *Wprowadzenie do CAD. Podstawy komputerowo wspomaganego projektowania*, Wydawnictwo Naukowe PWN, Warszawa 2009, 47.

[4] *BIM Curriculum*, Graphisoft, 2013, http://www.graphisoft.com/learning/training-materials/bim-curriculum/ (access: 21.03.2016).

[5] http://bimm-gmbh.de/en/ (access: 7.04.2017).

[6] http://www.archicad.pl/sprzet/flexijet/76-1/286-inteligentne-inwentaryzacje (access: 7.04.2017).

[7] Kurczyński Z., *Lotnicze i satelitarne obrazowanie Ziemi*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2000.

[8] https://www.archispace.pl/node/31231_Oferta (access: 7.04.2017).

[9] GRAPHISOFT BIMx Desktop Viewer User's Manual, GRAPHISOFT, 2016.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.17.187.7416

Adam Nadolny (adam.nadolny@put.poznan.pl)
Division of History, Theory and Heritage, Faculty of Architecture, Poznan University of Technology

Modern architecture of the 1960's of the XX century recorded in a movie as an element of heritage protection

Architektura modernistyczna lat 60. XX wieku zapisana w obrazie filmowym jako element ochrony dziedzictwa

Abstract

The article aims to present the ways in which feature films of the 1960s, with all their beauty, became a canvas of times so important for the history of Polish architecture. Writing about the past exceeds a simple description of a historical object. It is also an attempt to present it on a film reel. In my reflections, I wish to concentrate on the relations operating between a feature film and such an architectural object, which, over the course of time, became iconic for the discipline. I want to focus on considerations regarding the ways of defining film and architecture as an element of spatial and visual memory, in the context of Polish heritage protection of architectural modernism.

Keywords: modernism, heritage protection, film

Streszczenie

Tekst ma na celu ukazanie, jak obraz filmowy lat 60. XX wieku – ze wszystkimi jego walorami – stał się zapisem czasu tak ważnego dla historii polskiej architektury. Pisanie o przeszłości to nie tylko opisywanie obiektu historycznego, to także próba ukazania jego obrazu na taśmie filmowej. Chciałbym skupić się na rozważaniach dotyczących traktowania czy też definiowania filmu i architektury jako elementów pamięci zarówno przestrzennej, jak i wizualnej w kontekście ochrony dziedzictwa polskiej architektury modernistycznej.

Słowa kluczowe: modernizm, ochrona dziedzictwa, obraz filmowy

1. Preliminary considerations, heritage protection of the architecture of the second half of the XX century observed in a feature film

Protection of heritage is a very broad phenomenon in contemporary interdisciplinary studies. However, within the definition of the phenomenon, there are categories that can be applied to the Polish feature film covering the 60's of the 20th century[1]. In the case of heritage protection, in the broad meaning of the word, we can, on the one hand, treat the film as a medium in which the history of a particular architectural object or an entire urban concept is saved; on the other hand, feature film becomes for us a record of the era. Moreover, in many cases, it is a record of the times in which the object was created.

Unlike the static photographic image, a movie gives us an opportunity to watch, explore or travel through a particular object whose spatial form has been closed on a film tape and became part of the plot of a film. A motion picture enables to show an object in the same way as it has been seen by a user of the space, a resident, passer or an architect. A static, *a natura rei*, form of a building becomes a spectator's journey into the past. The question arises why exactly the architecture and Polish urban design of the 60's of the twentieth century is so valuable as an element of history that it should be promoted through a film. There are several answers to such a question. The first concerns the quality and some sort of a uniqueness of that period in the Polish history of architecture. Despite many constraints that derived from the circumstance of designing in a socialist political system, after 1956, Polish architects turned themselves back on the road of modernism, in both: European and international means. They were also aware of its limitations, but a passion for simple forms in the landscape of the city remained. The second factor that I would like to emphasize is the fact of the underestimation of scientific and study qualities of a film medium shared by Polish architectural historians and restores. As part of the material heritage of the era, it should be treated equally with such research materials as plans, projects, design documentation and photographs. In many cases, the archives of Polish design studios after 1989 were dispersed or liquidated. This has resulted in a kind of loss in historical sources for the purposes of studies of architecture of this period. For film experts, movie is a natural material for an inquiry, and for historians of architecture, not necessarily[2]. The third reason being the fact that in modern world of media, motion picture appeals to a mass audience much more than a static image.

---

1 This text is part of the author's research conducted within the research project "The architecture and the city in Polish feature film of the 60's, 70's and 80's of the twentieth century", stage III, funded within the statutory activity by Ministry of Science and Higher Education in 2016 at the Faculty of Architecture Poznan University of Technology.

2 Cooperation on research upon issues of architecture in the Polish feature film and documentary was launched in 2015 within the framework of a nationwide scientific conference "Architecture in Polish feature film and documentary of the 50's and 60's of the twentieth century", which the author has initiated and organized from the side of the Faculty of Architecture of Poznan University of Technology. The conference was organized in cooperation with the Department of Film, Television and New Media of Adam Mickiewicz University in Poznan.

2. Polish feature film from the 1960's as a source of research upon an architectural heritage protection of the second half of the XX century

Polish architecture of the 1960s, as we well know, has become a topic of research in recent decades. That undoubtedly picturesque and dynamic period in the history of Polish architecture has left its mark on many buildings of the epoch. Having that in mind, it is crucial to mention that, in Polish feature films of that period, we can see pictures of architectural works, which in retrospect are recognized as iconic for Polish post-war architecture. At the very beginning, I wish to present a selection of Warsaw examples, for which a film became the record of their history. A building, which appeared quite frequently in movies of the era, was Supersam. Although it ceased to exist, it continues to operate within public consciousness in such films as *Sam pośród miasta* [Alone among the city] (1965), *Lekarstwo na miłość* [Remedy for love] (1966) , *Dzięcioł* [Woodpecker] (1970) and others. Thanks to motion pictures, we can travel along with the characters of the movie throughout its interiors, observe the breath-taking design of its sales room or participate in their meetings under a distinctive sign of the façade.

Another object, which, on the one hand, has been firmly inscribed in the landscape of the city and simultaneously is also considered iconic, is a complex of buildings called The Eastern Wall. This composition, bearing the stylemarks of modernism, became a favourite *en plein air* of filmmakers in such movies as mentioned above *Lekarstwo na miłość* [Remedy for Love] (1966), *The Game* (1968) or *Człowiek z M3* [The man from M3[3]] (1968).

The object, which, for its scale and architectural quality, is perceived as a cult building, that appeared in movies of the era, was a so-called the CDT "Smyk" Department Store[4]. Its slender and characteristic form with horizontal windows, and a recognizable cafeteria suspended over the street appeared in such movies as *Wyrok* [The Sentence] (1961), *Lekarstwo na miłość* [Remedy for Love] (1966) and others.

The cinema called "Atlantic", with an expressive neon on the front facade, which was a synonym of fascination with electric light in the Polish architecture, appeared in the following films: *Do widzenia, do jutra* [Good bye, see you tomorrow] (1961) *I ty zostaniesz Indianinem* [And You Will Become an Indian] (1962). The object, which is also regarded as an achievement on the ground of Polish modernist architecture, is the Central Olympic Training Center in Warsaw, the hall of boxing and wrestling. The building with a characteristic hanging roof structure appears inter alia in the feature film *Hasło Korn* [Password Korn] (1968).

On the list of objects presented above, it can be deduced that only Warsaw was the scene of film shoots in the 1960s. This was not true, though. Along with large cities, which appeared on the screen, namely Kraków, Gdańsk, Łódź, also smaller centers have their place in the history of cinema. Bielsko-Biała is one such example – whose swimming pool "Panorama" appears in the movie *Jutro Meksyk* [Tomorrow Mexico] (1965). Following the trace of sport, one should mention a no longer existing complex of swimming pools, on Polanka Redłowska in Gdynia, which has been shown in the aforementioned film.

[3] According to PRL [PPR] nomenclature, M3 is a flat designed as a habitat for 3 people; its program, though, was limited to 2 rooms and a separate kitchen.

[4] Centralny Dom Towarowy = Central Department Store.

At this stage of reflection, I wish to draw attention to the fact that, in the case of the films of the 1960s, modernist architecture was presented as a kind of manifestation of modernity. In the motion pictures of the era, Polish architecture is perceived as very interesting and intriguing. In my view, this is another proof that, through film, we can try to preserve our architectural heritage of this period.

Fig. 1. Joanna (Kalina Jędrusik), main character of the film entitled „Lekarstwo na miłość" ['Remedy for Love'], against the background of the modern building of the Super Sam store in Warsaw, which no longer exists 1-F-2306-280

3. Polish feature film from the 1960's as a source of research upon the heritage protection of urban design of the second half of the XX century

The city with its modernistic ideas, proclaiming love for geometry, form and function, was in opposition to the traditionally perceived city with a winding network of streets and squares. Modernism promoted open space, less intimate than a historical city. The post-war wave of urbanization of Polish cities during the 1960s as we know proceeded in two directions. Firstly, there were still works concentrating on a reconstruction from the war damage of the historic centers of Polish cities. In the meanwhile, in the peripheral districts, spatial structures of housing estates were created – based on the principles of modernism.

The application of this style referred to newly constructed, both: fragments of downtowns, as in the case of The Eastern Wall in Warsaw, as well as new districts located outside of the city center. A result of that period in the Polish town planning are the numerous examples of housing areas erected in the so-called large panel system (constructed of pre-fabricated,

pre-stressed concrete). Such a not entirely realistic view of that new urbanism was depicted in the movie *Mocne uderzenie* [Hard strike] (1966). The main characters were captured on the background of a habitat Sady Żoliborz designed by Halina Skibniewska. This in a sense intimate area has become synonymous with modernity recorded in the movie frame.

Another important factor in Polish post-war town planning was the ideology according to which the new socialist society can and should in fact be modern. This modernity in the ideological sense were translated into urban aspects. Urbanization and modernity were very well attuned to the official language of the system, which promoted the image of a country, which was developing greatly, and at many levels. I would like to recall the words of Adam Kotarbiński, who in 1985 diagnosed in the following words the achievements of socialism in this area: "Realized (erected) achievements of urban design are reflected most clearly in the major complexes of buildings made publicly available as a finished whole, in the form of important urban systems, new structural and architectural objects along with orderly environment" (Kotarbiński 1985, p. 123).

An example of a motion picture in which we can grasp an idea of the changes in the centers of Polish cities is, for example, the film *Pieczone Gołąbki* [Baked Stuffed Cabbage] (1966), in which the camera captures the image of the newly created units on Bielańska Street in Warsaw. Another graphically valuable sign of modernity is the film *Gra* [The Game] (1968), through which we have the opportunity to see the process of demolishing the XIX century structure of the downtown of Warsaw for erection of a housing area of the Eastern Wall.

Urbanization of Polish cities in the spatial sense also carried an important social aspect. Modernism, to some extent, influenced the formation of a new society, new representatives of a post-war Poland. Fascinated by modernity, the otherness of the Modern City, the metropolitan character, the hubbub, as well as the play of light and shadow in space. In a sense, all these elements that were in use of the film of that period.

4. A reconstructed city in a feature film from the 1960's as an element of the heritage of the second half of the XX century

The reconstruction of the historical city centers in Poland in the 1960s proceeded towards the restoration of their spatial as well as visual and, in a sense, spiritual values – as a testimony of the past and national heritage. In many cases, the historical fabric of the city was complemented with modern forms, especially in these fragments, which did not have sufficiently preserved source materials.

However, in many cases, reconstruction proceeded as a result of assumptions, whose primary goal was to restore in a historical city (in example Warsaw, Gdansk or Poznan) as many historical buildings and historical parts of the downtown as it was possible. Thus, feature films produced in the 1960s can be treated in the means of cinematographic actions, thanks to which an information on the time of their reconstruction or the way of functioning is recorded, in the moment just after the process of their rebuilt.

As an example, I would use the film *Do widzenia, do jutra* [Good Bye, See You Tomorrow] (1960) where we have the opportunity to walk through the old town of the historic Gdansk, in

different stages of its reconstruction. Another movie, which provides the trace of documentation of the past of a reconstructed city, is *Małżeństwo z rozsądku* [Marriage of convenience] (1966). In that movie, the main characters live their adventures in Warsaw's old town.

Motion picture is, in my opinion, a very valuable comparative material for an architectural researcher. In many cases, it becomes a suitable complement of a story behind a particular work of architecture or urban design.

5. Summary

The research on the architecture and urban design of the 1960s can never be performed without an attempt to analyze an architectural and urban design of the period recorded in films. A great number of publications of urban as well as architectural works in journals such as, say, *Architecture, City* and other, give us the opportunity to trace the evolution of modern trends in design. From the standpoint of a researcher in the area of architecture, a film can be regarded as an important source of iconography. It surely is a great complement to archival materials and photographs.

One of the most natural features of an urban structure – similarly to an architectural object – is its stability in a city space. At the same time, it is spontaneously perceived by its user through walks and passing by. A static image of architecture, rendered on a photography, drawings and plans, does not always allow for an adequate transposition of all the values of an object that are possible to be depicted in the movie. Therefore, a feature film should be treated as an important element of the architectural and urban heritage of the second half of the XX century.

References

[1] Nadolny A., *Drawing as a record of the space in movie*, [in:] *Wyzwania XXI wieku. Rysować, malować czy korzystać z komputera*, (ed.) M.J. Żychowska, Wydawnictwo Politechniki Krakowskiej, Kraków 2015, 63-69.

[2] Kotarbiński A., *O ideowości i ideologii w architekturze i urbanistyce*, Arkady, Warszawa 1985.

[3] Nadolny A., *The image of modern architecture in the polish feature films of the 1960s – a photographic recording of modernity*, [in:] *Photography and Modern Architecture*, (eds.) A. Trevisan, M.H. Maia, C.M. Moreira, Centro de Estudos Arnaldo Arujo, Escola Superior Artistica do Porto, Porto 2015, 235-249.

[4] Nadolny A., *Changeability or stability: reflections on the issue of the modern city as shown by the polish film picture after the mid-20th century*, Technical Transactions, vol. 1-A/2014, 127-141.

[5] Nadolny A., Świt-Jankowska B., *Theatre space in Polish films of the 1960s. Records of designing concepts and theoretical discourse*, [in:] *Dramatic architectures. Places of drama – drama for places*, (eds.) J. Palinhos, M.H. Maia, Centro de Estudos Arnaldo Araújo, Escola Superior Artística do Porto, Porto 2014, 491-504.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.17.188.7417

Ernestyna Szpakowska-Loranc (ernestyna.szpakowska-loranc@pk.edu.pl)
Institute of Urban Design, Faculty of Architecture, Cracow University of Technology

Architectural narration in shin takamatsu's works

Narracja architektoniczna w twórczości shina takamatsu

Abstract

The author of the article investigates Shin Takamatsu's works in search of narration studied on the basis of the architect's buildings along with his own descriptions and texts interpreting his work – with particular reference to Felix Guattari's analysis. The use of metaphors and other devices of architectural poetics, such as architectural quotations, natural and mechanical inspirations, abstract and iconographic signs and cultural symbols in Takamatsu's structures equates them with literary works, which has been reflected in the architect's statements. The artist turns them into a coherent language with a particular narrative.

Keywords: Guattari, narration, metaphor, Takamatsu

Streszczenie

Autor artykułu poszukuje narracji prac Shina Takamatsu, przeanalizowanej na podstawie jego budynków, wraz z autorskimi opisami, oraz tekstów interpretujących twórczość architekta, ze szczególnym wskazaniem analizy Felixa Guattariego. Metaforyka struktur Takamatsu, zastosowanie środków „poetyki architektonicznej", takich jak cytaty z architektury, inspiracje formami naturalnymi i mechanistycznymi, wykorzystane abstrakcyjne i ikonograficznie znaki oraz symbole kulturowe, przybliża je do dzieł literackich, co potwierdzają słowa architekta. Tworzy z nich spójny język z określoną narracją.

Słowa kluczowe: Guattari, narracja, metafora, Takamatsu

1. Introduction

What I mean is that to tell a story, you must, first of all, construct a world, furnished as much as possible, down to the slightest details… The problem is to construct the world, the words will practically come on their own. Rem tene, verba sequentur, wrote Umberto Eco in the *Postscript* to *The name of the Rose*[1]. These sentences about creating novels may also refer to creating architecture; architecture interpreted as a story – i.e. each building forming a separate world, created with the use of metaphorical forms and other poetic means. A user of architecture penetrates this individual world when entering the building, but this world is also reflected by an external form of the architecture – the building in its context. The narratives of buildings co-create the story of the city.

The end of the 20th century and the 21st century tend to be defined as the time of narration; the period in which the concept of narrative occurs in an increasing number of domains[2]. Based on the discursive nature of human life and culture, as proclaimed by Anthony Giddens among others, it also enters the art of building. It has been said that man is a narrative animal – *homo narrativus* "writing" his own autobiographical narrative within the framework of the "I project"[3]. According to this narration, he creates a physical reality around him.

Shin Takamatsu's architecture is an example of creative work confirming Eco's words and enabling one to study narration understood as a story about a building-world; i.e. reading the architectonical text, the narrative created by denotational references[4] of the building – metaphors, expressions, etc. The realisations of the Japanese architect, who draws inspiration from the world of technology, nature, art and culture, create separate realities.

Accumulation and transposition of symbols and quotations have been a consistent manner of Takamatsu's work (from the 1980s to the present), regardless of the fact that he constantly changes the source of inspiration and stylistics of forms. His method of creating, however, is constant and very consciously built (presented by Takamatsu himself and analysed in the essays of Felix Guattari among others). The methodology may be illustrated as a triangle of elements: word, image, architecture. An observer of Takamatsu's buildings is therefore confronted with structures as "overdrawn" images, most of them "cleansed of abstraction" – most often with rich figurative features (as presented in conceptual sketches of the architect). Simultaneously, according to Takamatsu, these are structures that use words. The concepts reflected in images confirm the statement of the architect: *Architecture is generally unconnected to abstraction* [13, p. 142].

---

1 Eco U., *The Name of the Rose*, Houghton Mifflin Harcourt, Boston 1994, pp. 549-550.

2 Philosophy, historiosophy, psychology may be mentioned here, as well as "existential marketing" [7].

3 Umberto Eco coined the definition of man as a storytelling animal, whereas the term homo narrativus comes from Barbara Hardy, and autobiographical narration – "I project" is described in *Modernity and Self-Identity. Self and Society in the Late Modern Age* of Anthony Giddens [7, pp. 30, 33].

4 Eco writes about denotation of buildings, i.e. transferring their functions in a conventional way, by means of formal codes, in *The absent structure* [Eco, p. 209-212]. Denotations as an element of the "dictionary of references" of architecture were also described by Nelson Goodman [5, s. 664].

2. Structural narrative

Consistently used architectural quotes, natural and mechanistic forms as inspirations as well as iconographical signs and symbols used in buildings and their details can be observed in Shin Takamatsu's artwork throughout the years. These means of "architectural poetics" have been specified on the basis of analysis of the form of buildings and Takamatsu's descriptions of his own projects, where he characterizes the origin of each building.

The architectural quotes[5] include examples of: a group of pyramids at the Nima Sand Museum (Shimane 1989-1990), the fortified towers of the Business Center in Tbilisi (2001-2007), and the fortress of Pharaoh (Kyoto 1983-1984), the traditional Chinibu wall at the National Theatre (Okinawa 1998-2003), and Tō-ji pagoda in the Henjoto wooden tower project (Kyoto 2005). A pretext derived from modern architecture and used by Takamatsu in a contradictory manner at the Gotsu Community Center (Shimane 1992-1995) are multipurpose halls built in Japan from the 1960s. The architect himself suggests using the borrowed forms and transferring their meaning, so that they can be associated with architectural quotations as defined by Remei Capdevila-Werning, taking this phenomenon through the analogy from Nelson Goodman.

Takamatsu's inspirations with natural forms appeared e.g. as the sacred Buddhist mountain, Nose Myokensan, reflected in the tower of a worship hall in Hyogo (1995-1998) or various wings – of a butterfly as skylights in Earthtecture Sub-1 (Tokyo 1987-1991), a bird in Ela Tower Project (Tel Aviv 1995-1996), a swan as the structure of Tianjin Great Museum (Tianjin 2000-2004). In the last of these buildings (called Swanium) the bird is to be a symbol reflecting the glory of the city. An example of deriving a metaphor from the world of inanimate nature is Meteor Plaza (Shimane 1994-1995), where the oval body of the exhibition space takes the meteorite shape.

In the 80s and 90s of the twentieth century, a distinctive narrative motif in Takamatsu's artwork were machines. The *engine of trade and dynamism* [13, p. 116] may be recalled here – a symbolic form of Minatosakai Community Centre (Tottori 1994-1997), or Ark (Kyoto 1981-1983) – *a frozen locomotive* [13, p. 7]. However, mechanistic details or larger elements can be found in most realisations of that time.

Since the 80s, symbolic use of geometric solids and iconic signs has been apparent in Takamatsu's projects. This concerns an inverted cone wedged into an elliptical cylinder showing that *architecture has a symbolic obligation to the unavoidable strength of the site* [13, p. 96] at Nagasaki Ferry Terminal (Nagasaki 1993-1995), a fragment of a sphere as the reservoir at Tamayu Health Spa (Shimane 1993-1996), the letter O as the main motif of Omula Beauty College building (Fukuoka 1996-1998), or an elliptical 300 m high ring as a hotel in Ringdom Project (Tokyo 2006). It is also visible that the architect draws inspirations from other cultural elements, like the figures in Shoji Ueda's photograph titled *Four Girls Posing*, transposed into four exhibition spaces of Shoji Ueda Museum of Photography (Tottori 1993-1995). The geometry of the previously mentioned Meteor Plaza can also be added to this group.

[5] About quotations in architecture cf. [1].

The architecture of Shin Takamatsu may be perceived as an example of the narrative described by Nigel Coates as *binary narration*. This kind of narrative is defined by Coates as giving an object or an *architectural situation* a *parallel identity* – a particular *costume*, derived not from its function but from the so-called *trans-function*. Imagination, sublimation and transgression are used in the creation and reception of the binary narrative of architecture. An example of space created as an attempt to create illusion described by Coates (a Ferrari bed or an interior of a restaurant with eastern cuisine pretending to be far east) is probably the most obvious example of this type of narrative [4, p. 83]. The "architectural situation" discussed by Coates may be analysed as the accumulation of metaphors and quotations among others as well as the use of architectural poetics (in part formally derived from the literary language in a given building (or a group of buildings). Thus, the created narration may be defined as "structural" as the one referring to the structure – an architectural form (and giving it a specific costume) and devoid of the notion of an event ontologically related to time.

In spite of an inseparable relationship of architecture with time (e.g. according to Juhani Pallasmaa, apart from living in space, we also live in time [10], while Karsten Harries has portrayed building as a history of struggle with terror of time [Harries]), the "architectural" time is heading towards eternity[6]. Traditionally, the essence of architecture was its durability, although in the second half of the 20th century, the concept of an event began to appear more often with reference to the art of building[7]. While literacy narrative consists in creating stories as events that take place over time, the narrative of modern architecture may use the concept of an event or not. Eliminating this element, it operates only with the concept of time in the sense of "heading towards eternity" – a monumental time of a work of art, and not the time of a novel's events.

3. Takamatsu's architecture and language

In *Architecture and Language*, an essay by Masaru Kawatoko published in Takamatsu's collection of works, Kawakoto calls the architect's artistic goal *Architecture-ness*, a phenomenon that cannot be characterized with words, although it is words that begin every work of the Japanese architect. They construct an idea and thus enable one to encounter the genesis of architecture. According to Masaru Kawatoko, Takamatsu's words and sentences (sometimes described by critics as incomprehensible [6]) actually express the architect's impatience in articulating thoughts with *preexisting language*.

---

[6] The starting point for these considerations is the fact that architecture traditionally belongs to arts of space, and not arts of time. Karsten Harries wrote that "the language of beauty is essentially the language of timeless reality", and Juhani Pallasmaa quotes an artist Sandra Illiescu for whom experiential time in contact with a work of art differs from chronological time, linking past, present and future [10]. With regard to contemporary architecture, Mark Wigley considers buildings to be environments of events rather than single events [15].

[7] The concept of an event in architecture – "the world as a tissue of events", the basis of the city-event is characterised by Ewa Rewers among others [12, pp. 71-99].

Simultaneously, the commonly used language does not accurately describe the richness of Takamatsu's thought (as Kawakoto points out), which is a paradox for the fact that it is the words that construct the conceptual ideas. Consequently, the architect creates his own language, strengthening the existing meanings of the words used, and then introducing abstract elements into his designs – *making into languages, making into lines, making into figures* [13, p. 14]. According to the architect, it is impossible to describe the structures which are thus created by means of language. Takamatsu compares architecture to literature, understood specifically:

If literature is the creation of a yet unknown world through the command of known language, it's no exaggeration to say that this is extremely close to literature. To realize the existence of something of which no one is yet cognizant, through a command of known things in architecture – that is, materials and structural methods – and known ideas about architecture; I think that is the architect's creativity [13, pp. 14-15].

It can be assumed that the postulate of creating the world by means of literature brings to mind the above-quoted words of Umberto Eco. For the Japanese, however, the real architecture must exceed the language. Using words, he composes them in a way impossible in literature or poetry – *dramatic, overwhelming, unparalleled, unprecedented*. Architecture becomes a literary machine. Takamatsu constructs, transforms, destroys, and finally reconstructs.

4. Narrative devices

Felix Guattari characterised the stylistic means used by Takamatsu in the 1980s to create new worlds of references – *placing in existential suspension*. Guattari distinguished: disrupting symmetry (indicated by him in a near-obsessive use of vertical lines, yet distorted with the use of e.g. an oblique and transverse lines) and dicentric composition of forms, horizontal and vertical cuts as well as dividing buildings into two parts of different styles. He also pointed out openings and stairs leading into voids and the use of round forms [6, p. 135]. While these stylistic devices have evolved over the years, it should be noted that the language of Takamatsu's of architecture has such invariant features as deliberate separation of form and function, individual references to spatial and cultural contexts and specific use of quotations and metaphors.

Takamatsu rejects the literal interpretation of buildings' functions in their spatial forms. A clear example of such an action is the Origin I Building (Kyoto 1980-1981). The building got no function at the design stage; the investor decided to assign it to a completed facility only. Takamatsu considered working under such conditions as the opportunity to provide a visual answer to the question of what architecture is. The separation between form and function is also evident in Minatosakai Community Centre – a building with a multifunctional programme designed to support international trade. Takamatsu decided to symbolically to symbolically illustrate the message of the building in the form: the *engine of trade and dynamism* [13, pp. 24, 116].

The main façade of Origin I, with contemporary expression, mechanistic details and abstract shapes of openings, was juxtaposed with a longitudinal block of the building and

symmetrical distribution of spaces in its plan. The longitudinal block is similar to neighbouring buildings and the symmetry of the plan is reminiscent of some classical buildings. Thus, forms derived from spatial context and history were mingled with contemporary material – concrete – creating ambivalence: an amalgam of life, the richness of history and acceleration (movement into the future). According to Takamatsu, this ambivalence is the feature most required from architecture.

In another building, Pharaoh (a home and dental clinic), the architect uses the motif of a fortress as the genesis of the city. The "bulky" block was equipped with "embrasure windows", battlements and machicolations and "studs" in metal stripes along with small, round windows. The high traffic around the building resulted in the "defensive architecture", different inside and outside. The schizophrenic "microcosm" of the building and its immersion in the context of the place comprised the two worlds of architecture. This presents the Takamatsu's method of shaping architecture in the city – as two separate dimensions.

It can be considered that in the approach to function and context, Takamatsu's architecture fulfils the premise of Anthony Vidler's "third typology"[8] – typology based on a city. It is this metropolis that adds value to the narrative of architecture created by the Kyoto-based artist. Recalling two architectural models of creating it in relation to the environment that have emerged in history – Corbusieran immersion of the geometry of the designed forms in their context, and negation – buildings as separate works, detached from the neighbourhood (forms depending only on the quality of the structural objects themselves), Guattari finds the third way in Takamatsu's designs. His buildings are meant to be perfect aesthetical objects, simultaneously "hypersensitive" in relation to the surroundings [6, p. 133]. It is the neighbouring buildings that define the scale and geometry of the designed structure here. Such a phenomenon, described by Guattari, can be seen in e.g. Origin I, Kirin Plaza (Osaka 1985-1987). However, it is noticeable that Takamatsu also continued it later. Namba Hips (Osaka 2003-2007) may be an example here. The characteristic silhouette of the shopping and entertainment centre was designed by adjusting its height to the surrounding area. The cuboid of the structure was cut, creating a sharp wedge of about 1/3 of its height.

From fragments of a city Vidler draws three levels of meanings of new types used in contemporary works of architecture: meanings *ascribed by the past existence of the forms*, meanings of *the specific fragment and its boundaries*, and *re-composition of these fragments in a new context* – transformation of selected types into new entities where transferred meanings can constitute an interpretative key [14, p. 261]. From Kyoto, perceived by Takamatsu as a fractal organism, the architect also draws the means of formal symmetry and similarity of elements on a macro- and microscopic scale [6, p. 134].

Some of the transposed types were used by Takamatsu in a contradictory way. Multifunctional halls built in Japan since the 1960s, which acted as a pretext for Gotsu Community Centre, are perceived by the architect as architecture devoid of expression, thus emphasizing the lack of any characteristics in his design. The completed building (which is

[8] Vidler based his theory on the works of Italian rationalists, but his ideas had been published two years before Origin I was created, and this fact allows one to compare them with Takamatsu's projects, especially the ones coming from that time.

Fig. 1. Ark (Kyoto 1981-1983). Source: [13]

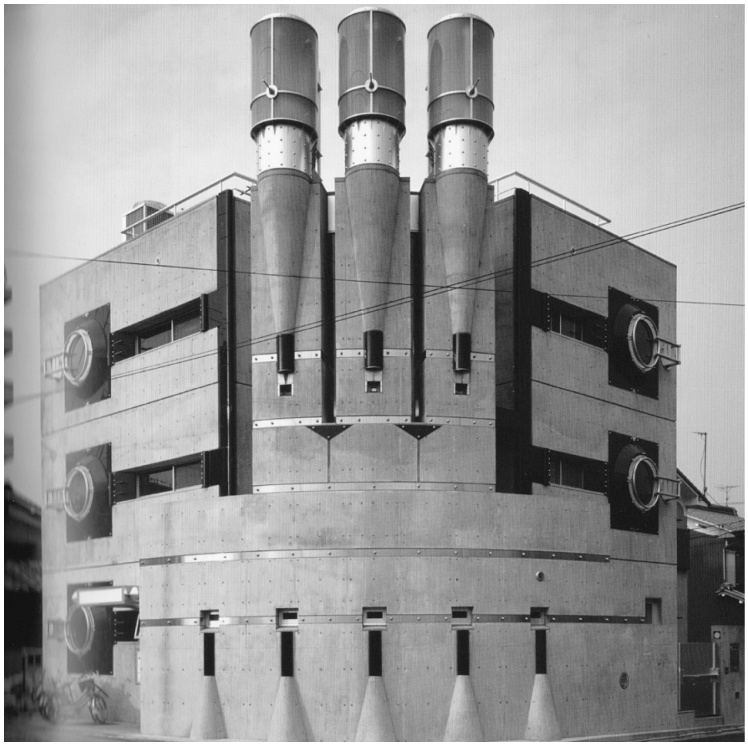


Fig. 2. Pharaoh (Kyoto 1983-1984). Source: [13]

only a fragment of the complex designed with a library, a vast plaza and a rectangle of high and low greenery), constituting a set of orthogonal volumes of different heights combined with an empty frame element, is to create *absence that induces presence*. *As a result, whatever purpose it is used for, the design is for a hall like a pure white canvas that necessitates a great deal of energy*, writes the architect [13, p. 92].

The dental clinic Ark, in turn, refers to the surrounding in a non-binding manner, drawing a metaphor of a locomotive from the vicinity of a railway line. This is architecture inspired by industrial revolution – expression of the dynamism of machines, *dynamics of form*. As Takamatsu says, from the moment of designing Ark, he feels more like an engineer than an architect [13, p. 42]. At the same time, as one of Takamatsu's mechanistic implementations from the 1980s, the building, clearly fits in with the style described by Charles Jencks as the second machine aesthetics, creating *mechanical monsters with bulging eyes*, presenting "bones and joints" – a skeleton as a building's exterior. Unlike the first machine aesthetics (the avant-garde of Mies Van der Rohe, Le Corbusier and others), it was oriented towards lightness, versatility, dynamism, electronics and self-regulation. The hyperbole of force, or a striking figure in contrast to the surrounding, represents the structure as more important than its content. Function here is also just "supported". *Unnecessary expression* and exaggeration form the architectural hyperbola. Frozen movement is also visible, manifesting itself in the presence of tubes, capsules, lifts, visualised in the form of the building [6].

The overview of Takamatsu's constructed buildings and designs from the 1990s show that at that time the architect replaced the stylistics of machine with platonic solids and other geometric shapes (using concrete, glass and metal as main materials), to draw from machines again in the second half of the 1990s – and, additionally, natural forms (the world of fauna and flora) and cultural signs, introducing more conspicuous light structures and diverse materials. The already mentioned Minatosakai Community Centre and Meteor Plaza come from that period.

The latter is a structure with cosmic metaphors situated in a small Japanese village where a meteorite fell in 1992. The building is multifunctional: it houses an exhibition space, a ferry terminal, an auditorium with 500 seats, a swimming pool and additional facilities (like a library, a restaurant and a reception room for private events). Each of these functions was placed in a different geometric volume: the exhibition in an elliptical one, the private events in a cone, the ferry terminal and the auditorium in two prisms covered by a common roof. The "cosmic solids" in a shiny metal cladding (anodized aluminum with glass and concrete), with an element of the exhibition space reflecting the meteorite, extend here from the mechanistic body – sculpture. Symbolically raised, placed over the undulated roof, it presents the reason the building was created for from a distance.

Nose Myokensan Worship Hall (Hyogo 1995-1998), a training building for Buddhist monks, is a solid composed of "a podium" and "a tower". The eight-arm star-shaped tower has a wooden structure, complemented by steel and glass elements (like walls and mezzanine floor inside). The structure is made of the trunks of surrounding trees. Nestled in the woods at the top of the sacred Nose Myokensan mountain, the architecture becomes a metaphor for the place – the tower-chapel as the top of the sacred mountain.

Nose Myokensan Worship Hall may also be referred to as an example presenting the essence of Takamatsu's metaphorical design – selecting the main idea, concept or image for each project and transferring it into the realm of architecture. Metaphor in architecture was defined by Marcin Charciarek as a *mental and formal phenomenon of the architectural landscape, a tropical figure that breaks rules of commonly understood and fixed syntax, abstracts through analogy and multiple associations from traditionally understood shapes, codes, contents, and architectural meanings* [2, p. 165]. It should be noted that for Takamatsu it is an element expanding architectural language – the original concept, the "keyword" which – expanded – creates a building from a conceptual design to its implementation. For example, in the building "Lumiere" at Omula Beauty College, the letter O as the school's initials became the basis of the project, "exhausting its theme" to such an extent that the subsequent design process merely consisted in (according to the architect) solving technical and structural problems, and therefore continuing a process which Takamatsu called a *momentary design* [13, p. 138].

In addition to metaphors, Takamatsu uses architectural quotations in his buildings. They can be found in the already described Pharaoh (transfers from the architecture of fortresses) or National Theatre Okinawa, where the traditional *chinibu* wall, transposed from residential architecture, forms a transparent partition for the *kumiodori* dance theatre, originally held outdoors rather than inside.

The Takamatsu's quotations are indirect quotations; therefore, they paraphrase the quoted elements, sharing some of their features. In Henjyoto Project (Kyoto 2005) the architect used the highest Tō-ji pagoda in Japan, dating back to the 8th century, to create a wooden structure named *Henjoutou* by the highest priest of Tō-ji; Hentou tower means "enlightenment", symbolizing the Buddha.

By expanding the language of architecture, applying metaphors with a broad spectrum of references, Takamatsu has reached a point where he discusses giving it any meaning at all. Among his works, one can thus find projects in which Takamatsu departed from conferring meanings, and which in fact serve to create new theorems.

Kirin Plaza (Osaka 1985-1987) is an example of architecture symbolizing corporations, but without using any corporate symbols. At the site, described by Takamatsu as the most inordinate place in the whole Japan, the architect decided to create *architecture symbolising architecture* – embedding it in the spatial context and simultaneously detaching it from it, intensified in architecture. According to Waro Kishi, four towers of Kirin Plaza symbolised the centre of the world; they formed four identical façades and an interior, which, occupying about half of the building's volume, constituted a substantial architectural space. The mechanistic details in the building were characteristic for Takamatsu – the frames of the openings, corners of the solid and its smaller elements. On the basis of this building Takamatsu passed on his theory of architecture: *architecture has a presence unlike a sign or index* [13, p. 58].

In turn, Namba Hips is a *secession from meaning* to create architecture independent from the later vicissitudes of life. The characteristic silhouette of the building with an abstract symbol on the front – round shapes of the opening with a balcony and glazing – *makes architecture impossible to be referenced* [13, p. 172] to anything familiar. Namba Hips is supposed to be the only symbol in the city over-larded with meanings.

Fig. 3. Kirin Plaza (Osaka 1985-1987). Source: [13]

Fig. 4. Meteor Plaza (Shimane 1994-1995). Source: [13]

Among the later designs, reflecting Takamatsu's design idea in its full spectrum, the most interesting seems to be Doshisha International Academy (Kyoto 2008-2011) – the building of a school with a slightly unpredictable programme. Takamatsu's brief comment on the building does not seem to exhaust the references contained therein. According to the architect's words, the unpredictability of the building's programme was reflected in the *unpredictability* of its architecture. The school was planned as one promoting international education and Takamatsu designed it as white geometrical solids with colourful accents in Mondrian's tones. The solids, geometric shapes of openings and details along with colour scheme give the impression of loose metaphors and quotations from the 20th-century occidental architecture. Reminiscences of the International Style as well as the later round forms are evident here. Takamatsu writes that the forms of the building are to be *vestiges of syntagmatic sentence structures that have lost their nouns and subjects*, which will gain their meanings during the functioning of the building. Such filling of the work with meanings – a dialogue between architecture and its users – can be considered as another characteristic feature of Takamatsu's architecture, which the architect himself describes in the following way: *Anything that possesses the capability is called a symbol* [13, p. 186].

5. Conclusions: private and institutionalized language

Masaru Kawakoto presents Takamatsu's architecture as an illusion on three levels: *personal*, *paired* (form as language) and *communal*, in which architecture occurs as a common illusion – created by a work of art presented to society. Perceived on the third level, architecture returns to its observer-user as an individual illusion. *Architecture-ness* [13, p. 14] is Takamatsu's private illusion. The artist forms structures creating new content in the architectural language.

Mark Rakatansky introduces a similar systematics in architectural narratives, divided into private and institutional ones. The institutional narratives are not related to the individual experience of space (different each time), but somehow arbitrarily accepted – combined with the specific purpose of a given building. Knowing its function, users of each space can imagine it before they enter it, even for the first time [11]. This situation is nevertheless different it the case of Takamatsu's architecture. By superimposing an individual language of the *personal illusion* on the *communal* one, the architect changes commonly accepted institutional narratives, often in a controversial manner. Not without reason, Felix Guattari described Takamatsu as *the most provocative* Japanese architect of his time [6, p. 132] (e.g. mentioning, for example, the extreme feelings and comments evoked by the construction of Ark resembling an engine – the building was even compared to a crematorium). In turn, Takamatsu's buildings were assigned the role of *inhuman subjects* by Guattari – *subjectivity machines* collaborating with human subjects both individual and collective ones [6, p. 135].

In contemporary building space, where the most common method of adapting one's designed buildings to the surrounding context is using quotations and copying scale, materials and forms of the neighbourhood, Takamatsu creates his own language. It is the reality interpreted as a text and cumulative heritage – recomposed, full of metaphorical references,

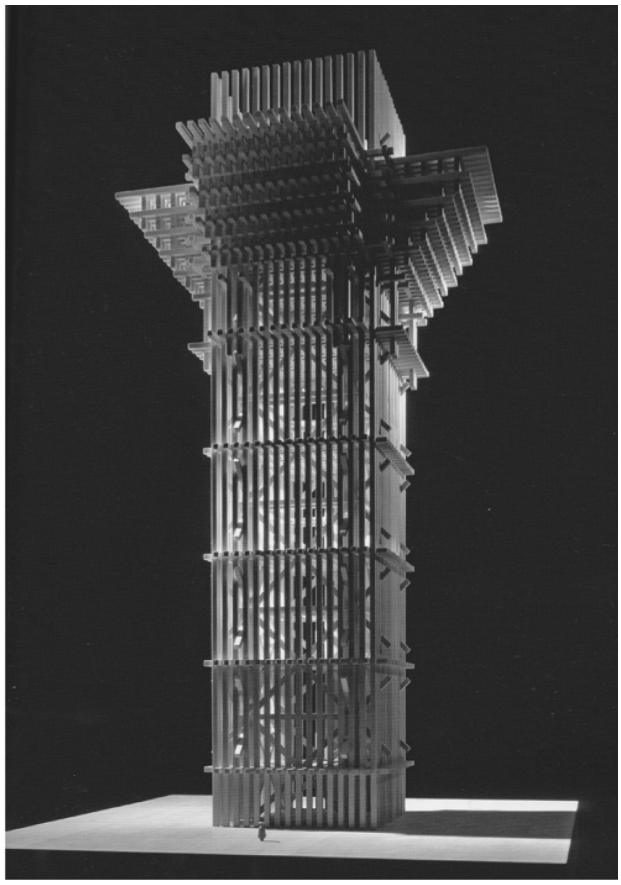


Fig. 5. Henjyoto Project (Kyoto 2005). Source: [13]

Fig. 6. Doshisha International Academy (Kyoto 2008-2011). Source: [13]

indirect quotations and abstracted, exaggerated elements. It confirms the idea of Christian Norberg-Schulz from *Intentions in architecture*, according to which perception of forms has a cultural background. According to Schulz, new forms can neither directly refer to the past, nor completely break with it. They are based on new systems of symbols: *we should conserve the structural principles of tradition rather than its motives* [9, p. 34]. Therefore, it is the language of forms rather than direct quotations that occur in architecture. It is in this language that Norberg-Schulz seeks contemporary architectural symbolism.

Takamatsu's architecture, an individual method of creating metaphors and architectural expressions, proves that not only do contemporary buildings constitute structures that meet the requirements of the Vitruvian triad (mainly functions and durability), but they also carry further contents and symbolic meanings in their forms. Such action is not only proper for historical temples, palaces and ceremonial roads. On the contrary, in contemporary architecture, metaphor may be considered as the *main theme, the essence of actions, materialising the sense of the world of architecture of this century* and *one of the more important stylistic functions, recognized as a mental and formal phenomenon of the architectural landscape* [2, p. 173].

Takamatsu's designs and their implementations confirm the idea from the thesis of Marcin Charciarek *O metaforze w architekturze współczesnej* (*On metaphor in contemporary architecture*): *it seems that the demonstrated artificiality and fictitiousness of these forms, enforced by poetic intentions, will be one of the factors that qualify architecture for a wide spectrum of multivalued art and thus will put it over utilitarianism of building and construction* [2, p. 173]. According to the above words, Takamatsu's works can be perceived not only as purely utilitarian forms but as works of art, which – through the parallel between art and death – is also confirmed by Felix Guattari. The philosopher sees the fear of annihilation with the simultaneous fascination with death in Takamatsu's architecture [6, p. 140].

Comprehending Takamatsu's buildings as examples of coherent narrative giving them timeless and ambiguous values is confirmed by the words of Waro Kishi: *This is not architecture as a simple project, but an expression of the objective existence of architecture, and of the intentions of the architect who created it* [13, p. 7]. In the discussion, conducted for years within the field of architectural theory, on providing buildings with meaning or treating them as "silent" spaces – functional "containers" with a certain aesthetic value – Takamatsu's works are arguments for adopting the first thesis. The "architectural poetics" of Takamatsu is nothing but a distinct language, which creates the author's own story. This story may appear overwhelming and incomprehensible for the recipient, but it cannot be ignored.

References

[1] Capdevila-Werning R., *Can buildings quote?*, The Journal Aesthetics and Art Criticism, No. 1, 2011, 115-124.

[2] Charciarek M., *O metaforze w architekturze współczesnej*, typescript of a doctoral dissertation, written at the Faculty of Architecture of Cracow University of Technology under the supervision of D. Kozłowski, Kraków 1999.

[3] Chow Ph., *Meteoric rise*, Architectural Review, No. 5, 1997, 44-49.
[4] Coates N., *Narrative Architecture*, Wiley, Hoboken 2012.
[5] Goodman N., *How buildings mean*, Critical Inquiry, No. 4, 1985, 642-653.
[6] Guattari F., *Les machines architecturales de Shin Takamatsu*, http://www.revue-chimeres.fr/drupal_chimeres/files/21chi14.pdf (access: 01.03.2017).
[7] Jakubowski P., *Pułapki tożsamości. Między narracją a literaturą*, Universitas, Kraków 2016.
[8] Jencks Ch., *Architektura późnego modernizmu i inne eseje*, Arkady, Warszawa 1989.
[9] Norberg-Schulz Ch., *Intentions in architecture*, [in:] *Theories and Manifestoes of Contemporary Architecture*, (ed.) Ch. Jencks, K. Kropf, Academy Press, London 2006, 33-34.
[10] Pallasmaa J., *The Space of Time*, Oz, No. 20, 1998, http://newprairiepress.org/cgi/viewcontent.cgi?article=1324&context=oz (access: 21.09.2017).
[11] Rakatansky M., *Spatial Narratives*, [in:] *Strategies in Architectural Thinking*, (eds.) J. Whiteman, J. Kipnis, R. Burdett, MIT Press, Cambridge 1988, 198-222, http://www.haussite.net/haus.0/SCRIPT/txt1999/05/TEXT1.HTML (access: 14.04.2013).
[12] Rewers E., *Postpolis. Wstęp do filozofii ponowoczesnego miasta*, Universitas, Krakow 2005.
[13] *Shin Takamatsu, with essays by / con saggi di Waro Kishi, Masaru Kawakoto*, Montadori Electa, Milano 2012.
[14] Vidler A., *The third typology*, [in:] *Theorizing a New Agenda for Architecture. An Anthology of Architectural Theory 1965-1995*, (ed.) K. Nesbitt, Princeton Architectural Press, New York 1996, 258-265.
[15] Wigley M., *The Architectural Cult of Synchronization*, The Journal of Architecture, No. 4, 1999, 409-435.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.17.189.7418

Justyna Tarajko-Kowalska (justarajko@tlen.pl)
Institute of City and Regional Planning, Faculty of Architecture, Cracow University of Technology

Fuctions and basic principles for spatial organization of educational and recreational farms with animals in european cities

Funkcje i zasady organizacji przestrzennej farm edukacyjno-rekreacyjnych ze zwierzętami w miastach europejskich

Abstract

A rapid development of technology, increasing urbanization, life in urban agglomerations as well as the degradation of the natural environment are constantly alienating humans from the world of plants and animals, depriving them of the possibility of using the therapeutic properties of nature. A specific antidote on the so-called "Nature Deficit Disorder" (NDD) in a big metropolis may turn out to be a city farm with animals, which, in addition to its main educational role, plays equally important functions: therapeutic, recreational and socio-cultural. The article presents the idea of such farms and analyzes various functions they perform in the structure of contemporary European cities. Their program and basic principles of spatial organization have also been characterized.

Keywords: city farm, educational farm, children farm, NDD – Nature Deficit Disorder, AAT – Animal-Assisted Therapy, AAE – Animal-Assisted Education

Streszczenie

Szybki rozwój techniki, postępująca urbanizacja, życie w środowiskach wielkomiejskich i degradacja środowiska naturalnego coraz bardziej oddalają współczesnego człowieka od świata roślin i zwierząt, pozbawiając go możliwości korzystania z terapeutycznych właściwości przyrody. Swoistym antidotum na tzw. Zespół Deficytu Natury w obrębie dużych miast może stać się farma miejska ze zwierzętami, która poza nadrzędną rolą edukacyjną pełni równie ważne funkcje terapeutyczne, rekreacyjne oraz społeczno-kulturowe. W artykule przedstawiona została idea takich założeń oraz przeanalizowane różnorodne funkcje, jakie pełnią one w strukturach współczesnych miast europejskich. Scharakteryzowano także ich program i podstawowe zasady organizacji przestrzennej.

Słowa kluczowe: farma miejska, farma edukacyjna, farma dziecięca, Zespół Deficytu Natury, AAT – terapia kontaktowa ze zwierzętami, AAE – edukacja wspomagana przez kontakt ze zwierzętami

1. Introduction – preface, objectives, methodology of work, scope of the article

1.1. Preface

A rapid development of technology, increasing urbanization, life in urban agglomerations as well as the degradation of the natural environment are steadily alienating humans from the world of plants and animals, depriving them of the therapeutic properties of nature [1]. Richard Louv [15], an American writer, essayist and journalist, coined the term *Nature Deficit Disorder* (NDD) to describe this phenomenon, having often negative consequences on human beings. In his book: Last Child in the Woods: Saving our Children from Nature Deficit Disorder, he argues, that much of modern human developmental dysfunction in the psychosomatic sphere may originate from a very limited or complete absence of humans' (especially children's) contact with nature. It is because the direct experience of nature has been replaced by an intermediary relation, easily accessible through media and electronics, but limited just to two senses: sight and hearing [15]. Prevention of the NDD may be implemented, among others, by alternative forms of outdoor education and lectures outside the classrooms, which not only help to reconstruct the lost contact with nature, but also allow to shape the appropriate sensitivity of people to their habitat. A specific antidote on the humans' alienation from the world of flora and fauna within the city metropolis, can become **city farm with animals which, in addition to its main educational role, plays equally important functions: therapeutic, recreational and socio-cultural.**

1.2. Objectives of work

The main objective of the article is to analyze diverse functions performed by educational and recreational urban farms with animals in the structure of contemporary European cities. It is also important to define a program of such farms and basic principles of their spatial organization, including opportunities and threats related to their functioning in urban areas.

1.3. Methodology of work

Analysis of the literature and online sources as well as case studies and on-site research of the city farms functioning in Europe.

1.4. Scope of the article

This article is dedicated to the city farms with animals that are set up in European cities for educational, therapeutic, social and entertainment purposes. The main criterion here is function – the discussed farms are not only a recreational mini-zoos, but they perform the above-mentioned functions with leading role of social and environmental education. The second criterion for selection is location – the analyzed farms are situated within the urban zone and form a separate unit, not being e.g. a part of a zoo.

2. Idea of educational city farms in Europe

Educational city farms are environmental and agricultural projects, intentionally established in urban areas, both in their central and peripheral zones. These are usually community initiatives, that involve people of all ages and from every social group into contact, integration, and work with animals and cultivation of plants. Their aim is to strengthen local social bonds and broadly understood environmental education, which builds awareness of the important role of agriculture and livestock in urban communities. City farms form a link between urban and rural life and make it easier for city dwellers to keep a relationship with livestock and food crops near their homes. For people who, for various reasons, cannot visit farms in the countryside, this is often the only opportunity to learn about livestock farming and the origin, usage and protection of different breeds, as well as food production, which enhances their ecological and consumer awareness. Farms are also places for meetings, recreation and therapy or daycare activities. They are a special form of integrative, interspecific spaces in cities, as well [9].

Local, organic products (e.g. eggs, milk, cheese, vegetables, fruits), natural cosmetics and bee products can be also purchased on farms. However, due to sanitary regulations related to animal husbandry in urban areas, which are excluded from agricultural production, city farms usually do not carry out large-scale production activities.

In different countries, these developments function under different names: **city farm** (*Urban Farm, Community Farm*), **educational farm** (*Granja Escuela, Ferme Pédagogique, Quinta Pedagogica, Fattorie Didattiche, 4H Gård*), **children farm** (*Jugendfarm, Kinderboerderij, Children's Farm, Youth Farm, Ferme d'Enfants*), **animation farm** (*Ferme d'Animation, Cascina di Animazione*) as well as **active playground** (*Aktivspielplatz, Abenteurspielplatz*).

Most common animals on the city farms are so-called "livestock", which according to the Polish law act[1] are, among others: equidae: horse (*Equus caballus*) and donkey (*Equus asinus*); domestic cattle (*Bos taurus*); poultry – birds of the species: hen (*Gallus gallus*), duck (*Anis platyrhynchos*), goose (*Anser platyrhynchos*) and turkey (*Meleagris gallopavo*); pigs (*Sus scrofa*), sheep (*Ovis aries*), goats (*Capra hircus*), honey bees (*Apis mellifera*); fur animals: chinchillas (*Chinchilla lanigera*) and rabbits (*Oryctolagus cuniculus*) [26]. On farms, there are also often present: ponies (e.g. Shetland), pigs of small breeds, hamsters and guinea pigs, ornamental poultry (e.g. guinea fowls, quails) as well as alpacas (*Vicugna pacos*) and llamas (*Lama glama*).

Although the history of educational farms in Europe dates back to the beginning of the twentieth century, their dynamic development started as late as in the 1960s. After World War II many European countries entered the road of intensive development and urbanization. This resulted in the rapid growth in the number of city inhabitants[2] and reduction of the

[1] Ustawa z dnia 29 czerwca o organizacji hodowli i rozrodzie zwierząt gospodarskich (Dz.U. 2007, Nr 133, poz. 921 z późn. zm.), art. 2, ust. 1, pkt 1 [26].

[2] In 2015, the average urbanization rate for Europe grew to 73% (http://www.statista.com/statistics/270860/urbanization-by-continent/).

agricultural sector. These changes have contributed to the stretching of the gap between humans and nature, environment and animals. A number of government programs were launched in the 1960s to bring residents of rapidly expanding cities (especially children and youth) closer to nature, as well as to farm life and livestock. At the same time, informal movement of the creation of city farms began simultaneously in many European countries [27]. The first city farms in which school's curricula were implemented were established in the Scandinavian countries of **Norway, Sweden and Denmark**. The main inspiration for that idea was **4H youth movement**[3], which was founded in the United States in 1914. 4H is short for four English words: head, heart, hands, health – understood as: open head, hot heart, skilful hands and good health. In Scandinavian countries, city farms are affiliated with organizations that act as part of 4H clubs. In Norway, it is **4H-gård Norge** (since 1995)[4] and in Sweden **Riksförbundet Sveriges 4H**[5].

In Germany, the idea of city farms is closely linked with the concept of active or adventure playgrounds (*Aktivspielplatz*, *Abenteurspielplatz*). In the 1960s the German government introduced the program **"Aktivspielplätze"** (Active Playgrounds). The first German farm – Jugendfarm Elsental was founded in the 1960s in Stuttgart. At the same time, first adventurous park in Berlin was opened, and in 1972 the German Federation **Bund der Jugendfarmen und Aktivspielplätze e.V** was founded, as well[6].

At the same time, as in Germany also **in the Netherlands**, in cooperation with the Ministry of Agriculture the project **"City farm"** was launched. **The Vereniging Samenwerkende Kinderboerderijen Nederland (vSKBN)** was established in 1988 and now associates more than 300 city farms[7].

In the 1970s, the movement of educational farms reached also **Belgium**, a country where the urbanization rate increased to 98%[8]. In the 1980s, two organizations supporting farming activities were founded: **Federatie Kinder, -Jeugd- en Gezinsboerderijen**[9] in the Dutch-speaking Flemish region and **Fédération Belge Francophone des Fermes d'Animation**[10] in Wallonia – the French-speaking part of Belgium.

The first **British** educational farm – Kentish Town, was established in London in 1972. Also, the first community gardens and farms linking animal husbandry with plant cultivation were organized in the UK. In the United Kingdom, city farms have also become one of the ways of the revaluation of undeveloped urban areas. Established in 1980, the **Federation of City Farms and Community Gardens (FCFCG)** joined now more than 150 city and school farms and over 1000 community and allotment gardens[11].

3 http://4-h.org/.
4 www.4h.no.
5 www.4H.se.
6 www.bdja.org.
7 www.kinderboerderijen.nl.
8 https://esa.un.org/unpd/wup/Publications/Files/WUP2014-Highlights.pdf.
9 www.kinderboerderijen.com.
10 www.fermedanimation.be.
11 www.farmgarden.org.uk.

In the 90s of the twentieth century, the idea of educational farms reached the Mediterranean basin, among others Italy. However, in some European countries, such as Italy, France, Switzerland, Austria, as well as Poland, these developments are most often located in rural areas – and thus go beyond the scope of this article.

As a result of joint meetings of delegates from various European countries, **the European Federation of City Farms (EFCF)**[12], was established in 1990 as a part of the European Environmental Bureau (EEB). The main aim of the EFCF is: "Promoting the interests and mutual co-operation of Kinderboerderijen, Jeugdboerderijen, Gezinsboerderijen, Fermes d'Enfants, Fermes d'Animation, Jugendfarmen, Aktivispielplätze, City Farms, 4H-Farms and similar organizations that actively promote the equal access and involvement of children, young people and adults through practical experience in a wide range of educational, environmental, recreational, social and economic activities focused around farming, empowering people to improve their own lives and environment in peaceful co-existence" [4].

3. Functions performed by city farms

city farms with animals can perform a variety of functions. Their main task is the **social and environmental education of the youngest**. The urban farm is a source of knowledge about farm animals, their habits, needs and ways of care. This is an alternative place of education, in which the curriculum is oriented primarily on practical work and workshop exercises. Under the supervision of the tutor, children have the opportunity to actively participate in the feeding and care of animals raised there (Fig. 1). Through the contact with animals (animal-assisted therapy) and gardening (horticultural therapy) urban farms also perform **therapeutic functions**. Farms are also often desirable green enclaves and **recreation areas** for children and adults, which provide a viable alternative to the sedentary lifestyle of many urban dwellers, thanks to their various leisure activities [30]. As places of integration and activation of local communities, they also have important **social functions**. City farms are often the places where **rare and native breeds of domestic animals are preserved**.

In the large cities, farmland areas are an **important link in the green infrastructure network**. They are also a way of the development of the wastelands or areas which, due to their unfavorable location, are not suitable for the other functions such as housing or public services. City farms also play an important role in the shaping of the sustainable development in European cities by promoting an eco-friendly and active lifestyle, healthy nutrition, local products, building a consumer awareness, promoting renewable energies, recycling, ecological farming and animal husbandry considering animal welfare principles [5].

[12] On the federation's website: www.cityfarms.org there are many valuable articles, presentations and educational materials, brochures, leaflets and good practice catalogs from all over Europe. Every year, scientific conferences devoted to the city farms are held in one of the Member States, as well.

a) b)

Fig. 1. Direct contact with animals is one of the tangible benefits of visiting a city farm: a) Children during riding lessons at the Vienna Kids Farm in Vienna, Austria; b) Children scour the sheep at the Vienna Kids Farm, Austria (photos by the author, 2015)

a) b)

Fig. 2. The most popular animals on city farms are sheep and goats: a) Male goat on the run at the Vienna Kids Farm in Vienna, Austria; b) Sheep waiting for feeding – Vienna Kids Farm in Vienna, Austria (photos by author, 2015)

3.1. Educational function

The main function of the city farms is **social and environmental education**, which is one of the most important areas of the integrated teaching of preschool and school children. Its main purpose is to educate pupils' cognitive abilities, to familiarize them with the life of nature and to raise their sensitivity to it, as well as to trigger, inspire and maintain a sense of respect for life in all its forms, including the attainment of self-care and own security [1]. It is also a conscious shaping of ecological attitudes and instilling knowledge about the importance of sustainable development to improve the quality of people's lives [9]. Animal husbandry plays a very important role in this process. Animals associate man in cities from the very beginning and their vital role in human daily life is well known and undisputed [10, 11]. On the farm children should have the opportunity not only to observe the animals and follow their

growth, but also to get to know their species name, appearance, structure and lifestyle, the rules of behavior towards them, their useful role for the human being and the environment and finally, be able to stroke, feed and cherish them. It is also important to develop a positive and caring attitude towards animals and to learn how to observe and play with them, without tormenting or scaring [1]. On urban farms, various curricula for pre-school and school groups are organized. The selection of topics related to the animal husbandry is broad. These can be, for example, programs to educate the origin of food based on specific products, such as: "milk path", "egg path", "path of meat" or "honey path", as well as "wool path" – presenting the process of wool origin and ways of using it or "bee path" – teaching about the functioning of the bee family [12]. Education is often done with help of sheep or goats, which are considered to be one of the most multifunctional animals in this regard [27] (Fig. 2).

3.2. Therapeutic function

Another very important role performed by urban farms is the **therapeutic function**. Along with the development of knowledge about the therapeutic role of animals in human life, there has also been an increase in the number of corrective and therapeutic programs directed primarily to children and youth, known as **Animal-Assisted Intervention (AAI)**. Animal therapy is now recognized as a natural method of supportive treatment, rehabilitation and resocialization, based on the direct contact with animals. Numerous studies have confirmed the positive impact of this method on the wellbeing of older people, as well [16].

AAI provide motivational, educational and/or therapeutic benefits. It can be run in different environments by trained professionals and/or volunteers, involving animals that meet certain criteria. There are various forms of animal therapy that can be realized with help of animals such as: dogs (kynotherapy), horses (hippotherapy), cats (felinotherapy), donkeys (onotherapy), alpacas (alpacotherapy) or bees (apitherapy). Pet Partners, a UK-based organization, has distinguished three main types of the AAI: **Animal-Assisted Activities (AAA)**, **Animal-Assisted Therapy (AAT)** and **Animal-Assisted Education (AAE)**[13]. All these methods – of course to varying degrees and extent – can be implemented on the urban farms. Some of the farms are even developed as specialist therapeutic or therapeutic-residential units, such as centers of hippotherapy, kynotherapy or alpacotherapy, where professional rehabilitation and sometimes resocialization programs are implemented. However, due to the extensive program and big land demand, most of such centers are located in suburban and rural areas and will therefore not be discussed in detail in this paper.

On a typical urban farm, the therapeutic function is primarily based on contact with animals and the measurable beneficial effect they have on the human body, both in the physical and mental sphere, additionally performing educational and motivational functions [2, 7]. City farms may also offer support to physically and intellectually disabled people by engaging them in farming activities involving cultivation or animal husbandry. For example, autistic persons, endowed with specific receptivity to reality, feel and function better in the

[13] https://petpartners.org/learn/terminology/.

farm environment – closer to nature. On the farm, there are also opportunities for work which is more understandable and satisfying for people with disabilities than in case of typical occupational therapy [13].

3.3. Recreational and socio-cultural functions

Other important roles performed by the urban farm are **recreational and socio-cultural functions**. A farm, located in cities' central or suburban areas, is often a popular resting place for local communities. Activities offered on farms can become an antidote to the NDD and related civilization diseases. The farm area is a place to locate children's playgrounds, picnic areas and even sports grounds. Various events are organized there: birthdays, bonfires, festivities that build and strengthen social bonds, enabling the integration of people of all ages, from different social and cultural groups [17]. Daily stay on a farm is also one of the least expensive forms of recreation for the whole family, comparing costs to other urban entertainments such as: cinema, theater or café [28].

The function of "learning by play" is particularly present in the farms developed in Germany, where so-called "active playgrounds" (*Aktivspielplätze*) provide active recreation for children of different age groups. They aim to develop autonomy and creativity in children and youth as well as to build their ecological awareness and social abilities[14].

The urban farms are often a unique green oasis in the city, providing the possibility of direct contact with nature and animals – even wild. Being in line with the trend of sustainable urban development, they provide the opportunity to recreate the lost links between the urban and rural environment, becoming an important part of the green infrastructure system, as well.

3.4. City farms as places for rare and local breeds' protection and survival

Urban farms also function as places for keeping rare, native animal breeds. Old livestock breeds, created over the centuries by humans in adaptation to local environmental conditions, are an important part of the cultural heritage of a given region. However, the future of many domestic animal breeds is becoming increasingly endangered, due to their lower economic efficiency and changes in agricultural production methods. Therefore, it is important to protect them from extinction. One of the popular forms is protection *in situ*, which is based on the typical ways of usage of protected populations [14]. **The British Rare Breeds Survival Trust (RBST)**[15], recognizes educational urban farms as an ideal place to secure the future of rare and indigenous breeds of farm animals [27]. An additional advantage is the possibility of using them in the education of visitors. Some of them, e.g. the German Schleswig horse, also have some special character traits, which designate them to therapeutic work with disabled people [29]. However, the most important to ensure the sustainability of rare, local breeds is

[14] http://www.bdja.org/files/bdja_englisch.pdf.
[15] www.rbst.org.uk.

the promotion of products derived from them [14]. This goal can be successfully implemented on a farm, where it is possible to sell and popularize locally produced food products such as milk, cheese, eggs, honey and meat.

4. Basic principles of the spatial organization of city farms[16]

City farms with animals in Europe are usually located in urban core areas, peripheral residential zones or in suburbs. The plots selected for the farms are often wastelands, taken over or leased from the city by local communities or foundations. Farms are also created in areas which, due to unfavorable conditions, such as the proximity of busy traffic arteries, railways, airports or other noise sources, are not suitable for other functions. They are also built on former rural farmland sites that operated on land gradually absorbed by the expanding cities.

The size of European urban farms is very different – from small ones: up to 0.5 hectares, medium: 1-5 hectares, to large: over 5 ha[17]. The research shows that the optimal size, where all program components of the farm are possible to organize is the area of ~ 2 ha [6].

Regardless of the size or location of an urban farm, its development should include designated areas and adequately equipped buildings and facilities (Fig. 3). One entrance/exit from the farm area is recommended for safety and hygiene purposes. Most of the farms have a multifunctional main building, which houses educational activities, livestock buildings, animal runs and gardens: ornamental, vegetable and/or orchard. Playgrounds for children are also located on the farms and, in the absence of space, even their standard equipment such as sandpits or climbing ropes.

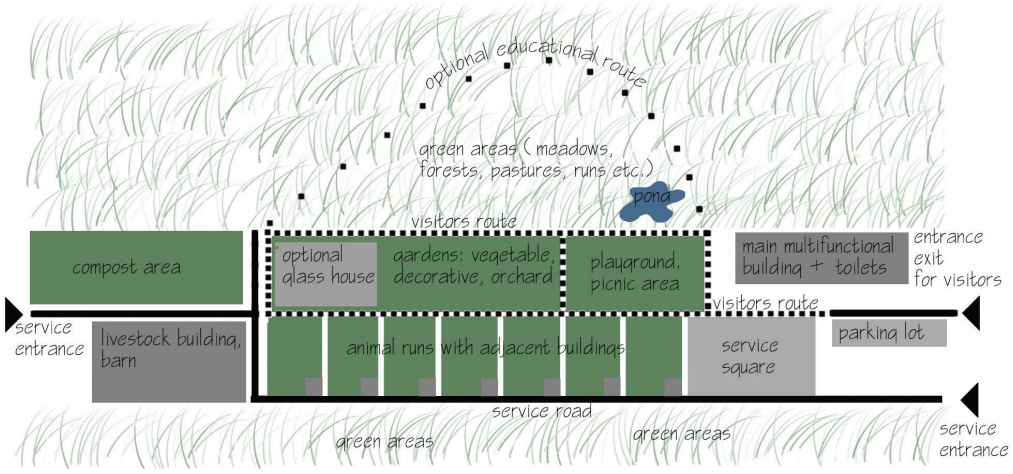


Fig. 3. Scheme illustrating spatial organization of functions in a city farm, according to the basic requirements and principles (drawing by the author, 2017)

[16] This chapter is compiled on the basis of code of practice: *Preventing or controlling ill health from animal contact at visitor attractions* [18] and case studies.

[17] One of the biggest urban farms in Europe is Mudchute Park and Farm located in London, with area of 13 hectares (mudchute.org).

Many farms also have a cafe or picnic area as well as a shop with their own products, sweets and ice cream. If there are horses on the farm, the development also includes outdoor or indoor riding halls and round pens. On bigger farms, the amount of green areas is most often increased by including pastures for animals, meadows or even forests.

The main criterion for the placement of particular functions on the farm is to prevent potential hazards. The most serious of them is the **risk of infecting visitors with zoonoses** [8], which can occur during direct contact with animals, their feces or raw materials of animal origin. Another threat is **potential pollution of land, surface water and groundwater** due to improper storage of animal feeds, waste, and excrement. The problem may also be the **potential nuisance of the farm for neighboring properties through generated noise or odors**.

Preventing these threats requires from the farm owners to fulfill certain sanitary, epidemiological and veterinary regulations, including: hygienic practices, animal testing and maintenance of cleanliness and order on farms and livestock buildings, proper storage of feed and storage of animal manure for agricultural use as well as a regular disposal of faeces and residues of feed or litter left on premises for shared use. Visitors are required to comply with the rules of personal hygiene (especially hand washing) and to follow the applicable regulations of the farm[18].

4.1. Zoning, communication routes, information and signage

Appropriate zoning is very important to ensure visitors' safety and to avoid functional collisions on the farm premises. First of all, it is necessary to designate the zones accessible for visitors, available for them under certain conditions (dictated for example by means of proper hygiene) and inaccessible ones, as well as zones of contact with animals and where such contact will not be possible. Access to places open for visitors should be convenient, and pedestrian paths together with communication routes should be free from pollution and properly marked [12]. Zones with direct contact with animals must be kept clean – free from feces and other impurities. In these zones eating, drinking and smoking should be prohibited. Visitors should not have unlimited access to livestock buildings and animal pens and runs, due to the increased risk of zoonotic infections through contact with excrements and bedding. Visitors may only have access to specially designed and regularly disinfected pens and enclosures.

The lower parts of animal fencing should be solid to protect the area just before them from uncontrolled leakage of liquid dirt. To contact with visitors should be directed only animals that are completely healthy and free from injuries. There should be no newly born animals in the contact zones as well as those that have recently given birth. Contact with animals should always be carried out under the supervision of workers. This is not only to ensure visitor safety, but also to protect the animals staying there.

[18] In order to ensure the health and safety of urban farms' visitors in the United Kingdom, a code of practice has been developed, titled: *Preventing or controlling ill health from animal contact at visitor attractions* [18]. It contains a set of practical guidelines for the prevention and control of zoonoses in places where direct contact between humans and animals occurs. This Code of practice, updated on a regular basis, can be downloaded from the Farming & Countryside Education website (FACE) (www.face-online.org.uk).

Sanitary facilities and recreational, play and picnic places should be located only in areas where contact with animals is not possible. The area of urban farms should be organized in a way that facilitates mobility, independent sightseeing and additional activities of visitors. Selected objects, crops or places for animals should be marked with informational boards that contain both text and appropriate graphics.

In larger farms, it is possible to propose the combination of selected zones and elements of development into a form of didactic path. As a form of attraction for such paths can serve devices for independent experiments [12]. Trails for the visitors should not cross the service roads used by animals and the vehicles or machinery used on the farm. If this is not possible, these trails should be kept clean – free from animal waste, liquid leaks or litter. In places where the surface of the pavement is often contaminated, it is recommended to install wooden platforms called *duckboards*. Walking routes should be traced so that visitors do not have access to the farm's service areas, waste disposal, manure or fertilizer sites. These zones should be additionally secured with fences and appropriate information boards prohibiting entry.

4.2. Buildings, premises and spaces for animals and their management

The biggest attraction of educational urban farms are the animals. They serve not only for the entertainment of visitors, but, above all, they play important therapeutic and educational roles. It is therefore important to provide them with proper care and adequate living conditions, taking into account the standards, depending on their maintenance systems, species, age and physiological state (Fig. 4). Basic principles of livestock welfare, known as the Five Freedoms, have been developed by the UK's Farm Animals Welfare Council and are included in the Codes for the Welfare of Livestock [3][19]. According to the Code, livestock

a) b)

Fig. 4. On urban farms, animals are often kept in small wooden buildings with fenced enclosures: a) A farmhouse and a run for turkeys at the Landgut Wien Coblenz farm, Vienna, Austria; b) A shed and run with a pond for gooses at the Vienna Kids Farm in Vienna, Austria (photos by the author, 2015)

[19] The basic legislative act is Council Directive 98/58/EC of 20 July 1998 on the protection of animals kept for farming purposes (31998L0058) [3].

should be free from: hunger and thirst, discomfort, pain, injury and disease, fear and stress and free to express a normal set of behaviors.

Consequently, decisive for animals' welfare on the urban farm are, among others: living area per animal (area of the pen, pit, cage, runway), freedom of movement, type of floor (litter, litter-free), adjustment of building equipment to their needs, microclimatic conditions and ventilation efficiency.

In addition, to meet the conditions of animal welfare, animals should be kept on the farm in a manner that does not impair the health of humans and does not pollute the air, soil and water, which is provided through the fulfillment of sanitary, veterinary, construction, environmental protection and animal protection rules as defined by the law[20]. The farm should also provide: feed storage area, manure collection space and fertilizers, including compost storage. These premises should be located far away from visitors' open areas, preferably where there is a minimal risk of getting any contaminants.

4.3. Organization and arrangement of the educational space

As already mentioned, the supreme function of the urban farm with animals is education. It is therefore essential to provide adequate educational infrastructure on its territory. In practice, all available places in urban farms can be used for teaching activities – from farm buildings and barns through animal pens, farmyards, playgrounds, gardens: orchard and vegetable ones, to appropriately designed objects and spaces. The organization and equipment of educational space are determined by, among others, characteristics of the target groups and diversification of the didactic program on the farm. Even if education is based primarily on the outdoor activities, the farm should be equipped with a roofed space which will allow education even under unfavorable weather conditions. For this purpose, can serve adapted part of the farm building e.g. barn, a roofing or a specially designed building with classrooms. The basic equipment of such rooms is places to seat, tables, writing boards, books, models and other teaching aids [12].

Buildings for teaching and workshop purposes, oftentimes equipped with food and beverage facilities, can also become a meeting place for local communities.

4.4. Playing area and picnic zone

Both the picnic area and playground are places in which there should be no direct contact with animals (Fig. 5a). In order not to provoke any attempts to contact, it is recommended to separate those zones from the animal pens with the doubled fencing with distance (Fig. 5b).

[20] In Poland this is regulated by the law Acts: on the protection of animals [25], on the organization of breeding and reproduction of livestock [26], on the protection of animal health and the control of infectious diseases of animals [23] and Regulations: on the minimum living conditions of particular species of livestock [19], on the requirements and behavior of keeping livestock species, for which protection standards have been defined in European Union legislation [20] and other than those for which protection standards have been laid down in EU legislation [21].

Picnic places should be located far away from the main service and walking routes. Sanitary facilities should be provided at a short distance from the entrance to the picnic area and, by means of appropriate signage, remind of the need to wash hands before a meal.

Kiosk with sweets and ice cream should be located in the zone without contact with animals, preferably in the vicinity of the entrance to the farm.

The play area should foster children's imagination and creativity while ensuring the safety [12]. Playground equipment may vary and depend largely on farm size and its spatial organization. However, it should be designed and constructed so that it can be easy to clean in order to remove any possible contamination. As in the case of a picnic area, users should have an opportunity to use the nearby wash basins and sanitation facilities both before and after the end of their stay in the playground. Also, fields used as pastures can be utilized as playgrounds and/or picnic areas. However, it is recommended to have a 3-week interval since the last grazing. The grass should be mown and intensely maintained, and any mowing residue should be harvested.

a) b)

Fig. 5. Picnic area and playground are important elements of urban farm development: a) Picnic place under the canopy prepared for a birthday party at the Vienna Kids Farm in Vienna, Austria; b) Double fence at the run for white donkeys, Landesgartenschau 2016 in Bayreuth, Germany (photos by the author, 2015 and 2016)

4.5. Sanitary facilities

Very important for the proper functioning of the urban farm is the provision of appropriately located and equipped sanitary facilities. The best solution is to create separate toilets for women and men, with their quantity adjusted to the number of visitors. It is also recommended to adapt common sanitary rooms to the needs of children by providing them with footstools and small toilet seats [12]. Sanitary units should be located at the entrance/exit from the farm or directly adjacent to the zones of direct contact with animals. They should also be located as close as possible to the play and picnic areas. Near the exit from the farm, a place to clean shoes, wheelchairs and baby carriages can be also provided.

5. Summary

Contemporary educational and recreational city farms with animals in Europe vary widely in terms of their location, area, spatial development, activities as well as management or funding.

Farms can be visited as part of a multi-hour individual, group or school trips, one-day or overnight stay, family and group programs, as well as weekly workshops. Farms can also be visited repeatedly throughout the year by the same group of pupils with longer school programs.

Some urban farms employ paid workers, but most rely on volunteer work. Some of them work in partnership with local authorities, some are run by private foundations or associations, others operate at therapeutic centers, psychiatric clinics, schools or orphanages.

The spatial-functional program of the farm is also varied and depends mainly on the area occupied by it. The biggest attraction of urban farms are the animals, so they are usually mostly exposed and livestock buildings and farmsteads are the main elements of their development. The most common are farm animals, including poultry – chickens, ducks and often goose and turkey, sheep and goats, horses and donkeys as well as rabbits, cows and pigs. Farms also often maintain house pets as: dogs, cats and guinea pigs, hamsters, ferrets, chinchillas and rats. On urban farms are also located facilities for educational purposes, playgrounds for children, utilitarian and ornamental gardens, and in larger farms also orchards, meadows and pastures for animals as well as water reservoirs.

There is a number of risks and threats associated with the functioning and development of animal farms in urban areas. The most serious include: potential zoonoses, potential ground, surface and groundwater contamination, potential nuisances for neighboring properties, extensive hygienic and sanitary regulations and excessive hygiene practices among parents, lack of funds to run and maintain the farms against rising costs of land plots together with investment pressure on areas occupied by farms [5].

Despite many unfavorable factors, the future of educational farms looks bright. They perform very important functions: educational, recreational, therapeutic and socio-cultural. They are also a form of protection for traditional, native breeds of farm animals. Farms are also part of the green infrastructure network and serve as integrative, interspecific spaces in the cities [9].

Location of farms in urban wastelands and uninhabitable areas can be a guarantee of their continued use and will provide the opportunity for more sustainable and better-quality investment in their development and architecture.

The European Federation of City Farms (EFCF) and similar organizations associating urban farms in different European countries provide a platform for exchanging experiences and good practices, giving chances and opportunities for their future development.

References

[1] Budniak A., *Edukacja społeczno-przyrodnicza dzieci w wieku przedszkolnym i młodszym szkolnym*, Impuls, Kraków 2009.

[2] Chmiel K., Kubińska Z., Derewiecki T., *Terapie z udziałem zwierząt w rehabilitacji różnych form niepełnosprawności*, Problemy Higieny i Epidemiologii, 95(3), 2014, 591-595.

[3] Council Directive 98/58/EC of 20 July 1998 on the protection of animals kept for farming purposes (31998L0058).

[4] *EFCF Information Booklet 2016*, http://www.cityfarms.org/files/attachments/.921/EFCF_Information_Booklet_2016.pdf (access: 13.07.2016).

[5] EFCF, *Some of the threats and the opportunities facing city farming throughout Europe*, 2004, http://www.cityfarms.org/files/attachments/.337/Some_of_the_threats_and_the_opportunities_facing_city_farming_throughout_Europe_2004.pdf (access: 13.06.2016).

[6] Ginsberg O., *Aktivspielplätze, Jugend- und Stadtteilfarmen in Europa und ihr Beitrag zu einer nachhaltigen Stadtentwicklung*, Berlin–Kreuzberg, August 2000, http://www.bdja.org/oli/index.html (access: 10.07.2016).

[7] Girczys-Połedniok K., Pudlo R., Szymlak A., Pasierb N., *Zastosowanie terapii z udziałem zwierząt w praktyce psychiatrycznej*, Psychiatria, tom 11, nr 3, 2014, 171-176.

[8] Heuvelink A.E., Valkenburgh S.M, Tilburg J.J.H.C., Van Heerwaarden C., Zwartkruis-Nahuis J.T.M., De Boer E., *Public farms: hygiene and zoonotic agents*, Epidemiology & Infection, 135, 2007, 1174-1183.

[9] Kleszcz J., *Edukacja i zabawa poprzez kontakt ze zwierzętami, jako metoda kształtowania postaw zrównoważonych wśród najmłodszych*, [in:] Materiały Ogólnopolskiej Konferencji Naukowej i Wyjazdu Naukowo-Studialnego – Niemcy, Holandia „Natura – Technologia – Kultura", „Zrównoważone Środowisko Życia", 3-7 września 2015, http://aiu.uz.zgora.pl/files/kaiu-2016-02-22-83-92-kleszcz.pdf (access: 8.07.2016).

[10] Kleszcz J., *Wpływ zwierząt na formę współczesnych przestrzeni miejskich*, Kultura i Wartości, nr 1(9), 2014, 67-82.

[11] Kleszcz J., *Zwierzę w mieście – rola zwierząt domowych w kształtowaniu przestrzeni miejskich oraz w świadomości społecznej ich mieszkańców*, Zeszyty Naukowe Politechniki Śląskiej, z. 54, 2014, 91-99.

[12] Kmita-Dziasek E., *Wprowadzenie do zagadnień edukacji w gospodarstwie rolnym*, Centrum Doradztwa Rolniczego, Kraków 2015, http://ksow.pl/uploads/tx_library/files/01_Wprowadzenie_do_zagadnie%C5%84_edukacji_w_gospodarstwie_rolnym.pdf (access: 10.07.2016).

[13] Kossewska J., Perzanowska A., *Dzieci stają się dorosłe – farmy dla osób z autyzmem*, Konspekt, nr 18/19, 2004, http://www.wsp.krakow.pl/konspekt/18/autyzm.html (access: 2.07.2016).

[14] Litwińczuk Z., *Ochrona zasobów genetycznych zwierząt gospodarskich*, Biuletyn Informacyjny PAN Oddział Lublin, nr 8, 2003, http://www.pan-ol.lublin.pl/biul_8/art_808.htm (access: 17.07.2016).

[15] Louv R., *Ostatnie dziecko lasu*, Wydawnictwo Relacja, 2014.
[16] Ornacka K., Żuraw K., Miś L., *Zooterapia jako innowacyjna metoda pracy socjalnej wspierająca jakość życia seniorów*, Zeszyty Pracy Socjalnej, tom 18, nr 3, 2013, 175-186, http://www.ejournals.eu/ZPS/2013/Numer-3/art/4611/ (access: 8.07.2016).
[17] Piessens M., *Children's Farms: Extending Bridges between Ethnic Groups in the Netherlands*, Master thesis, Wageningen University, 2013, http://www.cityfarms.org/files/attachments/.598/Childrens_Farms_Extending_Bridges_between_Ethnic_Groups_in_the_Netherlands_2013_screen.pdf (access: 13.06.2016).
[18] *Preventing or controlling ill health from animal contact at visitor attraction*, Industry Code of Practice, 2015, http://www.visitmyfarm.org/component/k2/item/339-industry-code-of-practice (access: 2.07.2016).
[19] Rozporządzenie Ministra Rolnictwa i Rozwoju Wsi z 2 września 2003 roku w sprawie minimalnych warunków utrzymania poszczególnych gatunków zwierząt gospodarskich (Dz.U. 2003, Nr 167, poz. 1629, z późn. zm.).
[20] Rozporządzenie Ministra Rolnictwa i Rozwoju Wsi z dnia 15 lutego 2010 r. w sprawie wymagań i sposobu postępowania przy utrzymywaniu gatunków zwierząt gospodarskich, dla których normy ochrony zostały określone w przepisach Unii Europejskiej (Dz.U. 2010, Nr 56, poz. 344).
[21] Rozporządzenie Ministra Rolnictwa i Rozwoju Wsi z dnia 28 czerwca 2010 r. w sprawie minimalnych warunków utrzymywania gatunków zwierząt gospodarskich innych niż te, dla których normy ochrony zostały określone w przepisach Unii Europejskiej (Dz.U. 2010, Nr 116, poz. 778).
[22] Ustawa z dnia 11 marca 2004 r. o ochronie zdrowia zwierząt oraz zwalczaniu chorób zakaźnych zwierząt (Dz.U. 2004, Nr 69, poz. 625) – tekst jednolity.
[23] Ustawa z dnia 21 sierpnia 1997 r. o ochronie zwierząt (Dz.U. 1997, Nr 111, poz. 724) – tekst jednolity.
[24] Ustawa z dnia 29 czerwca 2007 r. o organizacji hodowli i rozrodzie zwierząt gospodarskich (Dz.U. 2007, Nr 133, poz. 921) – tekst jednolity.
[25] Wolters P., *A city farm as a bridge between city and countryside. Food education. Sheep education*, EFCF Guidance, 2014, http://www.cityfarms.org/guidances/view/50 (access: 2.07.2016).
[26] Wolters P., *Farms with mission*, EFCF, 2013, http://www.cityfarms.org/files/attachments/.326/Farms_with_a_mission_text_2013.pdf (access: 13.06.2016).
[27] Wolters P., *Rare breeds. Education about preservation and adaptation*, EFCF, 2011, http://www.cityfarms.org/files/attachments/.332/Text_Education_about_preservation_and_adaptation.pdf (access: 13.06.2016).
[28] Zaal M., *City farms: An opportunity to combat a range of important health and safety threats*, Maastricht University, 2014, http://www.cityfarms.org/researchitems/view/19 (access: 13.06.2016).

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ARCHITECTURE AND URBAN PLANNING
DOI: 10.4467/2353737XCT.17.190.7419

Agnieszka Wojtowicz (agnieszka.wojtowicz@awf.krakow.pl)
Psychology Department, Faculty of Physical Education and Sport, University of Physical Education in Krakow

Barbara Wojtowicz
Division of Descriptive Geometry, Technical Drawing & Engineering Graphics, Faculty of Architecture, Cracow University of Technology

The need for cognitive closure and the level of creative behavior in relation to the comprehension and design of complex spatial structures

Potrzeba poznawczego domknięcia i poziom zachowań twórczych a rozumienie i projektowanie złożonych struktur przestrzennych

Abstract

Although the relationship between the cognitive style and creativity has been the subject of many studies, the results have been inconsistent. Therefore, it seems reasonable to further scrutinize the issue, particularly in the context of professionals, whose jobs involve creativity. The aim of the study was to determine the relationship between the need for cognitive closure and the level of creative and reproductive attitudes with the comprehension and design of complex spatial structures. The study involved 111 first-year students of Architecture at the Faculty of Architecture of the Cracow University of Technology. The results revealed that individuals with a low level of skills in design using geometric constructs had higher levels of reconstructive attitudes in the area of conformity and a lower level of the need for cognitive closure in the area of decisiveness than the ones with a high level of design skills, using geometric constructions. It was also found that as the level of the need for cognitive closure increased, the level of creative behaviour decreased; however, in participants with a high level of design skills using geometric constructions, it showed an upward trend. The results also indicated that the respondents held an average level of creative behaviour and a high level of conformity associated with a reproductive attitude, which is a worrying result as far as the future architects are concerned.

Keywords: creativity, need for cognitive closure, geometry, architecture

Streszczenie

Związki pomiędzy stylem poznawczym a twórczością są tematem wielu badań, jednak nie przyniosły one do tej pory spójnych rezultatów. Zasadne więc wydaje się dalsze analizowanie tego zagadnienia, zwłaszcza w odniesieniu do osób, których zawód w dużej mierze związany jest z szeroko rozumianą twórczością. Celem badania było określenie związków potrzeby poznawczego domknięcia i poziomu postaw twórczych oraz odtwórczych z rozumieniem i projektowaniem złożonych struktur przestrzennych. W badaniu wzięło udział 111 studentów pierwszego roku kierunku architektura Wydziału Architektury Politechniki Krakowskiej. Przeprowadzone analizy wykazały, że osoby o niskich umiejętnościach projektowania wykorzystującego konstrukcje geometryczne miały wyższy poziom postawy odtwórczej w obszarze konformizmu oraz niższy poziom potrzeby poznawczego domknięcia w obszarze zdecydowania niż osoby o wysokich umiejętnościach projektowania wykorzystującego konstrukcje geometryczne. Okazało się również, że u tych osób wraz ze wzrostem poziomu potrzeby poznawczego domknięcia zmniejszał się poziom zachowań twórczych, natomiast u osób o wysokim poziomie umiejętności projektowania wykorzystującego konstrukcje geometryczne miał tendencję wzrostową. Wyniki wskazują również, że wśród badanych dominował przeciętny poziom zachowań twórczych i wysoki poziom konformizmu związanego z postawą odtwórczą, co stanowi wynik niepokojący w odniesieniu do przyszłych architektów.

Słowa kluczowe: twórczość, potrzeba poznawczego domknięcia, geometria, architektura

1. Introduction

Creativity and the ability to comprehend and design complex spatial structures are essential skills for architects. At the same time, research on creativity show that only creative abilities may be insufficient if an individual does not possess appropriate personality traits [5]. Due to the fact that the findings concerning the links between creativity and personality traits, features of cognitive functioning, or environmental variables are inconsistent, there is a need for further exploration in this area.

2. Creative behaviour

Creativity invariably is associated with the production of unique works or ideas by prominent individuals. Nowadays, however, besides an elite approach to creativity, an egalitarian approach has been developing. In this approach, creativity is defined as a personal trait, which manifests itself in everyday life, and a creative individual is the one who shows an active attitude towards the reality, is open to change, solves problems creatively, and seeks to develop themselves, often by crossing their own borders [1, 18]. In the last decades, there have been many reports and publications concerning the characteristics of creative people, which included cognitive abilities, creativity, and personality traits [16]. Most of the developed models are interactive in character, that is, they take into account both personal, and socio-cultural aspects. Popek [19, 20, 21] proposed a model involving the concept of interactive creativity. According to Popek, creative abilities result from the interaction of the properties of an individual (cognitive, motivational and emotional) and environmental conditions (stimulation derived from the social environment, hierarchy of values). Popek [19, 20, 21] defines creativity as a cognitive and characterological property of an individual, which can be described on a continuum, with creative attitudes on the one end, and reproductive on the other. A creative attitude consists of heuristic behaviour (cognitive component) and nonconformity (characterological component) [20]. The reproductive attitude includes an algorithmic behaviour and conformity [20].

According to Popek [20], in the cognitive area, creative or heuristic behaviours are associated with the independence of observation, divergent thinking, logic memory, intellectual flexibility, reflexivity, creative imagination, constructive creativity, and a potential to develop an artistic creation [20]. The opposite to a heuristic behaviour is an algorithmic behaviour, which is associated with mechanical memory, reproductive imagination, directed perception, intellectual stiffness, convergent thinking, cognitive passivity, low level of reflection, low efficiency in processing and constructing, as well as the lack of artistic, or technical creativity [20].

From a characterological point of view, a creative attitude is related to nonconformity which is determined by independence, action, originality, consistency, courage, self-reliance, spontaneity, openness, perseverance, self-criticism, tolerance, expressiveness, and resistance [20]. In contrast to a creative attitude – the conformity is expressed through an adaptive stiffness, submission, stereotypes, dependence, fearfulness, defensiveness, low resistance, intolerance and lack of criticism [20].

A variety of research referred to the way creative individuals function. Bernacka [1], for instance, pointed out that creative nonconformists demonstrated a higher level of cognitive and creative needs, and more often were characterised by a heuristic behaviour. Research carried out by scientists from IPAR (Institute for Personality Assessment and Research) showed that creative people are independent, non-conformist, open to new experiences and not afraid of taking risks [23]. Malkiewicz and Piskozub's [17] study revealed that creative individuals were more task-oriented, and exhibited coping styles based on emotions than reproductive individuals. Similar results were obtained from Strzalecki's [25] study, who described a positive relationship between a creative behaviour and ability to overcome difficulties and obstacles.

In line with definitions of creative and reproductive attitudes, an architect should demonstrate a high intensity of heuristic behaviour and nonconformity, and a low intensity of algorithmic behaviour and conformity.

3. Need for Cognitive Closure (NfCC) and its relation to creativity

The need for cognitive closure was first defined and described in order to explain individual differences in the context of cognitive functioning of individuals, who seek or acquired a relatively simple, factual knowledge in order to reduce their sense of insecurity.

Kruglanski [12, p. 6] defined the need for cognitive closure (NfCC) as a "desire for a firm answer to a question, any firm answer as compared to confusion and/or ambiguity." Two mechanisms underlie this need: the seize of information, and the freeze of information that has already been "caught" [13]. As a consequence, an individual does not need to cope with too much information, but only with the ones, which have been activated, or "seized" allowing them for as quick action. The transition from a "seize" to a "freeze" phrase initiates the crystallization of beliefs. At the same time, the "freeze" of information in the knowledge structures protects them against modification.

Individuals with high levels of NfCC strive for an immediate and consistent information closure [11, 14]. Thus, such people prefer a structured and predictable lifestyle, and avoid uncertainty and ambiguity [11], or cognitive complexity [26]. What is more, NfCC has been associated cognitive inhibition aiding individuals in coping with irrelevant information [10].

An overview of research on the need for cognitive closure conducted by Kossowska [8] showed that individuals with a high level of the need for cognitive closure, tend to reduce the scope of acquired data, and generate hypotheses, use the first available data and keep their opinion. These aspects are significant in the context of creative behaviour, but seem to stand in opposition to the need for cognitive closure, as they are linked to the search for new, surprising solutions and tolerance for ambiguity. Furthermore, a motivational aspect of creativity, involving the need for novelty and curiosity, opposes close-mindedness related to NfCC. Therefore, individuals with a high level of the need for cognitive closure are characterised by a lower intensity of creative behaviour, and contrary creative individuals-by lower intensity of need for cognitive closure. Jaworski [6] confirmed this notion; the students of artistic subjects scored lower on the scale of the need for cognitive closure than the other groups of students.

In contrast, Chybicka [3] pointed out that the need for closure is contextually determined and can affect the differences in creativity effectiveness, and in a proper contextual "opening" in individuals, NfCC does not inhibit creativity. Also, Sternberg [24] noticed that individuals, who prefer a cognitive style and pursuit goals according to settled standards and rules, can be also creative, due to the fact that some areas of their profession require various styles of thinking.

People, who aim at the quickest possible information closure and avoid uncertainty in novel and ambiguous situations, might feel uncomfortable, and as a result base their behaviour on previously developed ways of acting, and store information according to consistent and well-known schemata. Such behaviour contradicts the notion of creativity, being an essential condition for the process of architectural design. The need for cognitive closure may also manifest itself in difficulties to integrate a big amount of information. It might provide an explanation why people preferring a cognitive style experience more problems to integrate mathematical information [22], and cope with new, geometric construction tasks.

In order to understand the spatial structure of the complex geometrical construction, it is necessary to transfer it onto a drawing sheet. While solving geometric design tasks, it is impossible for students to only use their perceptual experience, but they are required to correctly read the projections of all elements of the space. They need to visualize the principles of projection of individual elements, which do not result from an intuitive, automatic processing of visual stimuli. People, who are exhibit a cognitive style, may also tend to oversimplify a mathematical and diminish geometric complexity, due to a shallow analysis of inflowing information [7]. This, in turn, can lead to irregularities in the integration of information and incorrect conclusions [22].

4. The purpose of the study and hypotheses

The purpose of the study was to identify relationships between the need of cognitive closure, the level of creative and reproductive attitudes and the comprehension and design of complex spatial structures. It has been assumed that people, who experience difficulty in the comprehension and design of complex spatial structures, exhibit a higher level of the need of cognitive closure and a higher level of reproductive behaviour than individuals who perform well on such tasks. It has also been asserted that the relationship between the need of cognitive closure and creative behaviour of individuals with different levels of abilities to understand and design complex spatial structures, differs.

5. Method

5.1. Participants

The study involved 111 first-year students of Architecture at the Faculty of Architecture of the Cracow University of Technology. The sample consisted of 83 women and 28 men (M-age = 19.18, SD = .79).

5.2. Tools

Kossowska's [7] Polish adaptation of the Need for Closure Scale [15] was used in the study. The questionnaire consisted of 32 statements, each marked on a 6-point scale, ranging from 1 – *I strongly disagree* to 6 – *I strongly agree*. The scale was divided into five subscales: Desire for predictability (e.g. "I don't like to go into a situation without knowing what I can expect from it"), Preference for order and structure (e.g. "I think that having clear rules and order at work is essential for success"), Discomfort with ambiguity (e.g. "I'd rather know bad news than stay in a state of uncertainty"), Decisiveness (e.g. "I usually make important decisions quickly and confidently"), Close-mindedness (e.g. "I do not usually consult many different opinions before forming my own view"). A sum of all subscales indicated an overall level of NfCC.

In order to measure creative behaviour, the Creative Behaviour Questionnaire KANH [20] was used. It consisted of 60 statements, marked on a scale ranging from 0 to 2, where 0 indicated a false statement for the respondent, 1 – the statement was partially true, and 2 – the statement was true. The questionnaire was divided into four subscales. The scales of non-conformity (N) and the scale of heuristic behaviour (H) comprised the creative attitude, and the conformity scale (C) and algorithmic behaviour (A) comprised the reproductive attitude. The difference in the level of non-conformity and conformity referred to a characterological area of creative attitudes, while the difference in the level of heuristic and algorithmic behaviour – referred to a cognitive sphere of creative attitudes.

A geometric task was used to measure the comprehension and design skills of complex spatial structures. The principles of construction of a normal axonometric representation of an object, which is constructed based on its orthographic views, have been discussed within the lecture that presented the properties of a parallel projection method [4]. On the same day as the lecture, the students were supposed to construct an abnormal axonometric view of an object, but with taking no advantage of the notes which were taken by while listening to the presentation. The normal axonometric view was supposed to be constructed starting from the given two-view representation of an object presented in an orthographic projection method (Fig. 1).

Three various aspects of the solving procedure have been evaluated while checking the solutions and the obtained results. These were as follows:

1. Student's ability to correctly „read" and interpret the two-view drawing.
2. Ability to apply the construction of a normal axonometry by execution of the subsequent "steps of construction" that have been presented during the lecture:
   - Step 1: Assigning the segment X'Y' in the top view (EU method) – which is a horizontal trace of an axonometric picture plane; the triangle of traces X'Y'Z' must always be an acute-angled triangle whose heights determine the three directions of axonometric axes x^n, y^n and z^n;
   - Step 2: Assigning point M' on the line X'Y' as the base (= a starting point) of a vertical axonometric axis z^n. Point M' is a normal projection of O onto X'Y';

Step 3: Determining the characteristic points of intersection between the sides of the object and the line X'Y', which are the traces of the object's base on the trace X'Y'; The axonometric view of the object's base will be drawn parallel to respective axes x' and y' through the determined points(I II x' 1,2 II y');

Step 4: Drawing a copy of the segment X'Y', together with the assigned traces, in a separate space of a drawing;

Step 5: As it has been mentioned above, the axonometric axes x^n, y^n and z^n are determined as the heights of an acute-angled triangle of traces X'Y'Z', thus the axis z^n must be drawn perpendicular to X'Y' and starting from M'. The centre O^n of the axes x^n, y^n and z^n have to be chosen on z^n ($O^nM^n$is less than O'M'). Directions of axes x^n and y^n are uniquely defined by $x^n = O^nX'$ and $y^n = O^nY'$. The base of an object can be now easily constructed by drawing parallels to the axes x^n and y^n through the earlier determined traces on X'Y';

Step 6: In order to determine the foreshortened heights of an object on the axis z^n, the rotation of the triangle $Z'O^nY'$ must be executed. True heights of an object can be now measured on the rotated axis z^o and then transformed back (= rotated back) onto the axonometric axis z^n. The foreshortened heights will be used to add the heights to the base of a constructed axonometry.

3. Ability to correctly assign visible and hidden edges of a three-dimensional object by using either the thick continuous line (visible edges) or thin continuous lines (hidden edges).

Students who correctly performed two or three elements of the task were assigned to a group with a high level of design skills using geometric constructs, while those who managed only one element correctly or none, were assigned to a group with a low level of design skills using geometric constructs.

5.3. Procedure

The participants took part in a lecture on axonometry, and then performed a geometric task (without support of their own notes). A week after, the respondents were presented with the Need for Closure Scale and Creative Behaviour Questionnaire KANH.

5.4. Statistical analysis

The analyses were performed using (statistical analysis programs) STATISTICA 12 and SPSS 21. Leven's test was used to determine the homogeneity of variance. The t-test (in the absence of group homogeneity, the t-test was used with a separate estimation of variance) and the moderation analysis were used for comparisons between groups with low and high levels of geometric design skills. Pearson correlation analysis was used to analyse the relationship between the creative behaviour and the needs of cognitive closure. The significance level of $\alpha = .05$ was assumed; however, the results for the significance level of $\alpha = .10$ were also described as they fell within a limit considered as a trend ($.05 < \alpha < .10$).

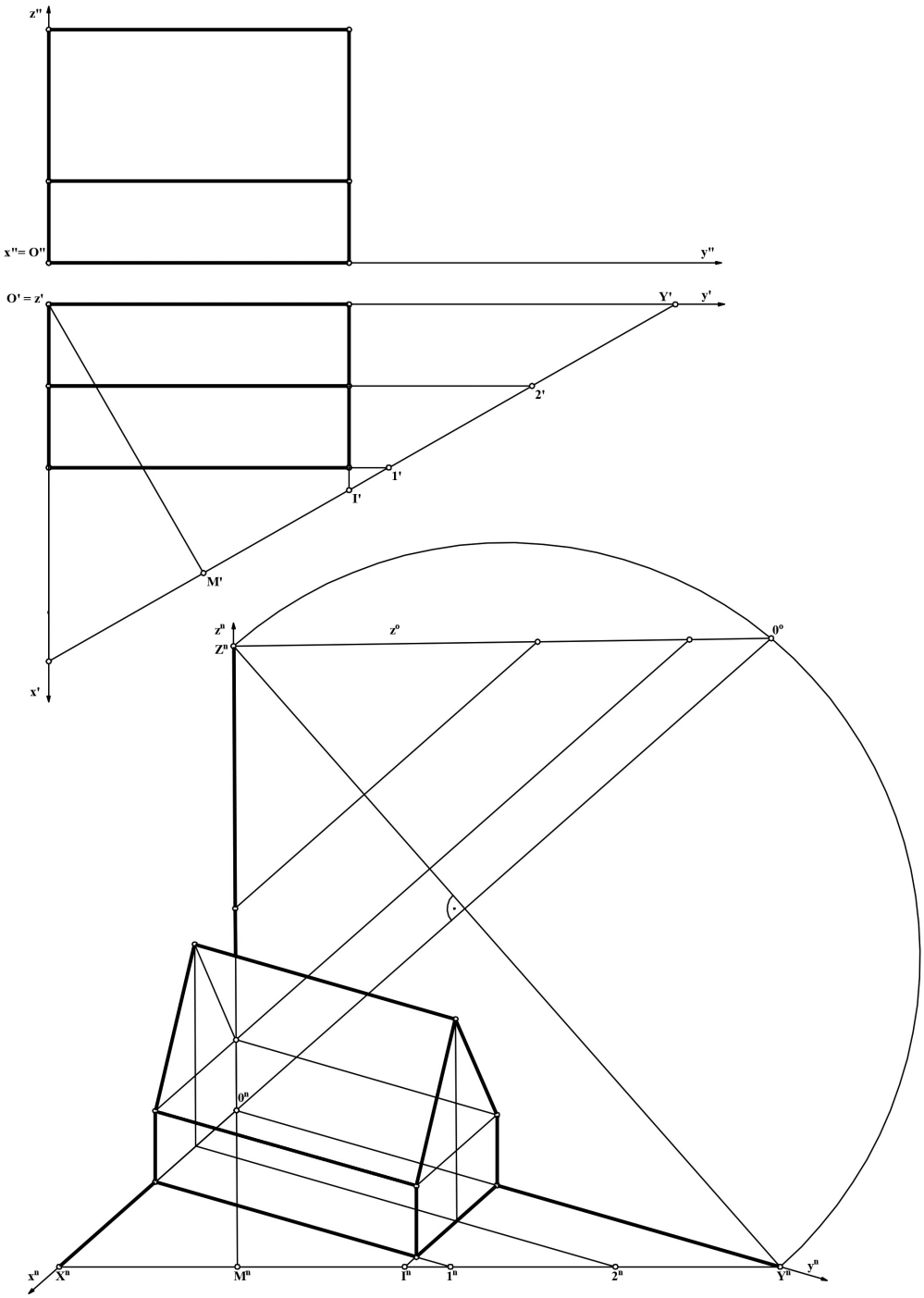


Fig. 1. Two views of a three-dimensional (3D) object together with a specified segment X'Y' where the axonometric picture plane cuts the horizontal picture plane and the normal axonometric image of the given in two views 3D object

6. Results

In order to determine whether the ability to comprehend and design complex spatial constructions was related to the level of the need of cognitive closure, the NfCC scores of students, who coped well and poorly with geometry task were compared (Tab. 1). It was observed that participants with a high level of design skills using geometric constructs showed a higher level of Decisiveness than students who performed a geometric task poorly, or not at all.

Table 1. Differences in level of the need of creative closure in individuals with low and high design skills using geometric constructs

	M_{All}	SD_{All}	M_{GL}	M_{GH}	t	df	p	n_{GL}	n_{GH}	SD_{GL}	SD_{GH}	F_L	p_L
Discomfort with ambiguity	26.42	3.20	26.38	26.41	-.03	108	.974	78	32	3.16	3.33	1.11	.700
Preference for order and structure	28.57	5.90	28.68	28.34	.27	108	.789	78	32	6.05	5.72	1.12	.745
Desire for predictability	29.14	5.90	28.59	30.50	-1.55	108	.125	78	32	6.09	5.35	1.30	.423
Close-mindedness	16.37	3.27	16.38	16.34	.06	108	.953	78	32	3.53	2.65	1.78	.075
Decisiveness	17.11	4.35	16.53	18.38	-2.42	85.24	.018	78	32	4.67	3.12	2.24	.014
NfCC	117.60	13.58	116.56	119.97	-1.19	108	.236	78	32	14.04	12.47	1.27	.467

p: p-value
GL: low level of design skills using geometric constructs
GH: high level of design skills using geometric constructs
NfCC: Need for Cognitive Closure
F_L and p_L: Leven's test of homogeneity

The measurement of creative behaviour, and transformation of the results onto a stenographic scale to derive the values of the general population, showed that in the examined group of future architects, only the level of conformity exceeded the average level for the population (7 sten); other results, also those related to the creativity, were exactly at the average level for the population (Tab. 2). In order to determine whether these traits were related to the handling of geometric constructions, the level of traits in participants who performed the geometric task well was compared with the level of traits in individuals who performed the geometric task poorly (Tab. 3). The comparison indicated that participants who did the geometric task well had a lower level of conformity than individuals who performed poorly, or not at all (p = .056). The other differences were statistically insignificant.

Table 2. The level of creative behaviour

	n	*M*	*SD*	Min	Max	Sten (*median*)
Conformity	110	12.50	4.21	4.00	22.00	7
Nonconformity	110	18.30	4.18	8.00	27.00	5
Algorithmic	110	13.52	2.99	7.00	23.00	5
Heuristic	110	17.97	3.59	10.00	26.00	5
Creative	110	36.27	7.02	20.00	52.00	-
Reproductive	110	26.02	6.16	13.00	41.00	-
SphereCH	110	5.80	6.62	-12.00	20.00	5
SphereP	110	4.45	4.92	-10.00	16.00	6

SphereCH: Characterological Sphere
SphereP: Cognitive Sphere

Table 3. Differences in levels of creative behaviour among individuals with a low and high level of design skills using geometric constructs

	M_{GL}	M_{GH}	*t*	*df*	*p*	n_{GL}	n_{GH}	SD_{GL}	SD_{GH}	*F*-Var	*p*-Var
Conformity	13.05	11.38	1.93	107	.056	77	32	3.97	4.50	1.28	.381
Nonconformity	18.25	18.63	-.43	107	.667	77	32	4.25	3.95	1.16	.659
Algorithmic	13.53	13.56	-.05	107	.962	77	32	3.19	2.49	1.64	.124
Heuristic	17.81	18.38	-.75	107	.455	77	32	3.65	3.53	1.07	.855
Creative	36.05	37.00	-.64	107	.523	77	32	7.13	6.82	1.10	.798
Reproductive	26.58	24.94	1.28	107	.203	77	32	6.17	5.97	1.07	.858
SphereCH	5.19	7.25	-1.48	107	.143	77	32	6.76	6.26	1.17	.646
SphereP	4.27	4.81	-.52	107	.606	77	32	5.24	4.18	1.58	.159

p: *p*-value
SphereCH: Characterological Sphere
SphereP: Cognitive Sphere
GL: low level of design skills using geometric constructs
GH: high level of design skills using geometric constructs

Several relationships between the level of creative behaviour and the need for cognitive closure were observed in the sample (Tab. 4), for instance, the higher the level of Preference for order and structure was, the higher the level of Non-Conformity, Heuristic behaviour, Algorithmic behaviour, as well as the global level of creative and reproductive attitudes became. As the level of Preference for predictability increased, the level of Conformity, Algorithmic behaviour and reproductive attitudes also increased, but the level of Nonconformist characterological attitude decreased. What is more, as the level of Decisiveness increased, the level of Non-conformity, creative attitude and the level of non-conformity sphere of characterological creative attitude increased, but the level of Conformity decreased. It was also observed that the higher the level of Close-mindedness, the lower the overall levels of creative attitude, heuristic thinking, Non-conformity and cognitive and characterological sphere of creative attitude. The analyses also showed that as the general level of the need for

cognitive closure increased, the level of Algorithmic behaviour and reproductive attitudes increased, but the level of the cognitive sphere of creative attitude decreased.

Table 4. The relationship between the need for cognitive closure and the level of creative behaviour

	Discomfort with ambiguity		**Preference for order and structure**		**Desire for predictability**		**Close-mindedness**		**Decisiveness**		**NfCC**	
	r	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Conformity	.08	.383	.06	.536	**.23**	**.014**	.01	.910	**-.20**	**.039**	.09	.367
Non-Conformity	-.13	.171	**.19**	**.042**	-.15	.111	**-.26**	**.006**	**.31**	**.001**	.03	.782
Algorithmic	.02	.857	**.34**	**<.001**	**.18**	**.064**	.09	.349	.09	.330	.28	.003
Heuristic	-.04	.670	**.20**	**.039**	-.07	.485	**-.38**	**<.001**	.16	.095	.01	.923
Creative	-.10	.302	**.22**	**.023**	-.12	.191	**-.35**	**<.001**	**.27**	**.005**	.02	.830
Reproductive	.07	.494	**.21**	**.031**	**.24**	**.010**	.05	.596	-.09	.355	.20	.039
SphereCH	-.14	.155	.08	.379	**-.24**	**.010**	**-.17**	**.076**	**.32**	.001	-.04	.691
SphereP	-.04	.675	-.06	.509	-.16	.103	**-.33**	**<.001**	.06	.535	-.17	.083

p: *p*-value
SphereCH: Characterological Sphere
SphereP: Cognitive Sphere
NfCC: Need for Cognitive Closure

It was also examined whether the relationship between the need for cognitive closure and creative behaviour were differentiated by the level of ability to comprehend and design of complex spatial structures, i.e. whether relationship between NfCC and creative behaviour would differ in participants with a low level of design skills using geometric constructs, and those with a high level of design skills using geometric constructs (Tab. 5).

Table 5. Differences in types of relationship between the need for cognitive closure and creative behaviour in participants with a low and high level of design skills using geometric constructions (Moderator: the level of ability to comprehend and design of complex spatial structures GH *vs* GL)

Predictor	**Dependent variable**	***ß***	***SE***	***t***	***p***	**Interaction**
1	2	3	4	5	6	7
Decisivness	SphereP	.13	.11	1.19	.238	
	SphereCH	.08	.11	.72	.475	
	Heuristic	.09	.11	.79	.432	
	Algorithmic	-.11	.11	-1.02	.311	
	Non-Conformity	-.02	.11	-.22	.826	
	Conformity	-.14	.11	-1.33	.187	
	Creative	.03	.11	.29	.774	
	Reproductive	-.15	.11	-1.39	.168	

1	2	3	4	5	6	7
Close-mindedness	SphereP	.08	.10	.80	.426	
	SphereCH	.05	.11	.51	.613	
	Heuristic	.11	.10	1.07	.287	
	Algorithmic	-.01	.11	-.05	.976	
	Non-Conformity	.08	.10	.79	.430	
	Conformity	-.003	.11	-.03	.976	
	Creative	.10	.10	1.03	.307	
	Reproductive	-.005	.11	-.05	.962	
Preference for order and structure	SphereP	-.14	.10	-1.20	.232	
	SphereCH	.02	.11	.22	.829	
	Heuristic	-.18	.10	-1.62	.109	
	Algorithmic	.10	.11	.09	.930	
	Non-Conformity	-.06	.11	-.52	.605	
	Conformity	-.01	.11	-.86	.391	
	Creative	-.13	.11	-1.14	.258	
	Reproductive	-.06	.11	-.55	.583	
Desire for predictability	SphereP	.09	.11	.85	.399	
	SphereCH	.21	.11	**1.96**	**.052**	**ß GL = -.37 (*p* = .001)** **ß GH = .05 (*p* = .795)**
	Heuristic	.04	.11	.39	.695	
	Algorithmic	-.10	.11	-.92	.359	
	Non-Conformity	.12	.11	1.07	.285	
	Conformity	**-.21**	**.11**	**-2.02**	**.046**	**ß GL = .37 (*p* = .001)** **ß GH = -.06 (*p* = .741)**
	Creative	.09	.11	.84	.405	
	Reproductive	-.20	.11	-1.03	.071	**ß GL = .36 (*p* = .001)** **ß GH = -.03 (*p* = .868)**
Discomfort with ambiguity	SphereP	.07	.12	.60	.552	
	SphereCH	**.23**	**.11**	**2.04**	**.044**	**ß GL = -.27 (*p* = .021)** **ß GH = .15 (*p* = .381)**
	Heuristic	.02	.12	.22	.829	
	Algorithmic	-.08	.12	-.72	.471	
	Non-Conformity	**.24**	**.11**	**2.11**	**.037**	**ß GL = -.25 (*p* = .029)** **ß GH = .18 (*p* = .289)**
	Conformity	-.13	.11	-1.11	.270	
	Creative	.16	.12	1.34	.182	
	Reproductive	-.13	.11	-1.11	.171	

Table 5 (cont.)

1	2	3	4	5	6	7
NfCC	SphereP	.06	.11	.51	.608	
	SphereCH	.17	.11	1.53	.129	
	Heuristic	-.003	.11	-.03	.979	
	Algorithmic	-.10	.11	-.91	.365	
	Non-Conformity	.08	.11	.71	.479	
	Conformity	**-.19**	**11**	**-1.74**	**.085**	**ß GL = .21 (*p* = .059)** **ß GH = -.17 (*p* = .368)**
	Creative	.05	.11	.41	.686	
	Reproductive	-.18	.11	-1.64	.104	

p: *p*-value
SphereCH: Characterological Sphere
SphereP: Cognitive Sphere
GL: low level of design skills using geometric constructs
GH: high level of design skills using geometric constructs

The results indicated that in the case of the relationship between Desire for predictability and creative behaviour in the participants with a low level of design skills using geometric, the level of Desire for predictability increased, as the level reproductive attitude, and Conformity increased, but the level of Characterological Sphere associated with creativity decreased. However, in the individuals who performed well on geometric constructions design, the relationship was reversed, though statistically insignificant. As far as the relationship between Discomfort with ambiguity and creative behaviour is concerned, it was observed that as the level of Discomfort with ambiguity increased, the level of nonconformity and entire sphere of characterological creativity decreased in the group of the respondents with a low level of design skills, whilst the relationship was reversed among the respondents with a high level of design skills using geometric construct, but it was statistically insignificant. Additionally, it was found that as the general level of NfCC increased, the level of Conformity rose in participants who performed poorly on the geometric task; however, the relationship was reversed among respondents who performed well on the geometric task, but statistically insignificant.

7. Discussion

The results showed that creative behaviour and the need for cognitive closure are closely linked; however, the direction of the relationship may be different depending on individual's level of design skills using geometric construct. The increase in the need for cognitive closure (desire for predictability, discomfort with ambiguity, and general level of NfCC) was associated with lower levels of creative behaviour and a higher level of reproductive behaviour in the group, who performed poorly on the geometric task. However, in the group of the respondents with a high level of design skills using geometric constructs, the direction was reversed, but statistically

insignificant. This may indicate that for individuals, who are able to adequately visualize, even complicated geometric constructions, the inhibition of creative behaviour does not occur, even at a higher level of the need of cognitive closure. Perhaps, in case of such individuals, the lack of discomfort with ambiguity facilitates the ordering of information according to one's own rules and not in a way enforced by others, which in turn promotes comprehension of complex spatial forms.

Also, the results concerning the relationship between the creative attitude and the need for cognitive closure seem to be interesting. It was found that the higher the intensity of Preference for order and structure was, the higher the level of creative behaviour occurred in the sample. This confirms the notion that in the case of individuals, who design new architectural forms, i.e. their activity is creativity-related, the realization of the task requires not only order and decisiveness, but besides artistic form, also technical and engineering skills.

The analyses showed that individuals with a higher level of design skills using geometric construct, also exhibited a higher level of decisiveness than individuals who performed poorly on geometric task. What is more, they presented a lower level of conformity. As far as the analysis of moderators is concerned, it is uncertain whether the characterological sphere of creativity is the most important here. Perhaps, as for construction tasks, they strive for order and predictability combined with a nonconformity, may enhance students' comprehension and solution finding. The low level of conformity might be necessary for architects to create rules and manage the design process on a two-dimensional surface, without having to rely on third parties.

Nevertheless, the findings concerning the level of creative and reproductive attitudes among surveyed students – the future architects seem alarming. None of the elements of creative attitude exceeded the average level for the population; however, a slightly higher than an average score was found in the intensity of conformity. Taking into account an earlier discussion of the link between the ability to comprehend and design geometrical constructs, and a low level of conformity, a serious concern arises as far as students' suitability for the profession of architect is concerned. The study of Wojtowicz and Wojtowicz [27] sheds a positive light on the problem. The results revealed that the grade obtained by students from an architectural design task was not significantly associated with the level of implementation of the geometric task. Nonetheless, the average level of heuristic behaviour, nonconformity, as well as the cognitive and characterological sphere of creative attitudes might reduce the future chances of pursuing architectural designs, which could, in some way, be indeed visionary. Even though architecture involves a conscious shaping of space to achieve its intended function in a well-defined form and structure [2], people who base on well-known and proven schemes only may end up with monotony and mediocrity on the subject of architecture.

References

[1] Bernacka R.E., *Konformizm i nonkonformizm a twórczość*, Wydawnictwo UMCS, Lublin 2004.

[2] Charytonow E., *Projektowanie architektoniczne*, Państwowe Wydawnictwa Szkolnictwa Zawodowego, Warszawa 1985.

[3] Chybicka A., *Otwarty umysł tworzy*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2004.
[4] Górska R., *Descriptive Geometry. Freshman Level Course Addressed to the Engineering Students*, Wydawnictwo Politechniki Krakowskiej, Kraków 2013.
[5] Feist G.J., Barron F.X., *Predicting creativity from early to late adulthood: Intellect, potential, and personality*, Journal of Research in Personality, 37/2003, 62-88.
[6] Jaworski M., *Polska adaptacja Skali Potrzeby Domknięcia Poznawczego*, Przegląd Psychologiczny, 41/1998, 151-163.
[7] Kossowska M., *Różnice indywidualne w potrzebie poznawczego domknięcia*, Przegląd Psychologiczny, 46/2003, 355-374.
[8] Kossowska M., *Umysł niezmienny… Poznawcze mechanizmy sztywności*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2005.
[9] Kossowska M., *Motives which can promote versus reduce prejudice*, Psychologia Społeczna, 02/2006, 13-22.
[10] Kossowska M., *The role of cognitive inhibition in motivation toward closure*, Personality and Individual Differences, 42/2007, 1117-1126.
[11] Kruglanski A.W., *Lay epistemics and human knowledge: Cognitive and motivational bases*, Plenum, New York 1989.
[12] Kruglanski A.W., *The psychology of closed mindedness*, Psychology Press, New York 2004.
[13] Kruglanski A.W., Webster D.M., *Motivated closing of the mind: "Seizing" and "freezing"*, Psychological Review, 103/1996, 263-283.
[14] Kruglanski A.W., Orehek E., Dechesne M., Pierro A., *Lay epistemic theory: Motivational, cognitive, and social aspects of knowledge formation*, Social and Personality Psychology Compass, 4/2010, 939-950.
[15] Kruglanski A.W., Webster D.M., Klem A., *Motivated resistance and openness to persuasion in the presence or absence of prior information*, Journal of Personality and Social Psychology, 65/1993, 861-876.
[16] Limont W., *Pedagogika twórczości, czyli edukacja ku twórczości*, [in:] *Pedagogika. Subdyscypliny i dziedziny wiedzy o edukacji*, Tom 4, (ed.) B. Śliwerski, Gdańskie Wydawnictwo Pedagogiczne, Sopot 2010, 263-289.
[17] Małkiewicz M.M., Piskozub M., *Różne nasilenie postawy twórczej a kompetencje społeczne i style radzenia sobie ze stresem u studentów pierwszego roku*, Studia Psychologica, 11/2011, 19-32.
[18] Nęcka E., *Psychologia twórczości*, Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2016.
[19] Popek S. (ed.), *Z badań nad zdolnościami i uzdolnieniami specjalnymi młodzieży*, Wydawnictwo UMCS, Lublin 1987.
[20] Popek S., *Człowiek jako jednostka twórcza*, Wydawnictwo UMCS, Lublin 2001.
[21] Popek S., *KANH – Kwestionariusz twórczego zachowania*, Wydawnictwo UMCS, Lublin 2000.
[22] Sarnataro-Smart S., *Personal Need for Structure: Indiscriminate Classification Systems As Barriers to Processing Mathematical Complexity*, Honors Thesis Collection, Paper 110, 2013.

[23] Simonton D.K., *Creativity: Cognitive, personal, developmental, and social aspects*, American Psychologist, 55/2000, 151-158.
[24] Sternberg R.J., *Thinking styles*, Cambridge University Press, Cambridge 1999.
[25] Strzałecki A., *Psychologia twórczości: między tradycją a ponowoczesnością*, Wydawnictwo Uniwersytetu Kardynała Stefana Wyszyńskiego, Warszawa 2003.
[26] Webster D.M., Kruglanski A.W., *Individual differences in need for cognitive closure*, Journal of Personality and Social Psychology, 67/1994, 1049-1062.
[27] Wojtowicz A., Wojtowicz B., *The personal need for structure as a factor affecting the understanding and projecting of complex spacial structures*, Technical Transactions, vol. 22/2015, 63-72.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
CHEMISTRY
DOI: 10.4467/2353737XCT.17.191.7420

Robert Grzywacz (robekk@gmail.com)
Jakub Szyman
Department of Chemical and Process Engineering, Faculty of Chemical Engineering and Technology, Cracow University of Technology

Mathematical model of a flat plate photocatalytic reactor irradiated by solar light

Model matematyczny płaskiego reaktora fotokatalitycznego, pracującego w świetle słonecznym

Abstract

The paper presents a model of a flat plate photocatalytic reactor under solar radiation. The model was based on convection and diffusive mass flux balances in two zones: thin liquid layer and pores in the layer of a porous catalysts. The flux of light intensity was described by Kubelka–Munk theory.

Keywords: modeling, photocatalysis, reactors, solar radiation

Streszczenie

W artykule przedstawiono model opisujący pracę płaskiego reaktora fotokatalitycznego, pracującego w świetle słonecznym. Model został oparty o bilans konwekcyjnych i dyfuzyjnych strumieni masy w dwóch strefach: cienkim filmie cieczy i porach warstwy katalizatora. Strumień natężenia światła został opisany za pomocą teorii Kubelka–Munka.

Słowa kluczowe: modelowanie, fotokataliza, reaktory, promieniowanie słoneczne

1. Introduction

Advanced oxidation processes are developed as an alternative for biological water treatment [1, 2] and air pollution removal [3, 4]. Several works also present different uses of these methods: hydrogen production from water splitting and the decay of hydrogen disulfide [5]. Photocatalysis is one of the most elaborated oxidation processes due to its potentially low-cost generation in the case of carrying the process under solar radiation.

The photocatalytic potential of oxidation is generated in semiconductor catalysts. An electron from the valence band may be excited to the conduction band if there a photon is absorbed, whose energy is higher than the difference of energy between the bands, which rely on the kind of semiconductor used in the photocatalytic process.

Photocatalysis is realized in photocatalytic reactors. At present, there are two general concepts of photocatalytic reactor design. The first kind is a flat surface coated with a semiconductor layer, where only a thin film of fluid makes contact with the immobilized catalysts. The volume where the reaction takes place is limited to the immobilized catalyst layer, where the flux of radiation penetrating the catalysts is significant. The second type may be defined as a suspension of semiconductor nanoparticles in an aqueous solution of the reacting substrate. Ideal mixing in reactors guarantees bigger than immobilized reactor volume where the reaction occurs and process kinetics is not limited by diffusion. The main problems of these devices are: limited concentration of particles due to their aggregation and a separation method that is heavy to realize in continuous mode, after which the catalyst's particles cannot be used again [6]. Several cases of scaling-up the flat plate type have been reported [7]. To optimize their work, it is necessary to create a mathematical model, obtaining efficiency of the process in a wide range of parameters.

2. Basic aspects of assumptions selection

2.1. Kinetics of photocatalytic process

Most of the works the describing kinetics of photocatalytic processes use Langmuir-Hinshelwood kinetics [8, 9]. It is assumed that the oxidation center is an electron hole located on the crystal's surface, so the reaction rate has to be correlated with the adsorption equilibrium:

$$r = k_{\mathrm{LH}} \frac{Kc}{1+Kc} \tag{1}$$

where:

k_{LH} – is a Langmuir–Hinshelwood constant describing the rate of surface reaction;
K – is an adsorption equilibrium constant;
c – is a concentration of degraded substrate.

Apart from the reports [10, 11] indicating inaccuracies in the theory of the Langmuir–Hinshelwood model with conclusions resulting from experiments, it has been still successfully

used in kinetic data regression [8, 9]. The Langmuir–Hinshelwood model may be simplified, if $K_{ads}c << 1$ kinetics will gain a pseudo-first order:

$$r = k_{app}c \tag{2}$$

where:

$k_{app} \sim Kk_{LH}$

It is reasonable when the concentration of the decayed substance is relatively low and such assumption will be made in this paper.

The relation of the light flux intensity on the reaction rate can be described by [10]:

$$k_{LH} = k\varphi^n \tag{3}$$

where:

k – kinetic constant, related to radiation intensity is 1 Wm^{-2};

φ – light intensity in Wm^{-2};

n – parameter.

Note that n is 1 for a low intensity level and could vary with different light flux, which is accounted on limiting kinetics by mass transfer resistance [10, 12]. In this work, the linear relation between the light intensity flux and the reaction rate is approached because mass transfer resistance is considered in a different way, so that the kinetic model is modified.

$$r = \frac{k_{LHodn}I}{I_{odn}} \frac{K_{ads}c}{1+K_{ads}c} \tag{4}$$

where:

k_{LHodn} – Langmuir–Hinshelwood constant, measured in I_{odn} light intensity;

I – light flux intensity in Wm^{-2}.

2.2. Light penetration through catalyst layer

Models describing the rate of degradation in flat plate type photocatalytic reactors [13, 14] do not consider that a photocatalytic reaction can also take part inside the catalyst layer, where the light flux penetrating the bulk catalyst can be still significant and has influence on the overall degradation effect. This work also takes into account degradation that takes place deep in catalyst pores; therefore, it is important to select the light penetration model. Recently, the Kubelka–Munk model was successfully used to describe light penetration in particulate materials [15]. The Kubelka–Munk theory considers that light is penetrating an endless plate on both sides and it changes by light absorption or scattering on penetrated media (Fig. 1). The process is described by a couple of equations, whose analytical solution is expressed as:

$$I(\lambda,\delta) = I_0(\lambda)\left[u(1-\beta)\exp(\alpha\delta) + v(1+\beta)\exp(-\alpha\delta)\right] \tag{5a}$$

$$J(\lambda,\delta) = J_0(\lambda)\left[u(1+\beta)\exp(\alpha\delta) + v(1-\beta)\exp(-\alpha\delta)\right] \tag{5b}$$

where:

I, J – light intensity flux, penetrating the mass of the penetrated substance on both sides in Wm^{-2};

I_0, J_0 – light flux intensity irradiated on both sides of the penetrated substance on its borders in Wm^{-2};

α – parameter expressed by equation (6a);

β – parameter expressed by equation (6b);

δ – depth for which the light flux intensity is expressed in m;

u, v – constants, calculated by equations (7a) and (7b);

α, β, u and v can be expressed by:

$$\alpha = \sqrt{\kappa(\kappa + 2\sigma)} \tag{6a}$$

$$\beta = \sqrt{\frac{\kappa}{\kappa + 2\sigma}} \tag{6b}$$

$$u = \frac{(\beta - 1)\exp(-\alpha\Delta)}{(1+\beta^2)\exp(\alpha\Delta) - (1-\beta)^2\exp(-\alpha\Delta)} \tag{7a}$$

$$v = \frac{(\beta + 1)\exp(-\alpha\Delta)}{(1+\beta)^2\exp(\alpha\Delta) - (1-\beta)^2\exp(-\alpha\Delta)} \tag{7b}$$

where:

κ – light absorption coefficient in m^{-1};

σ – light scattering coefficient in m^{-1};

Δ – depth of penetrated substance in m.

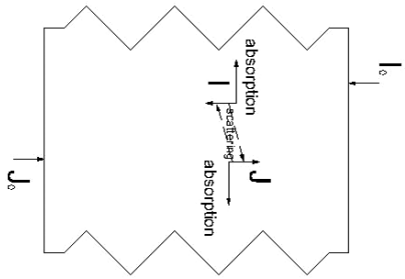


Fig. 1. Exposure of radiation stream, contributed to Kubelka–Munk equation

In this work, it is approached that a catalyst is irradiated by solar radiation only from the top, so $J_0 = 0$. As it is shown in equation (5a) and also because the light scattering and absorption coefficient can vary for different light wavelengths, one must knowthe spectral distribution of solar light. Work [16] presents the model based on the spectrum of solar light on a ground level. The solution may be used as I_0 in eq. (5a). Only for a part of the spectrum,

where the photon has higher energy, the excitation energy is balanced in this model. The photon energy is bonded with the wavelength by relation:

$$\lambda = \frac{hc}{E} \tag{8}$$

where:

λ – minimal wavelength to excite electron in semiconductor in nm;
h – Plack's constant in eVs^{-1};
c – velocity of light in nms^{-1};
E – excitation energy eV.

This means that only light, which has a wavelength lower than calculated from equation (8), can effectively excite an electron and generate a free electron – whole pair. Most of the semiconductors need radiationthat has a wavelength, which corresponds to near UV radiation or lower. It constitutes about 5% of the overall solar energy in a ground level [16]. The maximal amount of energy for TiO_2 atanase, obtained from that model, is about 55 Wm^{-2}. It is over twice more than the one measured in the experimental work [17], when the sky is clear, probably because they take into account the slope of the plate. Apart from that, the value in outdoor conditions should be higher, because most of the UV radiation is accumulated through high scattering from the ground [16]. It is hard to obtain that effect, so that I_0 could be approached as a value received from work [16].

The scattering and absorption coefficient can vary not only with the wavelength change and with the type of catalysts. Also, the method of catalyst's synthesis may be responsible for the disagreement of the experimental results [18]. Example values were measured in work [19].

The Kubelka–Munk model equation may be simplified. By calculating u and v from equations (7) in a boundary case, when , we can get:

$$\lim_{\Delta \to \infty} u = 0 \tag{9a}$$

$$\lim_{\Delta \to \infty} v = \frac{1}{1+\beta} \tag{9b}$$

These conditions change the form of the equation (5a) as:

$$I(\lambda,\delta) = I_0(\lambda)\exp(-\alpha\delta) \tag{10}$$

The overall light flux evaluated in the model is obtained by integration:

$$I(\delta) = \int_{\lambda=0}^{\lambda \min} I(\lambda,\delta)\,d\lambda \tag{11}$$

The light flux, which is related to the kinetic parameters obtained in an experimental way (I_{odn}), for sure cannot be measured experimentally, because it corresponds to the fictional value of the average light flux, which may quantitatively affect such a reaction effect and it can be calculated by:

$$I_{odn} = \frac{\int_0^{\Delta} I(\delta) d\delta}{\Delta} \tag{12}$$

where:

I_{odn} – a function, whose value can be obtained by equation (10), where I_0 equals the one measured in experiments;

Δ – depth of layer of semiconductor.

3. Mathematical model

The model of a flat-plate photocatalytic reactor is based on the mass balance of a degraded substrate. The mass streams included in the model are schematically presented in Figure 2. The equation describing the mass balance in a differential part of liquid film is defined as:

$$shdl\frac{dc_A}{dt} = shUc_A - shU\left(c_A - \frac{dc_A}{dl}dl\right) - sdl\varepsilon k_L\left(c_A - c_{Ar}\right) \tag{13}$$

where:

s – width of reactor plate;

h – height of liquid film;

U – mean velocity of liquid film;

c_A – concentration of degraded substance;

L – length of reactor;

ε – catalyst layer porosity;

k_L – overall mass transfer.

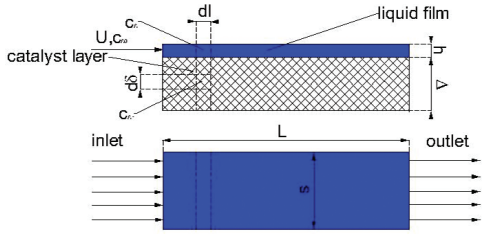


Fig. 2. Scheme of flat plate photocatalytic reactor, modeled in this work

The model also considers the equation of mass balance in an infinite volume of catalysts pores:

$$sdld\cdot\varepsilon\frac{\partial c_{Ar}}{\partial t}=-sdl\varepsilon D\frac{\partial c_{Ar}}{\partial\delta}+sdl\varepsilon D\left(\frac{\partial c_{Ar}}{\partial\delta}-\frac{\partial}{\partial\delta}\frac{\partial c_{Ar}}{\partial\delta}d\delta\right)-sdld\delta(1-\varepsilon)\frac{k_{\mathrm{LHodn}}I(\delta)Kc_{Ar}}{I_{\mathrm{odn}}} \quad (14)$$

where:

D – diffusion coefficient;

c_{Ar} – concentration of substrate in catalysts porous zone.

After introducing dimensionless variables, the model of the photocatalytic reactor operated under a steady state can be presented as:

$$\frac{\partial\alpha}{\partial z}=\frac{k_L\varepsilon\tau}{h}(\alpha-\alpha_r(0)) \quad (15)$$

$$\frac{\partial^2\alpha_r}{\partial z_{r^2}}=\frac{(1-\varepsilon)\Delta^2}{\varepsilon D_{\mathrm{AB}}}k_{\mathrm{LHodn}}K\frac{I(z_r\Delta)}{I_{\mathrm{odn}}}(1-\alpha_r) \quad (16)$$

where:

α – dimensionless concentration of substrate in film liquid zone;

α_r – dimensionless concentration of substrate in catalysts porous zone;

z – dimensionless length of reactor;

z_r – dimensionless depth of catalysts layer.

Values α, α_r may be obtained by:

$$\alpha=\frac{c_{A0}-c_A}{c_{A0}} \quad (17)$$

$$\alpha_r=\frac{c_{A0}-c_{\mathrm{Ar}}}{c_{A0}} \quad (18)$$

The initial condition is purposed as:

$$\alpha(0)=0 \quad (19)$$

For equation (16), boundary conditions (20) and (21) can be applied:

$$\frac{d\alpha_r}{\mathrm{d}z}(\delta=0)=\frac{k_L\Delta}{D_{\mathrm{AB}}}(\alpha_r-\alpha) \quad (20)$$

$$\frac{d\alpha_r}{\mathrm{d}z_r}(\delta=1)=0 \quad (21)$$

4. Discussion

The presented model can obtain the work of a flat plate photocatalytic reactor under solar radiation in a wide range of process parameters. It occurred through a sacrifice of its accuracy. Other works obtain the rate of a process as a function of the energy absorbed from the light

photon flux on a photocatalytic surface. The quantity of photons reaching the catalyst surface may be used in the model as the I_0 vector in the way of different calculation, for example, the local-area-specific rate of energy absorption (LASREA) [20] or the local volumetric rate of energy absorption (LVREA) [21], as they were used in the mentioned works. Note that such a calculation can have no influence on solution accuracy in the case of processing in changing atmospheric conditions, which are not accounted on the model and may have a stronger influence on it. The solution presented in this work measures the limitation of the process rate by diffusion and that aspect can vary with the size of the reactor. It is possible that the reaction rate has also been significantly underestimated instead of the diffusive transport of the decayed substrate and radiation, penetrated inside semiconductor layer was not considered. The model can be used as the basis for the evaluation of the reactor operation under the distribution of velocity of the liquid film and concentration gradient, generated through diffuse transport in catalyst pores may be needed.

Research may be also continuing by attaching other adsorption equilibrium models (BET isotherm or Dubinin equations, referred to microporous adsorption [22]).

References

[1] Parsons S., *Advanced Oxidation Processes (AOP) for water purification and recovery*, Catalysis Today, vol. 53(1), 1999, 51-59.

[2] Oller I., Malato S., Sanchez-Perez J.A., *Combination of Advanced Oxidation Processes and biological treatments for wastewater treatment decontamination – A review*, Science of The Total Environmental, vol. 409(20), 2011, 4141-4166.

[3] Oppenlander T., *Photochemical Purification of Water and Air: Advanced Oxidation Processes (AOPs)-Principles, Reaction Mechanism, Reactor Concepts*, Wiley, Darmstadt 2003.

[4] Zhao J., Yang X., *Photocatalytic Oxidation for indoor air purification: a literature review*, Building and Environmental, vol. 38, 2003, 645-654.

[5] Preethi V., Kanmani S., *Photocatalytic hydrogen production*, Materials Science in Semiconductor Processing, vol. 16, 2013, 561-575.

[6] Mukherjee P.S., Ray A.K., *Major Challenges in the Design of a Large-Scale Photocatalytic Reactor for Water Treatment*, Chemical Engineering & Technology, vol. 22(3), 1999, 253-260.

[7] Bahnemann D., *Photocatalytic water treatment: solar energy applications*, Solar Energy, vol. 77, 2004, 445-459.

[8] Behnajady M.A., Modirshahla N., Hamzavi R., *Kinetic study on photocatalytic degradation of C.I. Acid Yellow 23 by ZnO photocatalyst*, Journal of Hazardous Materials, vol. 133 (1-3), 2006, 226-232.

[9] Sauer T., Cesconeto Neto G., Jose H.J., Moriera R.F.P.M., *Kinetics of photocatalytic degradation of reactive dyes in a TiO_2 slurry reactor*, Journal of Photochemistry and Photobiologymistry, vol. 149(1-3), 2002, 147-154.

[10] Ramamurthy V., Schanze K.S., *Semiconductor Photochemistry and Photophysics*, Marcel Derrer, Inc., vol. 10, New York 2003.
[11] Lazar A.M., Varghese S., Nair S.S., *Photocatalytic Water Treatment by Titanium Dioxide: Recent Updates*, Catalysts, vol. 2, 2012, 572-601.
[12] Dionysiou D.D., Suidan M.T., Baudin I., Laine J.M., *Oxidation of organics contaminants in a rotating disk photocatalytic reactor: reaction kinetics in liquid phase and the role of mass transfer based on the dimensionless Damkolher number*, Applied Catalysis B: Environmental, vol. 38, 2002, 1-16.
[13] Al-Ekabi H., Serpone N., *Kinetic Studies In Heterogeneous Photocatalysis, 1. Photocatalytic Degradation of Chlorinated Phenols in Areated Aquerous Solutions over TiO_2 Supported on a Glass Matrix*, Journal of Physical Chemistry, vol. 92, 1988, 5726-5731.
[14] Khataee A.R., Fathinia M., Aber S., *Kinetic Modeling of Liquid Phase Photocatalysis on Supported TiO_2 Nanoparticles in a Rectangular Flat-Plate Photoreactor*, Industrial & Engineering Chemistry Research, vol. 49, 2010, 12358-12364.
[15] Ciani A., Goss K.U., Schawrzenbach R.P., *Light penetration in soil and particulate minerals*, European Journal of Soil Science, vol. 56, 2005, 561-574.
[16] Gates D.M., *Spectral Distribution of Solar Radiation at the Earth's Surface*, Science, vol. 151(3710), 1966, 523-529.
[17] Nogueira R.F.P., Jardim W.F., *TiO_2-fixed-bed reactor for water decontamination using solar energy*, Solar Energy, vol. 56(5), 1996, 471-477.
[18] Sclafani A., Palmisano L., Schiavello M., *Influence of the Preparation Methods of TiO_2 on Photocatalytic Degradation of Phenol in Aqueous Dispersion*, Journal of Physical Chemistry, vol. 94, 1990, 829-832.
[19] Cabrera M.I., Alfane O.M., Cassano A.E., *Absorption and Scattering Coefficients of Titanium Dioxide Particulate Suspensions in Water*, Journal of Physical Chemistry, vol. 100, 1996, 20043-20050.
[20] Zhang Z., Anderson W.A., Moo-Young M., *Rigorous modelling of UV-absorption by TiO_2 films in a photocatalytic reactor*, AIChE Journal, vol. 46(7), 2000, 1461-1470.
[21] Brandi R.J., Alfano O.M., Cassano A.E., *Modeling of flat plate photocatalytic reactor*, 5. World Congress of Chemical Engineering, San Diego, 14-18 July 1996, 1118.
[22] Do D.D., *Adsorption analysis: equilibria and kinetics*, Imperial College Press, Singapore 1998.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
CHEMISTRY
DOI: 10.4467/2353737XCT.17.192.7421

Krzysztof Skrzypek (krzysztofskrzypek93@gmail.com)
Robert Grzywacz
Department of Chemical and Process Engineering, Faculty of Chemical Engineering and Technology, Cracow University of Technology

The mixing hydrodynamics and efficiency of the Venturi jet mixer

Hydrodynamika i wydajność mieszania w mieszalniku Venturiego

Abstract

This article presents results of the numerical analysis of performance of the Venturi tube, used and applied as type of static liquid-liquid mixer. The calculations were performed for a binary acetone-benzene system of completely miscible Newtonian liquids. Several variables used to describe hydrodynamics of the Venturi were proposed. The influence of orifice location and inlet velocity on the hydrodynamics and mixing effectiveness is presented. Calculations were performed using CFD methods and the Fluent solver for 2D geometry.

Keywords: tank mixing eductor, Venturi jet mixer, liquid-liquid mixing, static mixing, passive mixing

Streszczenie

W artykule przedstawiono wyniki analizy numerycznej pracy ejektora, wykorzystanego jako statyczny mieszalnik ciecz-ciecz na przykładzie dwuskładnikowego układu aceton-benzen. Określono między innymi wpływ położenia otworków ssawnych na jego parametry hydrodynamiczne oraz prędkości wlotowej surowca na efektywność procesu. Na podstawie symulacji określono ogólną przydatność zaproponowanego urządzenia w zamierzonej dla niego roli na podstawie kilku wprowadzonych wielkości. Analizy dokonano metodami CFD przy użyciu solvera Fluent dla geometrii dwuwymiarowej.

Słowa kluczowe: mieszalnik zbiornikowy, mieszalnik Venturiego, mieszalnik statyczny, mieszanie pasywne, mieszanie cieczy

1. Introduction

Mixing is one of the most important industrial processes and one of the fundamental unit operations of Chemical and Process Engineering. It can be defined as both a process and its result that increases homogeneity of the system. Uniformity is crucial from the perspective of the product quality, but often of process efficiency as well, as macroscopic mixing can greatly supplement other transport phenomena. Despite its apparent simplicity, it remains one of the most complex operations and can be very difficult to carry out efficiently in terms of time and energy spent. There are several main reasons for this:

- The variety of substances that need to be mixed together and their individual and group properties;
- The multitude of ways mixing can be achieved, which raises the question of whether the method one intend to use is really the most efficient;
- The desired degree of product homogeneity, which is always a non-linear function of process time and conditions, and needs proper optimization.

This involves a vast amount of practical knowledge, as it is not possible to be covered purely by theory. Most of the mixing is performed in an individual apparatus through the use of rotating parts e.g. impellers or propellers, that mechanically agitate the fluids and promotes random particle relocation. Such a method is most often called active or dynamic mixing, as it needs an external source of energy to operate. In terms of results it is an efficient, but not always cost-efficient method, as keeping the engine working and agitator rotating continuously is considerable expense. The second group of mixing methods involve using the fluid's own internal energy and does not involve moving mechanical parts. Such techniques are known as passive or static mixing. In such apparatus, agitators are replaced with certain features of flow geometry that can lead to the development of hydrodynamic patterns that enhance both mechanical mixing and mass transfer. This is usually done by reorienting the flow, splitting the streams, joining them together and so on.

The advantages of static mixing are numerous, and are laid out below, as cited from [1]:

- Overall simplicity; static mixers are easy to construct and integrate with existing installations;
- Minimal maintenance; there are no moving parts that need to be serviced regularly;
- Low costs; static mixers are way cheaper to construct and operate than dynamic mixers;
- Process safety; ideal for handling flammable or fragile media, as they do not cause intense friction.

It should be noted that in a way, static mixers also use an external source of energy (pumps or fans). However, these are necessary anyway to operate any kind of chemical installation, and therefore it is not considered to be a downside. One of real disadvantages of static mixer is that each one needs to be specifically designed for an individual process, as there is minimum or even no option for automatic regulation. This encourages more reliance on simulations, which are way cheaper and faster than constructing several variants of apparatus and subsequent experiments.

2. Venturi jet mixer

Most of the common static mixers have the form of a pipe with helical or screw-like inserts of varying pitch, usually called Kenics mixers. The other kind is jet mixers, which use the high-velocity jet of one fluid, pumped into the environment of the other fluid. The device researched in this paper make use of an ejector, presented on the 2-D schematic below.

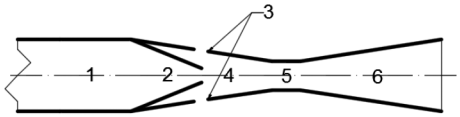


Fig. 1. Ejector: 1 – inlet, 2 – reducer, 3 – side orifices, 4 – mixing chamber, 5 – throat, 6 – nozzle

It performs its role thanks to the Venturi effect. The stream of one liquid, henceforth called the motive fluid, is pumped through the main inlet (1) and flows through the reducer of the ejector (2), where it increases its velocity to maintain continuity of flow. In consequence, static pressure decreases in accordance with the Bernoulli principle. This creates a zone of vacuum pressure and causes a suction effect that draws the secondary, "passive" fluid from the surroundings of the device through side orifices (3) and into the mixing chamber (4). Streams are collided there, accelerated in the throat section (5) and recompressed in the nozzle (6), through which resulting mixture is discharged outside. The apparatus analyzed in this paper consists of such an ejector inserted into the tank and submerged in the initial layer of one processed liquid.

In this paper, a device of this kind was combined with a tank and inlet pipe, thus constituting whole apparatus as presented in Figure 2. It is intended to work in a periodic way. First, pump P is turned on with V_1 valve opened (V_2, V_3 valves being closed) and the pure motive fluid is pumped into the tank and passes through the ejector R, up until the desired proportions between the substances is obtained. After that, V_1 valve is closed and V_2 is opened, so that the tank contents are sent through the recirculation loop for several more mixing cycles. After obtaining the desired uniformity, the pump is turned off, V_2 valve is closed and V_3 is opened. As a result, the whole mixture is discharged from the tank and the apparatus is ready for processing a new batch of substances. Inside the device, mixing essentially occurs during two distinct stages, first in the mixing chamber of the ejector and then in the tank itself.

Devices of this kind are already known in the industry under the name of tank mixing eductors, as found in several catalogues [2, 3], however the name applied in this paper was given by Sundararaj in his papers. In spite of their simplicity of construction and operation, such devices have not been the subject of such intense research as was the case in other kinds of mixers. Some specific data is known only by its industrial manufacturers and the articles by Sundararaj [4-6] remain as the only ones significant in the scientific field (at least concerning liquid-liquid mixing). To provide some more readily available info about such devices was the main motivation behind this paper. It should be noted however, that only preliminary research was conducted and it should be treated as a kind of guideline for future investigation and beginning of the research process, and not by any means its decisive conclusion.

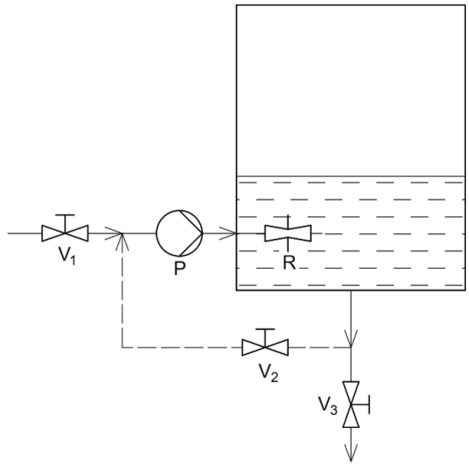


Fig. 2. Venturi mixer with supporting devices: V_1, V_2, V_3 – shut-off valves, P – pump, R – ejector; dashed line – recirculation loop, continuous line – inlet and outlet pipes, arrows – flow direction

3. Simulation setup details

Simulations were prepared and conducted using the ANSYS Workbench 17.2 components: Design Modeler, Meshing Tool and Fluent Solver, for 2-D planar geometry. The model used for calculations was the Standard k-ω, because of its simplicity and good convergence results. Species transport was also enabled, thus whole flow model consisted of six partial differential equations in the function of time and space.

a) Continuity

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \tag{1}$$

The first term stands for the accumulation of mass and second for its inflow and outflow.

b) Motion

The results shown in Table 2 exhibit a high convergence of the test outcome. The difference in force obtained by numerical tests in Abaqus relative to the actual sample is only 4.6%. The corresponding difference in Deform-3D is exactly 10.4%.

$$\frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) + \nabla p - \nabla \cdot (\bar{\bar{\tau}}) - \rho \vec{g} - \nabla \cdot (\rho \overline{v_x' v_y'}) = 0 \tag{2a}$$

$$\bar{\bar{\tau}} = \mu [(\nabla \vec{v} + \nabla \vec{v}^{\,T}) - \frac{2}{3} \nabla \cdot \vec{v} \cdot I)] \tag{2b}$$

The first term of Eq. 2a is accumulation of momentum per unit volume, the second addition through convection, the third and fourth through molecular transport, the fifth accounts for gravity and the last is turbulent momentum flux due to Reynolds stresses. The stress tensor is computed through Eq. 2b.

c) Turbulence equations for the k-ω model

$$\frac{\partial(\rho k)}{\partial t}+\nabla\cdot(\rho k\vec{v})-\nabla\cdot[(\Gamma_k\cdot\nabla\cdot k)]-G_k+Y_k=0 \tag{3}$$

$$\frac{\partial(\rho\omega)}{\partial t}+\nabla\cdot(\rho\omega\vec{v})-\nabla\cdot[(\Gamma_k\cdot\nabla\cdot\omega)]-G_\omega+Y_\omega=0 \tag{4}$$

With Eg.4, turbulence kinetic energy is computed, while Eg. 5 computes specific turbulence dissipation rate. They are analogous, with the first term of each equation standing for accumulation, the second for inflow or outflow through convection, the third for molecular transport and the final two terms, for generation and dissipation, respectively. The whole model information, along with default constants, as used in ANSYS® can be accessed in [7]. They were left unchanged for the purpose of simplicity of calculations.

d) Secondary phase volume fraction

$$\frac{\partial(\phi_p\rho_p)}{\partial t}+\nabla\cdot(\phi_p\rho_p\vec{v})+\nabla\cdot(\alpha_p\rho_p\vec{v}_{dr.p})-\sum(F_{M.qp}-F_{M.pq})=0 \tag{5}$$

e) Species conservation

Chemical species transport is solved for each phase by predicting the local mass fraction of each species, through a solution of convection equation. Diffusion terms are omitted based on solution assumptions.

$$\frac{\partial(\phi_q\rho_q Y_{s.q})}{\partial t}+\nabla\cdot(\phi_q\rho_q\vec{v}_q)-\sum(F_{M.qp}-F_{M.pq})=0 \tag{6}$$

4. Simulated case details

a) System properties

The liquids chosen for the simulations were acetone and benzene, with air as an additional component to fill the rest of the system. Their physicochemical properties were left as default in the Fluent database.

b) Initial conditions

Initially, the tank was filled with a stationary layer of passive liquid of height equal to 45 cm and air under atmospheric pressure. All variables were therefore equal to 0, safe from ω as being equal to 1.

c) Boundary conditions

Walls were assumed to be smooth and associated FLUENT settings were left as default. The inlet was specified as a velocity inlet and outlet as a pressure outlet. For the outlet, default options were chosen, for the inlet two values were specified: hydraulic diameter and turbulence intensity.

While in the case of the 2-D problem, the hydraulic diameter is simply the length of the line selected as inlet, turbulence intensity must be properly calculated. Theoretically, it is defined as the ratio of velocity oscillations to the mean, time-averaged value of velocity [8].

$$I = \frac{v'}{\overline{v}} = \frac{\sqrt{\frac{1}{3}\left(v_x'^2 + v_y'^2 + v_z'^2\right)}}{\overline{v}} \tag{7}$$

In this paper however, it is computed according to formula provided by Basse in [9]:

$$I = 0.227 \cdot \mathrm{Re}^{-0.1} \tag{8}$$

This is a strictly empirical equation, accurate only for clearly turbulent and fully developed flows in hydraulically smooth pipes. In no way should it be taken as measure of turbulence itself, it only describes the magnitude of fluctuations relative to the average velocity.

d) Variables introduced

This study makes use of several parameters that describe the various aspects of mixer performance, which were used for making plots and figures, as well as for general estimation of the process effectiveness.

Pumping coefficient, Π, is the total volume of mixture drawn into the mixing chamber per unit volume of the inlet motive liquid. Based upon a similar variable used by Sundararaj in [5].

$$\Pi = \frac{\sum F_{v,\mathrm{orifices}}}{F_{V,\mathrm{inlet}}} \tag{9}$$

In this formula, $F_{V,\mathrm{inlet}}$ is the volumetric flowrate through the main inlet, while the sum in the numerator denotes the combined volumetric flowrates through side orifices.

Pressure ratio, P^*, relates the pressure drop in the Venturi tube to the pressure drop that would lead to cavitation. In order to secure the process from vaporization of the stream, its value should not approach 1.

$$P^* = \frac{P_{in} - P_{\mathrm{mix}}}{P_{in} - P_v} \tag{10}$$

Where, P_{in} stands for pressure at the inlet, P_{min} being minimal pressure measured in the ejector, and finally, P_V as the saturated vapour pressure of acetone under process conditions. Acetone was chosen as a reference because it is more volatile then benzene. The value of P_V remain constant at 24 571,7 Pa, as indicated by calculator provided on the website of Dortmund Data Bank [10].

Mixing cfficiency, m_{eff}, will be used as a measure of process effectiveness, and the formula according to which it was calculated is presented below:

$$m_{\mathrm{eff}} = \left(1 - \frac{\left|\overline{w}_c - w_{c,\mathrm{ideal}}\right|}{\left|1 - w_{c,\mathrm{ideal}}\right|}\right) \cdot 100\% \tag{11}$$

When its value is 1, it means that mixture is perfectly mixed, while a value of 0 denotes that the component is pure and as a consequence, that means total segregation. As for the symbols in the formula, these are the mass fractions of the reference component – a measured one and one corresponding to the maximal uniformity state, respectively. For benzene as a reference component in acetone-benzene binary system of 0.5 volumetric ratio of benzene, the latter is equal to 0.5252.

Dimensionless length, z, is quite often used in analysis of the elongated objects like pipe or tower reactors, to make results appear more universal and intuitive.

$$z = \frac{x}{L} \tag{12}$$

In the formula for z, x denotes a position within an ejector (counted from its inlet along the central axis), with L being its total length. Therefore, z value ranges between 0 at the inlet and 1 at the outlet.

e) Geometry

Calculations were performed for the type of ejector presented in the schematic below, together with dimensions and corresponding dimensionless length.

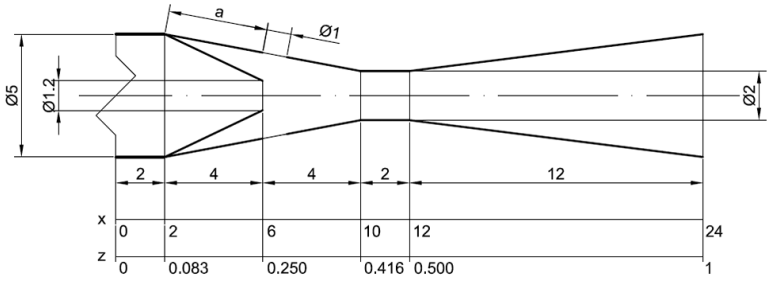


Fig. 3. Detailed ejector schematic, length unit – centimeter, a is a variable

It was used as the insert into the tank of height equal to 120 cm and diameter of 100 cm. An air outlet was also added in order to simplify the model and omit the compressibility. The meshed geometry of the ejector with its immediate surroundings is presented below.

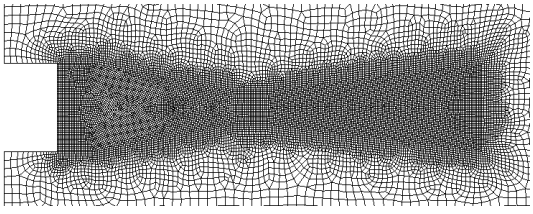


Fig. 4. Ejector mesh

As it may be observed, the mesh inside to the ejector is very fine, as the conditions here are predicted to be very turbulent, contrary to the volume of the tank where mesh was made far coarser because of its greater size.

5. Presentation of the results

Calculations were performed for several chosen values of *a*, as well as inlet velocities of 0.5, 1.0, 1.5 and 2.0 m/s, with acetone as motive liquid and benzene as passive liquid.

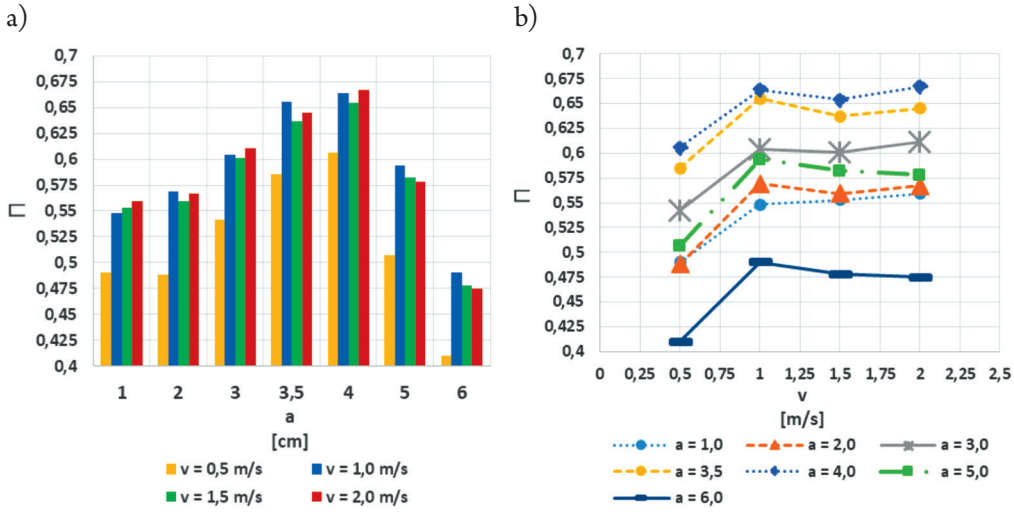


Fig. 5. a) Pumping coefficient vs inlet location; b) Pumping coefficient vs inlet velocity

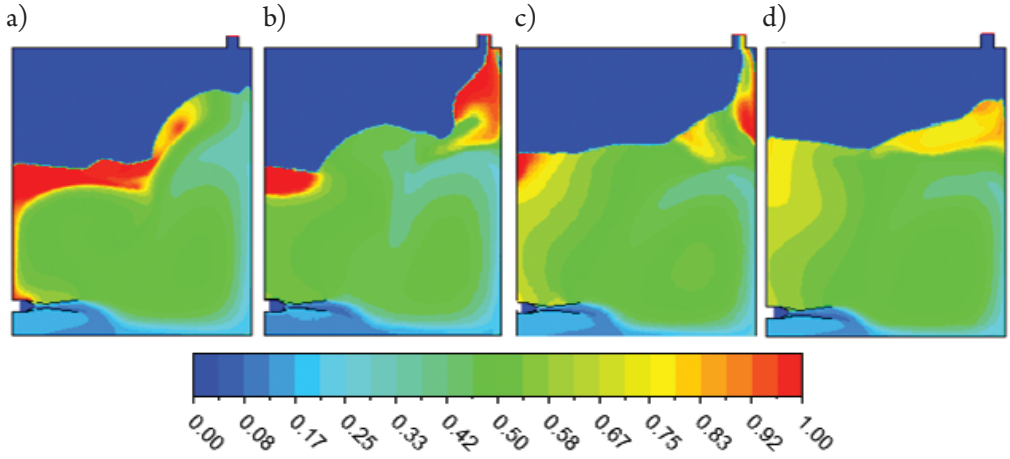


Fig. 6. Contour plots of benzene mass fraction for various inlet velocities: a) 2.0 m/s; b) 1.5 m/s; c) 1.0 m/s; d) 0,5 m/s

As it may be seen, the pumping coefficient is mainly dependent on the geometry and only a minuscule dependence on velocity was determined, visible only when comparing its lowest values with greater ones, for which it stays almost constant with apparently random oscillations.

It is easily visible on the plots above that some numerical artifacts are clearly present, caused perhaps by the simplicity of the flow model and the 2-D geometry, however the rest of the result

seems pretty promising. In all cases, acetone ended up as quite evenly distributed across the volume of the tank and finely mixed with the initial amount of benzene. Better results were obtained for lower injection rates, however it should be remembered that it is only first stage of the whole process, and it is intended to be repeated for several more mixing cycles. In all cases, the worst mixing occurred for the top and bottom layers of the fluid, especially for high injection rates. This was again caused by the limitations of 2-D geometry, as the bottom section of the tank was essentially isolated from the rest of the tank by the inflowing stream of acetone, while the top portions of the liquid phase were pushed up by its accumulation. This would not occur in a real 3-D process, in which some kind of circulation would be expected.

In addition to contour plots, average mixing efficiency was presented for different zones within the tank in the figure below. This was calculated on the basis of data taken from several representative points (three for region below the ejector, three near its outlet and up to nine points above it).

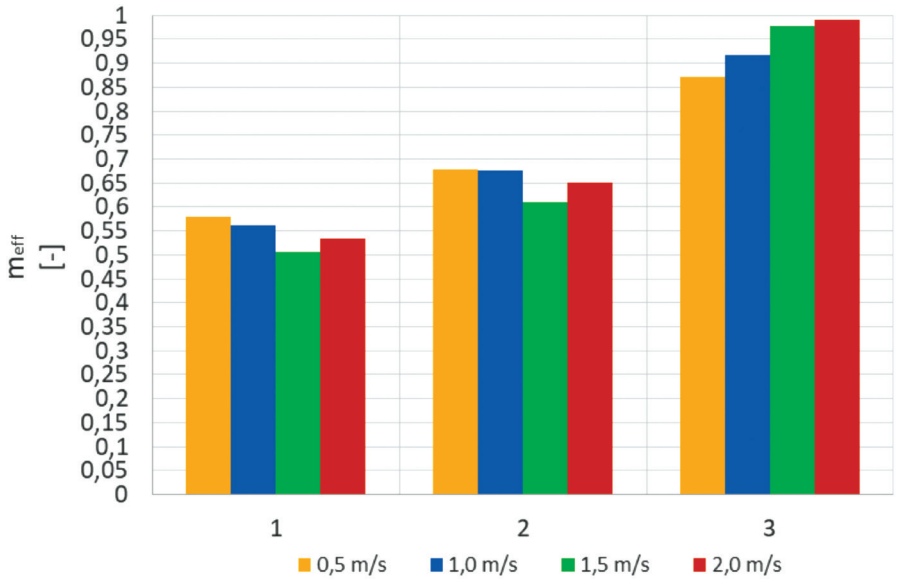


Fig. 7. Mixing efficiency for each zone: 1 – bottom, 2 – middle, 3 – top; a) 2.0 m/s; b) 1.5 m/s; c) 1.0 m/s; d) 0.5 m/s

The results presented in Figure 7 are further proof of what was observed on the contour plot. Liquids ended up as quite thoroughly mixed in the core of the tank, with lesser degree of mixing near the ejector and sections isolated by its presence. Introducing recirculation seems a reasonable and simple option and perhaps it will be a part of the next research case. In this paper however, only this first stage of the process was analysed.

The next results presented in the Figures 8 and 9, show the computed gauge pressure at the centreline of the ejector, with gauge pressure being essentially the deviation from the reference value of 1 Atm. Pressure losses and their magnitude are understood as crucial in operating such a device.

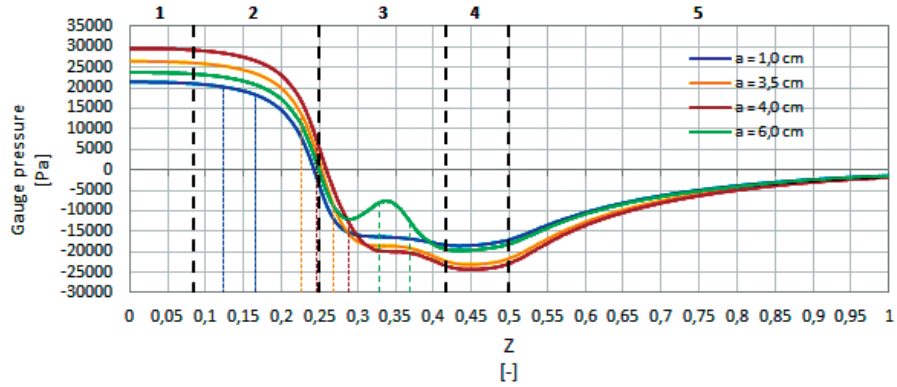


Fig. 8. Gauge pressure profiles along the ejector centreline for inlet velocity of 2.0 m/s and various locations of orifices (shown with colored dashed lines); thick dashed lines divide the plot according to Venturi geometry; 1 – inlet section, 2 – reducer section, 3 – mixing zone, 4 – throat section, 5 – nozzle section

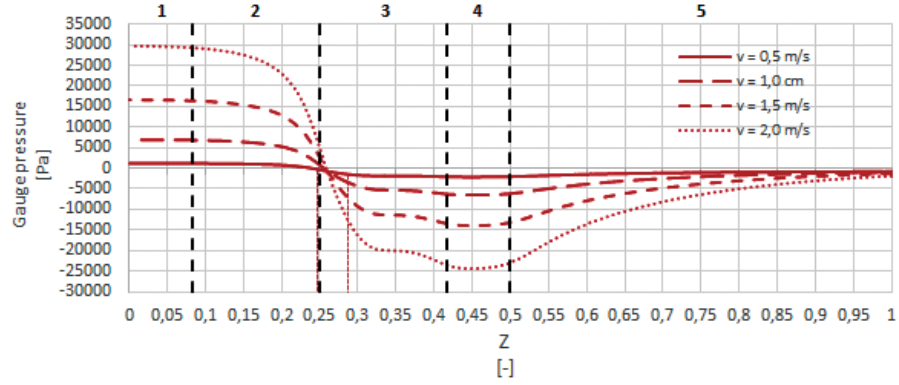


Fig. 9. Gauge pressure profiles along the ejector centreline for dimension a = 4.0 cm and various inlet velocities; thick dashed lines divide the plot according to Venturi geometry; 1 – inlet section, 2 – reducer section, 3 – mixing zone, 4 – throat section, 5 – nozzle section

It can be observed that greatest pressure drop in the analysed case, occurred not in the mixing zone nor reducer, but in the throat section of the ejector. This is fine, as it encourages the flow in the longitudinal direction of the device, while pressure drop in the mixing chamber is still reasonable. One thing to note is that locating orifices further from the reducer leads to a non-optimal profile, as indicated by the green peak on the Figure 8. The optimal location of orifices that contributes to greatest pressure drop seems to be just at the same coordinate as the reducer's outlet. Together with the analysis of the pumping coefficient results, this proves that its magnitude is indeed dependent on the pressure drop. While analysing Figure 9, it is clearly visible that increasing velocity contributes greatly to pressure drop, thus this should be one of the main concerns when designing or optimizing a device of such a kind. A proper compromise has to be achieved between the financial expenses, process time and mixing efficiency and finally, safety from cavitation.

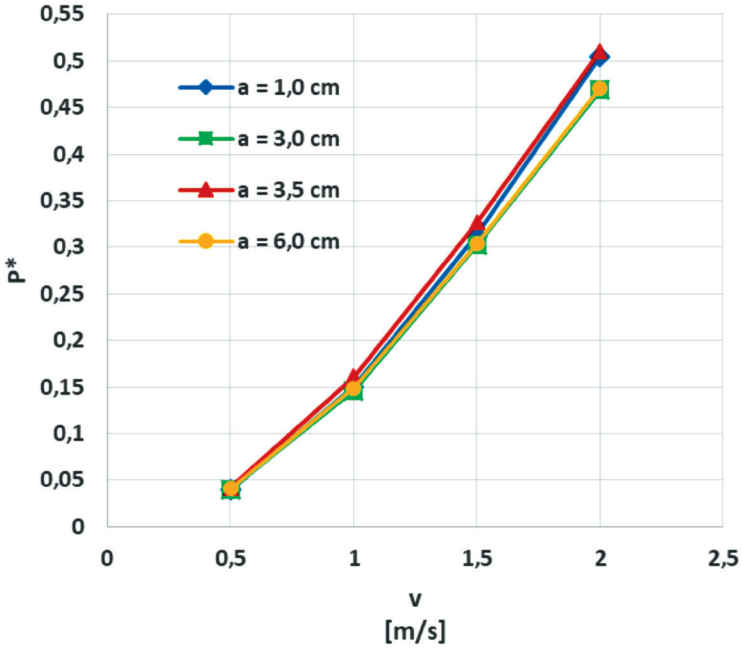


Fig. 10. Pressure ratio for various orifices locations and velocities

As indicated by Figure 10 and previous conclusions, the orifice location has only slight effect on the pressure ratio, which is dependent only on the velocity. Cavitation in the analysed process conditions would not occur, as its value is far from 1 for all cases of orifice location and inlet velocity. It must be noted however, that if greater increase in inlet velocity were needed or demanded, its limiting value must be properly established for every kind of mixed substances and process conditions.

6. Summary

As determined before, the calculations performed indicate that a Venturi mixer can potentially be very a efficient, yet simple way of relatively quick and proportional mixing of miscible liquids, like various organic solvents. It can be a decent alternative for traditional tank mixers with propellers or agitators. Of course, it must remembered that because of the limitations of the 2-D model, the results obtained should be taken qualitatively, not quantitatively. Further research is recommended, especially for fully fledged 3-D cases with introduced recirculation and more accurate and realistic assumptions. A stricter optimization of the ejector geometry is recommended, as well as evaluation of the best process conditions in terms of efficiency and expenses. With such questions in mind, perhaps ones for future consideration, this paper is hereby concluded.

References

[1] Static mixers: efficient mixing with no moving parts, http://www.entec-international.com/static-mixer.html (access: 10.05.2017).

[2] Eductors & syphons: Tank mixing eductors, https://www.s-k.com/eductors-syphons/tank-mixing-eductors.cfm (access: 10.05.2017).

[3] A Guide to Optimizing In-Tank Agitation and Mixing Using Eductors, www.spray.com/literature.../B635_Tank_Mixing_Eductors.pdf (access: 10.05.2017).

[4] Sundararaj S., Selladurai V., *The effects of arbitrary injection angle and flow conditions on Venturi-jet mixer*, Thermal Science, 16.1, 2012, 207-221.

[5] Sundararaj S., Selladurai V., *Flow and Mixing Pattern of Transverse Turbulent Jet in Venturi-Jet Mixer*, Arabian Journal for Science and Engineering, 38.12, 2013, 3563.

[6] Sundararaj S., Selladurai V., *Numerical and Experimental Study on Jet Trajectories and Mixing Behavior of Venturi-Jet Mixer*, Journal of Fluids Engineering, 132.10, 2010, 101104-1-9.

[7] ANSYS Workbench documentation, Release 17.2, Help System, Fluent, Theory Guide.

[8] Munson B.R., Young D.F., Okiishi T.H., *Fundamentals of Fluid Mechanics*, John Wiley & Sons, 1998.

[9] Basse N., *Turbulence intensity and the Friction Factor for Smooth- and Rough-Wall Pipe Flow*, Fluids, 2, 2017, 30.

[10] Saturated Vapour Pressure Calculator, http://ddbonline.ddbst.de/AntoineCalculation/AntoineCalculationCGI.exe (access: 1.05.2017).

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ENVIRONMENTAL ENGINERING
DOI: 10.4467/2353737XCT.17.195.7424

Anna Hajduk (anna.hajduk@uwm.edu.pl)
Marcin Dębowski
Marcin Zieliński
Paulina Rusanowska
Department of Environment Engineering, Faculty of Environmental Sciences, University of Warmia and Mazury in Olsztyn

The utilisation of zeolites for the reduction of ammonium concentration in biogas processes

Wykorzystanie zeolitów do ograniczenia stężenia azotu amonowego podczas produkcji biogazu

Abstract

The aim of this study was to determine the impact of the utilisation of zeolites in the removal of ammonium from systems of biological the transformation of organic substrates to biogas during methane fermentation. The results showed that the highest efficiency of removal of ammonium from digestate was achieved with a 10 g/dm^3 dose of zeolite. This efficiency was from 37.87 ± 0.54% to 46.01 ± 0.8%. The experiment demonstrated a linear relationship between the dose of zeolite in the range from 1 g/dm^3 to 10 g/dm^3 and the efficiency of the sorption and the final concentration of N-NH_4. The highest sorption of N-NH_4 was observed in a variant with the zeolite dose of 1 g/dm^3 and it was from 161.74 ± 2.01 mg N-NH_4/g to 132 ± 4.7 mg N-NH_4/g in the digestate and from 112 ± 8.54 mg N-NH_4/g to 122 ± 12.90 mg N-NH_4/g in the effluent from digestate.

Keywords: methane fermentation, nitrogen ammonium, zeolites, digestate, effluent from digestate

Streszczenie

Celem zrealizowanych prac badawczych było określenie wpływu zastosowania zeolitów w procesie usuwania azotu amonowego z systemów biologicznej transformacji substratów organicznych do biogazu w procesie fermentacji metanowej. Udowodniono, że najwyższą efektywność usuwania amoniaku z osadu beztlenowego uzyskano przy zastosowaniu dawki zeolitu wynoszącej 10 g/dm^3. Zawierała się ona w zakresie od 37,87 ± 0,54% do 46,01 ± 0,8%. Eksperyment udowodnił liniową zależność między dawką zeolitu w zakresie od 1 g/dm^3 do 10 g/dm^3 a efektywnością sorpcji N-NH_4 oraz stężeniem końcowym tego parametru. Największa jednostkowa ilość N-NH_4 sorbowana była w wariancie, w którym zastosowana dawka zeolitu wynosiła 1 g/dm^3. W odniesieniu do osadu pofermentacyjnego mieściła się w zakresie od 161,74 ± 2,01 mgN-NH_4/g do 132 ± 4,7 mgN-NH_4/g, natomiast w przypadku odcieku mieściła się w przedziale 112 ± 8,54 mgN-NH_4/g do 122 ± 12,90 mgN-NH_4/g.

Słowa kluczowe: fermentacja metanowa, azot amonowy, zeolity, osad fermentacyjny, odciek z osadu fermentacyjnego

1. Introduction

Methane fermentation is a biochemical process during which bacteria decompose organic matter and produce biogas [1]. This process can be disturbed or even hampered due to a variety of reasons. These include operator errors, technical problems, and the presence of inhibitors in substrates and biogas. The inhibitors, even in a low concentration, cause the incorrect operation of a biogas plant. The inhibitors can be delivered to a digester with a substrate or can be formed during the anaerobic fermentation inside a digester [2-4].

Ammonium is a nutrient for anaerobic bacteria, but its high concentrations inhibit fermentation. In digesters, ammonium occurs during the fermentation of nitrogen compounds such as urea and proteins [5]. Two forms of ammonium nitrogen can be distinguish. The first is the ionic form, ammonium ion NH_4^+, which is less toxic than the other form, ammonia NH_3. During fermentation of livestock slurry and manure, considerable amounts of ammonium are produced – this adversely affects the development and activity of methanogenic microorganisms. Therefore, the presence of ammonium in the digester reduces the efficiency of methane production. It has been proven that ammonium concentrations in the range from 1.7 to 14 gN-NH_4/L reduces methane production by up to 50% [5]. The concentration of ammonium is influenced by the pH and the temperature of methane fermentation. Higher temperatures of the process accelerate metabolism and thus produce higher concentrations of ammonium.

Thus far, the removal of ammonium from digesters has been based on physicochemical and biological processes that do not always give satisfactory results; therefore, extensive studies on the use of natural zeolites (mainly clinoptilolite) for this purpose have been initiated. The removal of ammonium through the use of zeolite takes place by ion exchange or by adsorption in the pores of the aluminosilicate skeleton [6]. The introduction of zeolites into the fermentation chambers reduces the content of ammonium – this significantly limits the negative effect of this factor on anaerobic bacteria and directly affects the efficiency of biogas production. Such minerals placed in the fermentation chamber also bind other cations, including heavy metals, and enable efficient hydrogen sulphide binding, which directly promotes higher levels of biogas purity and higher methane content [1]. It seems that the efficiency of methane fermentation can be improved by using a suitable number of zeolites in the digesters.

One of the basic directions of development and use of digestate is its agricultural or natural use. Zeolites in digestate act as fertiliser carriers and after introduction into the soil, they enable the systematic and slow release of ammonium nitrogen [7]. This is a very desirable and important agrotechnical and ecological factor that protects the environment and crops from uncontrolled nutrient losses due to surface run-off and erosion. These phenomena have a negative impact both on crop yields and the accelerated eutrophication of surface water bodies.

This article presents the effect of a sorbent on the content of ammonium nitrogen, not only in the sludge supernatant (effluent from digestate) but also in the digestate. So complete studies are hard to find in other publications. It should be emphasised that although the mechanism of bonding of ammonium by zeolites is already known and has been repeatedly

described, the use of such technology in biogas production systems is pioneering. According to the authors knowledge, this leads to a reduction in the concentration of ammonium, heavy metals and hydrogen sulphide in the fermentation chambers, which increases the efficiency of biogas production and improves its quality. The utilisation of zeolites also contributes to the development of fertilisers with modified properties that are justified both in agrotechnical and ecological terms. All the above aspects of zeolite application in biogas production systems will be developed in subsequent experimental work.

2. Materials and methods

2.1. Organisation of the study

In the experiment, the fermentation chambers were filled with sludge and plant biomass and different doses of zeolite. The study was divided into three series, depending of theplant biomass. Three plants were tested, in the 1 series – *Helianthus tuberosus*, in the 2 series – *Helianthus tuberosus*, in the 3 series – *Miscanthusgiganteus*. In each series the different doses of zeolite were used. According to dose of zeolite, series were divided into variants. The following doses of zeolite were used: 0 g/L (1 variant), 1 g/L (2 variant), 5 g/L (3 variant), 7.5 g/L (4 variant), 10g /L (5 variant), 12.5 g/L (6 variant). The studies focus on changes of ammonium concentration in digestate and effluent from digestate.

2.2. Materials

Fresh biomass was tested in the experiments. The biomass was obtained in November (week 46 of the year) from farmlands in Bałdy (53° 36′ 026″ N, 20° 36′ 625″ E). The biomass was crushed to a 3 mm fraction in a knife mill. The process was carried out at a rotational speed of 2,800 rpm for 5 minutes. The specifications of the tested plant biomass are presented in Table 1.

Table 1. The specifications of the plant biomass

Parameter	Unit	Helianthus tuberosus		Helianthus annuus L		Miscanthusgiganteus	
		Amount	Standard deviation	Amount	Standard deviation	Amount	Standard deviation
dry mass (d.m.)	[g d.m./g]	0.4109	0.0021	0.3512	0.0029	0.3594	0.0012
mineral mass	[g/g d.m.]	0.0910	0.0021	0.0445	0.0022	0.0313	0.0018
organic dry mass (o.d.m.)	[g/g d.m.]	0.9090	0.0021	0.9555	0.0022	0.9687	0.0023
total nitrogen	[g/g d.m.]	0.643	0.09	0.661	0.11	0.364	0.04
ammonium	[g/g d.m.]	0.482	0.027	0.351	0.022	0.273	0.015

The inoculum was anaerobic sludge obtained from digesters in a biogas plant in Łęguty (53° 45' 16" N, 20° 09' 52" E) with corn silage and pig manure applied as substrates (OLR – 3.2 kg o.d.m./m^3 x d, HRT – 40 d, temp – 42°C). The specifications of the anaerobic sludge are presented in Table 2. The particle size of the clinoptilolite used in the experiments was 3-5 mm (*Andalusia Sp z o.o.*). The specifications of the zeolite used in the study are presented in Table 3.

Table 2. The specifications of the anaerobic sludge

Parameter	Unit	Amount	Standard deviation
dry mass (d.m.)	[g d.m./g]	0.044	
mineral mass	[g/g d.m.]	0.386	0.00012
organic dry mass (o.d.m.)	[g/g d.m.]	0.614	0.0004
total nitrogen	[g/g d.m.]	0.387	0.078
ammonium	[g/g d.m.]	0.365	0.037

Table 3. The specifications of zeolite (provided by the manufacturer)

Chemical composition		
Parameter	**Unit**	**Amount**
Na_2O	[%]	2.468
Al_2O_3	[%]	14.431
SiO_2	[%]	74.620
CaO	[%]	2.186
K_2O	[%]	2.870
Fe_2O_3	[%]	2.030
MgO	[%]	0.850
others	[%]	0.545
Mineralogical composition		
Parameter	Unit	Amount
clinoptilolite	[%]	87
plagioclase	[%]	3
clay	[%]	3-4
crystobalite	[%]	6
rutile	[%]	0.1-0.2

2.3. Experimental setup

The experiments were performed using the manometric OxiTop Control system. The manometric device consists of a 0.5 L glass bottle provided with and a pressure transducer located in a measuring head. In each of the variants of the experiment, 0.1 L of anaerobic sludge with the prepared lignocellulosic biomass and a dose of zeolite was introduced to the glass bottles. In all variants, initial organic loading was established at a level of 5.0 g o.d.m./L. In order to ensure anaerobic conditions within glass bottles, nitrogen was purged to replace the air present inside. A complete set of measurement was placed in a thermostatic cabinet with hysteresis that did not exceed ±0.5°C. Measurements were carried out at 36°C. The duration of the experiment was 20 days.

2.4. Analytical and statistical methods

Analyses of the content of ammonium in the digestate and effluent from digestate were conducted. The effluent from the digestate was obtained through centrifugation. The concentration of the ammonium was determined by direct distillation (PN-EN 14671:2007). The dry organic matter and mineral matter in the sludge and biomass was determined with Polish Norm PN:EN 12880:2004. The total nitrogen content was determined through use of the Kiejdahl method. The measurements were performed after 20 days of methane fermentation.

The hypothesis on the distribution of each analysed variable was verified using the Shapiro-Wilk W test. The significance of the differences between the variables was stated with the use of a one-way analysis of variance (ANOVA). The homogeneity of variance in the groups was established using the Levene's test. The significance of differences between the analysed variables was determined with the RIR Tukey test. Differences were found to be significant at $\alpha = 0.05$.

3. Results and discussion

The final $N\text{-}NH_4$ concentrations in the digestate (6 variants) were 743 ± 11.13 mg/g in Series 1, 823 ± 28.13 mg/g in Series 2 and 721.33 ± 16.54 mg/g in the Series 3 (Fig. 1). The observed differences were statistically significant ($p \leq 0.05$). The comparative efficiency of the removal of $N\text{-}NH_4$ was noted in the effluent from digestate and depending on the series, the final concentrations of $N\text{-}NH_4$ ranged from 674.8 ± 20.89 mg/g to 809 ± 21.54 mg/g (Fig. 2). Results of the experiments show a linear relationship between the dose of zeolite in the range from 1 g/dm^3 to 10 g/dm^3 and the final concentration of $N\text{-}NH_4$ (Figs. 1-2). Further increases in the amount of zeolite introduced into the digesters no longer had a statistically significant ($p \leq 0.05$) effect on the removal efficiency of $N\text{-}NH_4$ from the digestate and effluent from digestat. The efficiency of $N\text{-}NH_4$ removal increased with increases in the doses of zeolite in both digestate and effluent from digestate. In digestate, the lowest $N\text{-}NH_4$ sorption efficiency

was 9.39 ± 0.06% and it was noted in the first variant of Series 1. The highest N-NH_4 sorption efficiency was 48.66 ± 0.19% and it was noted in the sixth variant of Series 3 (Fig. 3). The removal efficiency of N-NH_4 from effluent from digestate ranged from 8.69 ± 0.39% in the second variant of Series 2 to 49.8 ± 1.41% in the sixth variant of Series 3 (Fig. 4).

It was noted that as the zeolite dose increased, the efficiency of the sorbent capacity reduced. The largest sorption of N-NH_4 was in the second variant irrespective of the series (Figs. 5-6). Regarding the digestate, it ranged from 161.74 ± 2.01 mg N-NH_4/g to 132 ± 4.7 mg N-NH_4/g (Fig. 5). The observed differences were statistically significant ($p \leq 0.05$). In the case of effluent from digestate, the largest sorption of N-NH_4 ranged from 112 ± 8.54 mg N-NH_4/g to 122 ± 12.90 mg N-NH_4/g (Fig. 6). The observed differences were statistically significant ($p \leq 0.05$).

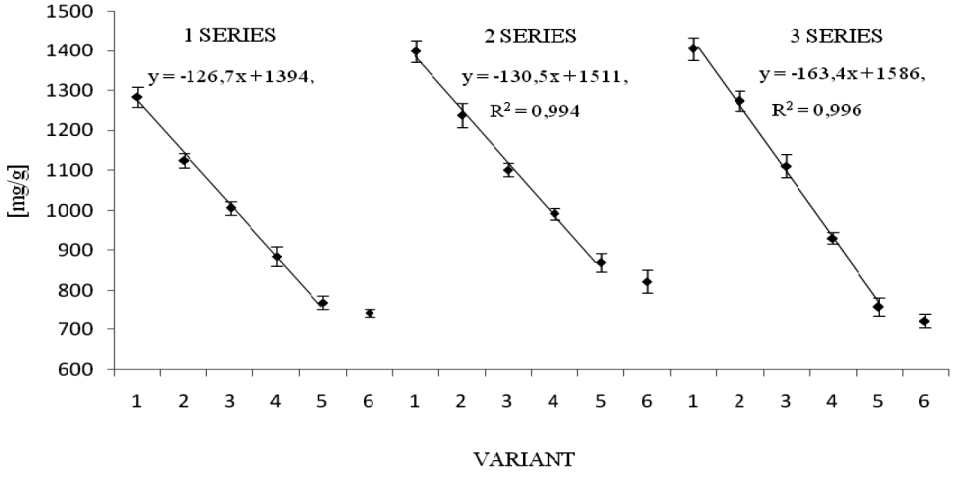


Fig. 1. The concentration of N-NH4 in digestate

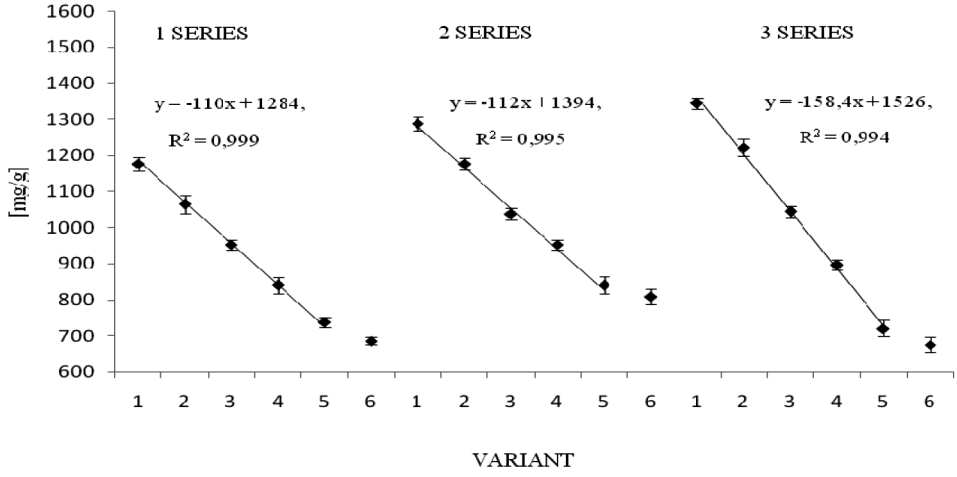


Fig. 2. The concentration of N-NH4 in effluent from digestate

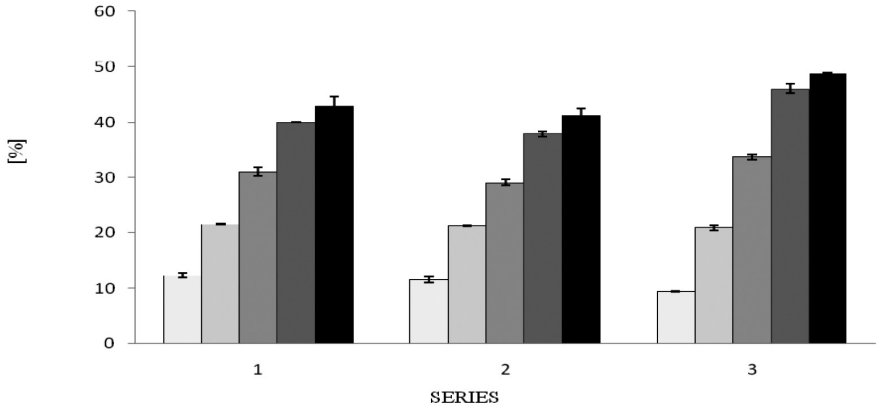


Fig. 3. The efficiency of N-NH4 removal in digestate

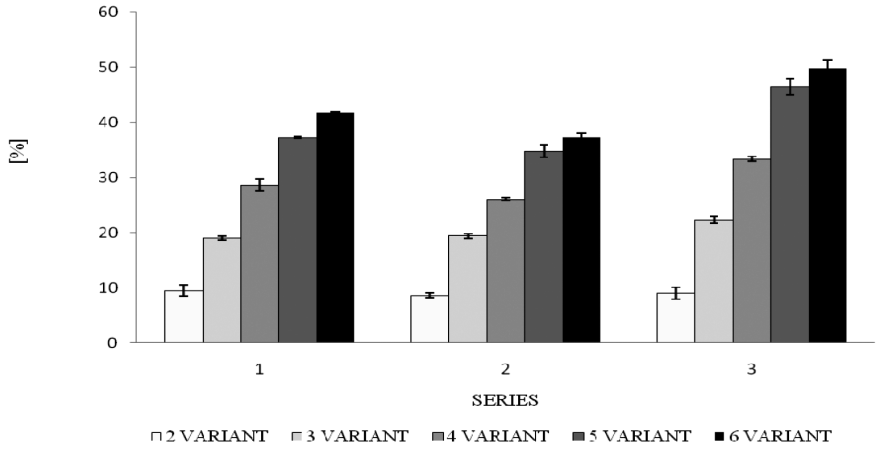


Fig. 4. The efficiency of N-NH4 removal in effluent from digestate

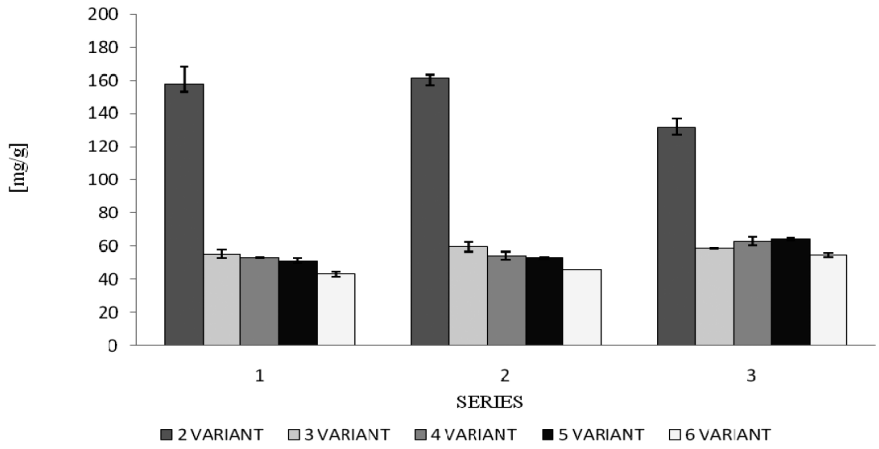


Fig. 5. The sorption of N-NH4 from digestate per gram of sorbent

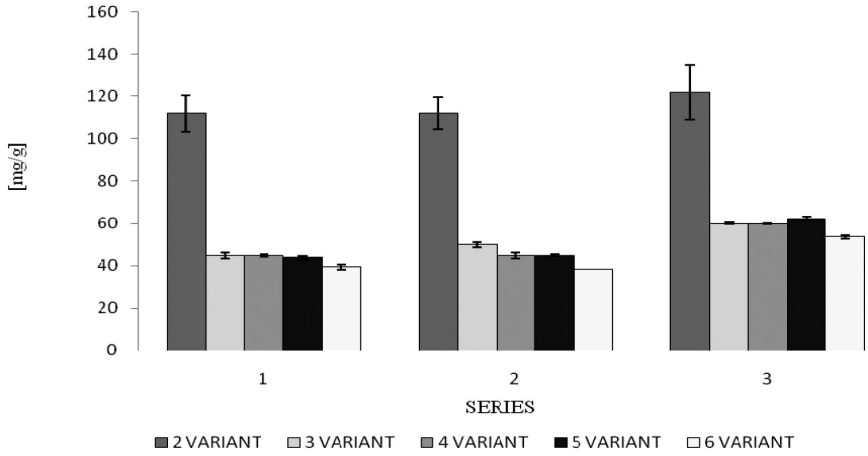


Fig. 6. The sorption of N-NH4 from digestate per gram of sorbent

Wiśniowska et al. (2015) performed a study regarding the use of zeolites for the removal of N-NH_4 on a laboratory scale. The authors used a condensate from the high temperature drying of sewage sludge and sediment liquids mixed with synthetic municipal wastewater. As sorbents, the authors used horticultural zeolite and bentonite (in the form of commercially available cat litter). The authors proved that zeolite was more effective for ammonium removal. The efficiency of ammonium removal was 7.80 mg N-NH_4/g of sorbent and it was of about 11% higher than the efficiency of ammonium removal with bentonite. These removal results are in the range obtained by other authors for natural zeolites, which were 0.4-25.5 mg N-NH_4/g of sorbent [8-10]. In the present study, the obtained ammonium removal efficiency was much higher.

The digestate obtained after the anaerobic digestion of organic wastes has a nutrient content which is favourable for use as a crop fertiliser. The most cost-effective method for soil fertilising is direct application on the land. However, digestate is not easy to handle due to the fact that land fertilisation in appropriate periods needs a large storage capacity for digestate. Storage of digestate is expensive and the transportation of large volumes is ineffective from a logistical and economic point of view. However, the utilisation of zeolite material as a fertiliser requires the consideration of certain aspects. A total of 500 kg of clinoptilolite would be needed in a 50 m^3 fermentation chamber to remove around 50% of the ammonium from digestate. In order to make the treatment process economically attractive, a greater increase in the concentration or a reduction in weight would be required; however, economy depends on both the cost of the ion exchanger/adsorbent and the transportation as well as the fertiliser properties regarding the nutrient bound to clinoptilolite, which may be more stable and easier to manage with regard to storage, distribution and application to soil [11].

Further removal of remaining ammonium might be removed in second-stage. Spreczyńska (2016) investigate removal N-NH_4 efficiency in a zeolite filled columns. The author removed N-NH_4 from reject water in columns with internal diameter of 3.4 cm filled with 436.7 g of zeolite. In the experiment, the author first identified the possibility of using natural zeolite to

remove of ammonium from solutions with a filtering speed of 5.0 m/h. Than evaluated the effect of filtration velocity on the removal efficiency of ammonium from reject water. The experiments conducted by the author showed that the adsorption/ion exchange process on a zeolite-filled bed is an effective method of removing ammonium from aqueous solutions. The ammonium removal efficiency from model solutions, containing only ammonium chloride, was 80-90%. This is in line with the literature. The ammonium removal efficiency from sludge supernatant was lower than in model solutions. The efficiency of ammonium removal dependent on the velocity of filtration through the zeolite bed. The efficiency of ammonium removal was 20-52% with filtration velocity 12.72 m/h. Threefold reduction of the filtration velocity lowered the efficiency of ammonium removal to 30-70%. The lowering of the efficiency of ammonium removal was caused not only by filtration velocity but also composition of sludge supernatant (concentration of organics and fine suspension) [7].

4. Conclusions

The efficiency of N-NH_4 removal increased linearly with increasing doses of zeolite (up to a dose of 10 g/dm^3) in the range from 9.39 ± 0.09% to 48.66 ± 0.19% in digestate and from 9.09 ± 1.10% to 49.8 ± 1.41% in effluent from digestate. Increasing the dose above 10 g/dm^3 did not improve the effect of ammonium removal. The highest sorption capacity was observed in the variant with a zeolite dose of 1 g/dm^3. In this variant, the sorption was 161.74 ± 2.01 mg/g in the digestate and 122 ± 12.90 mg/g in the effluent from digestate. There was no significant effect of the used plant biomass on the N-NH_4 removal efficiency from the digestate and the effluent from the digestate.

References

[1] Pilarska A., Pilarski K., Krysztofiak A., Dach J., Witaszek K., *Impact of additives on biogas efficiency of sewage sludge*, Agricultural Eng. Vol. 3 (151), 2014, 139-148.

[2] Dinamarca S., Aroca G., Chamy R., Guerrero L., *The influence of pH in the hydrolytic stage of anaerobic digestion of the organic fraction of urban solid waste*, Water Sci. Technol. Vol. 48, 2003, 249-254.

[3] Ahring B.K., *Perspectives for anaerobic digestion*, Adv. Biochem. Eng. Biotechnol. Vol. 81, 2003, 1-30.

[4] Vavilin V.A., Fernandez B., Palatsi J., Flotats X., *Hydrolysis kinetics in anaerobic degradation of particulate organic material: an overview*, Waste Manage. Vol. 28, 2008, 939-951.

[5] Chen Y., Cheng J.J., Creamer K.S., *Inhibition of anaerobic digestion process: a review*, Bioresour. Technol. Vol. 99, 2008, 4044-4064.

[6] Baykal B.B., Guven D.A., *Performance of clinoptilolite alone and in combination with sand filters for the removal of ammonia peaks from domestic wastewater*, Water Sci Technol. Vol. 35(7), 1997, 47-54.

[7] Spreczyńska E., *Wykorzystanie zeolitu do usuwania jonów azotu amonowego z cieczy osadowej*, Engineering and Protection of Environment Vol. 19(3), 2016, 391-399 (in Polish).
[8] Wiśniowska E., Karwowska B., Sperczyńska E., *Interwencyjne wykorzystanie zeolitów w oczyszczaniu ścieków*, Zeszyty Naukowe Inżynieria Środowiska Vol. 160(40), 2015, 56-63 (in Polish).
[9] Gupta V.K., Sadegh H., Yari M., Shahryari Ghoshekandi R., Maazinejad B., Chahardori M., *Removal of ammonium ions from wastewater a short review in development of efficient methods*, Global J. Environ. Sci. Manage. Vol. 1(2), 2015, 149-158.
[10] El-Shafey O., Fathy N.A., El-Nabarawy T., *Sorption of ammonium ions onto natural and modified Egyptian kaolinites: kintic and equilibrium studies*, Adv. Phys. Chem. Vol. 2014, 2014, 12.
[11] Kocatürk-Schumacher N.P., Bruun S., Zwart K., Jensen L.S., *Nutrient recovery from the liquid fraction of digestate by clinoptilolite*, Clean Soil Air Water Vol. 45(6), 2017, DOI: 10.1002/clen.201500153.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ENVIRONMENTAL ENGINERING
DOI: 10.4467/2353737XCT.17.196.7425

Zygmunt Kowalski
Mineral and Energy Economy Research Institute, Polish Academy of Sciences
Agnieszka Makara (amak@chemia.pk.edu.pl)
Katarzyna Fela
Institute of Chemistry and Inorganic Technology, Faculty of Engineering and Chemical Technology, Cracow University of Technology
Agnieszka Generowicz
Institute of Water Supply and Environmental Protection, Faculty of Environmental Engineering, Cracow University of Technology

The concentration of animal blood plasma using membrane methods that allow its recycling and reuse

Zatężanie plazmy krwi zwierzęcej metodami membranowymi umożliwijace jej recyklig i wtórne zużycie

Abstract

Animal blood plasma contains ~91% water and requires concentration prior to it been dried. Our studies concerned the suitability of membrane technology. The efficiency of the investigated filtration process for membranes was stable over time, and displayed no tendency to clog. The protein level in obtained filtrates ranged from 0.15-0.26%. Increased content of protein in the concentrate is a function of the degree of plasma concentration. For the raw material containing 6.05% proteins, with 2.3 times the concentration, it was 13.84% and with 3 times the concentration, it was 18.24%. The filtration efficiency increases with the temperature of the process but decreases with increased levels of concentration. For both of the investigated membranes the permeability was similar; however, the 0.07 μm membrane had higher filtration efficiency by an average of 30%.

Keywords: blood plasma, membrane techniques, ultrafiltration, concentration

Streszczenie

Plazma krwi zwierzęcej zawiera ok. 91% wody, dlatego wymaga zatężenia przed właściwym suszeniem. Przedstawione wyniki badań wykazały przydatność technik membranowych. Zawartość białka we wszystkich otrzymanych filtratach wynosiła od 0,15 do 0,26%, niezależnie od jego stężenia w plazmie. Stężenie białka w koncentracie wzrasta wraz ze stopniem zatężania. Dla surowca zawierającego 6,05% białka, przy 2,3-krotnym zatężeniu było równe 13,84%, zaś 3-krotnym 18,24%. Wydajność procesu filtracji membranowej jest stabilna w czasie. Membrany nie wykazują tendencji do zatkania się. Wydajność filtracji wzrasta z temperaturą i maleje wraz ze wzrostem stopnia zatężania. Dla obu użytych membran przepuszczalność osocza jest porównywalna, jednak membrana 0,07 μm ze względu na wyższą skuteczność filtracji (30%) jest preferowana do stosowania w skali przemysłowej.

Słowa kluczowe: plazma krwi, techniki membranowe, ultrafiltracja, wstępna obróbka plazmy

1. Introduction

One of the by-products of animal slaughter is blood and due to its composition and properties, it is interesting material which can be processed into food ingredients or feed [1-3]. For this purpose, only blood which has been subjected to veterinary testing to ensure that it is free from viral diseases can be used; this applies whether the blood is bovine or porcine. The ban on the use of meat and bone meal, which was introduced after the European countries due to BSE over 10 years ago, does not apply to feed from the processing of blood, such as blood meal, haemoglobin feed or dried blood plasma. Feed containing blood meal is not only a rich source of protein, but also promoter's resistance suitable in the feeding of monogastric animals [4-6].

From a physicochemical point of view, blood is slurry consisting essentially of the cellular components and serum of plasma, wherein the cells are suspended. The liquid part of blood is plasma; it is comprised of around 91% water, 8% organic compounds and 1% inorganic compounds which are responsible for the maintenance of acidic-alkaline balance of the animal's body. The organic compounds are mainly proteins and lipoproteins, fatty acids, cholesterol, triglycerides, hormones, glucose, vitamins, the products of protein metabolism (urea, amino acids) and haem (bilirubin and urobilinogen). Inorganic compounds include carbon dioxide and the salts of sodium, calcium, magnesium, potassium and anions containing chlorine, carbon and phosphorus [7, 8].

Blood is a product commonly obtained in an animal slaughterhouse which shows high impermanence, ease undergo haemolysis and provides an excellent base for the development of pathogenic bacteria. Therefore it requires rapid treatment or disposal. One of the main methods of processing is the separation of the blood into plasma and haemoglobin [9]. Due to differences in the specific gravities of blood cells and plasma, it is possible to perform their separation using centrifugation. Plasma obtained through centrifugation typically contains approx. 8-9% of dry mass. Unfortunately, the opportunity to use the thus obtained raw material in liquid form is minimal due to the rapidly progressing biochemical processes which result in the loss of nutrional properties of any food product that it is used to produce. Therefore, the plasma is further processed in order to obtain a product in which both the stability and the properties are improved. There are two main ways of processing, the first is drying which produces powdered plasma, and the second is freezing which, depending on the specific method used, produces either plasma blocks or flakes. Both dried plasma and frozen plasma are products that can be stored for use over a much longer time than liquid plasma [10, 11].

Dried blood plasma has high protein content, depending on the technological drying process, its content ranges from 70 to 80%. Within the major protein fraction of the blood plasma, the following components can be identified: albumin (approx. 50%); α-globulins (15%); β-globulins (15%); γ-globulin, which includes the most valuable immune globulin (15%) – this naturally stimulates the functioning of the immune system of animals. The obtained product, however, had lower concentration of sulphur amino acids, which deficiency should be completed during the preparation of feeds. This product is also rich in minerals, and the macro- and micronutrients which are necessary for the proper functioning of animal organisms [12, 13].

In the production processes of dried blood plasma, water plays important role. The higher the amount of water in the centrifuged plasma, the greater is the costs which must be incurred for its further processing. Thus, an initial plasma concentration is used to minimise the expense and time of drying or freezing. At present, there are two leading techniques used for the concentration of plasma – membrane processes and concentration by evaporation [8, 10, 11].

The paper presents research results on the suitability of membrane technology for the concentration of animal blood plasma and the production of plasma solutions with higher contents of protein and dry matter. The obtained results showed that this method can be successfully used for this purpose.

2. Experimental procedure

Membrane processes [8, 14, 15] exploit the properties of the materials which the membrane is made of to retain certain components of the mixtures during the separation process. The driving force of mass transfer across the membrane is the potential difference which occurs on each side of the membrane. It may be caused by the concentration difference, temperature or pressure. Membrane techniques which can be of use in the concentration of plasma include ultrafiltration, nanofiltration and reverse osmosis. The parameter differs these techniques is selectivity. Diagrams showing the differences between the different techniques of the treatment of animal plasma are presented below in Figure 1.

Concentration of the plasma solution using evaporators is a technique utilising differences in the boiling points of the two liquid components of the solution which causes the solvent (in the case of the plasma, it is water) to be removed through the process of evaporation during heating of the concentrate (protein solution). Evaporators used in the process of plasma concentration usually operate in a vacuum in order to give the plasma a boiling point of 36°C – this keeps the protein denaturation in the concentrated solution. By using the evaporating apparatus it is possible to obtain concentrated plasma with a dry mass content of 25-27%. Two types of evaporation devices are mainly used for the concentration this type of product. The first is the Centritherm type of evaporator, which consists of a number of heated rotating cones, these evaporate the dosed liquid within. In this solution, the plasma goes to the evaporator through a pipe system and is distributed over the heated surface of the cone through nozzles (one for each cone). The centrifugal force spreads the product over the entire heated surface in a very thin layer (approx. 0.1 mm). These cones rotate at a speed of 600 rpm and the plasma is transferred from the centre to the edge in about one second. The concentrated product is collected on the outer edge of the cone and leaves the apparatus. The resulting steam is transferred to the centre of the cone and then through a system of outlet tubes, it is removed to an external condenser. The heated steam is passed to the interior of each of the cones by a plunge pin. The advantage of using Centritherm evaporators is a very short time of contact with the hot surface of the apparatus – this prevents excessive denaturation of the protein [11].

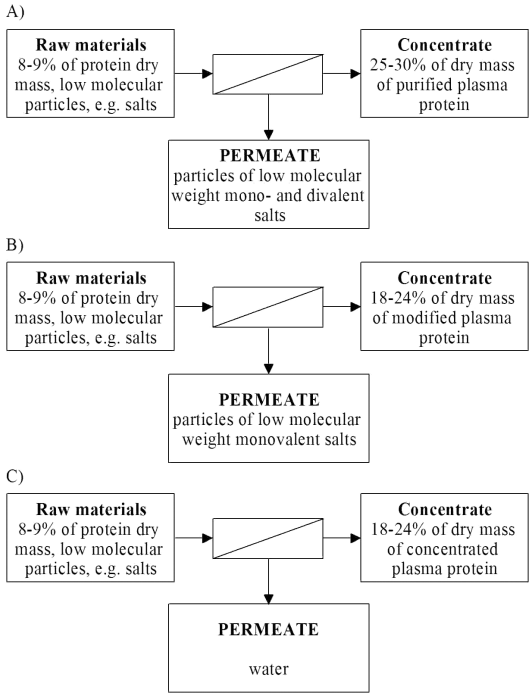


Fig. 1. Flow charts of the plasma concentration processes using A) ultrafiltration, B) nanofiltration, and C) reverse osmosis

The second type of devices used in the evaporative concentration of plasma are film evaporators in which liquid and vapour fall through a system of parallel, vertical tubes surrounded by a heated jacket. Supplied to the plasma apparatus is preheated and next followed by distribution system located at the head of the evaporator, flows down the inner periphery of the tube in the form of a thin film from where water vapour are evaporated. After leaving the calandria, concentrate is collected in the bottom of the evaporator and fumes are drawn into the side chamber to be ejected. Thin-film evaporators can be operated with a very small temperature difference between the heating medium and the boiling liquid, they also have a very short contact time, usually only a few seconds for a single run. This results in thin-film evaporators that are particularly useful for temperature-sensitive products, which includes animal plasma [11, 14].

3. Results and discussion

Below, the effects of using ultrafiltration on two pilot units, Alfa Laval and Intermasz, for the concentration of porcine blood plasma are compared. The conditions under which these processes are realised are discussed as is the efficiency of the processes based on the results of testing the obtained products. The first test was performed on membranes from Alfa Laval. Figure 2 shows a flow diagram of the Alfa Laval experimental unit.

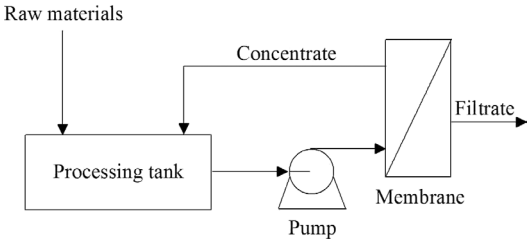


Fig. 2. Flow diagram of the Alfa Laval experimental unit

In the study, a spiral type membrane DSS-GR70PE-3838/80 with an area of 7 m^2 was used. The processing tank was filled with a quantity of porcine plasma prior to commencement of the tests. The process parameters were determined after the pump was switched on and the pressure differential was set. The process was conducted at differential pressures of 1 and 1.2 bars to obtain the assumed values, as measured by the change in volume of the solution introduced to the ultrafiltration unit.

The raw materials used and products obtained during the tests were sampled on the input and output, respectively and analysed in order to evaluate the content of protein and dry matter within the sample – these results are shown in Table 1. Figures 3 and 4 present dependences of the growth in protein content and dry weight, and the degree of concentration in the plasma solution.

Table 1. The results of analyses of the concentration of plasma solution with the use of membranes from Alfa Laval

No.	Differential pressure [bar]	Concentration degree	Raw material		Product	
			Protein [%]	Dry mass [%]	Protein [%]	Dry mass [%]
1	1.0	1.35	7.43	9.46	10.05	12.45
2		1.43	6.79	9.55	10.58	14.02
3		1.47	5.77	8.02	10.70	13.73
4		1.72	6.70	8.89	12.02	14.79
5	1.2	1.67	7.40	9.86	11.44	13.98
6		1.89	6.36	8.98	11.91	15.33
7		2.50	6.67	9.29	12.52	16.20
8		2.86	6.89	9.47	14.39	17.87

Ultrafiltration tests were then carried out on membranes from TAMI Industries using an experimental unit from the Intermasz Company, the schematic diagram of which is shown in Figure 5. In the study, two 23-channel ceramic microfiltration membranes of TAMI Industries were used, each with a filter area of 0.35 m^2 and with separation limits of 300 kD and 0.07 microns. The raw material for the filtration was periodically taken from the plasma storage tank and pumped to the plasma process tank of the experimental unit. The filtrate was continuously discharged outside the system. Concentrate comprising substantially all the proteins, remained in the filtration system.

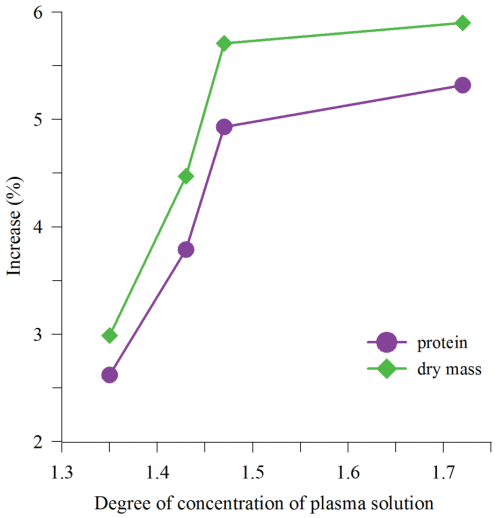


Fig. 3. Dependence of increase of the protein and dry matter in the plasma concentrate and the degree of concentration of the raw material at a pressure difference of 1 bar

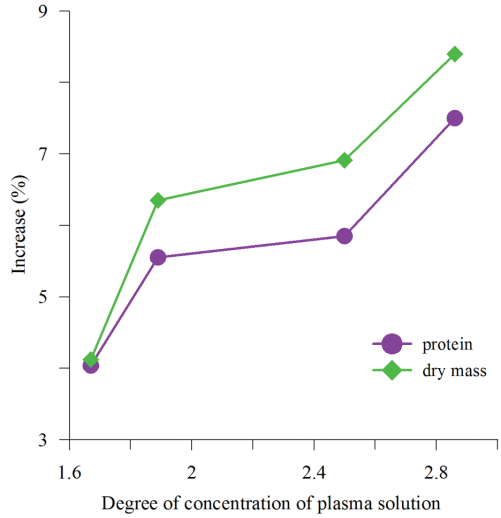


Fig. 4. Dependence of increase of the protein and dry matter in the plasma concentrate and the degree of concentration of the raw material at a pressure difference of 1.2 bar

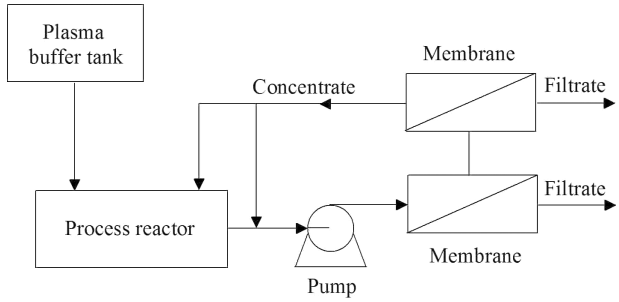


Fig. 5. Flow diagram of the Intermasz experimental unit for the concentration of animal blood plasma using the ultrafiltration method

Two tests were conducted. Filtrate and concentrate samples collected during the testing of raw materials were analysed to determine the amount of protein and dry matter in the filtrate and concentrate. Test 1 was carried out for duration of 3 hours 45 min, in the temperature range 22-40°C. The ultrafiltration of 70 dm^3 of plasma produced 20 dm^3 of concentrate and 50 dm^3 of filtrate. The output had approx. 3.5 times higher density than the input raw material. In the period from 0 to 1 hour 25 min, the efficiency of the process for the 300 kD UF membrane increased from 16 to 31 $dm^3/h\ m^2$ and for the 0.07 μm membrane, the increase was from 20 to 36 $dm^3/h\ m^2$. These resulted from the increase in temperature and followed the systematic increase in the level to 1.35 times the concentration. Fixed transmembrane pressure equal to 2 bars and a temperature range of 35-40°C was used later in the study. A systematic decrease in performance for the 300 kD membrane of 31 $dm^3/h\ m^2$ at 1.35-fold concentration to a level of approx. 4 $dm^3/h\ m^2$ to 3.4-fold concentration was observed. There

was also a steady decrease in performance for the 0.07 μm membrane with 36 $dm^3/h\ m^2$ at 1.35-fold concentration of up to approx. 7 $dm^3/h\ m^2$ to 3.4-fold concentration.

Test 2 was carried out for duration of 6 hours 10 min. 60 dm^3 of plasma was concentrated to obtain 20 dm^3 of concentrate and 40 dm^3 of filtrate. The process of plasma concentration was interrupted several times due to adding recycled filtrate into process tank and work on a few selected levels of concentration. Finally, a ~ 3-fold higher density than the input raw material was obtained.

The results of analyses of the samples of filtrate and concentrate are summarised in Table 2.

Table 2. Results of plasma solution concentration on membranes of TAMI Industries

Object of analysis	Dry mass [%]	Protein [%]
Test 1		
Plasma before concentration	8.99	6.10
Concentrate (3.5-fold concentration)	24.62	19.94
Test 2		
Plasma for filtration	8.99	6.10
Filtrate – membrane 0.07 μm after 20 minutes	2.79	0.25
Filtrate – membrane 300 kD after 20 minutes	2.75	0.15
Filtrate – membrane 0.07 μm (2.3-fold concentration)	2.93	0.26
Filtrate – membrane 300 kD (2.3-fold concentration)	3.08	0.23
Concentrate (2.3-fold concentration)	17.72	13.84
Filtrate – membrane 0.07 μm (3-fold concentration)	2.95	0.23
Filtrate – membrane 300 kD (3-fold concentration)	3.19	0.19
Concentrate (3-fold concentration)	22.65	18.24

In the described test, the process was performed at four selected concentration levels – 1.5-fold, 2-fold, 2.25-fold and 3-fold. After obtaining each of these concentration levels, the concentration process was stopped and the filtrate was recycled to the process reactor and the filtrate was once again concentrated. The dependence of filtration efficiency for the two applied types of TAMI Industries membranes and the degree of plasma concentration are presented in Figure 6.

After reaching the level of 1.5-fold concentration, the filtration efficiency for the 300 kD membrane was 27 $dm^3/h\ m^2$ at 32°C. The increase in temperature to 40°C after 2 h resulted in an increase in the filtration efficiency to 33 $dm^3/h\ m^2$. The process was conducted for a further 55 min while maintaining the previously achieved level of filtration efficiency. After reaching 1.5-fold concentration, filtration efficiency for the 0.07 micron membrane was 37 $dm^3/h\ m^2$ at 32°C. The increase in temperature after 2 h to 40°C resulted in an increase of filtration efficiency of 45 $dm^3/h\ m^2$. The process was conducted for a further 55 min while maintaining the previously achieved level of efficiency.

Filtration efficiency of the 300 kD membrane at 2-fold concentration remained at a stable level of 19 $dm^3/h\ m^2$, and at the 2.25-fold concentration, it decreased to 16 $dm^3/h\ m^2$. For the target 3-fold concentration, the efficiency was 8.5 $dm^3/h\ m^2$. Filtration efficiency of the

0.07 micron membrane at 2-fold concentration remained at a stable level of approx. 27 $dm^3/h\ m^2$, and at the 2.25-fold concentration, it decreased to approx. 24 $dm^3/h\ m^2$. For the target 3-fold concentration, the efficiency amounted to approx. 12 $dm^3/h\ m^2$.

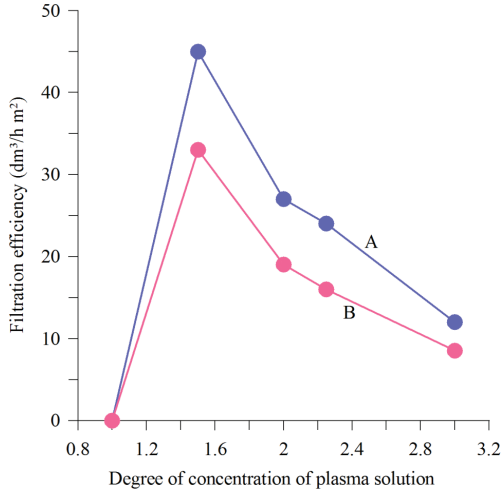


Fig. 6. Dependence of filtration efficiency of TAMI Industries membranes and degree of plasma concentration: A – membrane 0.07 μm; B – membrane 300 kD

4. Conclusions

Studies on animal blood plasma ultrafiltration with an Alfa Laval membrane were treated as preliminary tests in order to assess the suitability of membrane technology for the plasma concentration to produce plasma solutions with higher contents of protein and dry matter. The results obtained showed that this method can be successfully used for this purpose.

Tests carried out on the Intermasz unit were realized for a more detailed characterization of the concentration process especially to determine the influence of selected process parameters on the concentration efficiency. The results showed that the tested membrane functions meet the expected results in terms of protein retention on the concentrate side. The protein level in all obtained filtrates ranged from 0.15-0.26%, regardless of the plasma concentration in the system. Increase in the level of protein in the concentrate is a function of the plasma concentration in the system. Assuming that the content of protein in the raw material amounted to approx. 6.05%, compared protein level in concentrate was higher by 2.3 times (13.84% protein) and 3 times (18.24% protein). These allow to state that close 100% of processed proteins were concentrated.

Comparisons between the levels of dry mass in the filtrates obtained at different levels of concentration enable us to state that the tested membranes are characterised by similar levels of permeability to other components of the plasma (from a slightly higher permeability of the 0.07 μm membrane).

The efficiency of the filtration process for both membranes was stable over time, the membrane showed no tendency to clog. The filtration efficiency was dependent upon the process temperature (positive correlation) and the degree of concentration (negative correlation). Filtration efficiency of the 0.07 μm membrane used in Test 1 was 20-25% higher than that of the 300 kD membrane, while for the 0.07 micron membrane in Test 2, the filtration efficiency was higher at approx. 40%. For both membranes, the permeability of the plasma was similar; however, the 0.07 μm membrane had an average 30% higher filtration efficiency than the 300 kD membrane and is more suitable for use in any industrial plant.

The obtained concentrate of blood plasma can be recycled (off-process recycling) and re-used as a valuable component in the production of foodstuffs.

References

[1] Bah C.S.F., Bekhit A.E.-D.A., Carne A., McConnell M.A., *Slaughterhouse Blood: An Emerging Source of Bioactive Compounds*, Comprehensive Reviews in Food Science and Food Safety, vol. 12(3), 2013, 314-331.

[2] Lynch S.A., Mullen A.M., O'Neill E.E., García C.Á., *Harnessing the Potential of Blood Proteins as Functional Ingredients: A Review of the State of the Art in Blood Processing*, Comprehensive Reviews in Food Science and Food Safety, vol. 16(2), 2017, 330-344.

[3] Silva V.D.M., Silvestre M.P.C., *Functional properties of bovine blood plasma intended for use as a functional ingredient in human food*, Lebensmittel-Wissenschaft & Technologie, vol. 36(7), 2003, 709-718.

[4] King R.H., Campbell R.G., *Blood meal as a source of protein for grower/finisher pigs*, Australian Society of Animal Production, vol. 12, 1978, 149.

[5] Lafarga T., Wilm M., Wynne K., Hayes M., *Bioactive hydrolysates from bovine blood globulins: Generation, characterization, and in silico prediction of toxicity and allergenicity*, Journal of Functional Foods, vol. 24, 2016, 142-155.

[6] Przybylski R., Firdaous L., Châtaigné G., Dhulster P., Nedjar N., *Production of an antimicrobial peptide derived from slaughterhouse by-product and its potential application on meat as preservative*, Food Chemistry, vol. 211(15), 2016, 306-313.

[7] Toldra M., Elias A., Pares D., Saguer E., Carretero C., *Functional properties of a spray-dried porcine red blood cell fraction treated by high hydrostatic pressure*, Food Chemistry, vol. 88(3), 2004, 461-468.

[8] Konopka M., Kowalski Z., Fela K., Klamecka A., Cholewa J., *Otrzymywanie plazmy metodą wirowania krwi – charakterystyka procesu*, Czasopismo Techniczne, 1-Ch/2007, 67-74.

[9] Ockerman H.W., Hansen C.L., *Animal By-Product Processing and Utilization*, CRC Press, Boca Raton, Florida 2000.

[10] Kowalski Z., Makara A., Banach M., *Krew zwierzęca, metody jej przetwarzania i zastosowanie*, Czasopismo Techniczne, 2-Ch/2011, 87-105.

[11] Kowalski Z., Makara A., Banach M., *Blood plasma and hemoglobin production process*, Chemik, vol. 65(5), 2011, 466-475.
[12] Lafarga T., Hayes M., *Bioactive protein hydrolysates in the functional food ingredient industry: Overcoming current challenges*, Food Reviews International, vol. 33(3), 2017, 217-246.
[13] Bah C.S.F., Bekhit A.E.-D.A., Carne A., McConnell M.A., *Composition and biological activities of slaughterhouse blood from red deer, sheep, pig and cattle*, Journal of the Science of Food and Agriculture, vol. 96(1), 2015, 79-89.
[14] *Membrany i membranowe techniki rozdziału*, (ed.) A. Narębska, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń 1997.
[15] Bodzek M., Bohdziewicz J., Konieczny K., *Techniki membranowe w ochronie środowiska*, Wydawnictwo Politechniki Śląskiej, Gliwice 1997.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ENVIRONMENTAL ENGINERING
DOI: 10.4467/2353737XCT.17.197.7426

Małgorzata Kryłów (gosiak@wis.pk.edu.pl)
Piotr Rezka
Institute of Water Supply and Environmental Protection, Faculty of Environmental Engineering, Cracow University of Technology

Sources of endocrine-disrupting compounds and their migration to the environment

Źródła związków zaburzających funkcjonowanie układu hormonalnego i ich migracja w środowisku

Abstract

The aim of this article is to present issues related to the presence and pathways of bisphenol A emission and its migration to wastewater and the environment. Bisphenol A (BPA) is an organic compound mainly used in the production of plastics. It is classified as an endocrine disrupting compound (EDC) and should therefore not be emitted to the environment. This paper presents basic information on bisphenol A, its applications and potential sources of emission to the environment. A wide review of literature confirming the occurrence of bisphenol A in sewage, sediments, natural waters, drinking water and the atmosphere is performed. Effective wastewater treatment and neutralisation of bisphenol A in sewage sludge could partially reduce the levels of BPA pollution in the aquatic environment.

Keywords: bisphenol A (BPA), wastewater, water, sediments, air

Streszczenie

Celem artykułu jest przedstawienie zagadnień dotyczących obecności i źródeł występowania bisfenolu A w ściekach i środowisku. Bisfenol A (BPA) jest związkiem organicznym, stosowanym przede wszystkim do produkcji tworzyw sztucznych. Należy do związków wykazujących negatywne oddziaływanie na układ hormonalny (***endocrine disrupting compounds***, EDC), w związku z czym nie powinien być emitowany do środowiska naturalnego. W artykule przedstawiono podstawowe informacje dotyczące bisfenolu A oraz potencjalne źródła emisji tego związku do środowiska. Dokonano przeglądu literatury potwierdzającej obecność bisfenolu A w ściekach, osadach oraz wodach naturalnych. Efektywne oczyszczenie ścieków i neutralizacja bisfenolu A obecnego w osadach ściekowych pozwoliłoby na częściowe ograniczenie zanieczyszczenia środowiska naturalnego.

Słowa kluczowe: bisfenol A (BPA), ścieki, woda, osady, powietrze

1. Introduction

The presence of organic contaminants in the environment has been studied for many years. Pesticides, solvents, substances used in industry and, relatively recently, pharmaceuticals have also become one of the most discussed topics in this context in scientific publications worldwide. Another group of environmental pollutants are endocrine-disrupting compounds or chemicals (EDCs) which affect the endocrine (hormonal) systems of living organisms. These compounds are characterised by an extremely negative impact on the environment and human health at very low concentrations.

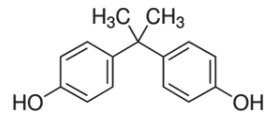
Natural and synthetic oestrogens have adverse effects on aquatic organisms at concentrations as small as 0.1 ng/l [24], they are toxic to humans and pose a risk through the consumption of fish and other animals living in water and exposed to these hormones. The other substance from the EDC group, bisphenol A, is a compound used on a very large scale and people are exposed to its harmful effects at work, in stores, on walks or even at home. Indirectly, logistics processes, such as transportation, waste disposal and management of waste are sources of environmental pollution from this compound.

The purpose of this paper is to present bisphenol A as a commonly occurring and harmful pollutant, its sources of emissions to the environment and the results of global research showing the presence of bisphenol A in wastewater, natural water, sediment and air samples.

2. Characteristic

Bisphenol A (BPA) is an organic chemical compound belonging to the phenol group. Its solid state is most commonly a white powder or flakes. It has a low level of solubility in water, but is soluble in metal hydroxides from alkali metal and organic solvents such as acetone, methanol, ethanol and diethyl ether. Table 1 presents basic information on BPA, such as molar mass, molecular formula and structure, and half-life in the human organism.

Table 1. Basic parameters of bisphenol A

Compound	Formula	Structure	Molecular mass [g/mol]	CAS no.
Bisphenol A (BPA)	$C_{15}H_{16}O_2$		228.29	80-05-7

The first method of preparation of BPA was its synthesis by two molecules of acetone and phenol in acidic conditions. This method was efficient (under the excess of phenol), and had an important advantage in the fact that the only by-product was water; however, the commercial production of BPA requires the large-scale distillation of a mixture of many by-products.

Bisphenol A is classified as an endocrine-disrupting substance which has a negative impact on the hormonal system of humans and animals mimicking the effect of female hormones (oestrogens); furthermore, the exposure of foetuses to BPA can trigger subsequent physical and neurological problems. It is also suspected of carcinogenicity and because of that, bisphenol A was classified as a compound which is extremely hazardous to humans and the environment.

3. Potential sources of emission

Figure 1 illustrates the potential sources of EDC (including BPA) release to the environment. Bisphenol A is widely used in the production of plastics, especially polycarbonate resins e.g. it is an essential monomer for epoxy resin. Such materials are used in the manufacture of a wide range of products, these include polycarbonate plastics, water bottles, baby bottles, sports equipment, CDs and DVDs, medical devices, dental sealants for teeth fillings, optical lenses and lining for water pipes. Epoxy resins containing BPA are used as linings for cans of food and beverages. Bisphenol A is also used as an antioxidant and an inhibitor in the production and processing of polyvinyl chloride, in the manufacture of car tyres, as flame retardant, in electronics, within the construction industry, in moulding and in the production of thermal paper for receipts [8]. This means that bisphenol A and materials produced with BPA are in widespread use in industrialised and developing countries.

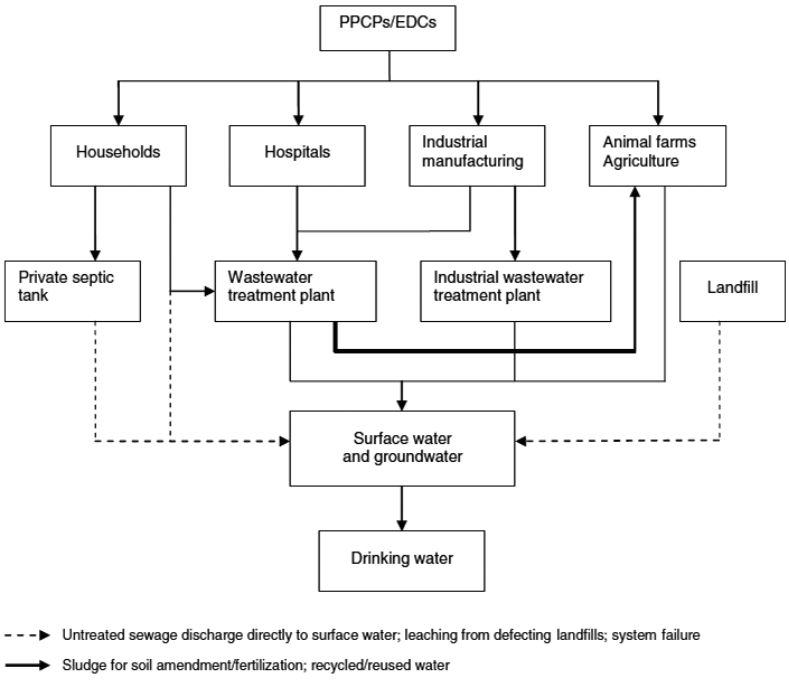


Fig. 1. Sources of endocrine-disrupting compounds (EDC) in the environment [7]

The main modes of human exposure to BPA are its ingestion, inhalation and contact with skin. Most BPA is ingested along with food, drinking water and beverages in cans and bottles, and tap water (due to the linings of water pipes containing BPA). Penetration into the body through skin may occur during normal use of products containing BPA, especially in the case of thermal paper receipts in which BPA is present in its free state (not in the form of polymer product) – this constitutes an increased risk of exposure to BPA for cashiers and customers. The penetration of bisphenol A into the body via inhalation may occur particularly while inhaling urban air polluted with BPA, being near a cash register during the printing of receipts and potentially during the recording of CDs and DVDs because the laser heats the plate to sometimes even up to 700°C. BPA is excreted primarily in urine, which is why household sewage is a significant source of BPA in raw wastewater within wastewater treatment plants [10].

The transportation, storage and disposal of waste containing BPA also poses a risk of environmental pollution. The presence of materials containing BPA in car tyres creates a continual risk of pollution through the abrasion of the tyres. Tyre particles may be blown into neighbouring roads and meadows and in the right conditions, it can also lead to BPA penetration into soil. Transportation or storage in landfills of inadequately protected materials that contain BPA also poses a risk of wastes spreading across the environment. Emissions to soil and groundwater can occur by eluting residues of BPA in precipitations from products containing it e.g. thermal paper. The combustion of materials containing bisphenol A also contributes to the pollution of the environment, in this case the atmosphere; however, the primary source of BPA in air samples is the production of BPA itself.

4. Occurrence in the environment

Bisphenol A present in the wastewater fed to the municipal wastewater stream comes mainly from wastewater produced in households and from industrial effluents that have been pre-screened and then discharged to the municipal treatment plant. Table 2 presents the confirmed cases of BPA detection in sewage treatment plants (raw and treated wastewater), natural waters and sewage sludge and sediments. Insufficient degradation or elimination of BPA during wastewater treatment processes results in the emission of BPA with the discharge of effluent into rivers or reservoirs and contamination of the aquatic environment.

The presence of BPA in the aqueous environment is not only related to effluent discharged from urban wastewater treatment, but also to pre-treated wastewater discharged directly from the factories and industrial wastewater treatment plants. If polluted wastewater is discharged into a body of water which is a tributary to other watercourses, bisphenol A will migrate and contaminate another river or even groundwater. The data reported in Table 2 confirms the ability of BPA to migrate into groundwater through, among other routes, permeable layers. BPA can also partially accumulate by passing through soil and river sediments.

Table 2. Concentrations (min-max or mean) in ng/l or ng/g dry mass for sediments of BPA; b.l.q. – below level of quantification

Influent	Effluent	Surface water	Groundwater	Sediments
13-2140[15] 60-600[43] 360-1620[18] 378-890[9] 416-2050[23] 1800[40]	2-44[43] 30-1100[15] 35-86[23] 110-300[18] 700[40]	1.2-1900[1] 2.1-87[13] 2.2-4230[21] 6-34[23] 6-68[11] 6-500[26] 6-881[22] 7.5-334[12] 55-162[27] 192-215[15] 460-4800[25]	b.l.q.-494[17] b.l.q.-7000[19] b.l.q.-9300[28] 1-1136[20] 1-11[4] 79-2550[15] 600-11000[14]	0.17-1.25[2] 0.58-36700[8] 1.1-43[42] 4.3-130[25] 10-530[23] 53-196[43] 0.32[6]

Fu and Kawamura [5] showed that bisphenol A is also present in air samples. In the agricultural areas of China, its concentration in the air does not exceed 240 pg/m3, but the air samples in urban areas are more contaminated (20-2.340 pg/m^3). Definitely, more air pollution bisphenol was found in India, where its concentration reached 9,820 and 17,400 pg/m^3. Air samples collected in coastal areas around the world contain trace amounts of BPA that does not exceed 32 pg/m^3. These results show how large the scale of bisphenol A pollution is in the environment – it is not only water that is contaminated but also the air and soil. Considering how large is the current need for BPA (despite attempts to withdraw from the production and use of bisphenol A in Canada), constantly increasing levels of environmental pollution can be expected.

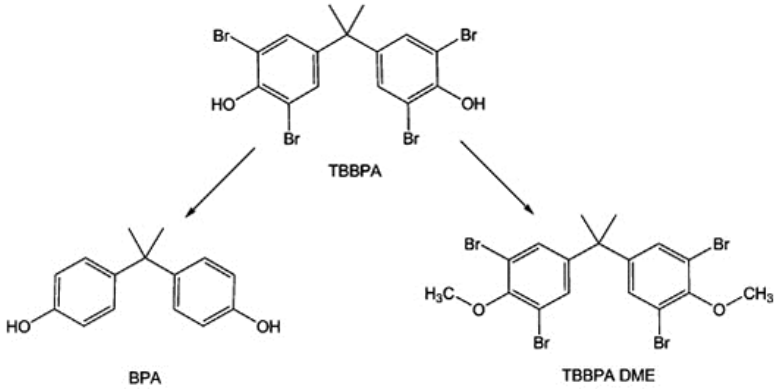


Fig. 2. Microbial-induced transformation of TBBPA to BPA and TBBPA DME [16]

It should be kept in mind that the presence of BPA in the environment can be linked not only with BPA itself, but also with more complex compounds that may transform into BPA. One of the most widely used brominated flame retardants, tetrabromobisphenol A (TBBPA) – depending on the conditions under the influence of bacteria present in the environment, this can be transformed into BPA (reductive debromination in anaerobic conditions),

TBBPA DME (dimethyl ether) and TBBPA monomethyl ether (O-methylation in aerobic conditions) (Fig. 2). This may explain the very high concentrations of BPA in groundwater (anaerobic conditions).

5. BPA in other matrixes

Previously mentioned data provides a picture of the wide migration of bisphenol A occurring in the aquatic environment. Bisphenol emitted to sewage or directly to surface water may accumulate in sewage sludge or river sediments; the use of such sludge for agricultural purposes may contribute to even greater levels of BPA pollution of soils.

If bisphenol A is present in the source water of a drinking water treatment plant, there is a risk that it can penetrate to the finished water despite the use of modern methods of treatment. There are known cases of bisphenol A detection in drinking water in China between 15-63 ng/l and 38.9-55.8 ng/l [8]; in Canada, 100 ng/l; in the USA, up to 24 ng/l [15]; also in other countries at concentrations not exceeding 100 ng/l (0.5-99 ng/l) [1].

Ingestion of drinking water polluted with bisphenol A is an additional route of exposure of the human body to BPA. Studies have shown that BPA was detected in the analysed samples of blood serum and urine at concentrations of 0-2,500 ng/l in serum and 110-3,200 ng/l in urine [41]. Also in Japan, BPA was detected in 23 samples of breast milk at concentrations of 280-970 ng/l [29]. This confirms the suspicion of the possibility of exposure of infants and young children to the harmful effects of BPA.

6. Conclusions

Bisphenol A is ubiquitous in the environment, not only in effluents and natural waters, but also in soil and the atmosphere. Its wide use hinders any control over the level and scale of pollution and calls into question any plans to withdraw the substance from production. Due to its advantages, in many countries bisphenol A is and will be the main raw material for the production of polycarbonate plastic used in the production of baby bottles. The first step should be the removal of BPA from everyday life where it is not needed and poses a direct threat to the body, especially for infants and young children (replacing linings in pipes and removing the material from the production of baby bottles). The next step should be to protect the air in workplaces (BPA withdrawal from the production of thermal paper) and areas adjacent to factories producing or using bisphenol A.

This does not mean that we need not care about the environment itself. Inefficient wastewater treatment and inadequate protection of landfill result in pollution of water and soil – this could ultimately mean exposure to human health through repeated contact with BPA. The application of membrane processes using reverse osmosis (RO) and nanofiltration (NF) could be used for efficient BPA removal. Experiments conducted by Dudziak and Bodzek [3] showed 85% and 70% reductions of bisphenol A from water solutions through

the use of RO and NF, respectively. Contaminated agricultural land (and consequently crop yields) and drinking water containing bisphenol A are additional sources of penetration of the harmful compound to the human body.

References

[1] Benner J., Helbling D.E., Kohler H.-P.E., Wittebol J., Kaiser E., Prassem C., Ternes T.A., Albers C.N., Aamand J., Horemans B., Springael D., Walravens E., Boon N., *Is biological treatment a viable alternative for micropollutant removal in drinking water treatment processes?*, Water Research, Vol. 47, 2013, 5955-5976.

[2] Dong B., Kahl A., Cheng L., Vo H., Ruehl S., Zhang T., Snyder S., Saez A.E., Quanrud D., Arnold R.G., *Fate of trace organics in a wastewater effluent dependent stream*, Science of the Total Environment, Vol. 518-519, 2015, 479-490.

[3] Dudziak M., Bodzek M., *Removal of Estrogenic Micropollutants from Water Solutions by High-Pressure Driven Membrane Processes*, Ochrona Środowiska, Vol. 31 (2), 2009, 33-36.

[4] Felix-Canedo T.E., Duran-Alvarez J.C., Jimenez-Cisneros B., *The occurrence and distribution of a group of organic micropollutants in Mexico City's water sources*, Science of the Total Environment, Vol. 454-455, 2013, 109-118.

[5] Fu P.Q., Kawamura K., *Ubiquity of bisphenol A in the atmosphere*, Environmental Pollution, Vol. 158, 2010, 3138-3143.

[6] Gardner M., Jones V., Comber S., Scrimshaw M.D., Coello-Garcia T., Cartmell E., Lester J., Ellor B., *Performance of UK wastewater treatment works with respect to trace contaminants*, Science of the Total Environment, Vol. 456-457, 2013, 359-369.

[7] Grassi M., Rizzo L., Farina A., *Endocrine disruptors compounds, pharmaceuticals and personal care products in urban wastewater: implications for agricultural reuse and their removal by adsorption process*, Environmental Science and Pollution Research, Vol. 20, 2013, 3616-3628.

[8] Huang Y.Q., Wong C.K.C., Zheng J.S., Bouwman H., Barra R., Wahlstrom B., Neretin L., Wong M.H., *Bisphenol A (BPA) in China: A review of sources, environmental levels, and potential human health impacts*, Environment International, Vol. 42, 2012, 91-99.

[9] Jiang J.Q., Yin Q., Zhou J.L., Pearce P., *Occurrence and treatment trials of endocrine disrupting chemicals (EDCs) in wastewaters*, Chemosphere, Vol. 61, 2005, 544-550.

[10] Kang J.-H., Kondo F., Katayama Y., *Human exposure to bisphenol A*, Toxicology, Vol. 226, 2006, 79-89.

[11] Kasprzyk-Hodern B., Dinsdale R.M., Guwy A.J., *The removal of pharmaceuticals, personal care products, endocrine disruptors and illicit drugs during wastewater treatment and its impact on the quality of receiving waters*, Water Research, Vol. 43, 2009, 363-380.

[12] Kim J.W., Jang H.S., Kim J.G., Ishibashi H., Hirano M., Nasu K., Ichikawa N., Takao Y., Shinohara R., Arizono K., *Occurrence of Pharmaceutical and Personal Care Products (PPCPs) in Surface Water from Mankyung River, South Korea*, Journal of Health Science, Vol. 55, 2009, 249-258.

[13] Kleywegt S., Pileggi V., Yang P., Hao C., Zhao X., Rocks C., Tchach S., Cheung P., Whitehead B., *Pharmaceuticals, hormones and bisphenol A in untreated source and finished drinking water in Ontario, Canada – occurrence and treatment efficiency*, Science of the Total Environment, Vol. 409, 2011, 1481-1488.
[14] Lapworth D.J., Baran N., Stuart M.E., Ward R.S., *Emerging organic contaminants in groundwater: a review of sources, fate and occurrence*, Environmental Pollution, Vol. 163, 2012, 287-303.
[15] Luo Y., Guo W., Ngo H.H., Nghiem L.D., Hai F.I., Zhang J., Liang S., Wang X.C., *A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment*, Science of the Total Environment, Vol. 473-474, 2014, 619-641.
[16] McCormick J.M., Paiva M.S., Haggblom M.M., Cooper K.R., White L.A., *Embryonic exposure to tetrabromobisphenol A and its metabolites, bisphenol A and tetrabromobisphenol A dimethyl ether disrupts normal zebrafish (Danio rerio) development and matrix metalloproteinase expression*, Aquatic Toxicology, Vol. 100 (3), 2010, 255-262.
[17] Meffe R., de Bustamante I., *Emerging organic contaminants in surface water and groundwater: A first overview of the situation in Italy*, Science of the Total Environment, Vol. 481, 2014, 280-295.
[18] Moral A., Sicilia M.D., Rubio S., Perez-Bendito D., *Determination of bisphenols in sewage based on supramocelular solid-phase extraction/liquid chromatograpy/fluorimetry*, Journal of Chromatography A, Nr 1100, 2005, 8-14.
[19] Musolff A., Leschik S., Moder M., Strauch G., Reinstorf F., Schirmer M., *Temporal and spatial patterns of micropollutants in urban receiving waters*, Environmental Pollution, Vol. 157, 2009, 3069-3077.
[20] Osenbruck K., Glaser H.-R., Knoller K., Weise S.M., Moder M., Wennrich R., *Sources and transport of selected organic micropollutants in urban groundwater underlying the city of Halle (Saale), Germany*, Water Research, Vol. 41, 2007, 3259-3270.
[21] Pal A., He Y., Jekel M., Reinhard M., Gin K.Y.H., *Emerging contaminants of public health significance as water quality indicator compounds in the urban water cycle*, Environment International, Vol. 71, 2014, 46-62.
[22] Peng X., Yu Y., Tang C., Huand Q., Wang Z., *Occurrence of steroid estrogens, endocrine-disrupting phenols and acid pharmaceutical residues in urban riverine water of the Pearl River Delta, South China*, Science of the Total Environment, Vol. 937, 2008, 158-166.
[23] Petrie B., Barden R., Kasprzyk-Hodern B., *A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring*, Water Research, Vol. 72, 2015, 3-27.
[24] Rezka P., Balcerzak W., Kryłów M., *Occurrence of synthetic and natural estrogenic hormones in the aquatic environment*, Technical Transactions, Vol. 3-Ś/2015, 47-54.
[25] Salgueiro-Gonzalez N., Turnes-Carou I., Besada V., Muniategui-Lorenzo S., Lopez--Mahia P., Prada-Rodriguez D., *Occurrence, distribution and bioaccumulation of endocrine disrupting compounds in water, sediment and biota samples from a European river basin*, Science of the Total Environment, Vol. 529, 2015, 121-130.

[26] Santhi V.A., Sakai N., Ahmad E.D., Mustafa A.M., *Occurrence of bisphenol A in surface water, drinking water and plasma from Malaysia with exposure assessment from consumption of drinking water*, Science of the Total Environment, Vol. 427-428, 2012, 332-338.
[27] Stasinakis A.S., Mermigka S., Samaras V.G., Farmaki E., Thomaidis N.S., *Occurrence of endocrine disrupters and selected pharmaceuticals in Aisonas River (Greece) and environmental risk assessment using hazard indexes*, Environmental Science and Pollution Research, Vol. 19, 2012, 1574-1583.
[28] Stuart M., Lapworth D., Crane E., Hart A., *Review of risk from potential emerging contaminants in UK groundwater*, Science of the Total Environment, Vol. 416, 2012, 1-21.
[29] Sun Y., Irie M., Kishikawa N., Wada M., Kuroda N., Nakashima K., *Determination of bisphenol A in human breast milk by HPLC with column-switching and fluorescence detection*, Biomedical Chromatography, Vol. 18, 2004, 501-507.
[30] Tadkaew N., Hai F.I., McDonald J.A., Khan S.J., Nghiem L.D., *Removal of trace organics by MBR treatment: The role of molecular properties*, Water Research, Vol. 45, 2011, 2439-2451.
[31] Vanderberg L.N., Hausen R., Marcus M., Olea N., Welshons W.V., *Human exposure to bisphenol A (BPA)*, Reproductive Toxicology, Vol. 24, 2007, 139-177.
[32] Vethaak A.D., Lahr J., Schrap S.M., Belfroid A.C., Rijs G.B.J., Gerritsen A., de Boer J., Bulder A.S., Grinwis G.C.M., Kuiper R.V., Legler J., Murk T.A.J., Peijnenburg W., Verhaar H.J.M., de Voogt P., *An integrated assessment of estrogenic contamination and biological effects in the aquatic environment of The Netherlands*, Chemosphere, Vol. 59, 2005, 511-524.
[33] Yu Y., Wu L., Chang A.C., *Seasonal variation of endocrine disrupting compounds, pharmaceuticals and personal care products in wastewater treatment plants*, Science of the Total Environment, Vol. 442, 2013, 310-316.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
ENVIRONMENTAL ENGINERING
DOI: 10.4467/2353737XCT.17.198.7427

Paweł Wilk (pawel.wilk@imgw.pl)
Paulina Orlińska-Woźniak
Joanna Gębala
Section of Modelling Surface Water Quality, Institute of Meteorology and Water Management, National Research Institute

Mieczysław Ostojski
Institute of Meteorology and Water Management, National Research Institute

The flattening phenomenon in a seasonal variability analysis of the total nitrogen loads in river waters

Zjawisko wypłaszczenia podczas sezonowej zmienności azotu ogólnego w powierzchniowych wodach płynących

Abstract

This article shows the results of analyses conducted of the seasonal variability of nitrogen concentrations and loads, depending on plants growing season, this doesn't seem to work, consider changing to something like 'dependent upon the stage of the plant growing season,' if that is what you mean as well as river flow and the precipitation levels in the basins of Middle Warta, Reda and Rega. The Macromodel DNS/SWAT, implemented for three basins, has been calibrated in calculations profiles of three river basins this is vague and unclear, the repetition of 'three (river) basins' is particularly confusing. The analysis confirmed the significant impact of cover crops on the retention of water and nutrients. The phenomenon of periodic decreases and stabilisation of nitrogen concentrations and loads in the surface waters flowing through the profiles of the enclosing water body was observed and analysed. The flattening phenomenon analysis of the phenomenon to me, the repetition of 'phenomenon' is confusing, you could clarify this by defining the second use of 'phenomenon' by, for example, inserting 'of these decreases and stabilisation' here can be used to assess the state of the ecosystem in the basin area.

Keywords: water pollution, environmental analysis, nutrients, bioaccumulation, nonpoint sources

Streszczenie

W artykule przedstawiono wyniki analiz sezonowej zmienności azotu ogólnego w zależności od sezonu wegetacyjnego, a także przepływu rzeki i wielkości opadów występujących na zlewni środkowej Warty, Redy i Regi. Do tego celu został wykorzystany Makromodel DNS/SWAT, który został skalibrowany dla profili ujściowych wybranych zlewni. Przeprowadzona analiza potwierdziła znaczący wpływ roślin okrywowych na retencję wody i składników odżywczych. Zaobserwowano okresowe zmniejszanie się stężenia azotu w powierzchniowych wodach płynących. Zaobserwowane zjawisko może zostać wykorzystane do oceny stanu ekosystemu w obszarze dorzecza, a także umożliwia ocenę rzeczywistego wpływu zarządzania zlewnią na cykl azotu.

Słowa kluczowe: energetyczne wykorzystanie biogazu

1. Introduction

Studies concerning the sources of nutrient emissions in Europe have confirmed that the largest share of these pollutants in most countries, including Poland, comes from diffuse sources [21]. Agricultural land occupies approximately 60% of the total area of Poland and has a significant impact on the levels of nitrogen discharged into surface waters [13]. Agricultural activity is one of the primary forms of human activity affecting the environment. Increased agricultural anthropopressure, i.e. human agricultural activities affecting the natural environment in the river basins, causes quantitative and qualitative changes in the aquatic environment [18]. The impact of agricultural pollution on the aquatic environment is directly related to the intensity of fertiliser economy (intensity of fertilization) and soil use. Fertiliser economy and agricultural practices which are beneficial with regard to crop production are not always beneficial to the environment and lead to dynamic changes in the amount of nitrogen in surface waters over time. The size of the nitrogen run-off into surface waters depends, inter alia, on land use, weather conditions, plant cover and soil properties [19, 49]. The responsibility for the study of surface water quality as well as the processing and dissemination of information on the aquatic environment in Poland belongs to the State Environmental Monitoring Department [11]. At present, mathematical models based on detailed data relating to land use, soil types, climatic and meteorological data are widely used. *A study* concerning the analysis of variability in concentrations of biogenic compounds has been carried out by the Institute of Meteorology and Water Management – National Research Institute in Poland, in 'Modelling of Water Pollution Section from 2012. The mentioned studies were initiated by IMGW-PIB due to problems that occurred during the model calibration of nutrients with the use of the mathematical model Macromodel DNS (discharge nutrient sea) [32]. The use of Macromodel DNS enabled us to obtain the full range of data with daily time steps of the analysed parameters at selected control points. This enabled the carrying out of a wide range of analyses, including an analysis of the seasonal variability of nutrients over a multi-year period. The seasonal variability of nutrients, i.e. seasonal fluctuations in the amount of nutrients in surface waters, depends, inter alia, on the hydrological regime, the biological activity of the river, and land use in the basin.

Nutrient load fluctuations occurring in surface waters over time led us to carry out research concerning the relationship between the nitrogen uptake of plants and its subsequent quantity in surface waters. The aforementioned seasonal variability of pollution is the subject of many research studies [35, 41, 42, 44, 46]. Processes such as the uptake of nutrients by plants, surface and point pollution run-off from the area as well as the impact of nutrients on the biological elements of the ecosystem, are analysed in detail. In publication [39], it is also stated that during the growing season, there are changes in the amount of nutrients discharged into surface waters as a result of their uptake by plants from the environment. The participation of plants in the nitrogen cycle in the environment is important and cannot be ignored in the analysis of the seasonal variability of nutrients in surface waters; the same is true of the stage of the season of plant growth, this can also have a significant impact on the amount of nutrients in surface waters throughout the year and this relationship was analysed.

Knowledge concerning seasonal variability can provide a good basis for determining, inter alia, the validity of multiple scenarios and action plans developed by the state administration and research units [6, 51], it is therefore important to continue research.

2. Materials and Methods

2.1. Objectives of study

The main objective of this work was to analyse the seasonal variability of the total nitrogen loads in the calculation profile. Analyses of the amount of the total nitrogen load are crucial for the proper assessment of nutrient balance in the basin; therefore, it is important to identify the seasonal fluctuations of loads in the river calculation profiles In order to determine the sum of the total daily nitrogen load at the calculation profile of the basin in over a period of several years, a cumulative mass curve was used. Analyses were carried out for three pilot basins: the Rega, the Middle Warta and the Reda. Macromodel DNS [32] was used to simulate the total daily loads of nitrogen at the rivers calculation profiles.

In order to determine the total nitrogen load flowing out of a basin area at the calculation profile basin over a given period of time, a mass curve was used. The mass curve of the size of the total nitrogen load is a curve in which the ordinate of each point indicates how the total volume of load flowed from the beginning of the period until a specified abscissa of this point in the calculation profile of the river. The values for the establishment of the curve are the total daily loads of total nitrogen treated as daily averages the mean. The curve is obtained by totalling up the required intervals of daily loads of total nitrogen to give the ordinates of the curve. By applying the size of the load a picture of the rising curve during the considered period is obtained. The line graph is a constantly rising curve with a number of inflection points [5, 33, 47]. Constructed in this manner, the curve helps to indicate, in the later stages of the work, seasonal variations in the total nitrogen load in the basin calculation profile during the analysed years. The main preconditions to construct a correct curve are aggregated daily items of data relating to the total nitrogen load.

2.2. Macromodel

In order to obtain daily data at selected control points in the study, the Macromodel DNS (discharge nutrient sea) mathematical model with SWAT (soil and water assessment tool) module was used. Macromodel DNS/SWAT was designed by the Institute of Meteorology and Water Management-National Research Institute for the analysis of processes taking place in a basin, such as water and matter cycles [30, 31]. It enables the simulation of the long-term impact of land use on water quality and the impact of pollutants discharged to surface waters. The SWAT module uses the hydrological transport model which is based on meteorological and hydrological data, the extent of surface run-off and the amount of fertiliser *contained within* in order to analyse phenomena and processes related to the transport of nutrients in the watershed [30]. The general scheme of Macromodel DNS is shown in Figure 1.

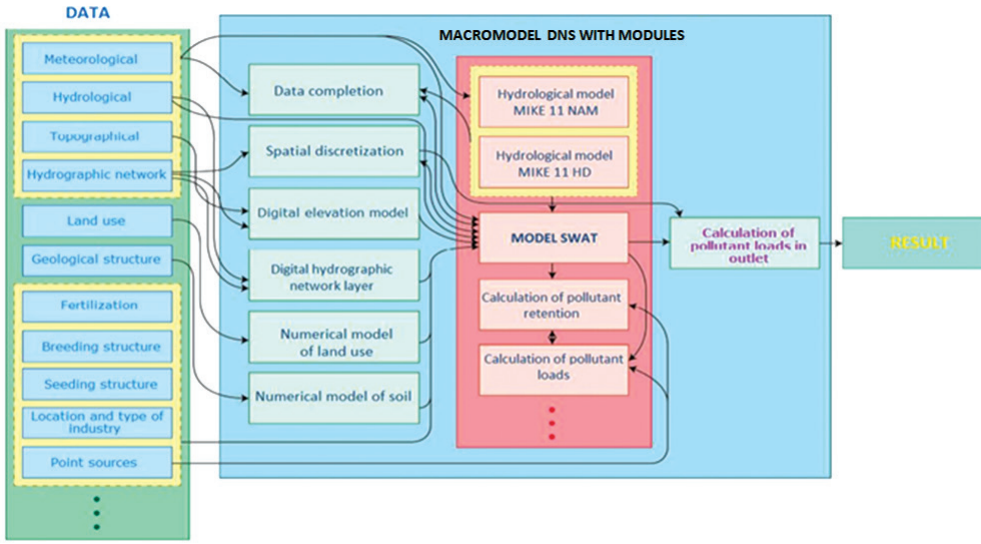


Fig. 1. The general scheme of Macromodel DNS [32]

The SWAT module is an element of Macromodel DNS and is used to analyse the processes of water cycles and organic matter in the basin [30, 31]. This allows us to carry out simulations of the long-term impact of land use management on water quality and to examine the amount of pollutants discharged from particular surface water bodies (SWB) to other surface water bodies. This module uses a hydrological transport model, which is based on, inter alia, meteorological data, the quantity of surface run-off and the amount of soil fertilisation; it also enables us to carry out analyses of phenomena and processes connected with the transportation of nitrogen loads in the sub-basin [31]. With the use of the Macromodel DNS/SWAT, all the elements form a homogenous, numerical sub-basin model that enables us to analyse different scenarios of sub-basin exploitation in different meteorological and hydrological conditions. The Macromodel DNS/SWAT can be used to analyse the efficiency of the proposed measures aimed at reducing the quantity of total nitrogen loads discharged into surface waters [17, 18, 32].

The *validity* of the use of Macromodel DNS/SWAT to study the variability of the total nitrogen load was confirmed by numerous scientific works carried out with the use of the mentioned model, including work concerning the modelling of the discharge of nutrients into the Baltic Sea [32], *analyses* concerning the impact of agricultural anthropopression on surface water quality [18] and analyses of the river absorption capacity [48].

For the analysis of the size of the total nitrogen load on the river estuary and its seasonal variability in the calculation profile, three pilot basins were selected.

2.3. Descriptions of the Basins

The Middle Warta basin (Fig. 2a) constitutes a part of the Warta basin and is closed by two profiles – Nowa Wieś Podgórna and Oborniki. The acreage of the basin is 6,039 km^2 and constitutes 11% of all the acreage of the entire Warta basin (around 54,500 km^2). There are

a few tributaries on the examined section of the river of which the most important are the Lutynia river, the Mosiński Canal and the Mogilnica river. The analysed section of the basin is characterised by a significant proportion of the area being exposed to nitrogen pollutants of agricultural origin. The largest agglomeration of the basin is the city of Poznań. The parent rocks of the basin area are post-glacial sediments, mainly sandy and loamy soils, the majority being brown and podzolic soils. The long-term observation studies of the Warta river indicate that the water quality is *varies* in particular sections. The major source of pollution is the constant and seasonal discharges of domestic, economic and industrial sewage from cities located near the river, and surface run-offs from agricultural areas [8, 38].

The basin of the Rega river (Fig. 2b) covers an area of 2766.8 km^2, and the length of the watercourse from the spring in Połczyn-Zdrój district to the Baltic Sea constitutes 187.57 km (35.4 miles), which makes it the fourth longest of the rivers flowing into the Baltic Sea (after the Wisła, the Odra and the Pasłęka rivers). There are six wastewater treatment plants in the basin: Świdwin, Gryfice, Łobez, Dobra, Resko, Węgorzyno. Within the region, podzolic soils comprised of sands and gravels as well as brown soils comprised of loamy sands and glacial tills dominate. The Rega basin is agriculturally dominant – agricultural areas constitute 54.5% of the acreage. The average size of an individual farm is > 15 ha, agricultural holdings constitute 17.3% and 5.8%, respectively [40].

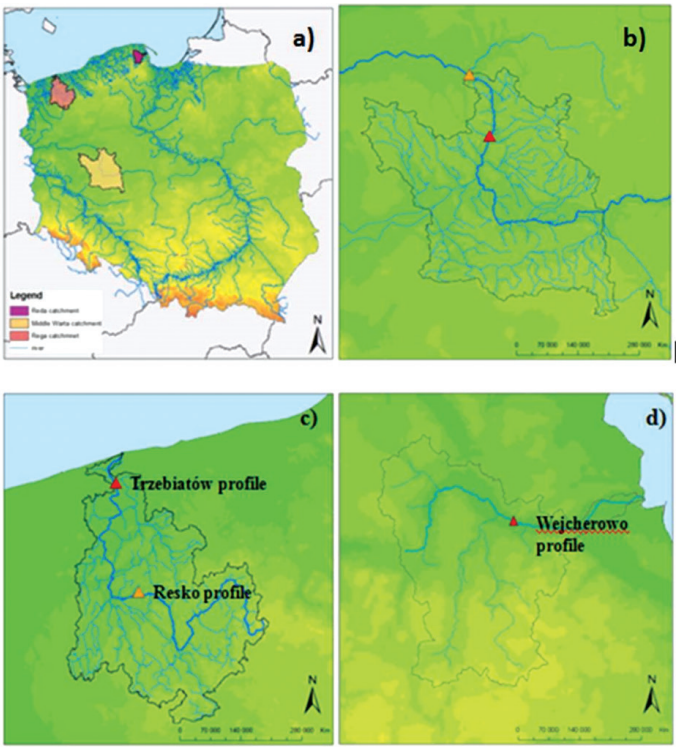


Fig. 2. a) Location of basins in Poland; b) the Middle Warta basin; c) the Rega basin; d) the Reda basin (indication of profiles for calibration and verification [red] and validation [orange]) [28]

The acreage of the Reda basin (Fig. 2c) is around 485.55 km^2. The length of the watercourse from its sources in Strzebielin to the Baltic Sea is 50.6 km (31.4 miles). The soils of The Reda and its tributaries are characterised by a significantly sandy terrain.. In the valleys, there are peat soils, very light clay soils and alluvial mud soils. The basin area is mainly used as agricultural land and grasslands. Moreover, special characteristics of the Reda tributaries are the relatively high river gradients that are typical for mountain rivers. The lakes of the basin are situated in the vicinity of the sources of the river tributaries. The total acreage of the lakes equals 3.187 km^2. In the Orle locality, there are artificial dammed reservoirs (the Nowe Orle and the Stare Orle), that were created from the exploitation of calcium deposits [22].

2.4. Model input data

For the use of the Macromodel DNS/SWAT, the following input data was prepared: digital elevation model (DEM); hydrology map; soil map; land use map; data concerning wastewater treatment plant; the daily meteorological and hydrological data; the amount of fertilisers. The gathered data formed the database that was required by the model [1, 42, 43].

2.4.1. Sub-basins

The Middle Warta river basin was divided into 70 sub-basins according to SWB, the Rega basin into 52 sub-basins and the Reda basin into 30 sub-basins according to the boundaries of the Surface Water Bodies, which are the basic units of water management in Poland, according to [11].

2.4.2. DEM

The DEM remains the national, central geodesic and cartographic level resource and is created on the basis of aerial photographs within a flat rectangular coordinate system labelled 'PUWG1992'. The terrain data corresponds to map sheets within the flat rectangular coordinate system '1992' on a scale of 1:10 000. The grid interval is assumed to be 10 to 50 m with an average error of 0.8 to 2 m. The data was based on aerial imaging and topographic maps. A triangulated irregular network map was used – this is a digital format of a continuous spatial data representation in which information about elevations are assigned to vertices of triangles and interpolated between them.

2.4.3. MPHP

The map of hydrographical divisions of Poland [28] is the basis for the information system of water management. The map which was used contains the details of river networks and bodies of water within the boundaries of the analysed catchments on a scale of 1:50 000.

2.4.4. Wastewater

Data concerning wastewater treatment plants located in the area of the analysed basins was obtained from the National Water Management Authority in Poland. The data contained detailed information including the geographical coordinates of any given wastewater treatment plant the amount of public wastewater treated within a year in thousands $m^3\ yr^{-1}$, total suspended solids (mg L^{-1}), total nitrogen (mg L^{-1}) and total phosphorus (mg L^{-1}).

2.4.5. Meteorology

Meteorological input data with a daily time step included solar radiation, wind speed, precipitation, relative moisture, and maximum and minimum temperatures. Table 1 shows the meteorological stations used in the models of the basins selected for analysis.

Table 1. Meteorological stations for the analysed basins

River basin	Number of weather stations used to build the model	Station names
Middle Warta	9	Koło, Piła, Gorzów Wlkp., Gniezno, Gorzyń, Pobiedziska, Nowa Wieś Podgórna, Poznań, Kalisz
Rega	7	Goleniów, Kołobrzeg, Łabędzie, Resko, Wierzchowo Pom., Starnin, Trzebiatów
Reda	9	Rozewie, Żelazno, Lębork, Wejherowo, Tepcz, Żelistrzewo, Rębiska, Gdynia, Gdańsk

2.4.6. Soil maps

Soil maps at a scale of 1:100 were obtained from the Institute of Soil Science and Plant Cultivation National Research Institute for the following types of soil: very light, light, average, heavy [23] (Tab. 2). Classification was made according to a granulometric soil group with data and system adopted by the IUNG-PIB.

Table 2. Percentages of the different soil classes found within the basin area [%][23]

Basin soil class	Middle Warta	Rega	Reda
very light	32.9	10.13	43.8
light	30.6	50.71	7.04
average	33.9	39.17	37.14
heavy	2.4	-	11.15

2.4.7. Land use

Land use maps of the Middle Warta and the Rega basin were created based on the CORINE LAND COVER information system [3, 9] which divides land use into six classes attributing relevant abbreviations to it which are acceptable and readable by the model (Tab. 3).

Table 3. Percentages of the different land use types found within the basin area [%] [9]

Basin land use types	Middle Warta	Rega	Reda
artificial surfaces	6.17	1.8	6.34
agricultural areas	72.82	54.5	38.91
forests	20.04	32.7	39.84
wetland areas	0.1	0.1	0.02
water bodies	0.85	1.2	0.71
meadows	-	9.7	14.18

2.4.8. Fertilisers

Input data used to calculate phosphorus loads from manure and mineral fertilisers were obtained from the Local Database [2] including information regarding livestock and the surface area of arable lands in hectares at the provincial level. The dominant crops with appropriate agricultural management operations are specified for each hydrological response unit (HRU) in the analysed basins. HRUs are comprised of unique land cover, soil and management combinations. Crops identified for basin areas are shown in Table 4. Each crop from Table 4 requires a different dose of fertiliser; however, doses were averaged for modelling purposes. The average dose of nitrogen fertilisers in the Middle Warta catchment was 156.6 kg N/ha, in the Rega catchment it was 85.6 kg N/ha, and in Reda catchment it was 105.9 kg N/ha.

Table 4. Crops in the analysed basins

Relevance for catchment			Crops
Middle Warta	Reda	Rega	
●	●	●	wheat
●	●	●	rye
●		●	barley
●	●	●	potatoes
●		●	sugar beets
●	●	●	rape
	●	●	oat

2.5. Sensitivity analysis and calibration, verification and validation processes

In further work, sensitivity analysis of the parameters in the model was carried out. The main purpose of applying sensitivity analysis was to define a set of parameters with the highest sensitivity, meaning those which have the greatest impact on the parameters affecting the flow and nitrogen load in the analysed calculation profile of the river. The method of sensitivity analysis in the ArcSWAT interface combines the Latin Hypercube (LH) and One-factor-At-a-Time (OFAT) sampling. A parameter is randomly selected and its value is changed from the previous simulation by a user-defined percentage for a number of LH loops. SWAT is run on the new parameter set and a different parameter is then randomly selected and varied.

Correct analysis of the modelling simulations is necessary in order to obtain reliable results. For this reason, it should be checked whether the modelling results fit the observed data during performance of the calibration, verification and validation processes, which are respectively defined as:

- adjustment of model parameters in order to obtain the greatest convergence of modelling results and observations conducted in calculation profile;
- checking at a measuring point whether the model is a good representation of the conceptual model – this is performed on independent data during the process of calibration conducted in calculation profile;
- final checking at a different measuring point to the calibration and verification point of whether the model is a good representation of reality by comparing the modelling results with observations conducted in a profile different from the calculation profile.

Observation data was obtained from State Environmental Monitoring (SEM) which, according to the law, is supposed to provide reliable information about the state of the environment. SEM leads analysis of water samples in selected monitoring sections of rivers in accordance with the scope and frequency specified in regulations [36, 37].

To evaluate the degree to which observations and simulations match, the following three statistical measures were applied: the coefficient of determination (R^2); the Nash-Satcliffe coefficient (NSE); the percent bias (PBIAS) [20, 29, 41]. The evaluation of the statistical measurements was conducted in accordance with the criteria shown in Table 5.

Table 5. Performance rating [27, 29, 32, 41]

Performance rating	NS	PBIAS		R^2
	flow/nutrients	nutrients	flow	flow/nutrients
very good	0.75 < NSE ≤ 1	< ±25	<± 10	0.64 < R^2
good	0.5 < NSE ≤ 0.75	±≤ 25 Pbias <± 40	±≤1 0 Pbias < ±15	0.49 < R^2 ≤ 0.64
satisfactory	0 < NSE ≤ 0.5	±40 ≤ Pbias <± 70	±15 ≤ Pbias< ±25	0.36 < R^2 ≤ 0.49
nonsatisfactory	NSE ≤ 0	Pbias ≥± 70	Pbias ≥ ±25	R^2 ≤ 0.36

For each of the selected pilot basins, a different body of data for the calibration, verification and validation. This was related to the availability of monitoring data over different periods

which was used to build models of the basin. In Table 6, information for each basin is shown regarding the analysis period, the monitoring point for which monitoring data was available from the SEM.

Table 6. Information regarding the analysis period, the monitoring point and the amount of available monitoring data for pilot basin [18, 48]

Phase Flow		The Middle Warta river		The Rega river		The Reda river	
		Total Nitrogen	Flow	Total Nitrogen	Flow	Total Nitrogen	Flow
Calibration	period of analysis	01.01.2000-31.12.2006		01.01.2001-31.12.2006		01.01.2002-31.12.2004	
	river monitoring point/ miles*	Poznań – Roch Bridge/127.45 miles		Trzebiatów/8.01 miles		Wejherowo/15.8 miles	
	total data quantity (TDQ)	2557		2191		1096	
	the data amount	2557	120	2191	157	1096	72
	percentage of TDQ	100%	5%	100%	7%	100%	7%
Verification	period of analysis	01.01.2007-31.12.2009		01.01.2007-31.12.2009		01.01.2005-31.12.2005	
	river monitoring point/ miles*	Poznań – Roch Bridge/127.45 miles		Trzebiatów/8.01 miles		Wejherowo/15.8 miles	
	total data quantity (TDQ)	1096		1096		365	
	the data amount	1096	25	1096	50	365	22
	percentage of TDQ	100%	2%	100%	5%	100%	6%
Validation	period of analysis	01.01.2003-31.12.2006		01.01.2002-31.12.2008		-	
	river monitoring point/ miles*	Oborniki/127.5 miles		Resko/33 miles		-	
	total data quantity (TDQ)	1461		2557		-	
	the data amount	1461	84	2557	48	-	-
	percentage of TDQ	100%	6%	100%	2%	-	-

In the case of the Reda river, validation of the model was not conducted due to a lack of observation data on another profile separate from the calculation profile (one river control profile).

3. Results and discussion

3.1. Sensitivity analysis results and the results of the calibration, verification and validation processes

For pilot basin models of the Middle Warta and Rega, within the functionality of SWAT, being in this case a DNS Macromodel module, a sensitivity analysis of the parameters associated with the flow and the total levels of nitrogen was conducted according to the description in Section 3.4. The results of this sensitivity analysis are presented in Table 7. For the Middle Warta basin, there are 16 parameters which are most sensitive and associated with the flow at the control point. For the Middle Warta basin, there are 16 parameters; for the Reda basin, there are 9 parameters; for the Rega basin, there are 11 parameters. For total nitrogen loads, from the range of parameters that may be manipulated during the calibration of the model, 9 parameters for the Middle Warta, 7 for Reda and 9 for Rega obtained the highest sensitivity. The parameters selected during the sensitivity analysis were used for the calibration of the model.

Table 7. The most sensitive parameters obtained from the sensitivity analysis in the SWAT model for the analysed basins [18, 48]

Relevance for basin			Parameter	Parameter description
Middle Warta	Reda	Rega		
1	2	3	4	5
	Flow parameters			
•	•	•	ALPHA_BF	base flow alpha factor [days]
•			BLAI	maximum potential leaf area index
•	•		CANMX	maximum canopy storage [mm H_2O]
•	•	•	CH_K(1)	effective hydraulic conductivity in tributary channel alluvium [mm/hr]
•	•	•	CH_K(2)	effective hydraulic conductivity in main channel alluvium [mm/h]
•	•	•	CN2	initial SCS runoff curve number for moisture condition II
•			EPCO	plant uptake compensation factor
•	•	•	ESCO	soil evaporation compensation factor
		•	GW_DELAY	ground water delay time [days]
•	•	•	GWQMN	threshold depth of water in the shallow aquifer required for return flow to occur [mm H_2O]

Tab. 7 (cont.)

1	2	3	4	5
●		●	GW_REVAP	groundwater 'revap' coefficient
	●	●	RCHRG_DP	deep aquifer percolation factor
●			SOL_ALB	moist soil albedo
		●	SOL_K	saturated hydraulic conductivity [mm/hr]
●		●	SURLAG	surface runoff lag coefficient
●	●		TIMP	snow pack temperature lag factor
	Total nitrogen parameters			
●			AL1	fraction of algal biomass that is nitrogen
●		●	BC1	rate constant of biological oxidation of NH_4 to NO_2 in the Reach at 20°C in well aerated conditions
●	●	●	BIOMIX	biological mixing efficiency
	●	●	CDN	rate factor for humus mineralisation of active organic nutrients (N and P)
●	●		CH_N2	Manning's n value for the main channel
	●	●	CMN	rate factor for humus mineralisation of active organic nutrients (N and P)
●	●	●	ERORGN	nitrogen enrichment ratio for loading with sediment
●	●	●	NPERCO	nitrate percolation coefficient
●	●	●	N_UPDIS	nitrogen up take distribution parameter
		●	RCN	concentration of nitrogen in rainfall
●			RS4	organic N settling rate coefficient
●		●	SDNCO	denitrification threshold water content

The results of the statistical measurement of the calibration, verification and validation processes are shown in Table 8. Due to the utilisation of three statistical coefficients, it is essential to define a hierarchy in the assessment of coefficients in order to introduce an unequivocal interpretation. Therefore, the coefficient of determination (R^2) was treated as a priority.

The analysis carried out with the use of three statistical measures showed that in the examples of the Middle Warta river and the Rega river, the assessment of matching the flow observations to simulations obtained from Macromodel DNS/SWAT, with regards to R^2, was very good in almost all cases in other words, there was a strong match between the observed and simulated flows. Only the calibration phase of the Reda river obtained a good value. For the remaining rivers, the flow provided good and satisfactory results for the NSE and PBIAS statistical measures. As far as total nitrogen loads are concerned, the strongest match between observations and simulations were obtained for the Middle Warta river and the Rega river, where for those rivers, the results oscillated between very good and satisfactory values. For The Reda river, the values for R^2 and PBIAS were respectively satisfactory and very good.

Only in the case of the NSE were results unsatisfactory for the calibration and verification phase – this was mainly due to the sensitivity of statistical measures on a small amount of available data. The results obtained thus confirm the high degree of agreement between the results of modelling and the data obtained through observation – this confirms that the model is a good representation of reality.

Table 8. Results for the model calibration, verification and validation phase for statistical measures with robust statistics for pilot rivers [18, 48]

Basin		Flow			Total nitrogen		
	Coefficient phases	R^2	NSE	PBIAS [%]	R^2	NSE	PBIAS [%]
Rega	Calibration	0.81	0.47	15.04	0.5	0	23.32
	Verification	0.87	0.75	-11.19	0.55	0.47	-13.06
	Validation	0.69	0	2.53	0.57	0.38	28.39
Middle Warta	Calibration	0.93	0.91	6.07	0.65	0.59	-0.44
	Verification	0.93	0.81	-0.84	0.81	0.57	0.14
	Validation	0.94	0.85	14.51	0.47	0.06	-0.58
Reda	Calibration	0.59	0.37	22	0.4	-1.13	12
	Verification	0.7	0.3	21	0.3	-2.37	23
	Validation	-	-	-	-	-	-

3.2. Analysis results

In order to analyse the size of the total nitrogen load flowing through the calculation profile of a given basin and the seasonal variability in the pilot basins which were calibrated, verified and validated with the Macromodel DNS/SWAT. This enables us to obtain coherent and comparable results for all three basins.

Cumulative curves of total nitrogen loads were produced (Figs. 3a-3c) in order to analyse the total amount of nitrogen compound that entered the calculation profile during the considered years, separately, as well as the average values for the analysed years which are marked with a black dashed line (Figs. 3a-3c).

The cumulative mass curve of the size of the total nitrogen load indicates changes in the total volume of the load which flowed from the beginning of the period in the calculation profile. As far as the Rega and the Middle Warta river are concerned, a reduction in the size of the total nitrogen load can be observed from March to April followed by a decrease and stabilisation of the total nitrogen load and again from November to December, followed by increased loads. In the case of the Reda river basin, these relationships are not so clear; however, a minimal reduction of the total nitrogen load in the months April to May and an increase after the months of October to November can be observed. In order to conduct an accurate analysis of changes in the total size of the average total daily nitrogen load within

the period 2002-2005 (Figs. 3a-3c, black dashed line), further detailed analyses were carried out. To this end, the average daily total nitrogen load from the 2002-2005 period was collated and analysed with the flow and precipitation in the growing season (in Poland, this lasts from around early March to around the end of October). Analyses of the period relating to the plant growing season is important because it has a significant influence on changes in the size of total nitrogen loads in surface waters. The analysis of seasonal changes in the total nitrogen loads for the analysed period showed that at the beginning of the plant growing season, the total nitrogen loads is the highest and decreases during the plant growing season until reaching the lowest value (Figs. 4-6). The highest value on the graph relating to the average daily total nitrogen load from the multi-year period is indicated by 'MAX' and the lowest value is indicated by 'MIN' for each of the analysed basins (Figs. 4-6). After the end of the growing season, the size of the total nitrogen loads in the calculation profile increases again. The size of the total nitrogen loads in the calculation profiles of the pilot rivers shows a decrease and stabilisation during the plant growing season. This phenomenon was observed for all analysed basins, and is called the *flattening phenomenon* this is defined as a periodic decrease and stabilisation of the size of the total nitrogen loads in the calculation profile during the plant growing season. The maximum and minimum loads of total nitrogen determines the precise boundaries of the *flattening*. These dependencies are illustrated in Figs. 4-6 with an indication of the average flow, precipitation, and a period of plant growing season in reference to total nitrogen load and total nitrogen concentration.

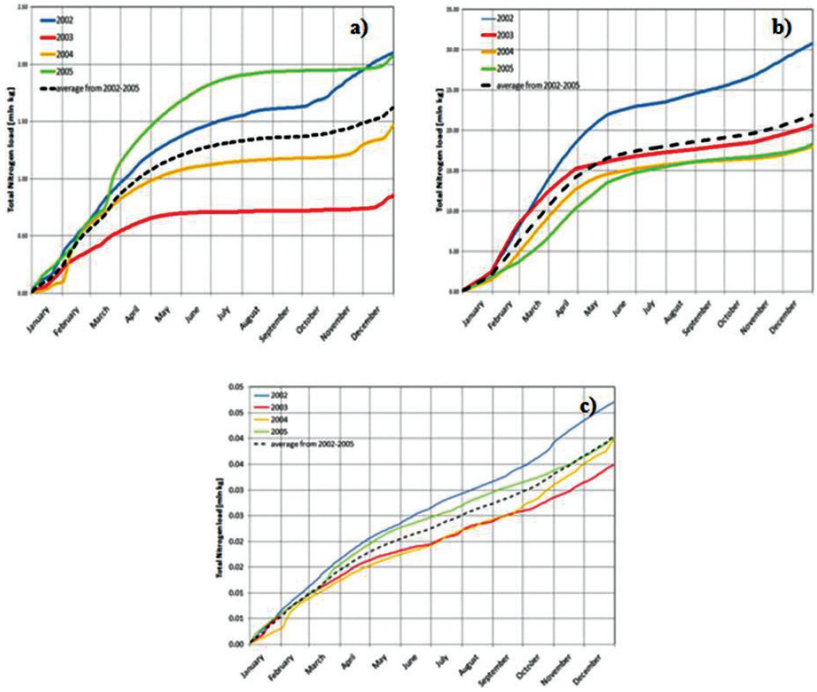


Fig. 3. Cumulative curves of the total nitrogen in the calculation profile of the Rega river (a), the Middle Warta river (b) and the Reda river (c) for separate years and the average cumulative curve (black dashed line) [own study]

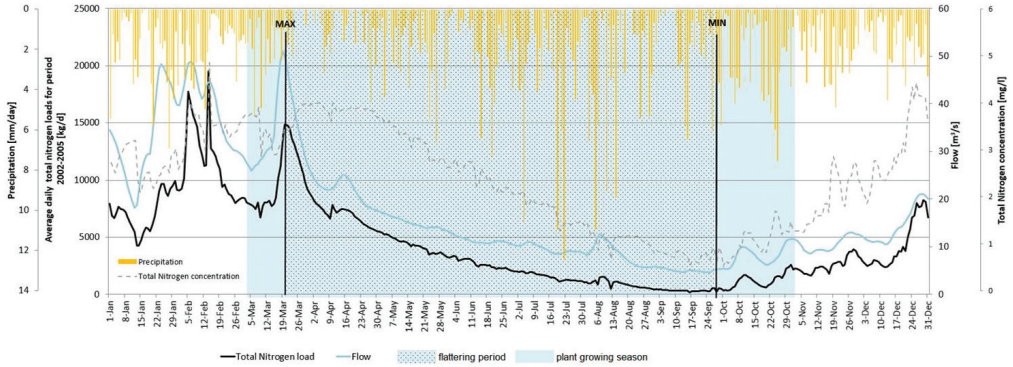


Fig. 4. Total nitrogen load, precipitation and flow rate with marked flattening period (MIN – MAX) and plant growing season in the Rega river basin for mean values for the period from 01/01/2002 to 31/12/2005 [own study]

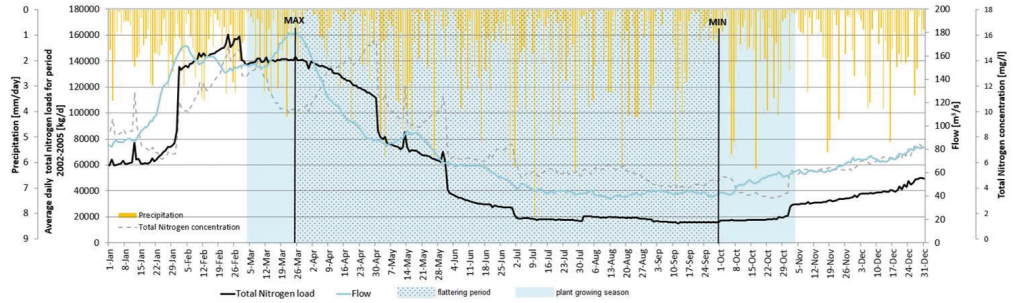


Fig. 5. Total nitrogen load, precipitation and flow rate with marked flattening period (MIN – MAX) and plant growing season in the Middle Warta river basin for mean values for the period from 01/01/2002 to 31/12/2005 [own study]

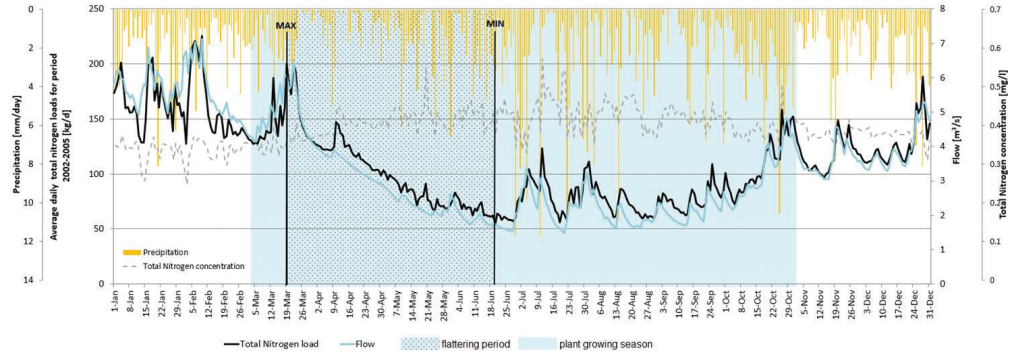


Fig. 6. Total nitrogen load, precipitation and flow rate with marked flattening period (MIN – MAX) and plant growing season in the Reda river basin for mean values for the period from 01/01/2002 to 31/12/2005 [own study]

Additionally, total nitrogen concentrations were analysed yearly. For the Rega and Middle Warta basins, the tendency observed for total nitrogen loads is similar for total nitrogen concentrations. The concentration of total nitrogen decreased during the vegetative season.

Consistent analysis of loads and concentrations proves the phenomena of the reduction of total nitrogen presence in nutrient balance during the vegetative season.

The study of Figures 4, 5 and 6 proves that the water retention process and the total nitrogen loads during the vegetative periods of plants is explicit for the Rega and Middle Warta rivers. In the Reda river basin, the *flattening phenomenon* based on loads analyses is less noticeable and shorter. Additionally, the total nitrogen concentration analysis shows that concentrations are nearly stable across the whole year. In the vegetative season, however, concentrations increase slightly in spite of the flow decrease.

Despite the fact that the total sum of precipitation for the periods of particular *flattening* between 01/01/2002 and 31/12/2005 is higher during the winter time (apart from the growing season) – the average flow in the rivers during this period is lower. In addition, the average amount of total nitrogen loads is lower during the *flattening* periods rather than outside this period as shown in Tables 9, 10 and 11. This has been acknowledged by the significant contribution of plants in the basin, including arable crops, in the retention of water and nitrogen compounds.

It is widely known that the surface run-off of nutrients from agricultural land depends on the retention capacity and erosivity of these areas. The retention capacity of agricultural areas is affected by cultivation and appropriate agricultural practices. In numerous studies [22, 35, 41, 42, 44], seasonal variation of loads discharged into the receivers was also observed. Nevertheless, analysis is required with regard to the cause of these relationships and for the determination of the cause of this variability in the context of changes in both hydrological and agricultural conditions.

Table 9. Comparison of average total nitrogen loads, total precipitation and average flow for the *flattening* period from 01/01/2002 to 31/12/2005 in the Rega basin (**flattening* period)

Period (dd-mm)	**Units**	**20/03 (MAX)-28/09 (MIN)***	**29/09-19/03**	**years/period 2002-2005**
Total precipitation	[mm]	363.7	345.2	**sum/total** 708.9
Flow	[m^3/s]	13.4	46.1	**average** 17.4
Nitrogen loads	[kg/d]	3389	11781.3	**average** 4425.2

Table 10. Comparison of average total nitrogen loads, total precipitation and average flow for the *flattening* period from 01/01/2002 to 31/12/2005 in the Middle Warta basin (**flattening* period)

Period (dd-mm)	**Units**	**25/03 (MAX)-29/09 (MIN)***	**30/09-24/03**	**years/period 2002-2005**
Total precipitation	[mm]	287.8	262.3	**sum/total** 556.6
Flow	[m^3/s]	69.1	198.7	**average:** 83.3
Nitrogen loads	[kg/d]	50 651.3	142691.1	**average:** 59 934.4

Table 11. Comparison of average total nitrogen loads, total precipitation and average flow for the *flattening* period from 01/01/2002 to 31/12/2005 in the Reda basin (**flattening* period)

Period (dd-mm)	**Units**	**19/03 (MAX)-20/06 (MIN)***	**21/06-18/03**	**years/period 2002-2005**
Total precipitation	[mm]	133.3	373.2	**sum/total** 739.1
Flow	[m^3/s]	2.9	9.3	**average:** 3.4
Nitrogen loads	[kg/d]	102	275.3	**average:** 110.6

Numerous studies conducted worldwide suggest ground-cover vegetation as a factor influencing the run-off of nutrients into surface water [4, 7, 15, 26, 50]. In countries located in areas of temperate climatic conditions, with clearly marked seasons and thus a relatively short vegetation period, the problem of increases in the concentration of total nitrogen in the surface run-off of arable land at the end of the vegetation period is an important issue to be taken into account when assessing the status of surface waters flowing in agricultural catchments.

The consumption of nutrients (total nitrogen) by the plants with regard to the size of the surface run-off of these nutrients was analysed in [12]. Scenarios of changes in size of trailing with respect to crops were created. In article [45], it was observed that the run-off of nitrogen compounds in the periods known as 'fallow' periods was 30% higher than in the cropping period; however, this was not referenced to the volume of loads on the calculation profiles.

Publication [46] shows an analysis of the nutrient run-off to the eastern Santa Barbara Channel. The authors evaluated the river loading and the dispersal of dissolved macronutrients and commented on the biological implications of these nutrient contributions. The objective of research on the biological impact nutrient loads was to analyse algal responses and upwelling. It was difficult to isolate the impact of algal uptake on nutrient loads because of the run-off increase caused by the influence of changing meteorological conditions. Further research was needed.

In 2005, a paper on modelling nutrient export was published. In the publication [39], the authors describe the relationship between the type of land use and the amount of nutrients discharged to surface waters. Analysis of the variability of nutrients during the year was conducted, depending upon the occurrence of stormy meteorological conditions. The article noted the phenomenon of load stabilisation, which appears in the summer months, but no attempt was made to analyse the reasons for its creation. Additionally, the presented research results clearly indicated large differences in the surface confluence of nutrients from land upon which the use varied depending upon the season.

Monitoring data from field studies conducted at other polish rivers also shows the occurrence of the described periodic lowering of the concentration of nitrogen in surface waters. In 2011, the Institute of Soil Science and Plant Cultivation carried out studies on, inter alia, the impact of cover crops (legumes) on nitrate leaching from the soil. Studies have

shown that the role of this type of plant was significant in terms of water quality and that their presence, to some extent, conditioned the concentration of nitrates in surface waters [16].

Decreases in the concentrations of nitrogen during the plant growing season and its subsequent rapid increases both in the soil and in nearby surface waters was also observed by the Institute of Technology and Life Sciences in Falenty during the analysis of selected watersheds [34].

The flattening phenomenon was also illustrated through the analysis of the content of nitrogen compounds in the waters of the Trzema river tributary of the Prosna river) [10] – collected monitoring data from several years confirms the occurrence of this phenomenon. At the same time in the indicated publication shows the differences between the behavior of total nitrogen and total phosphorus for which the flattening phenomenon does not occur. The significant influence of the plant growth period on the amount of nitrogen in the surface waters of the catchment areas dominated by agricultural land use was also observed on the San and Slina rivers [25]; however, it should be noted that these studies are based on small amounts of monitoring data and they do not specify either the scale of the decline in the amount of nitrogen concentration in surface waters or its duration – this information is crucial in terms of the possibilities of applying knowledge of the flattening phenomenon in, for example, the preparation of scenarios for action plans.

The described flattening phenomenon also confirms the need for change, which is used for environmental computations, for the flow of interest. At present, the so-called environmental flow should be used, which will take into account changes in the amount of pollutants depending on the seasons [24].

The *flattening phenomenon* is mainly a result of agricultural activities in the basin. At the beginning of the plant-growing season, high doses of fertiliser are applied on arable lands, this results in an increase of nutrients leaching to surface waters from areas without plant soil cover. During the plant growing season, a rapid increase of nitrogen uptake occurs – this is a consequence of the plants embedding it within their own cellular structures. Moreover, through the creation of soil cover, plants significantly reduce the nitrogen leaching which occurs through surface run-offs which result from intensive rains. In Table 7 (for the Rega basin) and in Table 8 (for the Middle Warta basin), the results show that during the *flattening* period (March to September) the total precipitation is higher than after the *flattening* period (October to March), and the size of the average flow is diminished by as much as 70%. This is confirmed by the total nitrogen concentration decrease. At the end of the vegetation period, the intense nitrogen uptake by plants decreases. At this time, nitrogen compounds located in the cells are redistributed inside the plant from leaves to stems and reproductive systems causing a continuous increase in biomass without up take from soils. In addition, harvesting crops from agricultural land (tantamount to biomass removal) deprives soil of plant cover – this causes an increase in surface run-off. At this time, a re-increase of nitrogen loads in surface waters is observed. These correlations have been confirmed in a separate dissertation [18], in which the effect of human agricultural pressure on the amount of nutrients and total suspended soils entering surface waters were analysed.

Analyses of the Reda basin show that the *flattening* occurs here as well but is not as visible as in the previously mentioned basins and lasts from mid May to mid July (Fig. 6). Although

additional analyses of concentrations proved that *flattening* does not occur. At this time, the size of total nitrogen loads in surface waters fluctuates, although total nitrogen concentration slightly increases. In order to find the reasons for this situation, the average content of mineral nitrogen in Polish soils [14] was compared with its content in the soils of the Reda basin (Tab. 12).

Table 12. Quantity of mineral nitrogen in the layers of Polish soils and in the Reda basin [14]

Soil profile	Average for Poland		Reda basin	
	Quantity of mineral nitrogen (NO3-N + NH4-N) [mg/kg]			
	Spring	Autumn	Spring	Autumn
0-30	9.0	11.6	16.7	19.4
30-60	6.7	6.9	8.2	11.6
60-90	5.8	5.1	8.9	10.2

It is noticeable that the soils in the Rega basin contain higher concentrations of mineral nitrogen than the rest of the country – this can have a significant impact on the size of the retention and thus the *flattening* effect. For both spring and winter, in all soil layers in the Reda basin, the amount of mineral nitrogen is much higher than the average values for Poland – by 70% on average. This situation might be caused by the long-term, excessive fertilisation of soils in the area, which might have caused the infiltration of nitrogen compounds into deeper soil layers. Nitrogen compounds accumulated in the deeper layers are relatively evenly leached to the river by subsurface run-off. Subsurface run-off, as opposed to surface run-off, is not as dynamic and is not associated with the time of the year and growing season – this can explain the lack of seasonal variability of nitrogen loads in the Reda river and the limited variability of nitrogen loads during the particular period. This observation shows that the absence of flattening can suggest an imbalance in the nutrient cycle in the basin.

4. Conclusions

The retention of vast amounts of precipitation waters by basin plants is vital as far as the diminishing quantities of total nitrogen loads in surface waters is concerned – this was confirmed by the results from the examined basins of the Rega and the Middle Warta rivers. When the growing season ends and the crops are removed from the fields, a period of increased total nitrogen concentration appeared at the analysed SWB calculation profiles in the Rega and the Middle Warta – this is mainly caused by surface run-off. The analyses indicate that crops have a crucial influence on the quantity of water retention and nitrogen compounds in the examined basin; simultaneously, they vastly influence the quantity of the surface run-off and the included nitrogen loads [18]. Water and nitrogen retention in the basin, even short term, is a desirable phenomenon since it is connected with the process of water regeneration. Withholding precipitation water in the soil and at the same time restricting the run-off of nitrogen originating from fertilisation, enables better use of fertilisers by plants. The *flattening phenomenon* does not perpetually occur and it greatly depends on the level of nitrogen

concentration present in the soil and the dynamics of the processes of its leaching to the waters. An example of a basin where nitrogen balance is disturbed by long-term fertilisation is the Reda basin. This case requires further analysis including the quantity of mineral nitrogen in the soils of the basin. The *flattening phenomenon* describes the dynamics of total nitrogen loads to surface waters and at the same time, it can be used to evaluate the conditions that make water eutrophication occur. Since the majority of total nitrogen in surface waters of Poland originates from farming [18] and is introduced by surface run-off, the analyses require information relating to the amounts of fertiliser used on the fields' crops. It has to be remembered that the temporary retention of total nitrogen in the soil does not lead to its total removal from the environment. It only helps plants to make better use of it on condition that the quantities of applied fertilisers are sufficient. The magnitude of *flattening* allows us to specify the relationship between the amount of nitrogen loads exported to waters in particular seasons, the occurrence of violent exceptional climatic events and periods of increase in the acquisition of these ingredients substances by plants. The lack of *flattening phenomenon* might prove that there is a strong correlation between the levels of nutrients consumed by plants and the leaching of pollutants to surface waters; therefore, it is very crucial to use adequate doses of fertilisers to meet the needs of plants, but this has to be preceded by an analysis of the chemical composition of the soil in order to establish the availability of nutrients. This enables us to estimate the doses of fertilisers that would allow a high *growth rate* of plants, and at the same time, reduces the excess of unused nitrogen leaching to surface waters and consequently limiting the eutrophication process. The *flattening phenomenon* can be used to assess the state of the ecosystem in the basin area and provides the opportunity to assess the actual impact of land use on the nitrogen cycle.

References

[1] Abbaspour K.C., *SWAT-CUP2: SWAT Calibration and Uncertainty Programs – A User Manual*, Department of Systems Analysis, Integrated Assessment and Modelling, Eawag, Swiss Federal Institute of Aquatic Science and Technology, Duebendorf 2008.

[2] BDL – Polish Local Database, http://stat.gov.pl/bdl/app/strona.html?p_name=indeks (access: 20.05.2015).

[3] Bossard M., Feranec J., Otahel J., *CORINE land cover technical guide – Addendum 2000*, Technical report No. 40, European Environment Agency, Copenhagen 2000.

[4] Burwell R.E., Timmons D.R., Holt R.F., *Nutrient Transport in Surface Runoff as Influenced by Soil Cover and Seasonal Periods1*, Soil Sci. Soc. Am. J., 39, 1975, 523-528.

[5] Byczkowski A., *Hydrologia*, vol. I, Warsaw University of Life Sciences, Warszawa 1999.

[6] Cai Y., Guo L., Douglas T., Whitledge T., *Seasonal variations in nutrient concentrations and speciation in the Chena River, Alaska*, Journal of Geophysical Research, Vol. 113, 2008.

[7] Causse J., Baurès E., Mery Y., Jung A.V., Thomas O., *Variability of N export in water: a review*, Critical Reviews in Environmental Science and Technology, 2015.

[8] Centralny Ośrodek Dokumentacji Geodezyjneji Kartograficznej – national databases (Centre of Geodesic and Cartographic Documentation), www.codgik.gov.pl (access: 6.06.2015).
[9] CORINE, CORINE LAND COVER, http://www.eea.europa.eu/themes/landuse/interactive/clc-download, July (access: 6.06.2015).
[10] Dabrowska J., *Ocena zawartości związków azotu i fosforu w wodach rzeki Trzemny*, Infrastruktura i ekologia terenów wiejskich, (07) 2008.
[11] Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy, 2000.
[12] Douglas C.L., King K.A., Zuzel J.F., *Nitrogen and Phosphorus in Surface Runoff and Sediment from a Wheat-Pea Rotation in Northeastern Oregon*, J. Environ. Qual., 27, 1998, 1170-1177.
[13] Fal B., *Zmienność odpływu z obszaru Polski w bieżącym stuleciu*, Institute of Meteorology and Water Management – notifications, 3, 1993, 3-19.
[14] Fotyma M., Kęsik K., Pietruch Cz., *Azot mineralny w glebach Polski*, Fertilizers and Fertilization, No. 38, Institute of Soil Science and Plant Cultivation – State Reasearch Institute, Puławy 2010.
[15] Gafur A., Jensen J., Borggaard O., Petersen L., *Runoff and losses of soil and nutrients from small watersheds under shifting cultivation (Jhum) in the Chittagong Hill Tracts of Bangladesh*, Journal of Hydrology, Vol. 274, 2003, 30-46.
[16] Gaweł E., *Rola roślin motylkowych drobnonasiennych w gospodarstwie rolnym*, Woda – *Środowisko* – Obszary Wiejskie, 11, 2011, 73-91.
[17] Gębala J., Orlińska-Woźniak P., Wilk P., *Zanieczyszczenia związkami azotu pochodzenia rolniczego wód powierzchniowych w Polsce – wybrane problemy oceny jakości wód*, Gospodarka Wodna, 11, 2013.
[18] Gębala J., *Method of assessing the impact of agricultural anthropopression on surface water quality on example of the Rega basin*, PhD thesis supervising by M. Ostojski, Institute of Meteorology and Water Management – National Research Institute, Warsaw 2015.
[19] Grabińska B., Kuc J., Glińska-Lewczuk K., *Zanieczyszczenia obszarowe w zlewniach rolniczych i leśnych – możliwości ograniczania. Wpływ użytkowania zlewni Narwi na zagrożenie wód związkami azotu*, 2004, 164.
[20] Gupta H.V., Sorooshian S., Yapo P.O., *Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration*, J. Hydrologic Eng., 4(2), 1999, 135-143.
[21] HELCOM, Stakeholder Conference on the Baltic Sea Action Plan, Eutrophication in the Baltic Sea, Draft HELCOM Thematic Assessment in 2006, Helsinki 2006.
[22] Hobot A., Banaszak K., Stolarska M., Serafin R., Stachura A., *Warunki korzystania z wód zlewni rzeki Redy – Etap 1 – Dynamiczny bilans zasobów wodnych*, The Regional Water Management Authority in Gdańsk, Gdańsk 2012.
[23] IUNG 2009, Mapy kategorii glebowych Ministerstwa Rolnictwa i Rozwoju Wsi, Institute of Soil Science and Plant Cultivation – State Reasearch Institute, http://www.susza.iung.pulawy.pl/index.html?str=mapkat (access: 1.06.2015).

[24] Kiryluk A., Rauba M., *Zmienność stężenia związków azotu w różnie użytkowanej zlewni rolniczej rzeki Ślina*, Woda – *Środowisko* – Obszary Wiejskie, 9, 2009, 71-86.
[25] Kuźniar A., Twardy S., Kowalczyk A., *Przyczyny zmian stężenia azotu i fosforu w wodach powierzchniowych górnej zlewni Sanu (po przekrój w Przemyślu) w latach 1990-2005*, Woda – *Środowisko* – Obszary wiejskie, 8, 2008, 185-196.
[26] Li J., Okin G., Alvarez L., Epstein H., *Quantitative effects of vegetation cover on wind erosion and soil nutrient loss in a desert grassland of southern New Mexico*, Biogeochemistry, Volume 85, Issue 3, 2007, 317-332.
[27] Maréchal D., Holman P., *Comparison of Hydrologic Simulations using Regionalised and Basin-Calibrated Parameter Sets for three Basins in England*, Institute of Water and Environment, Cranfield University, UK, 2004.
[28] MPHP, The Map of Hydrographical Divisions of Poland, Institute of Meteorology and Water Management – National Research Institute, Warszawa 2009.
[29] Moriasi D.N., Arnold J.G., *Model evaluation guidelines for systematic quantification of accuracy in watershed simulations*, American Society of Agricultural and Biological Engineers, 2007.
[30] Neitsch S.L., Arnold J., Kiniry R., Srinivasan R., Williams J.R., *Soil and Water Assessments Tool Input/Output File Documentation*, Blackland Research Center Texas Agricultural Experiment Station, 2004.
[31] Neitsch S.L., Arnold J., Kiniry R., Williams J.R., *Soil and Water Assessments Tool Theoretical documentation*, Blackland Research Center Texas Agricultural Experiment Station, 2005.
[32] Ostojski M.S., *Modelowanie procesów odprowadzania do Bałtyku związków biogennych na przykładzie azotu i fosforu ogólnego*, Wydawnictwo Naukowe PWN, Warszawa 2012.
[33] Ozga-Zielińska M., Brzeziński J., *Hydrologia stosowana*, Wydawnictwo Naukowe PWN, Warszawa 1994.
[34] Pietrzak S., *Priorytetowe środki zaradcze w zakresie ograniczania strat azotu i fosforu z rolnictwa w aspekcie ochrony jakości wody*, Instytut Technologiczno-Przyrodniczy, 2012.
[35] Przywara L., *Warunki i możliwości usuwania fosforanów i fosforu ogólnego ze ścieków przemysłowych*, Ph.D. desideration, Bielsko-Biała 2006.
[36] Regulation of the Minister of Environment on the forms and manner of monitoring of surface water bodies and groundwater bodies, Journal of Laws, No. 258/1550, 2011.
[37] Regulation of the Minister of the Environment on the classification of surface water bodies and environmental quality standards for priority substances, Journal of Laws, No. 257/1545, 2011.
[38] Report on the state of the environment in Western Pomeranian in years 2006-2007.
[39] Robinson H.T., Leydecker A., Keller A., Melack J.M., *Steps towards modeling nutrient export in coastal Californian streams with a Mediterranean climate*, Agricultural Water Management, 77, 2005, 144-158.
[40] RZGW – The Regional Water Management Authority in Szczecin, Charakterystyka ogólna, 2010, http://www.rzgw.szczecin.pl/ogolna-charakterystyka/index/id/129/print_site/1 (access: 2.06.2015).

[41] Sarma P.B., Delleur J.W., Rao A.R., *Comparison of rainfall-runoff models for urban areas*, Journal of Hydrology, Volume 18, Issues 3-4, 1973, 329-347.
[42] Srinivasan R., Hadley J., Uhlenbrook S., Van Griensven A., Holvoet K., Bauwens W., European SWAT summer school, UNESCO – IHE, Institute for Water Education, 2006.
[43] Srinivasan R., *Soil and Water Assessment Tool, Introductory Manual – teaching materials*, Texas, 2011.
[44] Sullivan D.M., Hart J.M., Christensen N.W., *Nitrogen Uptake and Utilization*, A Pacific Northwest Extension Publication, Oregon–Idaho–Washington 1999.
[45] Udawatta R.P., Motavalli P.P., Garrett H.E., Krstansky J.J., *Nitrogen losses in runoff from three adjacent agricultural watersheds with claypan soils*, Agriculture, Ecosystems & Environment, Elsevier, 2006.
[46] Warric J.A., Washbum L., Brzezinski M., Siegel D., *Nutrient contributions to the Santa Barbara Channel, California, from the ephemeral Santa Clara River*, Estuarine Coastal and Shelf Science, 62, 2005, 559-574.
[47] Watson C., Mills C., *Gross nitrogen transformations in grassland soils as affected by previous management intensity*, Soil Biology and Biochemistry, Vol. 30, 6, 1998, 743-753.
[48] Wilk P., *The method of calculating the absorption rate of the river as a tool to evaluate the physico-chemical state of surface water flowing*, Ph.D thesis supervising by M. Ostojski, Institute of Meteorology and Water Management – National Research Institute, Warszawa 2015.
[49] Youssef T., Skaggs M., Amatya R.D., *Temporal variations and controlling factors of nitrogen export from an artificially drained coastal forest*, Environmental Science & Technology, 46(18), 2012, 9956-9963.
[50] Zhang G., Liu G., Wang G., Wang Y., *Effects of Vegetation Cover and Rainfall Intensity on Sediment-Bound Nutrient Loss, Size Composition and Volume Fractal Dimension of Sediment Particles*, Pedosphere, Vol. 21, 2011, 676-684.
[51] Zwolsman J., *Seasonal Variability and Biogeochemistry of Phosphorus in the Scheldt Estuary, south-west Netherlands*, Estuarine, Coastal and Shelf Science, Vol. 39, Issue 3, September, 1994, 227-248.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
MATHEMATICS
DOI: 10.4467/2353737XCT.17.194.7423

Maciej Zakarczemny (mzakarczemny@pk.edu.pl)
Institut of Matchematics, Faculty of Physis, Mathematics and Computer Science, Cracow University of Technology

Lower and upper bounds for solutions of the congruence $x^m \equiv a \pmod{n}$

Dolne oszacowanie na największe i górne oszacowanie na najmniejsze rozwiązanie kongruencji $x^m \equiv a \pmod{n}$

Abstract

Let n, m be natural numbers with $n \geq 2$. We say that an integer a, $(a, n) = 1$, is the m-th power residue modulo n if there exists an integer x such that $x^m \equiv a \pmod{n}$. Let $C(n)$ denote the multiplicative group consisting of the residues modulo n which are relatively prime to n. Let $s(n, m, a)$ be the smallest solution of the congruence $x^m \equiv a \pmod{n}$ in the set $C(n)$. Let $t(n, m, a)$ be the largest solution of the congruence $x^m \equiv a \pmod{n}$ in the set $C(n)$. We will give an upper bound for $s(n, m, a)$ and a lower bound for $t(n, m, a)$.

Keywords: smallest solution, largest solution, upper bound, lower bound, congruence relation, residue class, n-th degree equation

Streszczenie

Niech n, m będą liczbami naturalnymi, takimi że $n \geq 2$. Powiemy, że liczba całkowita a, $(a, n) = 1$, jest m-tą resztą kwadratową modulo n, jeśli istnieje liczba całkowita x, taka że $x^m \equiv a \pmod{n}$. Niech $C(n)$ będzie grupą multiplikatywną zawierającą reszty modulo n, względnie pierwsze z n. Oznaczmy przez $s(n, m, a)$ najmniejsze rozwiązanie równania $x^m \equiv a \pmod{n}$ w zbiorze $C(n)$. Oznaczmy przez $t(n, m, a)$ największe rozwiązanie równania $x^m \equiv a \pmod{n}$ w zbiorze $C(n)$. Podamy górne oszacowanie na $s(n, m, a)$ oraz dolne na $t(n, m, a)$.

Słowa kluczowe: najmniejsze rozwiązanie, największe rozwiązanie, górne oszacowanie, dolne oszacowanie, kongruencja, klasa reszt, równanie wielomianowe

1. Introduction

Let n, m be natural numbers with $n \geq 2$. Let a be an integer, with $(a, n) = 1$. By $s(a, n, m)$, $t(a, n, m)$ we denote, correspondingly, the smallest and largest solutions of the congruence $x^m \equiv a \pmod{n}$, where $1 \leq x \leq n - 1$. We will give an upper bound for $s(n, m, a)$ and a lower bound for $t(n, m, a)$. Let $C(n)$ denote the multiplicative group consisting of residues modulo n, which are relatively prime to n (reduced set of residues modulo n).

Let $C_k(n)$ denote the subgroup of $C(n)$ consisting of k-th powers.

Denote $v_k(n) = [C(n): C_k(n)]$. Let n have prime factorization $n = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_r^{a_r}$, where $a_j \geq 1$. By [2]:

$$v_k(n) = v_k\left(p_1^{a_1}\right) \cdot v_k\left(p_2^{a_2}\right) \cdot \ldots \cdot v_k\left(p_r^{a_r}\right),$$

$$v_k(2) = 1,\ v_k(2^\alpha) = (k, 2)(k, 2^{\alpha-2}),\ \text{for } \alpha \geq 2.$$

If p is an odd prime and $\alpha \geq 1$, then $v_k(p^\alpha) = (k, \varphi(p^\alpha))$. Also $v_k(n) \leq 2k^r$.

Definition 1.1. Let

$$1 = g_0(n,k) < g_1(n,k) < \ldots < g_{v-1}(n,k), \tag{1}$$

be the smallest positive representatives of the $v = v_k(n)$ cossets of $C_k(n)$.

Definition 1.2. Let

$$w_0(n,k) < w_1(n,k) < \ldots < w_{v-1}(n,k) = n - 1, \tag{2}$$

be the largest positive representatives of the $v = v_k(n)$ cossets of $C_k(n)$.

By Norton [2] we have:

Theorem 1.3. If n, k are positive integers $0 \leq i \leq v - 1$, then

$$g_i(n,k) \leq 1 + \frac{n}{\varphi(n)}\left(2^{\omega(n)}\right)^{\frac{3}{2}}\left(\frac{iv}{v-i}\right)^{\frac{1}{2}} n^{\frac{1}{2}} \log n, \tag{3}$$

where $\omega(n)$ is the number of distinct prime divisors of n.

See [2].

Corollary 1.4. For each $\varepsilon > 0$

$$g_{v-1}(n,k) = O\left(n^{\frac{1}{2}+\varepsilon}\right), \tag{4}$$

where the implied constant depends only on k, ε and the number of distinct prime factors of n. See [2].

Corollary 1.5. If p, q, r are odd distinct prime numbers and α, β, γ are positive integers, then

$$g_{v-1}(n,k)<\begin{cases}1+3\sqrt{2}k\sqrt{n}\log n & \text{if} \quad n=p^{\alpha},\\ 1+24k\sqrt{n}\log n & \text{if} \quad n=2p^{\alpha},\\ 1+8\sqrt{2}k\sqrt{n}\log n & \text{if} \quad n=2^{\alpha},\alpha\geq 2,\\ 1+15k^{2}\sqrt{n}\log n & \text{if} \quad n=p^{\alpha}q^{\beta},\\ 1+35\sqrt{2}k^{3}\sqrt{n}\log n & \text{if} \quad n=p^{\alpha}q^{\beta}r^{\gamma}.\end{cases} \tag{5}$$

Proof. By theorem 1.3.

2. Theorems

The following theorem shows the relationship between $g_i(n,k)$ and $w_{v-1-i}(n,k)$.

Theorem 2.1. For $0 \leq i \leq v-1$, we have

$$g_i(n,k)+w_{v-1-i}(n,k)=n. \tag{6}$$

Proof. Let us note that

$$n-g_i(n,k)\in g_j(n,k)C_k(n) \quad \text{iff} \quad g_i(n,k)\in(n-g_j(n,k))C_k(n), \tag{7}$$

where $0 \leq i, j \leq v-1$.

We define a permutation $\sigma:\{0,1,\ldots,v-1\}\to\{0,1,\ldots,v-1\}$ by the relation

$$\big(n-g_i(n,k)\big)C_k(n)=g_j(n,k)C_k(n)=w_{\sigma(i)}(n,k)C_k(n). \tag{8}$$

Then by definition of $w_{\sigma(i)}(n,k)$ we have

$$n-g_i(n,k)\leq w_{\sigma(i)}(n,k). \tag{9}$$

On the other hand

$$\Big(n-w_{\sigma(i)}(n,k)\Big)C_k(n)=\Big(n-g_j(n,k)\Big)C_k(n)=g_i(n,k)C_k(n). \tag{10}$$

Hence by definition of $g_i(n, k)$ we get

$$g_i(n,k) \le n - w_{\sigma(i)}(n,k). \tag{11}$$

Therefore by (9), (11)

$$g_i(n,k) + w_{\sigma(i)}(n,k) = n. \tag{12}$$

Using (1) we obtain

$$w_{\sigma(\nu-1)}(n,k) < w_{\sigma(\nu-2)}(n,k) < \ldots < w_{\sigma(1)}(n,k) < w_{\sigma(0)}(n,k) = n-1, \tag{13}$$

hence by (2)

$$\sigma(i) = \nu - 1 - i, \tag{14}$$

and we are finished.

Using theorem 1.3 and theorem 2.1 we get the following lower bound on $w_i(n, k)$.

Theorem 2.2. If n, k are positive integers $0 \le i \le \nu - 1$, then

$$w_i(n,k) \ge n - 1 - \frac{n}{\varphi(n)}\left(2^{\omega(n)}\right)^{\frac{3}{2}}\left(\frac{(\nu-1-i)\nu}{i+1}\right)^{\frac{1}{2}} n^{\frac{1}{2}} \log n. \tag{15}$$

Proof. By theorem 2.1 and theorem 1.3.

It follows that

Remark 2.3.

$$n - w_0(n,k) = O\left(n^{\frac{1}{2}+\varepsilon}\right), \tag{16}$$

for each $\varepsilon > 0$, where the implied constant depends only on k, ε and the number of distinct prime factors of n.

Finally, in the proof of the following theorem, we will show how to reduce the problem of finding bounds for $s(a, n, m)$, $t(a, n, m)$ to the problem of finding bounds for $g_i(n, k)$ and $w_i(n, k)$.

Theorem 2.4. Let n, m be natural numbers such that $n \ge 2$. Let a be an integer relatively prime to n, which is m-th power residue modulo n. By $s(a, n, m)$, $t(a, n, m)$ we denote, correspondingly, the smallest and the largest solution of the congruence

$$x^m \equiv a \left(\mathrm{mod}\, n\right), \tag{17}$$

where $1 \le x \le n - 1$. Then

$$s(a,n,m)\le 1+\frac{n}{\varphi(n)}\left(2^{\omega(n)}\right)^{\frac{3}{2}}\left(v(v-1)\right)^{\frac{1}{2}}n^{\frac{1}{2}}\log n, \tag{18}$$

$$t(a,n,m)\ge n-1-\frac{n}{\varphi(n)}\left(2^{\omega(n)}\right)^{\frac{3}{2}}\left(v(v-1)\right)^{\frac{1}{2}}n^{\frac{1}{2}}\log n, \tag{19}$$

where $v=v_{\frac{\varphi(n)}{(\varphi(n),m)}}(n)$.

Proof. It is sufficient to consider equation $x^m = a$ in the group $C(n)$. Let $k=\frac{\varphi(n)}{(\varphi(n),m)}$. We may assume that there exist $0 \le i_0, j_0 \le v-1$ such that

$$s(a,n,m)\in g_{i_0}(n,k)C_k(n), \tag{20}$$

$$t(a,n,m)\in w_{j_0}(n,k)C_k(n), \tag{21}$$

since $s(a, n, m), t(a, n, m) \in C(n)$.

By definition of $g_{i_0}(n,k)$ and $w_{j_0}(n,k)$ we obtain

$$s(a,n,m)\ge g_{i_0}(n,k), \tag{22}$$

$$t(a,n,m)\le w_{j_0}(n,k). \tag{23}$$

On the other side

$$g_{i_0}(n,k)\in s(a,n,m)C_k(n), \tag{24}$$

$$w_{j_0}(n,k)\in t(a,n,m)C_k(n), \tag{25}$$

hence, there exist $\lambda, \theta \in C(n)$ such that

$$g_{i_0}(n,k)=s(a,n,m)\lambda^k, \tag{26}$$

$$w_{j_0}(n,k)=t(a,n,m)\theta^k. \tag{27}$$

But $(\varphi(n),m)|m$, thus by Euler's theorem we obtain

$$g_{i_0}(n,k)^m = s(a,n,m)^m\lambda^{km} = a\left(\lambda^{\frac{m}{(\varphi(n),m)}}\right)^{\varphi(n)} = a, \tag{28}$$

$$w_{j_0}(n,k)^m = t(a,n,m)^m\,\theta^{km} = a\left(\theta^{\frac{m}{(\varphi(n),m)}}\right)^{\varphi(n)} = a, \tag{29}$$

hence $g_{i_0}(n,k)$ and $w_{j_0}(n,k)$ are solutions of the equation $x^m = a$ in the group $C(n)$.

By definition of $s(a, n, m)$, $t(a, n, m)$, we get

$$s(a,n,m) \le g_{i_0}(n,k), \tag{30}$$

$$t(a,n,m) \ge w_{j_0}(n,k). \tag{31}$$

By (22), (23), (30), (31)

$$s(a,n,m) = g_{i_0}(n,k), \tag{32}$$

$$t(a,n,m) = w_{j_0}(n,k). \tag{33}$$

By theorem 1.3 and theorem 2.2 we get

$$s(a,n,m) = g_{i_0}(n,k) \le g_{v-1}(n,k) \le 1 + \frac{n}{\varphi(n)}\left(2^{\omega(n)}\right)^{\frac{3}{2}}\left(v(v-1)\right)^{\frac{1}{2}} n^{\frac{1}{2}} \log n, \tag{34}$$

$$t(a,n,m) = w_{j_0}(n,k) \ge w_0(n,k) \ge n - 1 - \frac{n}{\varphi(n)}\left(2^{\omega(n)}\right)^{\frac{3}{2}}\left(v(v-1)\right)^{\frac{1}{2}} n^{\frac{1}{2}} \log n. \tag{35}$$

Corollary 2.5. Under the assumptions of theorem 2.4 we have that

$$s(a,n,m) \le 1 + \frac{n}{\varphi(n)}\left(2^{\omega(n)}\right)^{\frac{3}{2}} v n^{\frac{1}{2}} \log n, \tag{36}$$

$$t(a,n,m) \ge n - 1 - \frac{n}{\varphi(n)}\left(2^{\omega(n)}\right)^{\frac{3}{2}} v n^{\frac{1}{2}} \log n. \tag{37}$$

Remark 2.6. If $m = \varphi(n)$, then $a = 1$, $k = 1$ $C_1(n) = C(n)$, $v_1 = 1$, $s(1, n, \varphi(n)) = 1$, $t(1, n, \varphi(n)) = n - 1$. In fact, we get optimal bounds using (18) and (19).

Remark 2.7. We may assume that $m \mid \varphi(n)$. Indeed, let d be a natural number such that $d \cdot \frac{m}{(\varphi(n),m)} \equiv 1 \pmod{\varphi(n)}$, we have equivalent congruencies

$$x^m \equiv a \pmod{n} \text{ if } x^{(\varphi(n),m)} \equiv a^d \pmod{n}. \tag{38}$$

Thus $s(a,n,m)=s\left(a^d,n,(\varphi(n),m)\right), t(a,n,m)=t\left(a^d,n,(\varphi(n),m)\right)$.
Note that the left-hand side of inequalities (18), (19) does not depend on a.

Remark 2.8. Let $n = p^\alpha$, where p is an odd prime and α is a positive integer. We may assume that $m|\varphi(n)$, (see remark 2.7). Then

$$v=v_{\frac{\varphi(n)}{(\varphi(n),m)}}(n)=v_{\frac{\varphi(n)}{m}}(n)=\left(\frac{\varphi(n)}{m},\varphi(n)\right)=\frac{\varphi(n)}{m}. \tag{39}$$

By corollary 2.5

$$s(a,n,m)\le 1+2\sqrt{2}\frac{n^{\frac{3}{2}}\log n}{m},\ t(a,n,m)\ge n-1-2\sqrt{2}\frac{n^{\frac{3}{2}}\log n}{m}. \tag{40}$$

Remark 2.9. Let $n = 2^\alpha$, where α is a positive integer greater or equal 2. We may assume that $m|\varphi(n)$. and $m<\varphi(n)=2^{\alpha-1}$ (see remarks 2.7 and 2.6). Then

$$v=v_{\frac{\varphi(n)}{(\varphi(n),m)}}(n)=v_{\frac{\varphi(n)}{m}}(n)=\left(\frac{\varphi(n)}{m},2\right)\left(\frac{\varphi(n)}{m},2^{\alpha-2}\right)=2\frac{\varphi(n)}{m}. \tag{41}$$

By corollary 2.5

$$s(a,n,m)\le 1+4\sqrt{2}\frac{n^{\frac{3}{2}}\log n}{m},\ t(a,n,m)\ge n-1-4\sqrt{2}\frac{n^{\frac{3}{2}}\log n}{m}. \tag{42}$$

We will now give an application of theorem 2.4.

Theorem 2.10. Let p be an odd prime number. For the congruence

$$x^{\frac{p-1}{2}}\equiv -1\pmod p, 1\le x\le p-1, \tag{43}$$

we have that:

1) the congruence (43) has a solution, i.e. –1 is $\frac{p-1}{2}$-th power residue modulo p,

2) the smallest solution $s\left(-1,p,\frac{p-1}{2}\right)$ is a prime number,

3) $s\left(-1,p,\frac{p-1}{2}\right)\le 1+4\frac{p}{p-1}p^{\frac{1}{2}}\log p$,

4) the largest solution $t\left(-1,p,\frac{p-1}{2}\right)\geq p-1-4\frac{p}{p-1}p^{\frac{1}{2}}\log p$.

Proof. If g is a primitive root modulo p (such primitive root exists, since p is a prime number), then $g^{\frac{p-1}{2}}\not\equiv 1\pmod p$ and $g^{p-1}\equiv 1\pmod p$. Hence $g^{\frac{p-1}{2}}\equiv -1\pmod p$ and 1) holds.

If $s\left(-1,p,\frac{p-1}{2}\right)$ were a composite number, it could be expressed as $s\left(-1,p,\frac{p-1}{2}\right)=s=ab$ where $a,b\in\mathbb{N},a,b>1$. Note that $a^{\frac{p-1}{2}}\not\equiv -1\pmod p$ and $b^{\frac{p-1}{2}}\not\equiv -1\pmod p$, since s is the smallest solution of the congruence (43). By Fermat's little theorem, we know that $a^{p-1}\equiv 1\pmod p$ and $b^{p-1}\equiv 1\pmod p$. Hence $a^{\frac{p-1}{2}}\equiv 1\pmod p$ and $b^{\frac{p-1}{2}}\equiv 1\pmod p$. Thus $s^{\frac{p-1}{2}}\equiv (ab)^{\frac{p-1}{2}}\equiv 1\pmod p$, a contradiction with definition of s.

Therefore the initial assumption, that s is a composite number, must be false. Hence 2) holds.

We have $k=\frac{\varphi(p)}{\left(\varphi(p),\frac{p-1}{2}\right)}=2, v=v_2(p)=(2,\varphi(p))=2, \left(2^{\omega(p)}\right)^{\frac{3}{2}}(v(v-1))^{\frac{1}{2}}=4.$

Thus 3) and 4) follows by theorem 2.4.

Example 2.11. For the congruence $x^{359}\equiv -1\pmod{719}$, we have

$$s(-1,719,359)=11,\ t(-1,719,359)=718, \tag{44}$$

in this case theorem 2.10, says that $s(-1,719,359)$ is a prime number and

$$s(-1,719,359)\leq 707,\ t(-1,719,359)\geq 12.$$

References

[1] Nathanson M.B., *Elementary Methods in Number Theory*, Vol. 195, GTM, Springer, New York 2000.

[2] Norton K.K., *k-th coset representatives modulo n*, Acta Arithmetica, XV, 1969, 161-179.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
MECHANICS
DOI: 10.4467/2353737XCT.17.199.7428

Zygmunt Dziechciowski (zygmunt.dziechciowski@mech.pk.edu.pl)
Laboratory of Techno-Climatic Research and Heavy Duty Machines, Faculty of Mechanical Engineering, Cracow University of Technology

Magdalena Kromka-Szydek
Institute of Applied Mechanics, Faculty of Mechanical Engineering, Cracow University of Technology

Gabriela Chwalik
student, Cracow University of Technology

The influence of changing the road pavement and the method of using a wheelchair on the vibration perception in accordance with ISO 2631

Wpływ zmiany podłoża oraz sposobu użytkowania wózka inwalidzkiego na odbiór drgań w oparciu o ISO 2631

Abstract

The aim of the study is to research the influence of pavement types and the position of wheelchair a user on driving comfort. The tests have been done for six different pavements and positions of the wheelchair user, as passive and as active. The assessment has been done on the basis of vibration measurements on the seat of a wheelchair for three perpendicular directions. The obtained characteristics for the selected frequency band have been compared to the criteria curves by ISO 2631. In the next stage of the study, the multiplicity of the exceedances of the vibration perception threshold and the vibration comfort have been calculated. The values of the multiplicity of exceedances have been used in the carried-out assessment process.

Keywords: wheelchair, wheelchair driving comfort, whole body vibration measurement, criteria curves of vibrations receiving

Streszczenie

Artykuł dotyczy badania wpływu rodzaju nawierzchni i rodzaju pozycji przyjmowanej przez użytkownika wózka inwalidzkiego na komfort jazdy. Badania przeprowadzono dla sześciu różnych nawierzchni i dwóch pozycji odpowiadających jeździe pasywnej i aktywnej. Ocenę przeprowadzono w oparciu o pomiar drgań na siedzisku wózka w trzech prostopadłych kierunkach. Otrzymane przebiegi dla odpowiednich pasm tercjowych porównane zostały do krzywych kryterialnych dla normy ISO 2631. Następnie obliczono krotności przekroczeń otrzymanych wartości w stosunku do granic odczuwania drgań oraz komfortu. Wartości krotności przekroczeń wykorzystano w przeprowadzonej ocenie.

Słowa kluczowe: wózek inwalidzki, komfort jazdy, pomiar drgań ogólnych, krzywe kryterialne oceny odbioru drgań

1. Introduction

The wheelchair is a wheeled vehicle driven by muscular force or by a motor. It is intended for people with physical disabilities. The main task of the wheelchair is to help a person, who lost the function of locomotion due to injury or congenital defect, adapt to life. The wheelchair has an influence on the stabilisation of the body, but it also allows a disabled person to move. A wheelchair is composed of basic modules, such as:

- body support system, i.e. items that are in direct contact with the user's body,
- drive system, i.e. allowing for movement (among others: pushrims and caster forks),
- tires and casters,
- frame, which connects all the components together [1, 2].

In Poland, there are about 110-150 thousand people moving in a wheelchair, and about 50% of them use universal manual wheelchairs [3]. Universal wheelchairs are designed for people who have temporarily lost their function of locomotion. A wheelchair is equipped with a push handles and cross braces for easy folding of the stroller. It is characterised by high stability and provides good damping of vibrations, affecting the ride comfort. The disadvantages of such wheelchairs include their large mass, large rolling resistance and that they are difficult to manoeuvre. It is important that the wheelchair user should not feel discomfort associated with everyday mobility and independent locomotion should not require much effort [1, 2].

The centre of gravity of the human-wheelchair system is changed together with the change of the body position (Fig. 1). During independent driving, as a result of an inclination of the thorax and the arms, the centre of gravity moves closer to the front wheels, thus the rolling resistance of the wheels increases.

In the case of active wheelchairs, about 80-90% of the total weight falls on the rear wheels, where as in universal wheelchairs, only 60 % of body weight falls on the rear wheel [2].

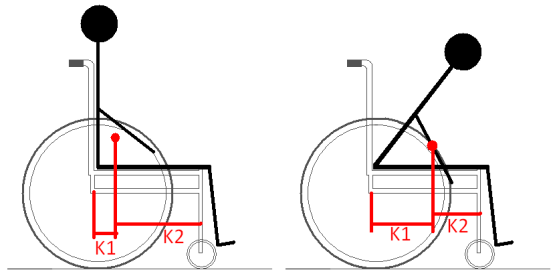


Fig. 1. Changing the centre of gravity depending on the inclination of the wheelchair user; K1, K2 – the distance between the centre of gravity and the axles of the wheels

When driving a wheelchair over uneven surfaces, vibrations are generated, which are transmitted from the seat through the pelvis bone and back into the user's body. Prolonged exposure of a user's body to vibration can cause permanent lesions in the skeletal system and the internal organs of that person. It can also reduce driving comfort and cause quicker tiring of the wheelchair user. A part of the vibrations is absorbed by the front and rear tires of the

wheels, and the wheelchair frame, but when driving on an uneven surface, a part of the energy is transmitted to the human body, where it is also partly absorbed [1, 2]. The type of surface also affects the rolling resistance, which has an influence on the energy expenditure of the wheelchair user when driving.

The influence of mechanical vibrations on the human body is evaluated on the basis of the vibration acceleration values. The majority of Polish and international standards are related to this parameter (ISO 2631-1 [4], ISO 26-31-2 [5], PN 88/B-02171 [6]). The Regulation of the Minister of Labour and Social Policy of 6 June 2014 on the Maximum Permissible Concentration and Intensity of Agents Harmful to Health in the Working Environment [7] states that the value of vibration acceleration is also accepted as a criterion of the assessment of harmful vibrations received by the human body. The reaction of the human body on mechanical vibrations depends on many factors i.a. the frequency value and the time of exposure to vibrations. It is also possible to refer the measurement results directly to the criteria set out experimentally and contained in the publications by Bekesy (1939) [8], Miwa, 1967 [9], McKay 1971 [10], Benson and Dilnot 1981 [11], Griffin and Parsons 1988 [12], Morioka and Griffin 2008 [13], Bellman 2002 [14], Ljunggren et al. 2007 [15]. Some of these studies became the basis of current standards.

In order to assess the measurement, vibration acceleration values in three mutually perpendicular directions shall be corrected, and then the dominant corrected vibration acceleration should be determined. In the case of short-term vibrations (which last less than 30 minutes), the corrected RMS acceleration value is determined in accordance with the formula:

$$a_{w,30\text{minmax}} = \max\{a_{wl1}, a_{wl2}, \ldots a_{wln}\}\ [\text{m/s}^2], \tag{1}$$

where:

- n – the number of performed operations during the exposure to vibrations,
- i – the number of operations,
- l – the direction of vibrations (x, y, z),
- a_{wli} – effective (RMS) frequency corrected vibration acceleration determined for direction l, after taking into account the appropriate direction coefficients $(1.4a_{wxi}, 1.4a_{wyi}, a_{wzi})$ [m/s^2].

The permissible values of mechanical vibrations, determined according to the procedures described in [16], are shown in the Minister of Labour and Social Policy of 6 June 2014 on the Maximum Permissible Concentration and Intensity of Agents Harmful to Health in the Working Environment [7].

According to the ISO [4], there are three characteristic states of the effect of vibrations on the human body, which determine appropriate working conditions, namely: comfortable, burdensome and harmful. The relevant criteria by which the states are separated from each other are shown there in the form of the so-called criteria curves. The criteria curves are referenced to the human vibration perception, which has been obtained on the basis of research.

In the assessment of vibrations in the workplace, the multiplicity of exceedances of the measured vibration acceleration value in relation to the limited (permissible) value is also administered. It can be calculated by the formula:

$$k_a = \frac{a_{meas}}{a_{perm}},$$

a_{meas} – the RMS value of vibration acceleration [m/s^2],
a_{perm} – the permissible values of vibration acceleration [m/s^2].

The occupational hazard is calculated based on the established values. When assessing occupational hazards, the principle described in the PN-N-18002 [17] can be accepted. According to the principle, it is assumed that a small risk and a medium risk are acceptable; a high risk is not acceptable. The following intervals of k_a index variability are assumed, as follows: $k_a \leq 0{,}5$ – small risk, $0{,}5 < k_a \leq 1$ – medium risk, $k_a > 1{,}0$ – high risk.

Not all the existing studies, which describe the problem of the determination of the vibration perception threshold, can be used directly in the research of vibration influence on the wheelchair user.

The researches carried out by Griffin & Parsons [12] and Morioka & Griffin [13] were performed for the case of vibration reception by a person in a sitting position without back support (the transmission of vibrations through the back on the thoracic and lumbar spine is not included). In these works, the criteria curves for the x and y axis (horizontal directions) have a different course, which differentiates them from the criteria given in ISO 2631-1 [4]. The authors concluded that this case is the most similar to the conditions that are met for a man during active driving in a wheelchair. The criteria curves, according to ISO [4], are more restrictive. Because both cases analysed in the work relate to the ride on a universal wheelchair (which – by definition – is used to ride with the support position back), the authors decided that referring the obtained results to the more general ISO standard is more preferred. The ISO standard also allows the determination of the limit of perception and comfort thresholds because, in some cases, it may be important (transport of people with various medical conditions and pregnant women). The obtained values corresponding to the thresholds of comfort, burden and harm (and intermediate values) are obtained by multiplying the acceleration value by appropriate coefficients (specified in the standard), which are formed by given criteria curves.

In the subsequent stages of the research, the reference of the obtained results to the criteria curves established by Morioka & Griffin [13] seems reasonable, especially with regard to the active position, in which there is no full contact between the backrest and the back of the wheelchair user.

2. The aim of the study

The aim of the study was to determine the effect of selected factors on driving comfort of a universal wheelchair. The study takes into account the impact of the pavement type and the change in load distribution on the wheels, which takes place during active and passive drives.

The authors have determined the extent to which the obtained results refer to the limit values of the characteristic quantities of mechanical vibrations. The authors also evaluate in

which of the frequency bands the values of vibrations, which are related to the perception threshold and comfort, were exceeded. The risks associated with the occurrence of mechanical vibrations as the multiplicity of exceedances of the measured values compared to the defined limit, which is determined by the limit of perception and comfort thresholds according to the ISO, has been determined.

On the basis of studies, it was undertaken to attempt to determine the impact of additional factors, such as driving technique and the position of a person moving by wheelchair associated with the body.

3. Methodology of research

The comfort of wheelchair use has been determined on the basis of vibration measurements. In the studies, a manual, universal wheelchair with the possibility of folding was used, which was produced by Meyra model Budget 9.050 (Figs. 2a and b). The wheelchair has a frame made of steel and a cross braces that provide strut support between the two side frames, along with arm rests and suspension of wheels without dampers. The seat and the back rest were made of textile. The analysed wheelchair had wheels with solid rubber tires. The total weight of the wheelchair was 18 kg.

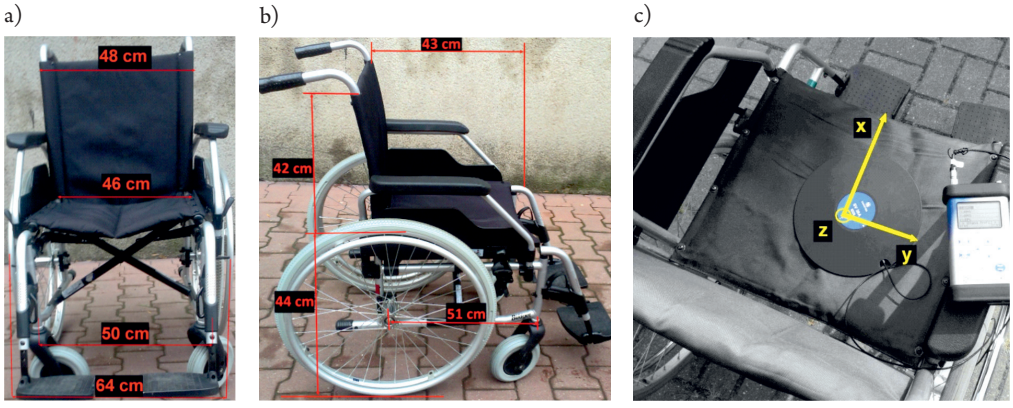


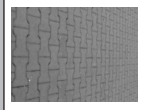
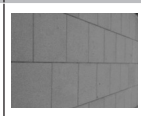
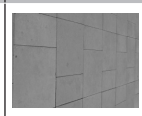
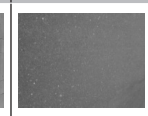
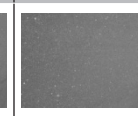
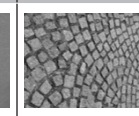
Fig. 2. The universal wheelchair used during measurement: a) a front view; b) a back view; c) location of 3-axis accelerometer

The measurement was carried out using a four-channel vibration analyser SVAN 958 SVANTEK. In the measurements, a three-axis seat accelerometer SV 39A/L was used for the whole-body vibration measurements. The vibration transducer was located in the middle of the wheelchair seat. The location of the accelerometer and the vibration directions are presented in Figure 2.

The measurement process was made for three directions of excitation, for 5 different pavements, on the campus of the Department of Mechanical Engineering at the Technical University of Cracow. The characteristics of particular pavements have been included in

Table 1. In each of the runs, the time was measured with an accuracy of +/- 1 [s]. During the measurement process, a constant speed of passage by the wheelchair was kept. The average values of speed on the various pavements are shown in Table 1.

Table 1. Characteristics of analysed pavement

Type of pavement	N-1 Behaton Pavers	N-2 Square paving 500 x 500 mm	N-3 Square paving 600 x 600 mm	N-4 Bituminous surface of the bicycle path	N-5 Bituminous surface of the roadway	N-6 The granite sett
						
The length of the route [m]	27	40	55	36	22	24
Average values of speed [m/s]	1.30	1.20	1.35	1.23	0.85	0.95

The wheelchair speed on the road pavement of N-6 was lower compared to the other surfaces because of the difficulty in moving along granite sett. A similar speed value has been reached on the pavement N-5, which is relatively easy to move (compared to N-6), which was included in order to compare the behaviour of the wheelchair in extremely different conditions. It was done in order to differentiate the behaviour of the wheelchair in extremely difficult conditions.

Two operable women aged 22 took part in the studies, identified in the measurements as follows: person 1 (58 kg, 158 cm tall) and person 2 (64 kg, 175 cm tall). For each of the analysed surfaces, two attempts to ride were made: the passage corresponding to passive drive and active drive:

- position A (Fig. 3a) – the user in an upright position, the wheelchair pushed by a second person,
- position B (Fig. 3b) – a user in an inclined position, the wheelchair driven by the user as an active driving.

Fig. 3. The position of the person's body during the tests: a) passive drive – position A; b) active drive – position B

The vibration analysis was conducted using SvanPC ++ software by SVANTEK. The results of the analysis have been presented as plots of changing RMS acceleration values for a particular 1/3 octave band, in the range of 1 Hz to 80 Hz. The frequency range of the analysis was determined by the frequency range of the applicability of ISO 2631.

The analysis does not take into account the initial and final stage of the ride. This was due to the occurrence of transients generated at the time of the initial and final movement of the wheelchair.

4. Results of the analysis

The obtained characteristics (1/3 octave band analysis in the range from 1 Hz to 80 Hz) for the directions *x*, *y* and *z* (Figs. 4-7) have been compared to the criteria curves by ISO 2631, and then the multiplicity of exceedances of the vibration perception threshold and the vibration comfort have been calculated (analogously to the formula (2)). The results obtained in this way are shown in Figures 8-11.

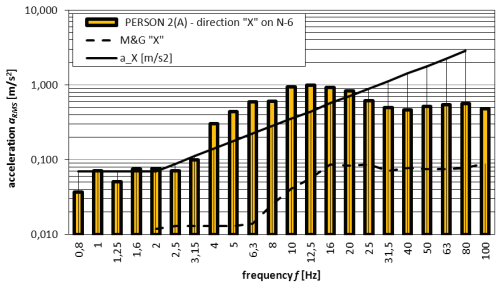


Fig. 4. The results of vibration measurement 1/3 octave band analysis for the pavement N-6 (a person 2, passive position, x-direction)

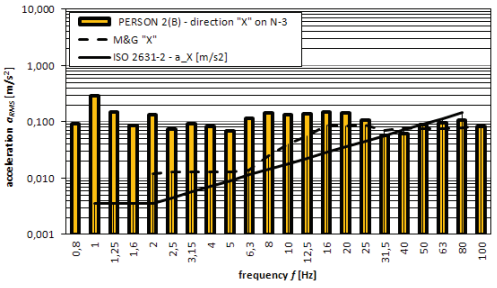


Fig. 5. The results of vibration measurement 1/3 octave band analysis for the pavement N-3 (a person 2, active position, x-direction)

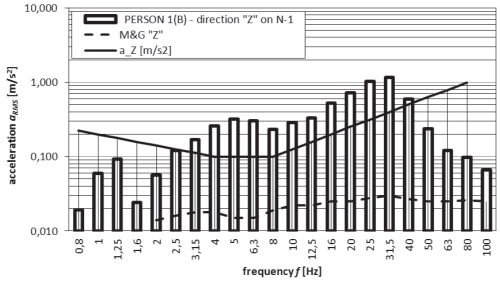


Fig. 6. The results of vibration measurement 1/3 octave band analysis for the pavement N-1 (a person 1, active position, z-direction)

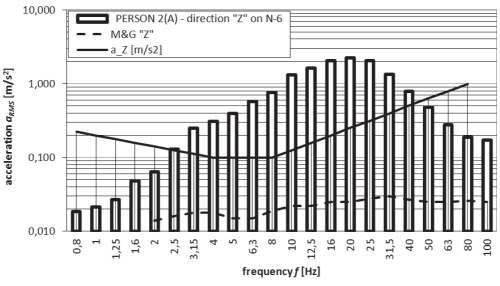


Fig. 7. The results of vibration measurement 1/3 octave band analysis for the pavement N-6 (a person 2, passive position, z-direction)

f [Hz]									f [Hz]									f [Hz]								
1	15	10	70	70	2	2	5	10	1	10	10	60	80	2	2	5	10	1	10	5	95	40	2	0	10	5
1,25	10	10	135	10	2	2	10	2	1,25	5	5	145	40	2	0	15	5	1,25	5	5	25	15	2	0	5	2
1,6	5	10	25	25	2	2	5	5	1,6	5	5	30	25	2	2	5	5	1,6	5	5	25	40	0	0	5	10
2	10	15	30	25	2	5	10	10	2	10	10	45	35	5	2	5	10	2	10	10	30	20	2	2	10	5
2,5	5	5	25	20	5	5	20	10	2,5	5	5	35	15	5	5	15	10	2,5	5	5	20	15	5	5	10	10
3,15	5	5	30	15	15	15	30	15	3,15	5	5	25	15	10	5	15	15	3,15	5	5	15	15	5	5	20	15
4	10	5	25	10	20	20	50	35	4	5	5	20	10	10	10	25	20	4	5	5	15	10	10	10	25	20
5	15	10	25	15	30	25	65	45	5	10	5	15	10	10	10	35	20	5	10	5	15	10	10	10	35	25
6,3	20	10	20	15	30	20	60	35	6,3	10	5	15	10	15	20	30	20	6,3	10	5	15	5	10	15	30	25
8	15	5	20	15	25	25	45	45	8	5	5	10	10	15	15	20	20	8	5	5	10	5	15	10	20	15
10	20	15	15	15	40	60	45	55	10	5	5	5	5	25	15	15	15	10	5	5	5	5	15	15	15	15
12,5	5	5	15	15	30	35	40	50	12,5	5	5	5	5	20	20	20	20	12,5	5	5	5	5	10	10	10	15
16	5	5	15	10	25	40	50	50	16	5	5	5	5	30	25	25	20	16	5	2	5	5	10	10	10	15
20	5	10	15	10	35	55	55	55	20	5	2	5	5	30	25	25	20	20	2	2	2	5	15	15	10	15
25	5	5	5	5	55	45	65	55	25	2	2	2	2	30	25	25	20	25	2	2	2	2	10	15	5	15
31,5	5	5	5	5	55	30	60	35	31,5	2	2	2	2	15	15	15	10	31,5	2	2	0	0	10	10	5	10
40	5	5	5	5	20	10	25	15	40	2	2	2	0	5	5	5	5	40	0	0	0	0	5	5	5	5
50	5	2	5	2	5	5	10	5	50	0	0	2	2	2	2	2	2	50	0	0	0	0	2	2	2	2
63	2	2	2	2	5	2	5	2	63	0	0	0	0	2	2	2	2	63	0	0	0	0	0	0	0	0
80	2	2	2	2	2	2	2	2	80	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0
	N-1/1A/D-X	N-1/2A/D-X	N-1/1B/D-X	N-1/2B/D-X	N-1/1A/D-Z	N-1/2A/D-Z	N-1/1B/D-Z	N-1/2B/D-Z		N-2/1A/D-X	N-2/2A/D-X	N-2/1B/D-X	N-2/2B/D-X	N-2/1A/D-Z	N-2/2A/D-Z	N-2/1B/D-Z	N-2/2B/D-Z		N-3/1A/D-X	N-3/2A/D-X	N-3/1B/D-X	N-3/2B/D-X	N-3/1A/D-Z	N-3/2A/D-Z	N-3/1B/D-Z	N-3/2B/D-Z
	N-1									N-2									N-3							

Fig. 8. The multiplicity of exceedances of the vibration perception threshold according to ISO for the pavements N-1, N-2 and N-3

f [Hz]									f [Hz]									f [Hz]								
1	15	15	80	50	2	2	5	5	1	10	5	135	15	2	0	10	5	1	20	15	135	65	5	2	10	10
1,25	10	10	110	70	2	2	10	5	1,25	5	5	70	10	0	0	5	2	1,25	15	15	60	20	5	5	10	5
1,6	5	10	40	30	2	2	5	5	1,6	5	5	30	35	2	0	5	10	1,6	20	15	40	25	5	5	5	5
2	10	15	40	30	2	5	10	10	2	10	10	60	20	2	2	15	5	2	20	20	45	40	10	10	20	15
2,5	5	5	30	15	5	5	10	10	2,5	5	5	35	15	2	2	15	5	2,5	15	20	35	25	20	20	20	20
3,15	5	5	25	10	5	5	25	15	3,15	2	2	45	15	5	2	30	10	3,15	20	35	40	25	45	45	40	40
4	5	5	20	10	10	5	35	20	4	5	5	40	10	5	5	40	20	4	40	80	40	25	90	60	70	70
5	5	5	15	10	15	10	40	25	5	5	5	25	10	10	10	45	25	5	50	75	45	35	75	80	85	90
6,3	5	5	15	10	10	10	35	20	6,3	5	2	10	10	15	10	35	20	6,3	50	60	60	50	115	115	75	95
8	5	5	10	10	10	10	25	15	8	5	5	10	10	10	10	25	10	8	40	40	55	55	155	150	110	140
10	5	5	5	5	10	10	20	10	10	5	5	5	5	10	10	15	10	10	50	50	45	50	185	215	135	160
12,5	5	2	5	5	10	10	15	10	12,5	5	2	5	5	10	15	15	10	12,5	45	40	40	40	170	205	165	160
16	2	2	5	5	15	10	15	15	16	2	2	5	5	10	15	15	10	16	30	35	30	35	160	205	145	160
20	2	2	5	5	15	10	20	15	20	2	2	5	5	20	15	20	15	20	25	25	20	20	175	180	155	105
25	2	2	2	2	20	10	20	15	25	2	2	2	2	30	20	25	20	25	15	15	10	10	130	130	115	75
31,5	2	2	2	0	15	10	15	10	31,5	5	2	2	2	30	15	25	15	31,5	10	10	5	5	70	70	55	50
40	2	0	2	0	5	5	10	5	40	5	2	2	0	15	10	15	15	40	5	5	5	5	35	30	25	25
50	2	2	2	2	5	5	5	2	50	5	2	2	2	10	5	5	5	50	5	5	5	5	15	15	10	10
63	2	2	2	2	2	2	2	2	63	2	2	2	2	5	5	5	5	63	5	5	5	5	5	5	5	5
80	2	2	2	2	0	0	0	0	80	5	2	2	2	2	2	2	2	80	5	5	5	5	5	5	5	5
	N-4/1A/D-X	N-4/2A/D-X	N-4/1B/D-X	N-4/2B/D-X	N-4/1A/D-Z	N-4/2A/D-Z	N-4/1B/D-Z	N-4/2B/D-Z		N-5/1A/D-X	N-5/2A/D-X	N-5/1B/D-X	N-5/2B/D-X	N-5/1A/D-Z	N-5/2A/D-Z	N-5/1B/D-Z	N-5/2B/D-Z		N-6/1A/D-X	N-6/2A/D-X	N-6/1B/D-X	N-6/2B/D-X	N-6/1A/D-Z	N-6/2A/D-Z	N-6/1B/D-Z	N-6/2B/D-Z
	N-4									N-5									N-6							

Fig. 9. The multiplicity of exceedances of the vibration perception threshold according to ISO for the pavements N-4, N-5 and N-6

f [Hz]									f [Hz]									f [Hz]								
1	1	0	4	4	0	0	0	1	1	1	0	3	4	0	0	0	1	1	1	0	5	2	0	0	0	0
1,25	1	0	7	0	0	0	1	0	1,25	0	0	7	2	0	0	1	0	1,25	0	0	1	1	0	0	0	0
1,6	0	0	1	1	0	0	0	0	1,6	0	0	2	1	0	0	0	0	1,6	0	0	1	2	0	0	0	0
2	1	0	2	1	0	0	0	0	2	0	1	2	2	0	0	0	1	2	0	0	2	1	0	0	0	0
2,5	0	0	1	1	0	0	1	0	2,5	0	0	2	1	0	0	1	0	2,5	0	0	1	1	0	0	0	0
3,15	0	0	1	1	1	0	2	1	3,15	0	0	1	1	0	0	1	1	3,15	0	0	1	1	0	0	1	1
4	0	0	1	1	1	0	3	2	4	0	0	1	1	1	0	1	1	4	0	0	1	1	0	0	1	1
5	1	0	1	1	1	0	3	2	5	0	0	1	0	1	0	2	1	5	1	0	1	0	1	1	2	1
6,3	1	0	1	1	1	0	3	2	6,3	1	0	1	1	1	1	1	1	6,3	0	0	1	0	1	1	2	1
8	1	0	1	1	1	0	2	2	8	0	0	1	1	1	1	1	1	8	0	0	0	0	1	1	1	1
10	1	0	1	1	2	5	2	3	10	0	0	0	0	1	1	1	1	10	0	0	0	0	1	1	1	1
12,5	0	0	1	1	1	0	2	2	12,5	0	0	0	0	1	1	1	1	12,5	0	0	0	0	1	1	1	1
16	0	0	1	1	1	0	3	2	16	0	0	0	0	1	1	1	1	16	0	0	0	0	1	1	0	1
20	0	0	1	1	2	5	3	3	20	0	0	0	0	2	1	1	1	20	0	0	0	0	1	1	0	1
25	0	0	0	0	3	0	3	3	25	0	0	0	0	1	1	1	1	25	0	0	0	0	1	1	0	1
31,5	0	0	0	0	3	0	3	2	31,5	0	0	0	0	1	1	1	1	31,5	0	0	0	0	0	0	0	0
40	0	0	0	0	1	0	1	1	40	0	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0	63	0	0	0	0	0	0	0	0	63	0	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0
	N-1/1A/D-X	N-1/2A/D-X	N-1/1B/D-X	N-1/2B/D-X	N-1/1A/D-Z	N-1/2A/D-Z	N-1/1B/D-Z	N-1/2B/D-Z		N-2/1A/D-X	N-2/2A/D-X	N-2/1B/D-X	N-2/2B/D-X	N-2/1A/D-Z	N-2/2A/D-Z	N-2/1B/D-Z	N-2/2B/D-Z		N-3/1A/D-X	N-3/2A/D-X	N-3/1B/D-X	N-3/2B/D-X	N-3/1A/D-Z	N-3/2A/D-Z	N-3/1B/D-Z	N-3/2B/D-Z
	N-1									N-2									N-3							

Fig. 10. The multiplicity of exceedances of comfort threshold in accordance of ISO for pavements N-1, N-2 and N-3

f [Hz]									f [Hz]									f [Hz]								
1	1	1	4	2	0	0	0	0	1	0	0	7	1	0	0	0	0	1	1	1	7	3	0	0	1	0
1,25	0	0	6	3	0	0	0	0	1,25	0	0	4	1	0	0	0	0	1,25	1	1	3	1	0	0	0	0
1,6	0	1	2	1	0	0	0	0	1,6	0	0	2	2	0	0	0	0	1,6	1	1	2	1	0	0	0	0
2	1	1	2	2	0	0	1	0	2	0	0	3	1	0	0	1	0	2	1	1	2	2	0	0	1	1
2,5	0	0	2	1	0	0	1	1	2,5	0	0	2	1	0	0	1	0	2,5	1	1	2	1	1	1	1	1
3,15	0	0	1	0	0	0	1	1	3,15	0	0	2	1	0	0	1	1	3,15	2	1	2	1	2	2	2	2
4	0	0	1	0	0	0	2	1	4	0	0	2	1	0	0	2	1	4	4	2	2	1	4	3	4	4
5	0	0	1	0	1	0	2	1	5	0	0	1	0	0	0	2	1	5	4	2	2	2	4	4	4	5
6,3	0	0	1	1	1	0	2	1	6,3	0	0	1	0	1	0	2	1	6,3	3	3	3	3	6	6	4	5
8	0	0	1	0	1	0	1	1	8	0	0	0	0	1	0	1	1	8	2	2	3	3	8	8	5	7
10	0	0	0	0	1	1	1	1	10	0	0	0	0	0	0	1	1	10	2	3	2	3	9	10	7	8
12,5	0	0	0	0	1	0	1	1	12,5	0	0	0	0	1	1	1	1	12,5	2	2	2	2	8	10	8	8
16	0	0	0	0	1	1	1	1	16	0	0	0	0	1	1	1	1	16	2	2	2	2	8	10	7	8
20	0	0	0	0	1	1	1	1	20	0	0	0	0	1	1	1	1	20	1	1	1	1	9	9	8	5
25	0	0	0	0	1	1	1	1	25	0	0	0	0	1	1	1	1	25	1	1	1	1	6	6	6	4
31,5	0	0	0	0	1	0	1	0	31,5	0	0	0	0	1	1	1	1	31,5	0	0	0	0	4	3	3	2
40	0	0	0	0	0	0	0	0	40	0	0	0	0	1	0	1	1	40	0	0	0	0	2	2	1	1
50	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	50	0	0	0	0	1	1	1	1
63	0	0	0	0	0	0	0	0	63	0	0	0	0	0	0	0	0	63	0	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0
	N-4/1A/D-X	N-4/2A/D-X	N-4/1B/D-X	N-4/2B/D-X	N-4/1A/D-Z	N-4/2A/D-Z	N-4/1B/D-Z	N-4/2B/D-Z		N-5/1A/D-X	N-5/2A/D-X	N-5/1B/D-X	N-5/2B/D-X	N-5/1A/D-Z	N-5/2A/D-Z	N-5/1B/D-Z	N-5/2B/D-Z		N-6/1A/D-X	N-6/2A/D-X	N-6/1B/D-X	N-6/2B/D-X	N-6/1A/D-Z	N-6/2A/D-Z	N-6/1B/D-Z	N-6/2B/D-Z
	N-4									N-5									N-6							

Fig. 11. The multiplicity of exceedances of comfort threshold in accordance of ISO for pavements N-4, N-5 and N-6

5. Discussion of results

The analysis of the arrays shown in Figures 4 and 5, containing the results of the measurements, allows the comparison of individual surfaces in terms of vibration comfort. Certain frequency ranges that these arrays show can be identified, for example, with various types of ailments occurring in humans. Forces that occur in the given frequency range may affect the comfort of the wheelchair user.

For the pavement (N-1), the maximum of the vibration perception threshold for the x-direction (on average approximately 15 times), were recorded for the 1/3 octave band from 1 to 20 Hz. For frequency ranges 1.0, 1.25, 2.0, and 3.15 Hz the exceedances are much higher, even 135 times for the active position of a user (position B). For the passive position (position A) and z-direction the multiplicity of exceedances of the vibration perception threshold (in the range of 1/3 octave bands from 40 Hz to 3.15 Hz) take high values (minimum 10 times, the maximum 60 times). For the active position the exceedances of the vibration perception threshold present in the frequency range of 1 to 50 Hz, (maximum value in the range of 10 to 25 Hz).

For the pavement N-2 and N-3 (square paving), in the x-direction, for the B user position, for the low frequency, are exceeded for the perception threshold (the 1/3 octave bands to 2.5 Hz – 145 times). For the z-direction, the exceedance are for 1/3 octave bands in the range of 3.15 Hz to 31.5 Hz, for passive position, and from 2 Hz to 31.5 Hz for active position. The maximum value of exceedances for the z-axis were recorded for the 1/3 octave bands 5 Hz and 6.3 Hz. The higher values of vibration were recorded for the pavement N-3. It can be caused by the method of square paving arranged. Above the frequency range of 63 Hz was not observed exceedances of the vibration perception threshold, for both pavement (N-2 and N-3) and for both vibration directions, for the active position and the passive position.

For bituminous surface N-4 and N-5 (bicycle path and the roadway), for passive position and vibration in the x-direction, maximum exceeding values of the perception vibration threshold are observed for the 1/3 octave band from 1 to 2 Hz. For the roadway the multiplicity of exceedances varies between 5 to 10 times, but for the bicycle path even 10 to 15 times. In the remaining 1/3 octave bands, for both bituminous surfaces, the exceeding are much lower (maximum up to 5 times). For the active position the wheelchair user were observed much more diverse values of multiplicity of exceedances of the vibration perception threshold. The highest exceeding are observed for the 1/3 octave band of 4 Hz. For the very low frequency ranges (1 and 1.25 Hz) is reported, even more than 100 times compared to the limit of perception limits. For the vibration in the z-direction and the position A – it can be seen similarity of exceeding value of vibration perception threshold for both bituminous surfaces. For this case the multiplicity of exceedances, for the frequency ranges from 4 to 40 Hz, are amounted to 10-15 times, but for the 1/3 octave bands from 20 to 31.5 Hz even 30 times. For the position B of user, higher values of exceeding (about 10-25 times) can be seen for the frequency range of 2 to 40 Hz. The maximum values (from 25 to 45 times) have been recorded for the user who had lower body weight.

For the case of the wheelchair moving on the pavement N-6 (granite sett) and the receiving of vibration for the x-direction, the highest exceeding was recorded for the passive

position for the 1/3 octave bands from 3.15 Hz to 20 Hz. For that frequency bands the values of multiplicity of exceedances are 25-80 times. For the active position the high values of the multiplicity of exceedances include a similar ranges of the 1/3 octave bands (the values of exceeding are lower and amount maximum up to 60 times). For the vibration in the z-direction, for the passive position, were observed significantly higher values of the exceeding of the vibration perception threshold in relation to the active position. For both positions, very high exceeding was observed for the 1/3 octave bands from 6.3 to 25 Hz (100 and 200 times), the lower values (40 times) was recorded for the 1/3 octave bands 2.5 Hz, 3.15 Hz and 31.5 Hz. For the 1/3 octave bands from 1 Hz to 1.6 Hz as well as 63 Hz and 80 Hz the values of multiplicity of exceedances are not more than 10 times.

For all tested pavements, for the vibration in the x-direction, was observed exceedance in the low frequency range (1/3 octave bands from 1 to 2 Hz). There was any exceedance above the frequency range of 40 Hz for both directions, and both methods of use of the wheelchair.

For the pavement N-1 (behaton pavers), during the studies, it was registered exceeding of the perception vibration threshold about 5 times for 1/3 octave bands 10 and 20 Hz. In the remaining frequency bands have not been found the exceeding of the perception threshold. It seems reasonable to check the influence of pavement unevenness or structure of the wheelchair. In the z-direction, for the position B (active drive), it can be observed the exceeding of 2-3 times for the 1/3 octave bands from 4 to 31.5 Hz.

For the pavements N-2 and N-3 the exceeding of comfort threshold according to ISO standard is observed for the x-axis, an active position, for 1/3 octave bands of 2.5 Hz. For the pavement N-3 the exceeding were higher. In the other frequency ranges, for both pavement, there were any exceeding. Only for the 1/3 octave bands from 5 and 6.3 Hz, for the direction z and active position of user, the exceeding of comfort threshold are twice.

For bituminous surfaces (N-4 and N-5), the differences of values of the multiplicity of exceedances of the comfort threshold have been observed for 1/3 octave bands from 1 to 2 Hz (for the x-axis) only. Additionally for the vibration in the z-direction were recorded that the exceeding of comfort threshold, for an active position and the frequency bands from 4 to 6.3 Hz, are twice.

For the pavement N-6 (granite sett), for the vibration excitation in the z-direction, the largest exceeding of comfort threshold (from 5 to 10 times) are for 1/3 octave bands from 8 to 20 Hz. The higher values of the multiplicity of exceedances of the vibration comfort threshold refer to the case of passive user's position (position A). The lower values the multiplicity of exceedances (from 2 to 6 times) were recorded for 1/3 octave bands from 3.15 to 6.3 Hz and from 25 to 40 Hz. For the vibrations perception in the x-direction, the maximum exceeding of the comfort threshold (from 2 to 4 times) appear for 1/3 octave bands from 4 to 16 Hz.

6. Conclusions

The major impact on the value of the vibrations, which are received by the wheelchair user, is had by the type of pavement on which the wheelchair is moving. It can be specified items, such as: the size of the plate, from which the surface is made of (arranged), the pavement

pattern (determined by the pavement elements), the size of the gap, surface roughness and its condition (in the case of new surfaces associated with the accuracy of the performance). These factors affect the change of rolling resistance. These factors are important, and in the case of a wheelchair, they are not entirely recognised, which is confirmed in other studies [18].

The highest values of vibration acceleration have been recorded for the granite pavement, which has the most uneven pavement of all the analysed types. This is due to both the technological performance of a single element, as well as its technology of laying. The method of its arrangement (granite fan pattern) could determine the frequency of the vibration excitation and values of the parameter that describes the vibration excitation. The lowest values of the vibration acceleration can be observed for surfaces made of concrete paving slabs with dimensions of 600 mm x 600 mm. In the range of above 31.5 Hz, for the x-axis, excessive vibrations are not observed.

When changing the load distribution of the wheelchair wheels, which is different for active and passive driving, a change of the value of the parameter describing the vibrations transmitted to the user's body has been observed. For movement carried out in the active position, the vibration acceleration values are higher. The relationship between the experience when using the wheelchair (driving technique) and the level of received vibration can be recognised. It should also be noted that the tests were carried out in a wheelchair, which had a solid tire that significantly affects the achieved results.

The tested wheelchair is a universal wheelchair, which means that, by definition, it is not intended to move actively on rough ground, but it is used to transport people on different surfaces, over short distances. Bearing in mind that the users of this means of transport are people with different diseases and, for example, pregnant women, it seems important to ensure adequate comfort of drive, the criterion of which may be the multiplicity of exceedances of the vibration perception threshold.

References

[1] Sydor M., Zabłocki M., *Wybrane problemy doboru i konfiguracji wózka inwalidzkiego z napędem ręcznym*, Fizjoterapia Polska, Vol. 6, 2(4), 2006, 172-177.

[2] Sydor M., *Wybór i eksploatacja wózka inwalidzkiego*, Wydawnictwo Akademii Rolniczej, Poznań 2003.

[3] http://stat.gov.pl/ (access: 21.11.2015).

[4] ISO 2631 – 1, Mechanical vibration and shock – Evaluation of human exposure to whole-body vibration – Part 1: General requirements, 1997.

[5] ISO 2631 – 2, Mechanical vibration and shock – Evaluation of human exposure to whole-body vibration – Part 2: Vibration in buildings (1 Hz to 80 Hz), 2003.

[6] PN-88/B-02171, Ocena wpływu drgań na ludzi w budynkach, 1988.

[7] Regulation of the Minister of Labour and Social Policy of 6 June 2014 on Maximum Permissible Concentration and Intensity of Agents Harmful to Health in the Working Environment, Journal of Laws of 2014, No. 817.

[8] Békésy G., *Über die Empfindlichkeit des stehenden und sitzenden Menschen gegen über sinusformigen Erschütterungen*, Akustische Zeitschrift, 4, 1939, 360.
[9] Miwa T., *Evaluation methods for vibration effect. Part 1. Measurements of threshold and equal sensation contours of whole body for vertical and horizontal vibrations*, Industrial Health, 2, 1967, 183-205.
[10] McKay J.R., *Human response to vibration: some studies of perception and startle*, Ph.D Thesis, University of Southampton, 1972.
[11] Benson A.J., Dilnot S., *Perception of whole-body linear oscillation*, Proceedings of U.K. Informal Group on Human Response to Vibration, Heriot-Watt University, Edinburgh 1981, 92-102.
[12] Parsons K.C., Griffin M.J., *Whole-body vibration perception thresholds*, Journal of Sound and Vibration, 121(2), 1988, 237-258.
[13] Morioka M., Griffin M.J., *Absolute thresholds for the perception of fore-and-aft, lateral, and vertical vibration at the hand, the seat, and the foot*, Journal of Sound and Vibration, 314, 2008, 357-370.
[14] Bellmann M.A., *Perception of Whole-Body Vibrations: From basic experiments to effects of seat and steering-wheel vibrations on the passenger's comfort inside vehicles*, Universitat Oldenburg, 2002.
[15] Ljunggren F., Wang J., Agren A., *Human vibration perception from single-and dual-frequency components*, Journal of Sound and Vibration, 300(1), 2007, 13-24.
[16] PN-EN 14253+A1:2011, Drgania mechaniczne. Pomiar i obliczanie zawodowej ekspozycji na drgania o ogólnym działaniu na organizm człowieka dla potrzeb ochrony zdrowia. Wytyczne praktyczne.
[17] PN-N-18002:2000, Systemy zarządzania bezpieczeństwem i higieną pracy – Ogólne wytyczne do oceny ryzyka zawodowego.
[18] van der Woude L.H.V., de Groot S., Janssen T.W.J., *Manual wheelchairs: Research and innovation in rehabilitation, sports, daily life and health*, Medical Engineering & Physics, 28, 2006, 905-915.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
MECHANICS
DOI: 10.4467/2353737XCT.17.200.7429

Michal Ptaszynski (ptaszynski@cs.kitami-it.ac.jp)
Fumito Masui
Department of Computer Science, Kitami Institute of Technology

Experiments with language combinatorics in text classification: lessons learned and future implications

Eksperymenty z zastosowaniem kombinatoryki językowej do klasyfikacji tekstu: dotychczasowe wnioski i implikacje na przyszłość

Abstract

This paper presents a meta-analysis of experiments performed with language combinatorics (LC), a novel language model generation and feature extraction method based on combinatorial manipulations of sentence elements (e.g., words). Along recent years LC has been applied to a number of text classification tasks, such as affect analysis, cyberbullying detection or future reference extraction. We summarize two of the most extensive experiments and discuss general implications for future implementations of combinatorial language model.

Keywords: language combinatorics, natural language processing, text classification

Streszczenie

W niniejszym artykule przedstawiono metaanalizę badań przeprowadzonych za pomocą kombinatoryki językowej (language combinatorics, LC), nowej metody generacji modelu języka i ekstrakcji cech, opartej o kombinacyjne manipulacje na elementach zdań (np. słowa). W trakcie ostatnich lat LC została zastosowana do wielu zadań z dziedziny klasyfikacji tekstu, takich jak analiza afektu, wykrywanie cyberagresji lub ekstrakcja odniesień do przyszłych wydarzeń. W niniejszym artykule podsumowujemy dwa z najbardziej obszernych doświadczeń i omawiamy ogólne implikacje dotyczące przyszłych zastosowań kombinatoryjnego modelu języka.

Słowa kluczowe: kombinatoryka językowa, przetwarzanie języków naturalnych, klasyfikacja tekstu

1. Introduction

Language modeling refers to a set of basic techniques in Natural Language Processing (NLP). It is crucial to most of NLP applications, including final word prediction [4], language identification [5], information retrieval [6], speech recognition [7], machine translation [8], part-of-speech (POS) tagging [9], or sentiment analysis [10].

However, despite such a wide applicability, there has been little progress within the language modeling field itself. There exist two to three general paradigms for language modeling, while most of the research still applies only the most basic ones, such as bag-of-words (BoW) model. Although modifications and some more sophisticated methods have been proposed (e.g., the skip-gram model), they too are bound with constraints hindering the thorough analysis of language phenomena. The more sophisticated methods for language modeling still represent a niche and are yet to be used more widely.

In this paper we analyze one of such methods called Language Combinatorics (LC) [11]. It addresses the limitations of previous models by defining a language pattern as any frequently appearing ordered combination of sentence elements. This flexible definition allows extracting from sentences all possible patterns, not limited to single words, as in the BoW model, or phrases, as in the *n*-gram model, but extends the extraction to sophisticated patterns with disjointed elements. To prove the advantage of the model, during recent several years we have extensively applied it to various experiments. This paper presents a meta-analysis of findings we drew from some of them.

The outline of the paper is as follows. We describe other research related to ours (section 2), and explain the general idea of combinatorial language model (section 3). Next we summarize (section 4) and analyze the experiments in which the model has been applied (section 5), draw general conclusions, and discuss future applications (section 6).

2. Related Research

The computationally simplest language model, the bag-of-words (BOW) model [12], considers a piece of text or document as an unordered collection of words, thus disregarding grammar and word order. Some researchers proposed improvements to BoW, e.g., by using semantic concepts instead of words (bag-of-concepts) [13], or adding word positions in sentences to the equation, thus retaining general information on word order (positional language model) [14]. Unfortunately, though one could use any general feature type to build a language model (e.g., concepts, parts-of-speech), order and longer element strings (e.g., phrases) will still be disregarded. Moreover, sentences can be of different length and any word can be preceded by another word. Thus the position of a word in sentence is not a constant value and makes the model strictly data-dependent and of limited practical use.

An approach retaining word order, based on n-grams [15], perceives an input (e.g., sentence) as a set of n-long ordered sub-sequences of elements (letters, words). Although retaining word order, n-grams allow only for simple sequence matching, while disregarding

deeper sentence structure. Again, instead of words one could use sequences of POS or concepts, however, ngrams still cannot cover more sophisticated patterns than word sequences.

An example of a language model aimed to go beyond BoW and n-grams, the skip-gram model (also known as skipped or distanced n-gram) [16], assumes that some words within an n-gram could be skipped over. In theory, this should allow extraction of most language patterns. However, to limit computational complexity of the model, skip-grams include a number of assumptions hindering the model, for example, that 1) skip-grams are generated from n-grams, not from the whole sentence, 2) a skip can appear only in one place, 3) the number of skipped elements is recorded separately for each gap (thus, two words separated by one word in between (1-skip-bigram) and by five words (5-skip-bigram) would necessarily be considered as different patterns), or that 4) the skip-length is predetermined, meaning that if the researcher chooses to extract only bigrams with only one skip in between, the same pattern, but with five skips will be disregarded from the beginning.

The above assumptions are counter-intuitive, since one can easily imagine the same sentence pattern appearing in two sentences of different length, or separated by gaps of different sizes. To illustrate this problem in Table 1 we compared which of the above-mentioned language models would discover particular patterns present in the two following sentences. The last right column represents a language model based on Language Combinatorics (LC).

1) John went to school today.
2) John went to this awful place, generously called by many school, not yesterday, but today.

Table 1. Comparison of capabilities of different language models to capture (○) or not (×) certain patterns from the corpus containing two sentences, (1) and (2)

pattern	language model			
	BoW	n-gram	skip-gram	LC
John	○	○	○	○
John went	×	○	○	○
John * to	×	×	○	○
John * school	×	×	×	○
John * to * today	×	×	×	○

Finally, in all previous research on skip-grams the model was studied only for up to 4-elements [17]. Only recent attempts used 5-element-long skip-grams [18], however, still generating them from n-grams, not the whole sentences.

Language Combinatorics is capable of dealing with any of the sophisticated patterns, by defining a pattern, or specifically, a sentence pattern, as any **ordered combination of sentence elements frequently occurring in a corpus**. This definition allows extraction of all possible meaningful linguistic patterns from unrestricted text. In our research so far, we have focused on applications of the method to various tasks from the areas of automatic pattern extraction and text classification.

3. Combinatorics-Based Language Modelling

Example: What a nice day !				
5-el. pattern:	4-el. patterns:	3-el. patterns:	2-el. patterns:	1-el. patterns:
What a nice day !	What a nice * !	a nice * !	What a	What
	What a nice day *	What a nice	What * !	a
	What a * day !	What a * !	nice * !	nice
no. of patterns: (1)	(5)	(10)	(10)	(5)

Fig. 1. Examples of various length (= number of elements) combinations extracted from one sentence

The text classification method applying combinatorial language modeling is composed of four steps: feature extraction, weight calculation, classification, and threshold optimization.

3.1. Feature Extraction with Language Combinatorics

To extract sentence patterns LC perceives sentences as bundles of ordered combinations of elements (words, etc.), and frequent combinations appearing in many sentences are defined as sentence patterns. As long as patterns are defined this way, they can be automatically extracted by generating all ordered combinations of such elements, verifying their occurrences within a corpus, and filtering out those combinations which appear only once.

In particular, in every n-element sentence there is k-number of combination clusters, such that $1 \leq k \leq n$. The number of k-element combinations is equal to binomial coefficient, represented in eq. 1. Here, all combinations for all values of k from the range of $\{1, ..., n\}$ are generated. Thus the number of all combinations generated for n-long sentence is equal to the sum of combinations from all k-element clusters of combinations, like in eq. 2.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{1}$$

$$\sum_{k=1}^{n}\binom{n}{k} = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + \ldots + \frac{n!}{n!(n-n)!} = 2^n - 1 \tag{2}$$

Moreover, all non-subsequent elements are separated with an asterisk ("*"), indicating some elements appeared between those two elements. Some examples of combinations extracted this way is represented in Figure 1.

3.2. Weight Calculation

After combinations are extracted, their occurrences O are calculated. Those combinations which appeared only once are discarded, and those which appeared more than once are considered as patterns j characteristic to the sentence collection from which they were

extracted. In research applying LC in binary text classification [2, 3], the occurrences were also calculated separately for the positive side O_{pos} and the negative side O_{neg}. Such occurrences of each pattern j are further used to calculate normalized weight w_j according to equation 3, fitting the weight in the range {1, ... , -1}.

$$w_j = \left(\frac{O_{pos}}{O_{neg} + O_{neg}} - 0.5 \right) * 2 \tag{3}$$

For purposes of a text classification experiment, previous research also modified the weight in several ways, considering what makes a pattern representative for a corpus. In particular, w_j was modified by multiplying it with:

- pattern length k_j, which provides a weight with awarded length w_{LA}, like in equation 4, or
- k_j and overall pattern occurrence (O_{pos}+O_{neg}), which provides a weight with awarded length and occurrence w_{LOA}, like in equation 5.

$$w_{LOA} = w_j * k_j * \left(O_{pos} + O_{neg} \right) \tag{4}$$

$$w_{LOA} = w_j * k_j * \left(O_{pos} + O_{neg} \right) \tag{5}$$

Moreover, when two collections of sentences of opposite features (such as "positive vs. negative") are compared, the generated pattern list will contain patterns appearing uniquely in only one of the sides or in both (later: *ambiguous patterns*, or **AMB**). A special type of an ambiguous pattern is the one appearing on both sides with the same occurrence, making its weight equal 0 (*zero-patterns*, or **0P**). Thus the list of originally generated patterns can be further modified by discarding either ambiguous patterns or zero-patterns.

3.3. Classification

In the classification, previous research applied a classifier function defined as a sum of weights of patterns found in a sentence (eq. 6).

$$\text{score} = \sum w_j, \left(1 \geq w_j \geq -1 \right) \tag{6}$$

It produces a score for each analyzed sentence. The score alone does not yet specify a class (e.g., positive or negative). The intuition suggests that the higher above or below zero is the score, the more it resembles a style of writing usually found in one of the sides. However, an intuitive rule of thumb, with zero as a universal threshold does not apply to pattern-based method, since even one word difference in a sentence can produce much larger number of patterns on one of the sides, causing an imbalance in the data. Therefore a threshold optimization of the classifier is performed to specify which threshold is optimal for the classification of provided data.

3.4. Threshold Optimization and Heuristic Rules

Data is never ideally balanced. The collections of sentences are usually biased toward one of the sides (more sentences on one of the sides, or sentences are longer, etc.). This could produce more patterns for one of the sides. To minimize the bias, instead of applying a fixed rule of thumb, a more effective way is to automatically optimize the threshold, by verifying the performance for each step and selecting the optimal one for the given data.

Finally, to deal with combinatorial explosion occurring during exhaustive combinatorial manipulations, two heuristic rules were applied. The procedure of pattern generation would (1) generate up to six elements patterns, or (2) terminate at the point where no more frequent patterns were found.

4. Applications

During recent years, LC was applied in a system for the support of experiments in text classification [19] and was used for a number of research in binary text classification. We summarize some of them below.

One of the research analyzed a small set of emotive (emotionally loaded) and non-emotive sentences. Ptaszynski et al. [3] performed a study of such sentences with the use of LC and found out that completely automatic approach to extraction of emotional patterns from sentences can give similarly good results to state-of-the-art tools developed manually and much better results than traditional classifiers (SVM).

The capability of Language Combinatorics to capture hidden patterns in language confirmed in the above research encouraged Ptaszynski et al. [2] to extract patterns of cyberbullying or Internet harassment. They analyzed a medium sized dataset containing such harmful entries, and confirmed that the LC-based method outperformed all compared previous methods for cyberbullying detection.

In another research Nakajima et al. [20] applied LC in analysis of future related expressions for trend prediction. The experiments showed that sentences referring to the future contain frequent patterns, while patterns in other sentences (present, past or other) are sparse. This proved that future-referring sentences can be analyzed as one separate kind of sentences.

5. Meta-Analysis of Experiment Results

Below we present meta-analysis of the results performed in previous papers. In the analysis we applied two datasets from the ones mentioned in section 4, namely, 1) small dataset containing emotive sentences, and 2) medium-sized dataset containing cyberbullying. We omitted the largest dataset [20] since experiments with it were performed only with one type of dataset preprocessing, while others used several kinds of preprocessing.

5.1. Datasets – Short Description

Emotive Sentences. Dataset used in [3] consists of 50 emotive and 41 non-emotive sentences collected originally by Ptaszynski et al. (2009) [21] for the needs of evaluating their affect analysis system. To collect the data they performed an anonymous survey on thirty participants of different age and social groups. Each of them was to imagine or remember a conversation with any person they knew and write three sentences from that conversation: one free, one emotive (emotionally loaded), and one non-emotive (neutral, or non-emotional). Additionally, the participants were asked to make the emotive and non-emotive sentences as close in content as possible, so the only perceivable difference was in emotional load.

Cyberbullying. The dataset used in cyberbullying detection [2] contains 1,490 harmful and 1,508 non-harmful entries. The original data was provided by the Human Rights Research Institute Against All Forms for Discrimination and Racism in Mie Prefecture, Japan (later: Human Rights Center) [22] and contains data from a number of informal school Websites from Mie Prefecture, Japan. The harmful and non-harmful sentences were manually labeled by experts, members of Internet Patrol, according to instructions included in an official governmental manual for dealing with cyberbullying [23].

5.2. General Setup of Experiments

Both analyzed research used similar experiment setup. The prepared datasets were used in a text classification experiment with the use of the proposed LC-based method, and other methods (previously developed systems and classifiers). In classification, researchers compared the performance of sophisticated patterns to more common n-grams, and BoW. Feature weights were calculated according to the equations explained in section 3.2. For classifiers based on BoW, a traditional weight calculation scheme was also applied, namely, term frequency (*tf*), and term frequency multiplied by inverted document frequency (*tf*idf*).

Dataset Preprocessing. Both datasets were in Japanese and were preprocessed in the following ways (• – features used in both experiments; ◦ – features used only in cyberbullying detection).

- **Tokenization**: All words, punctuation marks, etc. are treated as separate features (later: **`TOK`**).
- **Lemmatization**: Same as above but the words are represented in their generic (dictionary) forms, or "lemmas" (later: **`LEM`**).
- **Parts of speech (POS)**: POS are used instead of words (later: **`POS`**).
- **Tokens with POS**: Both words and POS information is included in one element (later: **`TOK+POS`**).
- **Lemmas with POS**: Same as above but with lemmas instead of words (later: **`LEM+POS`**).
- **Chunking**: Sentences are divided into sub-parts (chunks) by grammatical rules, such as noun phrases, verb phrases, etc. (later: **`CHUNK`**).

- **Dependency structure**: Same as above, but with information on grammatical relations between chunks (later: **DEP**).
- **Chunks with Named Entities**: Chunks with named entities (private names, numericals, etc.) annotated on sentences (later: **CHUNK+NER**).
- **Dependency structure with Named Entities**: Both dependency relations and named entities are used (later: **DEP+NER**).

Each kind of preprocessing (or feature set) represents a different level of generalization. Higher sentence generalization produces less unique patterns, but the produced patterns are more frequent. This can be explained by comparing a tokenized (low generalization) sentence with its POS representation (high generalization). For example, in the sentence from Figure 1 the phrase "nice day" is represented by POS as ADJ N. There will be more ADJ N patterns than nice day, because many word combinations can be represented as ADJ N. There are also more words in a dictionary (around ten thousand) than POS labels (about a dozen). Comparison of classification results for different preprocessing methods can help specify whether it is better to represent sentences as more generalized or as more specific.

In meta-analysis we re-analyzed the results of experiments to answer the following questions:

- Is LC better than simple language modeling methods (n-grams, BoW)?
- Which preprocessing method (feature set) was the best?
- Which classifier modification was the best? (see sec. 3.2)

To answer these questions we compared the highest achieved balanced F-score within the threshold span achieved by each feature set. We also checked the correlations between generalization level of features and performance of each classifier modification. We also looked at break-even points (BEP) of Precision and Recall, showing which version was more balanced.

5.3. Small Dataset: Emotive Sentences

5.3.1. F-score Comparison Between Feature Sets

The highest achieved F-score was obtained by parts-of-speech (.774) while tokenized dataset with POS scored as second (.769). Both lemmatized datasets achieved the lowest scores (.744 and .746 for lemmas alone and with POS, separately). The initial intuition would suggest that parts-of-speech were the optimal setting, while lemmatization decreased the results. Worse results also tended to have wider dispersion between Precision and Recall.

As for the performance of modifications, all of the best classifier versions always used length awarded (LA), with either all ambiguous patterns (AMB) or zero-patterns (0P) deleted from pattern lists. No straightforward answer was obtained whether it was more useful to use patterns or n-grams. Although three out of five highest-scoring settings were based on n-grams (POS, TOK+POS, TOK), patterns were always second best and the differences were not significant. The results showing the highest achieved F-scores were represented in Figure 2.

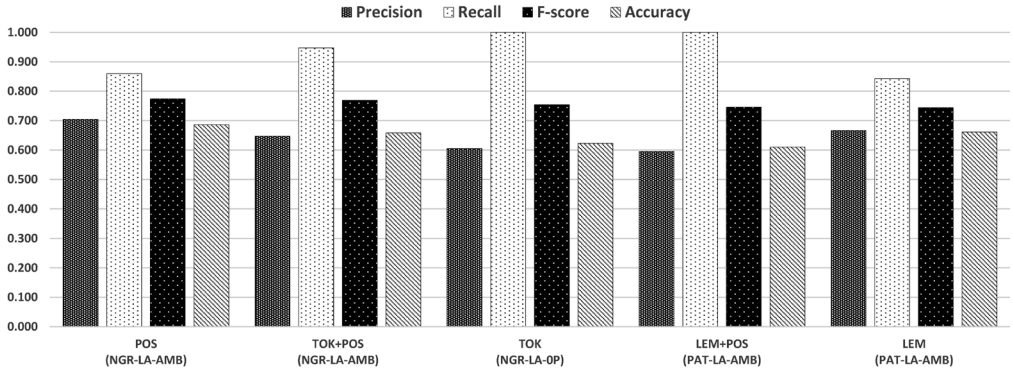


Fig. 2. Best F-scores for each preprocessing of emotive sentence dataset, ordered from left to right, with corresponding Precision, Recall and Accuracy. Classifier version that achieved the score – in brackets

Standard SVM-based classifier trained on Bag-of-Words language model and *tf*idf* weighting scored much lower, with the highest score of F1 = 73%, which indicates that simple BoW language model, even when used to train an efficient SVM classifier is not suitable for classification of emotive language.

Table 2. Comparison of Break-Even Points (BEP) of Precision and Recall for all classifier versions and preprocessing types for emotive sentences dataset. Best within each preprocessing group in bold type font. Best within each classifier type underlined

Feature sets	TOK	POS	TOK+POS	LEM	LEM+POS
PAT-ALL	0.650	0.701	0.713	0.649	0.632
PAT-0P	0.637	**0.710**	0.713	0.627	0.653
PAT-AMB	0.624	0.560	0.702	**0.664**	0.637
PAT-LA	0.679	-	0.722	0.551	0.651
PAT-LA-0P	0.676	-	0.722	0.626	0.659
PAT-LA-AMB	**0.688**	-	0.716	-	-
NGR-ALL	0.594	0.695	0.712	0.629	0.665
NGR-0P	0.609	**0.697**	0.712	0.633	0.665
NGR-AMB	0.595	0.668	0.664	0.610	0.628
NGR-LA	0.620	**0.680**	**0.723**	0.635	**0.682**
NGR-LA-0P	0.633	**0.680**	**0.723**	0.645	**0.682**
NGR-LA-AMB	**0.665**	-	**0.707**	0.652	0.655

5.3.2. Break-Even Point Analysis

In the BEP analysis we looked 1) which classifier version got the highest BEP, 2) which usually got the highest BEP for different dataset preprocessing, and 3) which preprocessing most often provided highest BEP.

The comparison revealed that TOK+POS dataset almost always performed best, achieving the highest BEP. This stands somewhat in contradiction to the results for F-scores, where

POS dataset obtained highest scores. However, detailed analysis revealed that, although POS achieved the single highest F-score, for other thresholds they scored similarly, or lower than other datasets. Moreover, even with POS as the highest, TOK+POS was still the second-best. While also being the most balanced (highest BEP for almost all cases), TOK+POS could be the optimal for analysis of emotive sentences.

This would also suggest that the method works better on more specific, less generalized features. Although the best BEP of all, with P = R = F = 0.723 was achieved by n-gram based classifier awarding pattern length in weight calculation, again, there was no clear answer whether it was better to use patterns, or *n*-grams. Comparison of all BEPs for all classifier versions and experiment settings is represented in Table 2.

5.3.3. Influence of Dataset Generalization on Results

Next we analyzed the influence of dataset preprocessing on the results. To achieve this we needed a quantifiable measure showing dataset generalization. A dataset is the more generalized, the fewer number of frequently appearing unique features it produces. Therefore to estimate dataset generalization level we decided to apply Lexical Density (LD) score [24]. It is a score representing an estimated measure of content per lexical units for a given corpus, and is calculated as the number of all unique words from the corpus divided by the number of all words in the corpus. However, since in our research we used a variety of different features, not only words, we will further call this measure Feature-based Lexical Density, or shortly, Feature Density (FD).

Table 3. Analysis of influence of dataset generalization for emotive sentences dataset

Dataset Preprocessing		**No. of unique unigrams**	**No. of all unigrams**	**Feature Density (FD)**	**Highest achieved F-score**	**Highest unmodified F-score**	**BEP**
Fetaure sophistication ←low high→	**TOK+POS**	311	821	0.3788	0.769	0.755	0.723
	TOK	306	821	0.3727	0.754	0.733	0.688
	LEM+POS	280	821	0.3410	0.746	0.733	0.682
	LEM	276	821	0.3362	0.744	0.733	0.664
	POS	12	819	0.0147	0.774	0.728	0.710
		unique 1ngr with			**FD with**		
		F1	**F1-unmod.**	**BEP**	**F1**	**F1-unmod.**	**BEP**
Pearson Correlation		-0.6018	0.5114	-0.3007	-0.6018	0.5114	-0.3007
Coefficient (*p*-value)		($p = 0.283$)	($p = 0.378$)	($p = 0.623$)	($p = 0.283$)	($p = 0.378$)	($p = 0.623$)
with statistical				**F1 & BEP**		**F1-unmod. & BEP**	
significance				*0.921		0.5735	
(p-value)				($p = 0.0265$)		($p = 0.312$)	

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

After calculating FD for all datasets we calculated Pearson's correlation coefficient (*p*-value) to see if there was any correlation between dataset generalization (FD) and the

results. Pearson's coefficient can achieve scores from 1.0 (perfect positive correlation), through 0.0 (no correlation) to -1.0 (perfect negative correlation). In comparison we used the highest achieved F-scores. However, since the highest overall F-scores were various classifier settings (all patterns, or zero-patterns deleted; with length awarded, or not, etc.), we also used an unmodified version of the classifier (**PAT-ALL**). As an equivalent set of results we also used BEPs. Finally, we verified whether the correlations were statistically significant.

Firstly, the highest achieved F-scores were significantly positively correlated with BEPs, which indicates that both measures show in general similar tendencies. The highest achieved F-scores indicated somewhat strong negative correlation with both unique unigrams as well as with Feature Density score, and mild negative correlation with BEPs. Interestingly, highest F-scores for unmodified dataset (all patterns) were somewhat positively correlated with unique unigrams and FD. The two results stand in contradiction, since the first one suggests that the less dense is the feature set the higher the results will get, while the second result indicates the opposite. This is most probably due to the least feature-dense POS-tagged dataset, which achieved the highest score. Unfortunately, neither of the correlations were statistically significant.

5.4. Medium-sized Dataset: Cyberbullying

5.4.1. F-score Comparison Between Feature Sets

The best F-score (.803) was achieved by lemmatization with POS information. Interestingly, while the winning settings showed high consistency between Precision and Recall, close to BEP (.802), for other preprocessing settings, the lower was the F-score, the wider was the gap between P and R. This is meaningful not only regarding the general performance, also provides insight into the influence of generalization on performance.

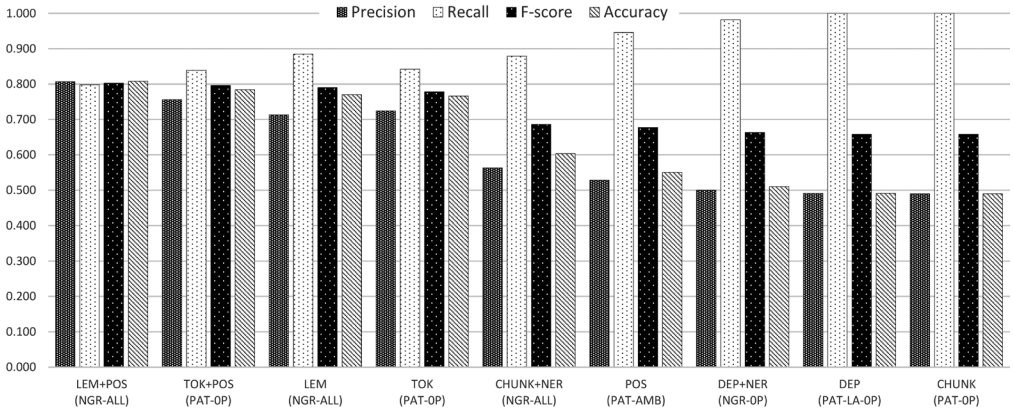


Fig. 3. Best F-scores for each dataset, ordered from left to right, with corresponding Precision, Recall and Accuracy. Classifier version that achieved the score – in brackets

Although the best score was achieved by n-grams, both settings were the highest interchangeably. For example, second best (F1 = 0.796) was pattern-based (TOK+POS/PAT-0P) third best was n-gram based, fourth – again – patterns, etc. This suggests that we need to perform more experiments, most desirably on a wider threshold span to choose whether patterns or n-grams are better. The most optimal classifier settings was the unmodified one, or the one with zero-patterns deleted. This suggests, that, in the case of cyberbullying messages, it is more effective to use ambiguous patterns in classification. The results can be clustered into two groups: with a small and with a wide gap between P and R This grouping is similar further in BEP analysis. Figure 3 shows the F-scores ordered decreasingly from left to right.

5.4.2. Break-Even Point Analysis

As for BEPs, the highest score of all (P = R = F1 = .802) was achieved by n-gram-based classifier. Lemmatized dataset combined with part-of-speech usually scored highest, differently to emotive sentence dataset, where this setting was one of the worst. The method usually performed better on more specific feature sets ((TOK, LEM, TOK+POS, LEM+POS), than for more generalized ones (POS, CHUNK, DEP, CHUNK+NER, DEP+NER). The results for best BEPs for all versions of the classifier were represented in Table 4.

Table 4. Break-even points for all feature sets on cyberbullying dataset

Feature sets	TOK	LEM	POS	TOK +POS	LEM +POS	CHUNK	DEP	CHUNK +NER	DEP +NER
PAT	0.761	0.751	0.613	0.785	0.781	**0.633**	0.566	0.603	0.510
PAT-0P	0.763	0.751	0.613	**0.786**	0.781	0.632	0.551	0.605	0.512
PAT-AMB	**0.770**	0.751	0.613	0.764	0.782	0.629	**0.591**	0.603	0.514
PAT-LA	0.729	0.748	0.613	0.726	0.781	0.632	0.568	-	0.505
PAT-LA-0P	0.729	0.737	0.596	0.726	0.760	**0.633**	0.549	-	0.505
PAT-LA-AMB	0.711	0.737	0.594	0.715	0.761	0.629	**0.591**	-	0.516
NGR	0.761	**0.784**	**0.614**	0.785	**0.802**	0.632	0.566	**0.655**	0.547
NGR-0P	0.762	**0.784**	0.613	**0.786**	**0.802**	0.632	0.551	0.652	**0.548**
NGR-AMB	**0.770**	0.767	0.570	0.764	0.777	0.612	**0.591**	0.610	0.526
NGR-LA	0.729	0.767	0.605	0.726	0.762	**0.633**	0.551	0.619	0.546
NGR-LA-0P	0.729	0.768	0.607	0.726	0.769	0.631	0.559	0.622	**0.548**
NGR-LA-AMB	0.711	0.762	0.596	0.715	0.750	0.613	0.589	0.589	0.529

5.4.3. Influence of Generalization on Results

Feature Density score revealed somewhat strong negative correlation (around –0.7) between the results and FD. This means that the results were better when the FD was low. The correlation was not ideal due to the fact that the dataset with the lowest FD (POS)

achieved one of the lowest results. Interestingly, preprocessing methods resulting in very high FD (dependency parsing, etc.) also achieved similarly low results. For the given datasets the performance is growing along with decreasing FD, until the lowest FD is reached (POS), which also obtained low results. Thus, in the future we plan to use the FD measure to find a preprocessing method with optimal feature density, resulting in even better results. The analysis of influence of dataset generalization on results is represented in Table 5.

Table 5. Analysis of influence of generalization on results for cyberbullying dataset

Dataset Preprocessing			No. of unique unigrams	No. of all unigrams	Feature Density	Highest achieved F-score	Highest unmodified F-score	BEP
Feature sophistication	←low high→	DEP	12802	13957	0.917	0.658	0.658	0.591
		DEP+NER	12160	13956	0.871	0.663	0.662	0.548
		CHUNK	11389	13960	0.816	0.658	0.658	0.633
		CHUNK+NER	10657	13872	0.768	0.686	0.684	0.655
		TOK+POS	6565	34874	0.188	0.796	0.795	0.786
		TOK	6464	36234	0.178	0.778	0.778	0.770
		LEM+POS	6227	36426	0.171	0.803	0.783	0.802
		LEM	6103	36412	0.168	0.790	0.764	0.784
		POS	13	26650	0.000	0.677	0.677	0.614
			unique 1ngr with			FD with		
			F1	F1-unmod.	BEP	F1	F1-unmod.	BEP
Pearson Correlation			-0.450	-0.453	-0.431	-0.735	-0.736	-0.706
Coefficient (*p*-value)			($p = 0.224$)	($p = 0.221$)	($p = 0.247$)	($p = 0.0242$)	($p = 0.024$)	($p = 0.0336$)
with statistical					F1 & BEP		F1-unmod. & BEP	
significance					0.9681		0.9595	
(*p*-value)					($p = 0.00002$)		($p = 0.00004$)	

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

6. Conclusions and Future Work

We presented a meta-analysis of experiments performed with Language Combinatorics (LC), a novel method for language model generation based on combinatorial manipulations of sentence elements. For the analysis we selected the most exhaustive experiments, namely, emotive sentence detection (small dataset) [3] and cyberbullying detection (medium-sized dataset) [2]. We briefly summarized the experiments and discussed general implications for future implementations of LC.

Meta-analysis revealed many contradictory results. For example, POS-tagged dataset obtained the highest F-score for the small dataset, but the worst for the medium-sized dataset.

Feature Density (FD) significantly correlated with F-scores and BEPs for mid-size dataset, but not for small dataset. Also, results for pattern list and classifier modifications were not consistent among the two datasets. For small dataset awarding pattern length resulted in better scores, usually boosted further by deleting ambiguous patterns. For medium dataset such modifications usually hindered the results.

As for similarities, awarding both pattern length and occurrence in classification was usually not effective. Therefore this option could be discarded in future experiments to reduce the overall time required for experiment.

In the future we plan to unify the meta-analysis. This time small dataset also had smaller number of preprocessing variations used in experiments, which could influence the correlations of Feature Density with results. Also, although there was no clear answer on whether patterns or n-grams were more effective, both always produced better results than Bag-of-Words model. To further confirm this result, we plan to extend the scope of patterns length from six elements to the maximal possible length, since the least frequent patterns will still be filtered out during the feature extraction procedure.

We plan to train other classifiers (SVM, Neural Networks, etc.) on the proposed pattern-based language model. This however will require much stronger hardware than was available at the time of writing. Finally, the results of meta-analysis could have been influenced by various differences in datasets – not only in their sizes, but also, e.g., the type of language. Therefore in the future we plan to repeat the experiments on size-unified datasets.

References

[1] Ptaszynski M., Masui F., Rzepka R., Araki K., *First Glance on Pattern-based Language Modeling*, Language Acquisition and Understanding Research Group Technical Reports, 2014.

[2] Ptaszynski M., Masui F., Kimura Y., Rzepka R., Araki K., *Extracting Patterns of Harmful Expressions for Cyberbullying Detection*, Proceedings of LTC'15, 2016, 370-375.

[3] Ptaszynski M., Masui F., Rzepka R., Araki K., *Subjective? Emotional? Emotive?: Language Combinatorics based Automatic Detection of Emotionally Loaded Sentences*, Linguistics and Literature Studies, Vol. 5, No. 1, 2017, 36-50.

[4] Bickel S., Haider P., Scheffer T., *Predicting sentences using n-gram language models*, Proceedings of HLT-EMNLP 2005, 2005, 193-200.

[5] Li Haizhou, Bin Ma, *A phonotactic language model for spoken language identification*, Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005, 515-522.

[6] Ponte J.M., Croft W.B., *A language modeling approach to information retrieval*, Proceedings of the 21st annual international ACM SIGIR Conference on Research and Development in Information Retrieval, 1998, 275-281.

[7] Brown P.F., Cocke J., Pietra S.A.D., Pietra V.J.D., Jelinek F., Lafferty J.D., Mercer R.L., Roossin P.S., *A statistical approach to machine translation*, Computational linguistics, Vol. 16, No. 2, 1990, 79-85.

[8] Mays E., Damerau F.J., Mercer R.L., *Context based spelling correction*, Information Processing & Management, Vol. 27, No. 5, 1991, 517-522.
[9] Kupiec J., *Robust part-of-speech tagging using a hidden Markov model*, Computer Speech & Language, Vol. 6, No.3, 1992, 225-242.
[10] Hu Y., Lu R., Li X., Chen Y., Duan J., *A language modeling approach to sentiment analysis*, Computational Science – ICCS 2007, 1186-1193.
[11] Ptaszynski M., Rzepka R., Araki K., Momouchi Y., *Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion*, International Journal of Computational Linguistics (IJCL), Vol. 2, No. 1, 2011, 24-36.
[12] Harris Z., *Distributional Structure*, Word, Vol. 10, N. 2/3, 1954, 146-162.
[13] Cambria E., Hussain A., *Sentic Computing: Techniques, Tools, and Applications*, Springer, 2012.
[14] Lu Y., Zhai C.X., *Positional Language Models for Information Retrieval*, 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2009, 299-306.
[15] Markov A.A., *Extension of the limit theorems of probability theory to a sum of variables connected in a chain*, Reprinted in Appendix B of: R. Howard, *Dynamic Probabilistic Systems*, Vol. 1: *Markov Chains*, John Wiley and Sons, 1971.
[16] Huang X., Alleva F., Hon H.W., Hwang M.Y., Rosenfeld R., *The SPHINX-II Speech Recognition System: An Overview*, Computer, Speech and Language, Vol. 7, 1992, 137-148.
[17] Guthrie D., Allison B., Liu W., Guthrie L., Wilks Y., *A closer look at skip-gram modelling*, Proceedings of LREC-2006, 2006, 1-4.
[18] Pickhardt R., Gottron T., Korner M., Wagner P.G., Speicher T., Staab S., *A Generalized Language Model as the Combination of Skipped n-grams and Modified Kneser Ney Smoothing*, Proceedings of ACL 2014, 2014, 1145-1154.
[19] Ptaszynski M., Lempa P., Masui F., *A Modular System for Support of Experiments in Text Classification*, Technical Transactions, vol. 7-B/2015, 229-243.
[20] Nakajima Y., Ptaszynski M., Honma H., Masui F., *Investigation of Future Reference Expressions in Trend Information*, Proceedings of the 2014 AAAI Spring Symposium Series, 2014, 31-38.
[21] Ptaszynski M., Dybala P., Rzepka R., Araki K., *Affecting Corpora: Experiments with Automatic Affect Annotation System – A Case Study of the 2channel Forum*, Proceedings of PACLING-09, 2009, 223-228.
[22] Human Rights Research Institute Against All Forms for Discrimination and Racism in Mie Prefecture, Japan, http://www.pref.mie.lg.jp/jinkenc/hp/ (access: 21.04.2017).
[23] Ministry of Education, Culture, Sports, Science and Technology (MEXT), *'Netto-jo no ijime' ni kansuru taio manyuaru jirei shu (gakko, kyoin muke)*, MEXT, 2008.
[24] Ure J., *Lexical density and register differentiation*, [in:] *Applications of Linguistics*, (eds.) G. Perren, J.L.M. Trim, Cambridge University Press, London 1971, 443-452.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
MECHANICS
DOI: 10.4467/2353737XCT.17.201.7430

Norbert Radek
Faculty of Mechatronics and Machine Design, Kielce University of Technology
Jacek Pietraszek (pmpietra@mech.pk.edu.pl)
Institute of Applied Informatics, Faculty of Mechanical Engineering, Cracow University of Technology
Agnieszka Szczotok
Department of Materials Science, Silesian University of Technology
Jozef Bronček
Faculty of Mechanical Engineering, University of Žilina

Production and properties of electro-spark deposited coatings modified via LBM

Wytwarzanie i właściwości powłok elektroiskrowych modyfikowanych obróbką laserową

Abstract

The paper is concerned with determining the influence of laser treatment process on the properties of electro spark coatings. The properties were assessed after a laser treatment by analysing microstructure and X-ray diffraction and measuring surface geometric structure and microhardness. The studies were conducted using WC-Cu electrodes produced by powder metallurgy of nanostructural powders. The coatings were deposited by means of the EIL-8A model and they were laser-treated with the Nd:YAG, BLS 720 model.

Keywords: electro-spark deposition (ESD), laser beam machining (LBM), coating, properties

Streszczenie

W artykule badano wpływ obróbki laserowej na właściwości powłok nanoszonych elektroiskrowo. Ocenę właściwości powłok po obróbce laserowej przeprowadzono na podstawie obserwacji mikrostruktury, analizy składu fazowego oraz pomiarów struktury geometrycznej powierzchni i mikrotwardości. Badania przeprowadzono, wykorzystując elektrody WC-Cu, które zostały wytworzone przez spiekanie nanostrukturalnych proszków. Do nanoszenia powłok elektroiskrowych użyto urządzenie model EIL-8A. Obróbkę laserową nałożonych powłok elektroiskrowych wykonano laserem Nd:YAG, model BLS 720.

Słowa kluczowe: obróbka elektroiskrowa, obróbka laserowa, powłoka, właściwości

1. Introduction

A number of modern surface processing methods use an energy flux. The examples include electro-spark deposition and laser treatment. Electro-spark deposition (ESD) is a cheap, high-energy process. Developed in the post-war period, this technology has been frequently modified. Its main advantages are the ability to select precisely the area to be modified, ability to select coating thickness, which may range from several to several dozen micrometers, good adhesion of a coating to a substrate, and finally, cheap and simple equipment for coating deposition. ESD has been known under several other terms such as spark hardening, electric spark toughening and electro-spark alloying (ESA).

The processes of coating formation on metal parts including electro-spark deposition involves mass and energy transport accompanied by chemical, electrochemical and electrothermal reactions [1]. Today, different electro-spark deposition techniques are used, which are suitable for coating formation and surface microgeometry formation [2-4].

Electro-spark deposited coatings have some disadvantages but these can be easily eliminated. One of the methods is laser beam machining (LBM); a laser beam is used for surface polishing, surface geometry formation, surface sealing or for homogenizing the chemical composition of the deposited coatings [5, 6].

It is envisaged that the advantages of laser-treated electro-spark coatings will include lower roughness, lower porosity, better adhesion to the substrate, higher wear and seizure resistance, higher fatigue strength due to the occurrence of compressive stresses on the surface and higher resistance to corrosion.

2. Materials and treatment parameters

The working electrode (a stationary) was made from C45 carbon steel. The elemental composition of the steel was as follows: (wt.%): C: 0.42-0.50, Mn: 0.50-0.80, Si: 0.10-0.40, P: 0.04, S: 0.04.

In the experiment, the coatings were electro-spark deposited using a WC-Cu (50% WC and 50% Cu) electrode with a cross-section of 4 x 6 mm (the anode) – onto samples made of carbon steel C45 (the cathode). The main characteristics of the powders used in this work are included in Table 1.

The powders were mixed for 30 minutes in the chaotic motion *Turbula T2C* mixer. The mixture was then poured into rectangular cavities of a graphite mould, each 6×40 mm in cross section, and consolidated by passing an electric current through the mould under uniaxial compressive load. A 3 minute hold at 950°C and under a pressure of 40 MPa permitted obtaining electrodes of porosity <10% and strength sufficient to maintain integrity when installed in the electrode holder.

The equipment used for electro-spark deposition was an EIL-8A model. Based on our previous work and instructions given by the equipment manufacturer, the following parameters were assumed to be optimal for ESD voltage U = 230 V, capacitor volume C = 150 μF, current intensity I = 0.7 A.

Table 1. Powders used to manufacture WC-Cu electrodes

Powder	Particle Size, μm	Producer
WC	~0.2*	OMG
Cu	~0.04*	NEOMAT

* measured using Fisher Sub-Sieve Sizer

Then, the coatings were treated with an Nd:YAG laser (impulse mode), model BLS 720. The samples with electro-spark deposited coatings were laser-modified using the following parameters: spot diameter $d = 0.7$ mm, power $P = 60$ W, laser beam velocity $v = 250$ mm/min, nozzle-workpiece distance $\Delta f = 6$ mm, pulse duration $t_i = 0.4$ ms, pulse repetition frequency $f = 50$ Hz, beam shift jump $S = 0.4$ mm, nitrogen gas shield $Q = 25$ l/min.

3. Results and discussion

3.1. Microstructure and X-ray diffraction analysis

A microstructure analysis was conducted for WC-Cu coatings before and after laser treatment using a scanning electron microscope Joel JSM-5400.

Figure 1 shows the microstructure of an ESD WC-Cu coating. It is clear that the thickness of the obtained layer varied from 36 to 60 μm, whereas the heat affected zone (HAZ) ranged from 20 to 30 μm into the substrate. Figure 1 also reveals a clear boundary between the coating and the substrate and pores within the coating. The ESD WC-Cu coatings were modified by laser treatment, which caused changes in their composition. The laser treatment homogenizes the coating chemical composition, causes structure refinement, and crystallization of non-equilibrium phases due to the occurrence of temperature gradients and high cooling rates.

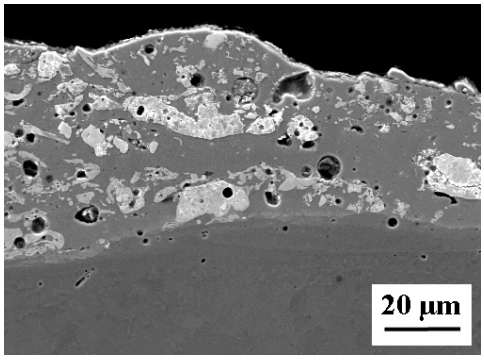


Fig. 1. WC-Cu coating microstructure after electro-spark alloying

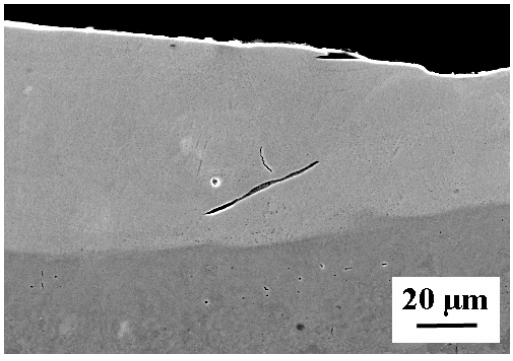


Fig. 2. Microstructure of the electro-spark alloying WC-Cu coating after treatment with an Nd:YAG laser

The laser-modified outer layer does not possess microcracks or pores (Fig. 2). There is no discontinuity of the coating-substrate boundary. The thickness of the laser-treated WC-Cu

coatings ranges from 40 to 62 μm. Moreover, the heat affected zone (HAZ) is in the range of 25 to 35 μm, and the content of carbon in the zone is higher.

Using the X-ray diffraction method, the analysis of the phase composition of the examined coatings was performed with Philips PW 1830 instrument. Cu Kα filtered radiation was employed. Tests were carried out for the 2θ angle in the range 30-60° and the scanning velocity of 0.05°/3 seconds.

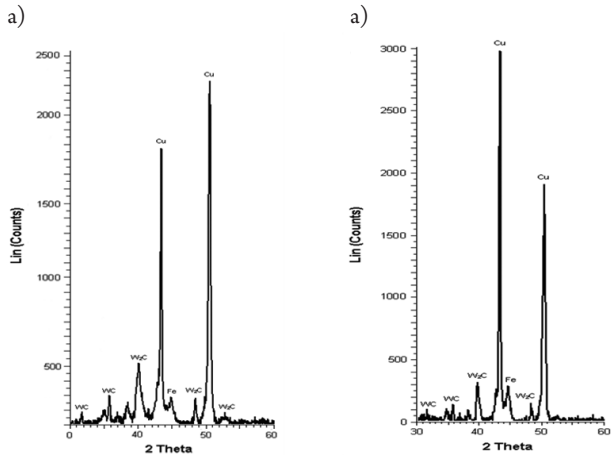


Fig. 3. X-ray diffraction pattern of the WC-Cu coating: a) before laser treatment; b) after laser treatment

The analysis of the phase composition of the WC-Cu coating (Fig. 3a) revealed that the surface layer of the coating consisted mainly of Cu and W_2C and a small amount of WC and Fe. Laser treatment did not cause the melting of the WC-Cu coating to penetrate into the substrate material (Fig. 3b). The surface layer of the WC-Cu coating after the laser treatment has the same composition as that of the coating before the treatment. The most intense peaks originate from Cu (Fig. 3a and 3b).

3.2. Measurements of the surface geometric structure and microhardness

Measurements of the surface geometric structure were carried out at the Laboratory of Computer Measurements of Geometric Quantities of the Kielce University of Technology. Those were performed using Talysurf CCI optical profiler that employs a coherence correlation algorithm patented by Taylor Hobson company. The algorithm makes it possible to take measurements with the resolution in the axis z below 0.8 nm. The result of measurements is recorded in 1024 x 1024 measurement point matrix, which for the x10 lens yields the 1.65 mm x 1.65 mm measured area and the horizontal resolution 1.65 μm x 1.65 μm.

Three-dimensional surfaces and their analysis with TalyMap Platinum software made it possible to precisely identify the geometric structure of the surfaces under consideration. Table 2 provides major parameters of the surface geometric structure of the examined specimens.

Table 2. Parameters of the surface geometric structure

SGS parameters	Coating	
	WC-Cu	WC-Cu + laser
Sa [μm]	4.02	6.95
Sq [μm]	5.24	8.48
Ssk	0.15	0.02
Sku	3.89	2.77
Sp [μm]	26.44	34.03
Sv [μm]	21.21	66.76
Sz [μm]	47.65	100.80

A greater value of the mean arithmetic deviation of surface roughness *Sa*, a basic amplitude parameter in the quantitative assessment of the state of the surface under analysis, was recorded for the specimen after the laser treatment. For the specimen before the laser treatment, the value of this parameter was almost 50% smaller.

A similar tendency is observed for the root mean square deviation of surface roughness *Sq*. Complementary information on how the surface of examined elements is shaped is provided by amplitude parameters, namely the coefficient of skewness (asymmetry) *Sku* and the coefficient of concentration (kurtosis) *Ssk*. Those parameters are sensitive to the occurrence of local hills or valleys, and also defects on the surface. *Ssk* parameter has a positive value for both specimens. The value is close to zero for the specimen before the treatment, which indicates the symmetrical location of the distribution of ordinates with respect to the mean plane. The values of kurtosis that were obtained are close to *Sku* = 3, which indicates that the distribution of ordinates for both specimens is close to normal distribution.

Before the laser treatment, the specimen had a random isotropic structure (*Iz* = 88.52%), whereas after the treatment, the structure became periodic, located in the transient area between isotropic and anisotropic structures (*Iz* = 55.32%). That is confirmed by the shape of the autocorrelation function of both surfaces. For the surface before the treatment, the shape is circular and symmetrical, whereas for the surface after the treatment, it is asymmetrical and elongated.

The microhardness was determined using the Vickers method (Microtech MX3 tester). The measurements were performed under the load of 0.4 N. The indentations were made in perpendicular microsections in three zones: the white homogeneous difficult-to-etch coating, the heat affected zone (HAZ) and the substrate. The microhardness of the substrate after the electro-spark deposition was 278 HV0.4; the same value was reported for the substrate before the process. There was a considerable increase in microhardness after depositing the WC-Cu coating. The microhardness of the WC-Cu coating was approx. 643 HV0.4. The microhardness of the WC-Cu coating in the heat affected zone (HAZ) after the electro-spark treatment was 58% higher than that of the substrate material. Laser treatment had a favourable effect on the changes in the microhardness of the electro-spark deposition of the WC-Cu coating. There was an increase of 122% in the microhardness of the WC-Cu coating.

4. Summary

1. Laser irradiation of coatings resulted in the healing of micro-cracks and pores.
2. Parameters of surface geometric structure of electro-spark coatings have lower values when compared to SGS parameters of coatings after the laser treatment.
3. The surface layer of the WC-Cu coating before and after the laser treatment consists mainly of Cu and W_2C and a small amount of WC and Fe.
4. Laser treatment caused a 9% decrease in the microhardness of the electro-spark alloying WC-Cu coatings.

References

[1] Galinov I.V., Luban R.B., *Mass transfer trends during electrospark alloying*, Surface & Coatings Technology, 79, 1996, 9-18.

[2] Antoszewski B., Evin E., Audy J., *Study of the effect of electro-spark coatings on friction in pin-on-disc testing*, Journal of Tribology – Transactions of the ASME, 3, 2008, 253-262.

[3] Chang-bin T., Dao-xin L., Zhan W., Yang G., *Electro-spark alloying using graphite electrode on titanium alloy surface for biomedical applications*, Applied Surface Science, 257, 2011, 6364-6371.

[4] Radek N., *Determining the operational properties of steel beaters after electrospark deposition*, Eksploatacja i Niezawodność – Maintenance and Reliability, 4, 2009, 10-16.

[5] Radek N., Bartkowiak K., *Performance properties of electro-spark deposited carbide-ceramic coatings modified by laser beam*, Physics Procedia, 5, 2010, 417-423.

[6] Radek N., Antoszewski B., *Influence of laser treatment on the properties of electro-spark deposited coatings*, Kovove Materialy – Metallic Materials, 1, 2009, 31-38.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
MECHANICS
DOI: 10.4467/2353737XCT.17.202.7431

Patryk Różyło (p.rozylo@pollub.pl)
Department of Computer Modelling and Metal Forming Technologies, Faculty of Mechanical Engineering, Lublin University of Technology

Łukasz Wójcik
Department of Machine Design and Mechatronics, Faculty of Mechanical Engineering, Lublin University of Technology

Comparative analysis of the experimental and numerical research of the stamping process of axially-symmetrical elements

Analiza porównawcza badań doświadczalnych i numerycznych tłoczenia elementów osiowo-symetrycznych

Abstract

The paper presents experimental and numerical results of the stamping process of an axially-symmetrical element. The subject of the study was a sample made of a 0.47 mm thick aluminium alloy Al-1100 sheet. Experimental studies were conducted on a universal testing machine, while numerical simulations were carried out on two different numerical software programs Abaqus® and Deform-3D. The results obtained from the numerical analysis allowed to develop a numerical model of the stamping process.

Keywords: stamping process, numerical and experimental analysis, Abaqus, Deform-3D

Streszczenie

W artykule przedstawiono wyniki badań eksperymentalnych i numerycznych procesu tłoczenia elementu osiowo-symetrycznego. Przedmiotem badań była próbka wykonana ze stopu aluminium Al-1100 o grubości 0,47 mm. Badania doświadczalne przeprowadzono na uniwersalnej maszynie wytrzymałościowej, natomiast symulacje numeryczne zrealizowano w dwóch różnych programach komputerowych Abaqus® oraz Deform-3D. Wyniki uzyskane z analizy numerycznej pozwoliły na opracowanie numerycznego modelu procesu tłoczenia.

Słowa kluczowe: proces tłoczenia, analizy numeryczne i doświadczalne, Abaqus, Deform-3D

1. Introduction

It is common to reduce costs and shorten the design time of metal sheet processing. The designing of sheet metal stamping is a complicated and time-consuming process. The limitations that negatively influence the optimisation of the element shaping processes with a non-expandable surface are complex experimental research, which generates high costs. The costs relate to the implementation of often complex tools. Therefore, methods are sought that would eliminate experimental research. The most commonly used are computer (numerical) techniques, assisting the design of metal forming processes (finite element method or finite volume method). The application of numerical calculations and simulation has allowed for costs and time reduction of the design and preliminary testing of stamping processes. Simulation software is mainly used to predict the direction of the flow of metal, of stress distribution analysis, deformation, temperature or possible defects.

The sheet stamping process is the most common process used to shape thin-walled components. A detailed classification of the stamping process is presented in [1]. The stamping process involves converting the flat sheet portion into a non-expandable surface element. In the stamping process, axially-symmetric [2], rectangular [3] and complex shaped drawpieces are produced.

During the stamping of thin metal sheets, radial stresses – and the peripheral tensile – compression occurs in the formed flanges. It is assumed that the process is carried out in a two-dimensional stress state.

The analysis of the stamping process is based on determining the occurring phenomena and interaction between the drawing and stretching areas. The analysis of the phenomena occurring during sheet metal stamping is described in publication [4-7].

The stamping process is burdened with two major constraints that negatively affect the shaping of the sheet. The most commonly occurring phenomena, restricting the stamping process, are the corrugating of the drawpiece flange and bursting of the cylindrical part of the drawpiece wall. The limitations causing drawpiece or tool damage are described by the researchers in publication [8].

2. Subject of the study

The subject of the study was a disc-shaped aluminium alloy sample. The test sample was characterised by a fixed diameter of 50 mm and a thickness of 0.47 mm. The graphical representation of the test sample is shown in Figure 1.

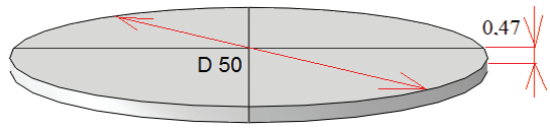


Fig. 1. Geometrical parameters of the specimen

The test samples were characterised by specific material properties, consistent with the definition of the elastic-plastic model. All experimentally determined material characteristics required the correct implementation of the dynamic issue in the FEM analysis, which are presented in Table 1.

Table 1. Material properties of the test sample (own research)

Al-1100 aluminium alloy material properties					
Density [kg/m^3]	**Young Modulus** [MPa]	**Poisson's ratio** [-]	**Yield point** [MPa]	**Tensile strength** [MPa]	**Elongation** [%]
2700	70000	0.33	85	115	2

3. Experimental studies

Experimental studies were carried out on the universal testing machine Instron 3369. In order to properly carry out the stamping process, the test specimen was suitably positioned in the lower die. A universal testing machine is equipped with a guideway at the upper part of the die, which provides axial guidance of the stamp. No initial pressure was applied to the test sample. As part of the research, load-displacement characteristics of the stamp have been determined. The process was carried out with a constant traverse speed of 1.67 mm/s. The duration of the analysis was 38.92 s, which was related to the displacement of the stamp by 65 mm. The test stand is shown in Figure 2.

The experimental studies allowed obtaining the drawpiece, which was compared with the numerical model. In regard to the conducted test, further validation with computer simulations based on finite element method in two independent environments was needed.

Fig. 2. Testing equipment set-up for the stamping process

4. Numerical analysis

The numerical analyses were conducted by the use of two independent software – Abaqus and Deform-3D. The key stage in the preparation of the process was the appropriate definition of the material model based on experimental material parameters. The material model had elastic-plastic properties. Geometrical parameters were mapped from the measurements obtained from actual test assemblies. The general contact between cooperating tangential and normal surfaces was determined. A permanent friction model corresponding to the contact conditions of aluminium alloy and steel was adopted. The boundary conditions relevant to the experimental study are shown in Figure 3.

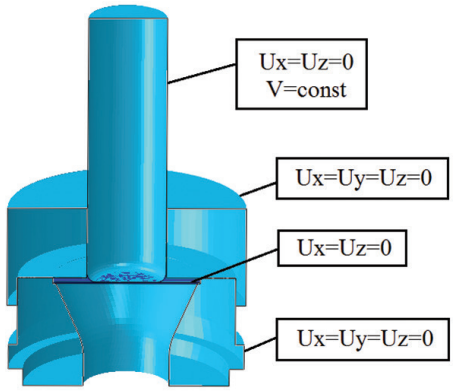


Fig. 3. Boundary conditions of the numerical model

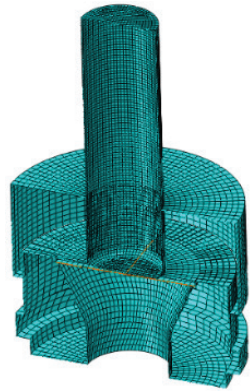


Fig. 4. Discrete model

The discretisation process was based on the use of two types of elements – the C3D8R (deformable) for the test sample in which simulation results were possible to obtain, and the R3D4 coating (non-deformable) for the remaining parts [9]. The number of finite elements of C3D8R type was equal to 25 000 (mesh density – 0.5 mm), while for R3D4 – 30000 (mesh density 1 mm). The adaptive mesh reconstruction algorithm was used. Adequate numerical methods were used for both software environments. The graphical representation of the discrete model obtained in the Abaqus program is shown in Figure 4.

Numerical analysis was based on the solution of the non-linear dynamic issue, under exactly the same conditions as the experimental study.

5. Testing results

The purpose of the numerical and experimental studies was to obtain the stamping process characteristics of the research system. The graphical presentation of the initial stages of the FEM simulation process is shown in Figure 5.

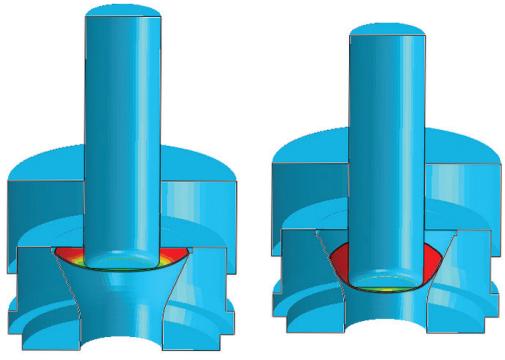


Fig. 5. Graphical presentation of the initial stages of the stamping process

As part of the obtained analysis, the actual form of the drawpiece has been obtained from the flat aluminium alloy sample, as well as from the performed simulations. The approximate drawpiece's shapes, presented in Figure 6, demonstrate the high quality of numerical models.

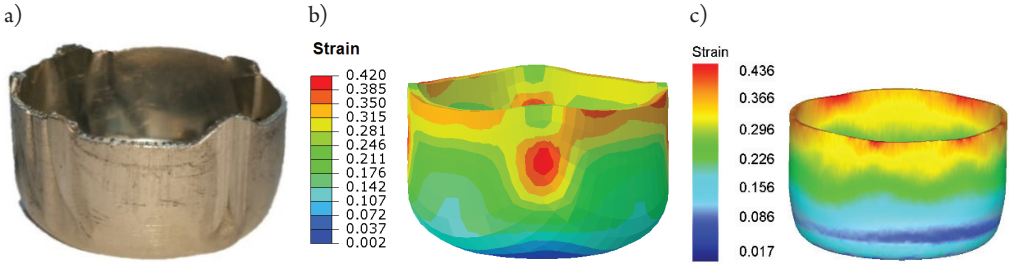


Fig. 6. Graphical representation of drawpieces: a) actual object; b) Abaqus numerical model; c) Deform-3D numerical model

The actual drawpiece is similar to the one obtained by the Abaqus numerical model. The drawpiece obtained from Deform-3D is slightly different from the other cases. Regarding the obtained results of strain level, a high convergence between Abaqus and Deform-3D software was demonstrated.

The conducted studies enabled to derive the necessary characteristics of the system, based on which the convergence of the research results was estimated. Figure 7 shows the stamping process characteristics.

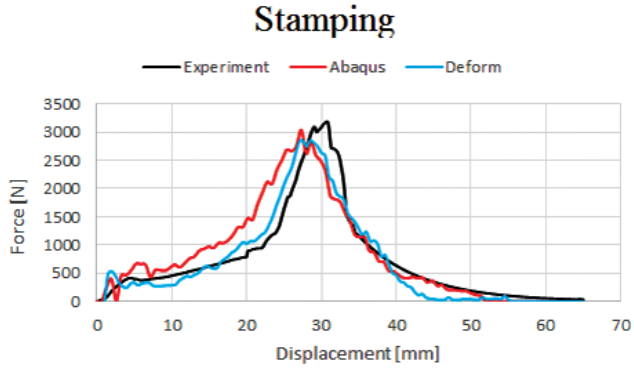


Fig. 7. Comparison of the characteristics of the stamping process

The obtained results, presented in the graph above, constitute about the high quality of the performed tests and prepared numerical models. Both experimental and simulation analysis have shown that the greatest force in the stamping process is approximately 30 mm in the displacement of the stamp. A detailed breakdown of the maximum force values is shown in Table 2.

The results shown in Table 2 exhibit a high convergence of the test outcome. The difference in force obtained by numerical tests in Abaqus relative to the actual sample is only 4.6%. The corresponding difference in Deform-3D is exactly 10.4%.

Table 2. Comparison of the results of maximum stamping forces

Experiment [N]	Abaqus [N]	Deform [N]	Difference Abaqus – EXP [%]	Difference Deform – EXP [%]
3185.4	3038.7	2853.7	4.6	10.4

6. Conclusions

The analysis of actual and numerical studies of the stamping process allowed determining the consistency of the obtained results. The results showed a high convergence between computer simulations and experimental studies. It has been demonstrated that there are advanced numerical computation programs that allow for a greater consistency of results – Abaqus and lower – Deform-3D within a given type of issue. The use of numerical calculations is supported by the use of several independent validation capabilities in regard to the actual experimental research. The Abaqus has generated better quality results than Deform-3D. The discrepancy between the Abaqus and the experiment is 4.6%, and between the Deform-3D and the experimental is 10.4%. The quality of the obtained results constitutes about properly prepared numerical model. The overall assessment of the quality of the research results constitutes a well-prepared numerical model.

References

[1] Romanowski W.P., *Poradnik obróbki plastycznej na zimno*, WNT, Warszawa 1976.

[2] Mori K., Nishijima S., Tan C.J., *Two stage cold stamping of magnesium alloy cups having small corner radius*, International Journal of Machine Tools and Manufacture, 49, 2009, 767-772.

[3] Chen F., Huong T., Chang C., *Deep drawing of square cups with magnesium alloy AZ31 sheets*, International Journal of Machine Tools and Manufacture, 43, 2003, 1553-1559.

[4] Pereira M., Rolfe B., *Temperature conditions during cold sheet metal stamping*, Journal of Materials Processing Technology, 214, 2014, 1749-1758.

[5] Vrolijk M., Lorenz D., Porzner H., Holecek M., *Supporting lightweight design: virtual modeling of hot stamping with tailored properties and warm and hot formed aluminum*, Procedia Engineering, 183, 2017, 336-342.

[6] Jin J., Wang X., *A single-step hot stamping-forging process for aluminum alloy shell parts with nonuniform thickness*, Journal of Materials Processing Technology, vol. 228, 2016, 170-178.

[7] Zhou J., Wang B., Lin J., Fu L., *Optimization of an aluminum alloy anti-collision side beam hot stamping process using a multi-objective genetic algorithm*, Archives of Civil and Mechanical Engineering, Vol. 13, 2013, 401-411.

[8] Isik K., Silva M., Tekkaya A., Martins P., *Formability limits by fracture in sheet metal forming*, Journal of Materials Processing Technology, 214, 2014, 1557-1565.

[9] Zienkiewicz O.C., Taylor R.L., *Finite Element Method*, Vol. 2: *Solid Mechanics*, 5th ed., Elsevier, 2000.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
MECHANICS
DOI: 10.4467/2353737XCT.17.193.7422

Dariusz Szwarkowski (szwarderek@gmail.com)
Jakub Zięba
Faculty Civil Engineering, Institute of Constructions Mechanics, Cracow University of Technology

A slope stability analysis of an open-pit mine to evaluate its suitability as a site for the intersection of the S-7 and S-52 expressways

Analiza stateczności kopalni odkrywkowej w oszacowaniu przydatności pod planowaną budowę trasy S-7 i północnej obwodnicy Krakowa trasy S-52

Abstract

This paper presents a comprehensive approach to the numerical modelling of the geotechnical issues related to the stability of the slopes of a former open-pit mine. The mine is located within short distance of the planned S-7 expressway route and the northern bypass of Krakow; therefore, there is a need of opinion as the project might have a significant impact on the surrounding area including structures planned nearby, as well as stability of slopes of former open-pit mine. The finite element method (FEM) was applied to the numerical analysis with the specific aim of assessing the risk of the movement of soil mass as far as the slopes of the Zesławice open-pit mine are concerned [4]. Field work and numerical analysis were conducted in reference to land reclamation plans of former Miocene clay mine located in the Carpathian Foredeep. The numerical modelling includes zoning plans. The numerical modelling was conducted with a terrestrial laser scanner application [12]. In addition, spatial distribution and the identification of the parameters of the subsoil layers was performed. A numerical soil model, based on Mohr–Coulomb theory, was also taken into consideration. shear reduction method (SRM) was applied to determine the slope stability; the areas at risk of mass movement were then identified on the basis of the slope stability ratio.

Keywords: numerical modelling, failure surface, slope stability, terrestrial laser scanning

Streszczenie

W artykule przedstawiono kompleksowe podejście do modelowania numerycznego zagadnień geotechnicznych, które dotyczą stateczności skarp dawnej kopalni odkrywkowej. Kopalnia zlokalizowana jest w pewnej odległości od planowanego przebiegu trasy S-7 i północnej obwodnicy Krakowa, konieczne jest zatem przeanalizowanie, czy projekt może mieć znaczący wpływ na otaczający obszar, planowane i istniejące obiekty budowlane oraz stateczność stoków byłej kopalni odkrywkowej. Metoda elementów skończonych (MES) została zastosowana do analizy numerycznej, przede wszystkim w celu dokonania oceny ryzyka ruchów masowych gruntów stoków nieczynnej kopalni odkrywkowej. Prace terenowe i analizy numeryczne zostały przeprowadzone w odniesieniu do planów rekultywacji obszaru kopalni odkrywkowej iłów mioceńskich zapadliska przedkarpackiego. Modelowanie numeryczne uwzględniło planowany obszar. Jest on wyznaczony na podstawie naziemnego skanera laserowego. Wykonany został również przestrzenny rozkład i identyfikacja parametrów warstw gruntów. Model numeryczny gruntu, bazujący na teorii Coulomba–Mohra, został również uwzględniony w analizie. Metoda redukcji wytrzymałości gruntu na ścinanie (SRM) została zastosowana w celu określenia stateczności skarp. Obszary zagrożone ruchami masowymi zidentyfikowano na podstawie współczynnika stateczności.

Słowa kluczowe: modelowanie numeryczne, powierzchnia zniszczenia, stateczność zboczy, naziemny skaning laserowy

1. Introduction

Zesławice clay mine is one of the oldest open-pit mines near Krakow see above note (District XVII – Krzesławickie Hills), minerals were exploited here from 1952. A brickyard was also located at the site of the mine, this supplied the inhabitants of Nowa Huta and the Tadeusz Sendzimir Steelworks with building materials. According to the zoning plan, a significant part of the mine area has been assigned for housing development (single family dwellings), while the other part has been assigned for the site of the intersection of the S-7 and S-52 expressways. The intersection of the S7 expressway and the northern bypass of Krakow – Expressway S7 as well as gen. Okulicki Street will undeniably be one of the routes around Krakow that carry the highest volumes of traffic. The investment plan to be completed according to [1], and it will be possible to use this route to travel between Krakow and Warsaw around the year 2021. Afterwards, the northern part of Krakow will be ready for further development. As a result, the intersection of Krakow – Mistrzejowice (former: Nowohucki) will be the next route where traffic congestion will increase. Moreover, it might have a negative influence on the surrounding area, in particular, the condition of subsoil.

Fig. 1. Exposed parts of Miocene clay in an open-pit mine

2. Open-pit clay mine characteristics

Zesławice clay mine has an area of 45 ha and covers a section of the geomorphological unit known as the Carpathian Foredeep, where Miocene clay deposits (Fig. 1) can be found. Clay, which is the base material for the building industry, is no longer extracted there due to the fact that a section of the S-7 expressway is to run through it. Moreover, the intersection of the S-7 and S-52 expressways will also run through it because, as stated in the zoning plan, there is no better location for this intersection – the decision to build it at this location is thus justified. However, the mine site remains home to some active landslides which have brought this place to ruin, and consequently, have brought operations to a halt. It is therefore important to conduct an analysis on how the new investment will affect the mine site and its surroundings, taking into consideration the Miocene clay deposits and the fact that this area is at risk of landslides. The route plan shown in Figure 2.

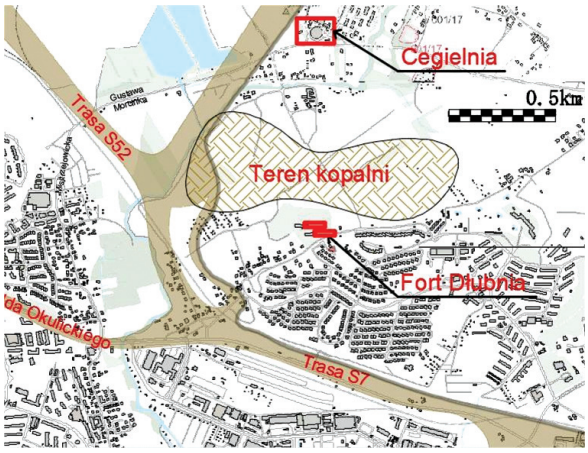


Fig. 2. Route plan of the S-7 and S-52 intersection on the former open-pit mine area in Zesławice

The other reason for the investment to be analysed is the presence of two historical structures which are in the proximity of the investment site. The first is the abovementioned brickyard, which is made of reinforced concrete frames and was built in 1952. The brickyard, abundant in raw materials like clay deposits and built alongside the clay, served as a source of building materials for nearby construction sites, e.g. property development and the Tadeusz Sendzimir Steelworks. The second structure is Dłubnia Fortification 49a, built from brick and stone and dating back to the nineteenth century. The historical significance of this fortification was one of the reasons for the open-pit mine to be shut down. The altered shape of the mine site due to earthworks, the susceptibility of clay to earthflows, the risk of landslides, accessible mine highwalls and floor and considerable sums of money spent to secure the site were among the other reasons for its closure.

It is crucial to analyse every stage of the investment project and its impact on the existing housing estates and relicts of the past. The study included four stages:

- analysis of the sensitivity of the slope to triggering mechanisms and the stability of the slope of the former Zesławice open-pit mine;
- assessment of soil consolidation and soil subsidence due to covering the basin of the work with anthropogenic soil parts;
- impact of housing development on the subsidence of anthropogenic soils on lands which underwent reclamation;
- analysis of the impact of the planned intersection on slope stability within the area of the mine, including analysis of the impact of traffic vibrations on the housing development that is close to Dłubnia Fortification.

3. Calculation assumptions

3.1. Slope stability evaluation

In evaluating the stability of road embankments, hillsides, and mine terraces, the standard methods of ground-stability evaluation based on the theory of limit equilibrium and the assumption of cylindrical slip surfaces (eg. Fellenius, Bishop, Janbu method) are utilised. However, nowadays there are engineering issues in which discrete methods based on the finite element method (FEM) or finite difference method (FDM) are more functional. The slope stability method is based on modified shear strength reduction (SRM). The SRM algorithm consists of reduction strength parameters of the soil (friction angle and cohesion) in each iteration step. The process of iteration is considered complete when the body reaches instability. The ratio of the output parameters resulting from the iterative procedure allows to determine the safety factor defined as:

$$\tau = \frac{\tau_f}{\tau} = \frac{\tan\varphi}{F}\sigma_n + \frac{c}{F} \tag{1}$$

Limit equilibrium methods are successfully used for homogeneous soil conditions for which it is founded flat cylindrical slip surface, which corresponds to the actual behavior of the soil medium. However, in most cases, the arrangement of geotechnical layers is more elaborated. In such cases, the use of numerical methods to evaluate the stability of the analysed medium is reasonable. This causes the shape and the surface not provided with failure in this case is limited to cylindrical and may have various forms of stability loss.

Engineering practice is dominated by an approach based on the designation of the slope stability under consideration in the plane state of deformations. In the case of linear infrastructure, railway and road embankments where the geological layers do not change the length of structure. This approach is justified. However, in particular situations, the designer should adjust the range of the computational model to the prevailing soil conditions. This determines that the designer needs to take into account the different models that are built on the basis of a greater amount of field tests, allowing the nature of the soil to be more accurately determined. Analysis of the plane state of deformations only allows the obtaining of information regarding the local safety factor and shape in range of the designated failure surface. Obtaining more accurate results, including the location and range of the actual slip surface, demands three-dimensional modelling and analysis [10].

3.2. Geotechnical parameters

Since 2010, the soil-structure of Cracow University of Technology has been continuously conducting research on the former clay mine in Zesławice. This work is related to determining the strength parameters of the Miocene clays under various loads. Over the years, a number of field studies and laboratory tests have been performed. Examples of these tests include: static CPT and CPTU tests using the Pagani mobile device; light and heavy probing with

SLVT and DPL, DPH and DPSH devices outdoor and several laboratory works including oedometric tests, and shear box tests, and triaxial stress tests The test procedures are consistent with European standards and guidelines [3, 16].

This paper attempts to assess the stability of slope terraces located in the south-eastern part of the discussed open-pit mine. A spatial arrangement of geotechnical layers is matched with the three-dimensional terrain of the mine. As a result of the 2010-15 field tests, the geotechnical profile of analysed area of the mine has been identified.

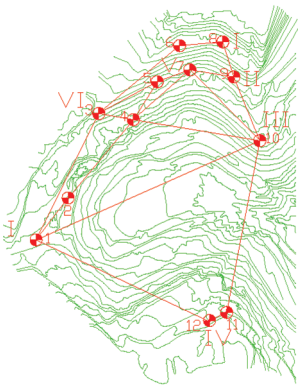


Fig. 3. Geotechnical profiles on the contour line map of the south-eastern section of the mine

The geotechnical prospection is performed in six different planes (Fig. 3). This enables the creation of a 3-D model of the subsoil, assuming that the change of the geotechnical layers occurs linearly along the identified profiles. Field studies have shown that a layer of silts (Si) with a thin layer of silty sands (siSa) are located immediately below the ground level (Fig. 4). The tertiary Miocene clays are located below the layer of sands.

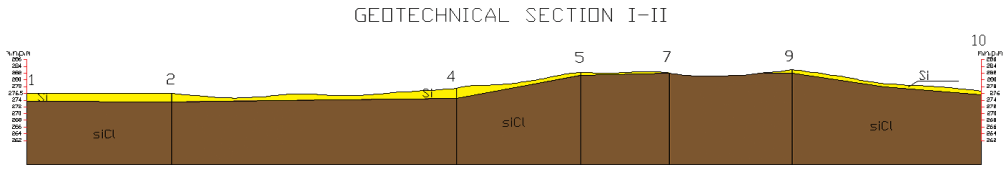


Fig. 4. Geotechnical intersection within I-II

Table 1 presents the parameters of identified there. The values of the internal friction angle ϕ and cohesion *c*, as well as volume weight γ, plasticity index I_L, Young modulus of the deformation *E*, and Poisson's ratio *ν* are obtained from the tests.

Table 1. Characteristic values of the geotechnical parameters for the layers included in the analysis

Soil name	**Φ [o]**	***c* [kPa]**	**γ [kN/m³]**	***E*0 [MPa]**	***ν*** **[-]**	**I_L [-]**
saSi	30.0	25.0	19.56	74	0.3	0.12
Cl	16.0	40.0	21.49	197	0.3	0

3.3. Spatial modelling

Elaborated numerical modelling includes the spatial character of the land surface in addition to the spatial distribution of the geotechnical layers. It is important in the slope stability analysis to accurately reflect the slope of the hill, affecting the general stability of massive ground. In the stability analysis for post-mining excavations with terraces comprising many slopes, accurate surface representation with slopes is obtained using a terrain model derived from terrestrial laser scanning.

The measurements campaign is made with precise the Riegl VZ-400 terrestrial laser scanner in 11 measuring points, enabling a comprehensive survey of the area of the south-eastern part of the open-cast mine. The measuring device is characterised by a scan range of 400 to 500 m, with a precision of 5 mm (Fig. 5). The process of laser scanning [9, 12] is based on the automatic measurement of existing buildings or land through the use of a high frequency sampling rate (i.e. f = 122 kHz). Due to connecting frame and linear scanning in the device from an analogue signal to the digital one tens of thousands of points per second are processed in real time.

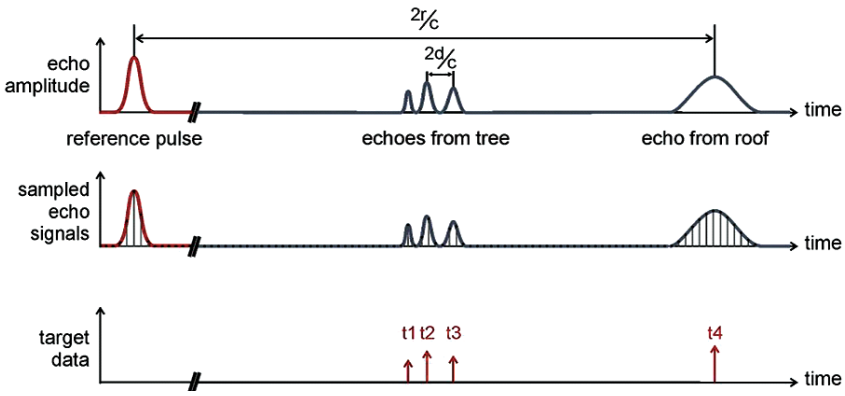


Fig. 5. Collecting data reflected by the measuring device [12]

During the measurements, terrestrial laser scanner generated a point cloud area. As a result of the signal processing by removing unnecessary objects in the measured area (i.e. plants, trees, buildings, etc.), a spatial hypsometric map is generated (Fig. 6) and is then approximated with a cloud of points in a triangular finite elements mesh.

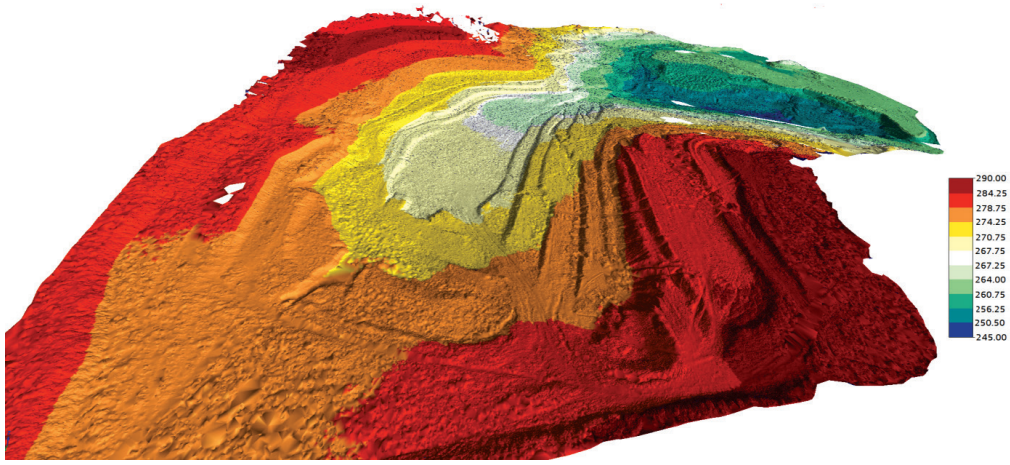


Fig. 6. Hypsometric map of the area of the spatial model

3.4. Numerical model

The modelling procedure of the analysed open-pit mine used information from literature [2, 14]. The analysed area is a three-dimensional medium and takes into account the spatial stress and strain distribution. In the modelling of the soil, the Mohr–Coulomb hypothesis was applied – this is widely used in geotechnical engineering; it provides sufficiently reliable results of the overall analysis of nonlinear soil behaviour (Fig. 7).

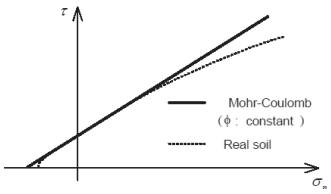


Fig. 7. Comparison of the Mohr–Coulomb hypothesis and the real behaviour of the soil [4]

The Mohr–Coulomb criterion can be written in the form of principal stress as:

$$\frac{\sigma_1 - \sigma_3}{2} = -\frac{\sigma_1 + \sigma_3}{2}\sin\varphi + c\cos\phi \tag{2}$$

For the purposes of the numerical model, Mohr–Coulomb presented in the form of the equation of invariants of the stress tensor I_1, J_2 and Lode angle θ [4].

$$\begin{aligned} f(I_1, J_2, \Theta) &= -\frac{1}{3} I_1 \sin\varphi + \sqrt{J_2}\left(\cos\Theta + \frac{1}{\sqrt{3}}\sin\Theta\sin\varphi\right) - c\cos\varphi = 0 \\ g(I_1, J_2, \Theta) &= -\frac{1}{3} I_1 \sin\psi + \sqrt{J_2}\left(\cos\Theta + \frac{1}{\sqrt{3}}\sin\Theta\sin\psi\right) - c\cos\psi = 0 \end{aligned} \tag{3}$$

The Mohr–Coulomb criterion is an irregular hexagonal pyramid with a line indicating the point of stress stress as shown in Figure 8. The shape of the deviator plane π $(\sigma_1 + \sigma_2 + \sigma_3 = 0)$ takes the form of an irregular hexahedron.

The analysis assumed a perfectly elastic-plastic model of the Mohr–Coulomb linear plasticity condition.

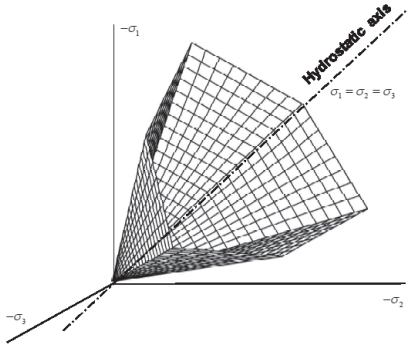


Fig. 8. Mohr–Coulomba criterion in the deviatoric space along the hydrostatic axis [4]

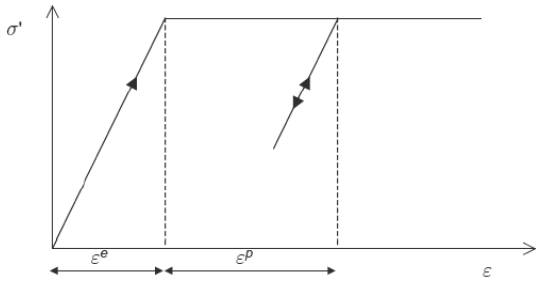


Fig. 9. Graph of the stress from the deformation of elastic-plastic model M-C [4]

A limitation of the Mohr–Coulomb model is the linear nature of the destruction. The friction angle does not change with pressure-limiting (hydrostatic pressure). In addition,

the linear nature of the destruction of the soil causes beyond the shear strength of formed permanent plastic deformation and the soil does not transfer stresses (Fig. 9). Figure 10 presents the hexahedral finite element used for the discretisation of digitised terrain.

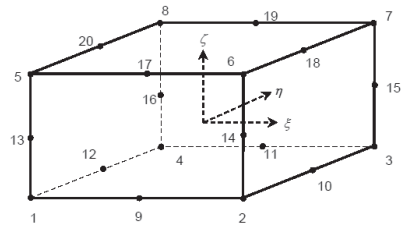


Fig. 10. Hexahedral element used for the discretisation model

4. Numerical analysis model

Numerical analysis is used to determine the spatial distribution of the stability factor of the analysed area and determine the location of the largest shear strains in the soil. The model is discretised with 297,000 finite elements of a higher-order hexahedral type. Boundary conditions are applied in the form of a sliding pivot about a vertical axis, preventing movement in the horizontal direction in two perpendicular directions along the x and y axes. The lower edge of the model is blocked in all three directions through the use of non-slip joints. Soil parameters used in computations are taken from Table 1. Figure 11 presents the digitised 3-D model. The computations are performed using a modified shear strength reduction method (SRM) for the dead weight loading.

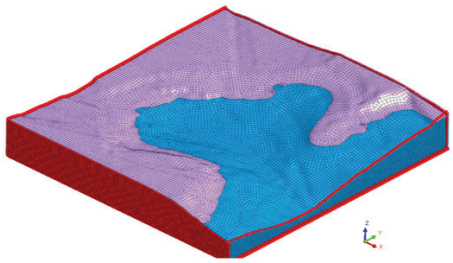


Fig. 11. Digitised three-dimensional model of the analysed area

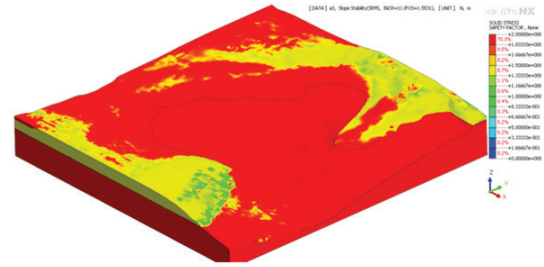


Fig. 12. Distribution of the spatial slope stability factor

Figure 12 shows the spatial distribution of the stability factor of the analysed area. The stability of the analysed area is characterised by a factor in the range 1.10 to 2.00. It mostly consists of areas with a stability factor equal to 2.00. The smallest values are in the range 1.1-1.3 (green and yellow) and occur in the western and eastern parts of the valley. The slopes and terraces of mine are quite stable. The first form of loss of stability may occur in the upper layers of the terraces and will be related to the loss of stability of the Quaternary layers of silts.

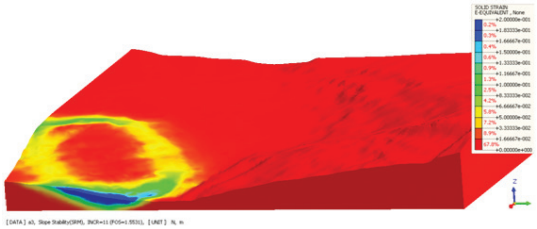


Fig. 13. Exhibits the form of stability loss of the western slope of the terrace of the mine. The soil slip surface covers the entire height of the cliff of the mine

5. Final remarks

The paper focuses on the slope stability analysis of an open-pit mine in the context of the gradual development of prospective S-7 and S-52 intersection. The issue raised in this article is based on the first stage of mine reclamation. The area definitely needs to be restored to an acceptable standard of use since the new investment plan for the site includes the development of housing and infrastructure. A modified shear strength reduction method was used in order to determine the spatial slope stability of the area of a former open-pit mine in Zesławice. The analysis results have been used to determine the actual slope stability of the high walls of the mine. Numerical analysis of the area revealed that slopes are formed of silt that is prone to mass movement. In the case of tertiary clay deposits, which are stable if the shear strength parameters decrease, the slope stability factor may significantly drop. The on-site investigation revealed minor landslide activity of slopes in the southern part of the mine due to water activity in the silt layer.

References

[1] Decyzja o środowiskowych uwarunkowaniach dla odcinka S7 o przebiegu Moczydło – Szczepanowice – Widoma – Zastów – Kraków (Ptaszyckiego/Igołomska) została wydana przez Regionalnego Dyrektora Ochrony Środowiska w Krakowie 15 stycznia 2015 r.

[2] Griffiths D.V., Lane P.A., *Slope stability analysis by finite elements*, Geotechnique, Vol. 49 (3), 1999, 387-403.

[3] Wysokiński L., *Rekultywacja terenów kopalni odkrywkowych i ich wykorzystanie budowlane*, Przegląd Geologiczny, t. 58, nr 9/2, 2010.

[4] MIDAS GTS NX, *Manual specifications*, 2016.

[5] Ministerstwo Budownictwa i Przemysłu, *Dokumentacja geologiczna złoża surowców ceramiki budowlanej „Zesławice" w kat. B+C1+C2*, Zakład Projektów i Dokumentacji Geologicznych w Katowicach – oddział w Krakowie, Kraków 1979.

[6] Obwieszczenie Marszałka Sejmu Rzeczypospolitej Polskiej z dnia 2 kwietnia 2004 r. (Dz.U. 2004, Nr 121, poz. 1266 z późn. zm.).

[7] Olesiak S., *Wykorzystanie sondy wkręcanej WST w badaniach mioceńskich iłów krakowieckich*, Górnictwo i Geoinżynieria, z. 1, 2009, 467-473.
[8] Pilecka E., Bazarnik M., *Application of terrestrial laser scanner for monitoring the railway infrastructure threatened by landslides*, Wydawnictwo Politechniki Krakowskiej, Kraków 2015, 171-190.
[9] Pilecka E., Szwarkowski D., *Wykorzystanie skanera laserowego RIEGL VZ-400 do badania deformacji na terenach górniczych GZW*, Budownictwo Górnicze i Tunelowe, R. 21, Nr 4, Katowice 2015.
[10] Pilecka E., Szwarkowski D., Manterys T., *Wpływ drgań drogowych na propagację deformacji wzdłuż czynnej powierzchni poślizg i stateczność nasypów na terenach osuwiskowych*, Autobusy: Bezpieczeństwo i Ekologia, nr 6/2016, Radom 2016.
[11] Polska Norma PN-EN-1997-2 Eurokod 7. Projektowanie geotechniczne. Część 2: Rozpoznanie i badanie podłoża gruntowego.
[12] RIEGL LMS, *Oprogramowanie Systemowe i Przetwarzania Danych RiSCAN PRO dla skanerów laserowych 3D firmy RIEGL LMS*, Austria 2009.
[13] Takahashi M., Hight D.W., Vaughan P.R., *Effective stress changes observed during undrained cyclic triaxial tests on clay*, Proceedings of the International Symposium on Soils under Cyclic and Transient Loading, Vol. 1, Swansea 1980, 201-209.
[14] Wiłun Z., *Zarys Geotechniki*, Wydawnictwa Komunikacji i Łączności, Warszawa 2010.
[15] Wrana B., *Lectures on soil mechanics*, Wydawnictwo Politechniki Krakowskiej, Kraków 2014.
[16] *Zasady oceny przydatności gruntów spoistych Polski do budowy mineralnych barier izolacyjnych*, Instytut Techniki Budowlanej, Warszawa 2007.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
MECHANICS
DOI: 10.4467/2353737XCT.17.203.7432

Bartosz Wieczorek (bartosz.wieczorek@put.poznan.pl)
Mateusz Kukla
Dominik Wojtkowiak
Chair of Basics of Machine Design, Faculty of Machines and Transport, Poznan University of Technology

Piotr Okoniewicz
Karolina Ostrowska
Institute of Machines and Motor Vehicles, Poznan University of Technology

Method of programming the nitinol springs in the space of the kiln chamber

Metoda programowania sprężyn z nitinolu w przestrzeni komory pieca

Abstract

This paper shows a method of programming the NiTinol springs by heating them up in the furnace chamber filled with technical nitrogen. A fully specified methodology, which consists of preparing the spring forms, the numerical analysis of the heating time and the description of the heating process are presented in this article. The effectiveness of the following method is confirmed using research of shape retention after a series of duty cycles performed by an activated NiTinol spring.

Keywords: NiTinol, Shape Memory Alloys (SMA), NiTinol programming, helical springs

Streszczenie

W artykule opisano sposób programowania sprężyn z NiTinolu z wykorzystaniem pieca z atmosferą azotu technicznego. W ramach artykułu opisano metodykę procesu. W jej skład wchodzi metoda przygotowania form ze sprężynami, obliczenia numeryczne początkowych nastaw pieca oraz opis procedury wygrzewania w atmosferze azotu. Praca została podsumowana badaniami odwzorowania kształtu sprężyny po procesie programowania.

Słowa kluczowe: NiTinol, materiały z pamięcią kształtu, programowanie NiTinolu, sprężyny

1. Introduction

Shape Memory Alloys (SMA) are a group of materials which can be classified as smart materials [1-3]. These materials have the ability to recreate a programmed shape as a result of certain external factors e.g. temperature, magnetic field or a previously applied stress. One of the most commonly used shape memory alloys, along with alloys based on copper and iron, is NiTinol. Originally it was invented by Bühler in 1962 [3] and patented along with Wiley, because of the characteristic mechanical memory, in 1965 [2]. NiTinol is an alloy of nickel and titanium (NiTiX). Depending on the source, the amount of nickel in an alloy is in a range of 53-57% wt. [1-3], or 53-54% wt. [2].

NiTinol has two stable phases: austenite with the cubic structure of the crystal (high-temperature phase) and martensite with the prismatic structure of the crystal (low-temperature phase). The martensite structure can be deformed without braking its atomic bond – this phenomenon is called twinning [1].

The effect of the unidirectional shape memory in this material is based on the single time preservation of the programmed geometry. The essence of this process is to anneal a formed element, made of NiTinol, which is currently in a high-temperature phase. The annealing is made in temperature 500°C within 30 minutes' time [1] and it is commonly called „programming" of the SMA. The parameters which can be used in the annealing process are also described in the following papers [6-8].

The heating time depends on the geometry of the programmed elements. It can be between 1 and 20 minutes. The exact value of the heating time can be determined thanks to the experimental studies based on the chemical composition of NiTinol alloy [5-8].

Overheating of the NiTinol alloy to temperatures higher than 600°C is not recommended due the loss of chemical stability. It also has a negative effect on the oxidation resistance, because of damaging the thin layer of oxides which covers an alloy in normal conditions [4]. Upon heating the NiTinol wire from temperature A_s (activation of the transformation) to A_f (end of a transformation), a transformation from martensite to austenite can be observed. Further increase of the temperature does not change the phase of an alloy, but causes the ordering of the crystal lattice and lowers the value of the internal stress. After the annealing process finishes, followed by a cooling process, the austenite phase stable in higher temperature with a crystal lattice of higher symmetry transforms into the martensite phase stable in lower temperatures, which is characterised with lower symmetry. In normal air temperature, a crystal lattice of the martensite phase is deformed due to twinning [1].

2. Methodology of the programming process

In order to obtain a desired shape of the NiTinol alloy, it is necessary to increase temperature over the austenite transformation temperature during the programming process. The next step is to anneal an alloy in approximately 500°C in a specified time.

In this research program, two stages of forming and programming the shape details of helical springs were considered. Because of the straight shape of the NiTinol wire provided by the distributor and used in this research, it was necessary to design a novel method of forming and securing the coiled springs. Otherwise they would lose their geometry during the heat treatment due to internal stress.

During the methodology studies, numerical analyses were made in order to determine the values of programming time and temperature. These calculations provided necessary data for selecting the proper heat treatment ramp and adjusting it to the performed programming process.

2.1. Shaping the helical spring

One of the most important stages of the programming methodology is shaping the NiTinol helical spring. In order to complete this task, the dedicated apparatus was designed and manufactured. The designed device allows to shape a helical spring as well as to secure its geometry during heat treatment. Its construction and the methodology used were later described in the patent application [4]. The schematic of a device for manually shaping NiTinol helical springs is presented in the Figure 1.

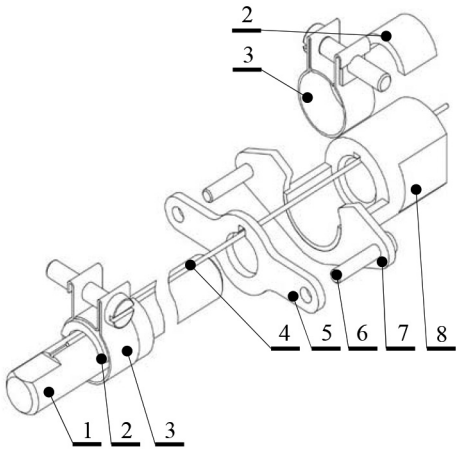


Fig. 1. Components of the device for manually shaping NiTinol helical springs [4]: 1 – core, 2 – clamping pad, 3 – U-bracket, 4 – NiTinol wire, 5 – spinning washer, 6 – blocking pins, 7 – pressure plate, 8 – base

The procedure of shaping NiTinol helical springs for the programming process in a hot air stream using the previously described device consists of several elementary steps. Firstly, it is necessary to mount the bearing set in the vice using flat surfaces in the base of the device 8. Before the threading of the NiTinol wire 4, U-bracket 3, which is open, must be embedded on the bearing set in such a way that its screw is placed in a relief in the base. Next, the wire is threaded through a bearing set and placed in the groove in the base. The spinning washer 5 is now mounted on the wire and then the remaining part of the wire is embedded in the

groove on the core 1. It is important to place it on the shorter shaft neck of the core, which is ended with the collar. To prevent movement of the wire, it is compressed using the clamping pad 2 and the U-bracket. After that, the spinning washer is fitted with pins 6, which block its rotational degree of freedom, and the wire lays in the grooves of axially mounted components. The core is translated through the centre holes in the spinning washer and the base until its collar has contact with the spinning washer, which is pressed against the pressure plate 7. The last stage is to put the clamping pads between the wire and the U-bracket, which is tightened up by a screw.

After the preparation stage, the coiling can start. First, the core is clutched with the proper tool by using flat surfaces on the core. Then the core is rotated along with pressing it with an axial load in the direction towards the bearing set. Each rotation creates one coil of a helical spring and causes the distance between the collar and the spinning washer to grow. When the spring is ready, it is necessary to clamp its end with the U-bracket, pull out the core from the bearing set and cut off the remaining wire.

The effect of the above-mentioned procedure is the NiTinol helical spring, which is properly shaped, protected and ready for the programming process by putting it in a furnace chamber. An example of the prepared specimen is presented in Figure 2. The NiTinol alloy used in the presented research had 55/45%wt Ni/Ti.

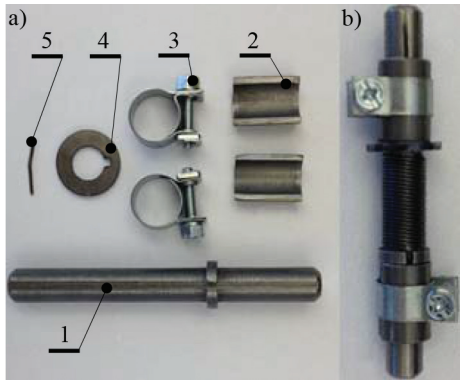


Fig. 2. Specimen ready for programming: a) components; b) form with coiled spring; 1 – core, 2 – clamping pad, 3 – U-bracket, 4 – spinning washer, 5 – NiTinol wire

2.2. Simulation analysis of the heating process

A convenient way to determine the heating time in which a specimen will achieve the specified temperature is to perform the simulation in Abaqus – educational version. The obtained results are not precise due to the limit of nods, but are adequate to design the programming process. An example of the results is presented in Figures 3 and 4 (red – 500°C, yellow – 350°C).

The characteristics of temperature of the NiTinol spring in the function of the heating time is presented in Figure 5. Based on the performed simulations it can be concluded that the heating time for programming the NiTinol spring at 500°C equals approximately 16 minutes.

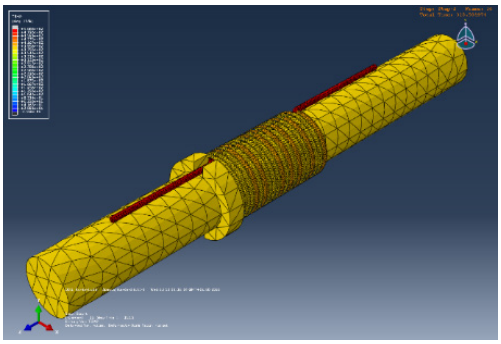


Fig. 3. Temperature distribution after 5 minutes of heating up in the furnace chamber at 500°C

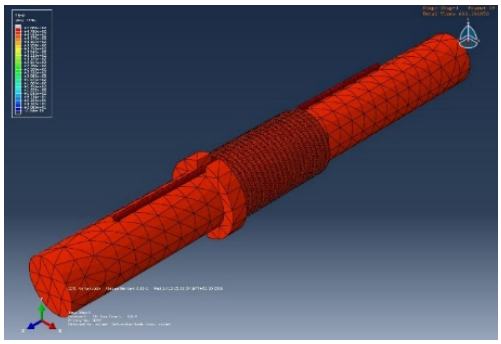


Fig. 4. Temperature distribution after 10 minutes of heating up in the furnace chamber at 500°C

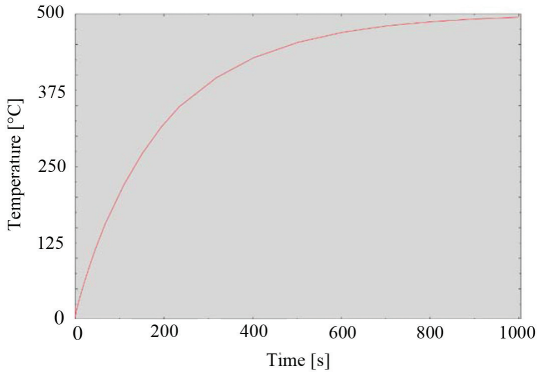


Fig. 5. Temperature characteristic of the NiTinol spring in the function of the heating time

2.2. Industrial furnace characteristics

The NiTinol alloy was heat-treated using the industrial furnace of VTR type shown in Figure 6. Multiple operations can be made in this furnace chamber e.g. tempering, annealing, heating and nitriding.

The work chamber of the furnace is made of heat resistant steel and has a cylindrical shape with vertical loading of the furnace feed. Circulation of gases inside the work chamber is forced by a fan, a baffle and a guide wheel. That is why a gas flows through the furnace similarly to the flow in the reversing chamber [1].

The industrial furnace is equipped with a resistance heating device. It's maximal operating temperature equals 650°C. It allows to realise thermal processes with a pressure ranging from 28 to 32 mbar. The allowable mass of components in the furnace cannot be higher than 200 kg and their dimensions cannot exceed 400 mm x 400 mm x 600 mm. The following gases can be used in the furnace chamber: nitrogen (neutral gas used for rinsing a chamber before and after nitriding), ammonia (used for nitriding) and air (activator of surfaces of treated elements). The measuring system makes it possible to control temperature of the furnace feed by a set of thermoelements (9 thrown-in thermocouples and 1 retort thermocouple). It is also possible

to analyse the chemical composition of the chamber atmosphere or check the pressure inside the furnace chamber during the heating treatment. The heating system makes it possible to define the heating ramp, which is used during annealing NiTinol springs.

Fig. 6. VTR type industrial furnace, Institute of Machines and Motor Vehicles, Poznan University of Technology

3. Characteristics of the programming process

In order to program NiTinol helical springs used in this research, a proper procedure of for the heating process was required. The performed process indicates the necessity to specify the heating velocity, temperature of the process and the annealing time for each type of specimens. By using the neutral protective atmosphere of nitrogen, it was possible to safely slide in and out a single specimen into the furnace work chamber in the exact determined heating time. For all samples, the heating velocity of the kiln was constant and equalled 8°C per minute and the procedure was similar. Only two variables were present in the process: annealing temperature and time. The comparison between each process performed during this research is presented in Table 1. The lowest temperature of programming was established at 400°C, because only in higher temperatures does NiTinol lose its initial straight shape and gains a new programmed shape of the helical spring coiled on the core.

Programming of NiTinol springs was divided into several stages. First, the work chamber of the furnace was emptied in order to get rid of the air inside. After the vacuum inside the chamber was generated, it was filled with nitrogen and the furnace was heated up to a desired temperature of 400°C, 500°C or 600°C. Total time needed to anneal a specimen is the sum of the following times: the period in which the kiln chamber attains the desired temperature, the time needed to heat up a specimen to the same value and the time of annealing a specimen in the desired temperature. The second value (16 minutes) was calculated based on the simulations in Abaqus, while the third time was chosen for analysis (2.5, 5 or 7.5 minutes). After finishing the annealing process, the final stage of cooling down to 70°C started. In order to achieve that, nitrogen was removed from the furnace chamber until reaching the pressure of 0.5-1 mbar. The arrangement of specimens inside the furnace is presented in the Figure 7.

Table 1. Parameters for programming (heat treatment) in an industrial VTR furnace

400°C		Temperature of the process		
		500°C	600°C	
Time needed to achieve desired temp. in a kiln chamber		2 h 01 min	1 h 55 min	1 h 53 min
Time needed to head a specimen to a desired temp. [min]*		16		
Annealing time [min]	Specimen no. 1 on a rod	5		
	Specimen no. 2 on a rod	5		
	Specimen on a feed basket**	33	31	30
* – time counted from the moment of achieving the desired temperature in the kiln chamber ** – time counted from the moment of achieving temp. lower by 5°C than the desired temp.				

Each process contained three NiTinol springs with two of them mounted on the rods in the sample uptake port (Figs. 7a, c) and one lying on the furnace feed basket (Fig. 7b). The sample uptake port was used in the programming process as a two-way canal lock. That is why it was possible to move specimens in and out of the furnace chamber before and after finishing the heating treatment.

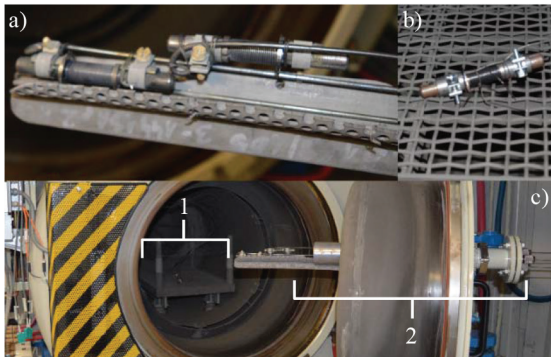


Fig. 7. Specimen distribution in the furnace during programming: a) specimens placed on a half cylindrical bonnet for a small feed; b) specimen placed on a feed basket; c) overall distribution; 1 – feed basket, 2 – canal lock in furnace

The task for the profile formed open bonnet (Fig. 7a) was to keep a proper path of motion for the specimen during putting it inside the kiln chamber. This bonnet prevented the sample from slipping outside the feed area during sliding it towards the outer rim of the kiln door. Thanks to that, it was possible to precisely slide the previously prepared NiTinol springs on the forms inside the canal lock in the furnace door. The samples were mounted on bended

rods using extra burnt non-galvanised steel wire, which provided the specimen with position stability during the high-temperature heat treatment.

The described methodology makes it possible to put specimens inside the VTR furnace chamber, heat it to a desired temperature, anneal a sample for a specified amount of time and then pull out two of programmed springs outside the furnace chamber, all during a continuous heat treatment.

In order to control the heating velocity of the Ni-Ti wire, a thermoelement was mounted on its surface as shown in Figure 8. The retort thermocouple measurement was responsible for the steering process of the furnace chamber temperature, as per its calculation signal. Its value depends on the heating ramp, defined before the start as 8°C/min. Only the temperature directly on a coil of the spring from the feed basket was monitored. These time-based charts are presented in Figure 9.

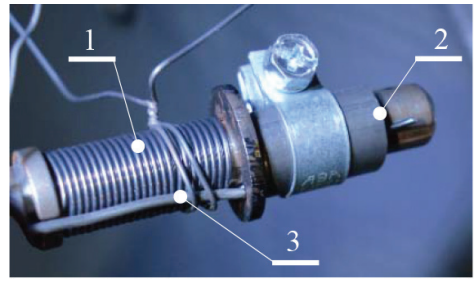


Fig. 8. Sample from the feed basket with a mounted thermocouple for temperature measurement: 1 – NiTinol spring, 2 – programming form, 3 – thermocouple

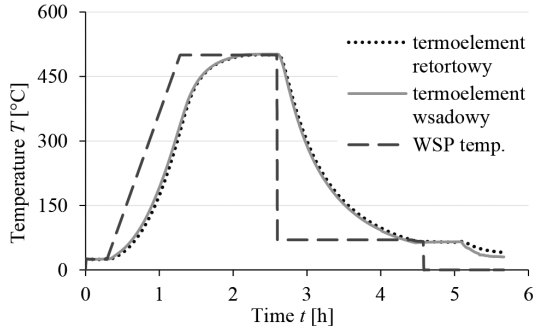


Fig. 9. Chart showing the progress of the NiTinol helical spring programming. Process parameters: 500°C/5min – specimens on rods in the uptake sample port and 500°C/31 min – the specimen rested on the feed basket

The specimen, which laid on the feed basket, was heated from 26°C to 495°C for 2 hours and 1 minute. The last 5°C before reaching 500°C was gained in additional 10 minutes' period. Then, the annealing process was performed for 16 + 5 minutes – according to the research plan. In the final stage, the sample was cooled down along with the whole furnace chamber to 70°C. It lasted 1 hour and 49 minutes.

In the starting stage of cooling down, the coating of the furnace work chamber was refrigerated with forced air circulation. Then, in the later phase, when the door was opened, the temperature was reduced by convection. It is clearly visible in the chart in Figure 6 for the period of 5 hours and 6 minutes. After removing the specimens from the furnace, they were continuously chilled by convection with air.

4. Method verification

The main goal of this research was to determine the influence of heat treatment parameters on the precision of the geometry of NiTinol helical springs. The quality of the programmed springs was checked right after the programming process. The specimens were tested by loading them 36 times and measuring their length after activation with the electric current. In order to evaluate the presented methodology of programming, the results were compared with the spring programmed in a non-professional way. The reference specimen was programmed with the propane-butane gas torch by heating up the core with the spring in such a way that the inner cone was moved along the outer surfaces of the spring with a uniform distribution of the heat source. The whole process lasted 5 minutes and the temperature of the Ni-Ti wire was established at about 480°C.

For the comparison of the reference sample with the spring programmed in the furnace chamber, similar parameters were chosen: the programming temperature 500°C and the annealing time 5 minutes. This spring was mounted on the feed rod in the sample uptake port. As soon as the temperature in the chamber reached a desired value, the sample was slid into the chamber, where it was heated up for 16 minutes and then annealed for 5 more minutes. Then, the spring on the core was hidden back into the sample uptake port, where the cooling process took place at a temperature of approximately 100°C. Both samples are presented in Figure 10 (the reference sample) and Figure 11 (the tested sample).

Fig. 10. The spring programmed with the propane-butane gas torch: 5 min annealing time, ~480°C temperature, 21 m length of the spring after programming

Fig. 11. The spring programmed in the furnace chamber in the atmosphere of technical nitrogen: 5 mins annealing time, 500°C temperature, 22 mm length of the spring after programming

Taking into account the geometry of the NiTinol springs removed from the cores, there is no difference between the reference and the test sample. Both ways of heating: in the furnace chamber and with the gas torch, reprogrammed the straight NiTinol wire into the helical spring. This means that, in both cases, recrystallisation was performed successfully. An identical shape of the springs must be the effect of properly prepared forms.

A more important question was: how will the geometry of both springs change after a few duty cycles? In order to check that, it was necessary to deform the spring and then activate it, so that the spring should go back to the previous state due to the unidirectional shape memory effect. The test stand used in this research was designed and manufactured by the authors and shown in Figure 12. To achieve the strain of the spring, one end was fixed while the other end was loaded with a specified force. The loading force was obtained using a set of the twine and the hook weight. After the spring was extended to a proper value, it was activated by resistive heating with the electric current flowing through the coils of the spring. The electric power delivered to the sample equals about 22 W, which increases its temperature to 60-70°C. This temperature increase caused the austenite transformation in the wire, which led to the regularisation of the crystal lattice and the spring was compressed to its nominal length. During the experiment, the extension caused by the temperature and the force was measured. The view from the thermovision camera shown in Figure 13 illustrates the shortening of the spring due to the temperature increase.

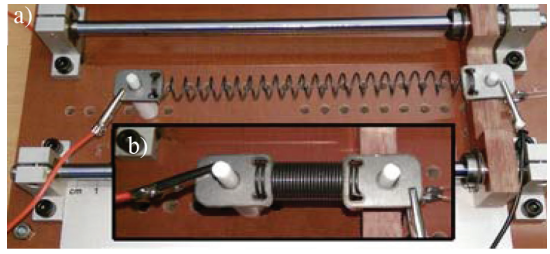


Fig. 12. Test stand used to check the shape retention after duty cycles: a) before activation; b) after activation

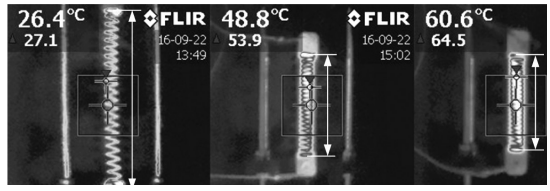


Fig. 13. Thermovision camera pictures during activation of the specimen loaded with 300 g and activated with 22 W power

Based on the series of performed experiments, elastic characteristics were drawn for two different methods of programming. They are presented in Fig. 14. The error for the displacement value was defined based on the standard deviation of the mean value of the estimator. When comparing the elasticity of both samples, it is clear that by using a furnace with nitrogen atmosphere inside the chamber to program the NiTinol wire into the helical spring, the product

will have higher elastic constant. Both characteristics are linear, like in conventional helical springs, and this is confirmed by the value of R^2 factor close to 1.

During the tests of the gas torch programmed spring, it was noticeable that its structure was destroyed due to the uncontrolled overheating while programming. It caused constant permanent deformations observed in the upper length of the spring after activation with the electric current. This means that the sample did not go back to its nominal length after taking the load off and final activation. The final form of the reference spring after 36 cycles of duty is presented in Figure 15a.

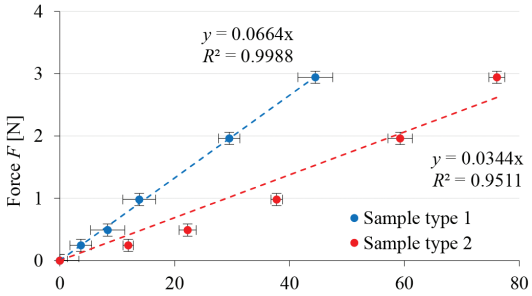


Fig. 14. Characteristics of the force in the function of displacement for the NiTinol spring after activation for different loads

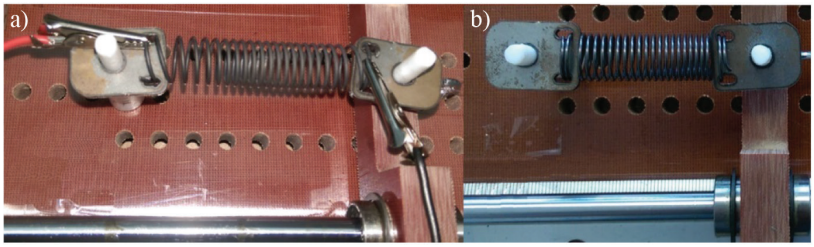


Fig. 15. Specimens after 36 cycles of loading and activation: a) spring programmed with the propane-butane gas torch, length before activation: 22 mm, length after 36 duty cycles and activation: 57 mm; b) spring programmed in the furnace chamber at 500°C for 5 mins, length before activation: 22 mm, length after 36 duty cycles and activation: 31 mm

The test sample failure occurred probably due to the simplified programming methodology. Programming with the gas torch is not a proper method. It results in an inhomogeneous distribution of heat in the various coils of the spring. Because of that, the crystal lattice of the alloy is unevenly reorganised. This can have a significant impact on the quality of the repeatability of the programmed shape. During the experiment, it was also noted that the electrical resistance of the test sample changes. The process progressed with the number of deformation cycles, which also can be explained with the inner structure failure.

The second test sample, which was annealed in the furnace chamber with protective nitrogen atmosphere had much better properties for shape preservation. The designed methodology provided the possibility for precisely steering the temperature and the programming time. Thus,

all the coils in the spring were properly annealed in the desired temperature thanks to preheating the specimen before the proper annealing. As the experiment shows, after 36 cycles of loading and activation, this specimen still maintains its initial shape after taking the load off (Fig. 15b).

5. Conclusions

The Industrial VTR furnace with the horizontal steel retort makes it possible to program the SMA Ni-Ti alloy in the protective atmosphere of technical nitrogen. By using the sample uptake port adapted to transport the springs coiled on the core with a specified geometry, it is possible to test a different annealing time in a single duty cycle of the furnace. The additional advantage of the designed methodology is the repeatability of the conditions in which the spring is programmed. It causes high stability and repeatability of the results obtained in post programming experiments. By using an industrial computer an ability to monitor and control the parameters of the performed heat treatment can be gained.

The reference sample programmed with the propane-butane gas torch revealed that this method is much less effective than the proposed methodology. Without the control of the programming parameters, it is impossible to obtain a unilateral temperature distribution in cross-sections of all coils of the NiTinol spring.

During the experiments on the springs programmed in the furnace chamber, the unidirectional shape memory effect was observable. This phenomenon can be easily inverted and, as the test shows, it remains stable even for a series of duty cycles. This can be the determinant for the validity of the designed programming methodology with the VTR furnace.

In future research, focus should be directed onto a further analysis of the influence of time and form of the heat treatment on the properties of the final product – the helical NiTinol spring.

References

[1] Dokumentacja Techniczno-Ruchowa przemysłowego pieca typu VTR.

[2] Kucharczyk W., Mazurkiewicz A., Żurowski W., *Nowoczesne materiały konstrukcyjne. Wybrane zagadnienia*, Wyd. 3, Politechnika Radomska, 2011.

[3] Makarewicz G., *Materiały inteligentne – zastosowanie w systemach aktywnej redukcji hałasu i drgań*, „Bezpieczeństwo pracy" nr 12/2015.

[4] Ćwikła A., *Medyczne zastosowania materiałów inteligentnych*, Scientific Bulletin of Chelm Section of Technical Sciences No. 1/2008.

[5] http://encyklopedia.naukowy.pl/Nitinol,vstrona_3.

[6] http://www.imagesco.com/articles/nitinol/03.html.

[7] Vojtech D., *Influence of heat treatment of shape memory NiTi alloy on its mechanical properties*, Metal 2010, 18. 18-20.5.2010, Roznov pod Radhostem, Czech Republic.

[8] http://smartwires.eu/index.php?id_cms=9&controller=cms&id_lang=1.

[9] Kłaput J., *Studies of selected mechanical properties of nitinol – shape memory alloy*, Archives od Foundry Engineering, vol. 10, Issue 3/2010.

TECHNICAL TRANSACTIONS 11/2017
CZASOPISMO TECHNICZNE 11/2017
PHYSICS
DOI: 10.4467/2353737XCT.17.204.7433

Arkadiusz Papaj (arkadiusz.papaj@gmail.com), Piotr Weszka (piotrweszka@gmail.com)
Marcin Bocheński (bochenski.marcin@outlook.com), Mateusz Michałek (mateusz.michalek@pk.edu.pl)
Paweł Karbowniczek (pkarbowniczek@pk.edu.pl), Radosław A. Kycia (kycia.radoslaw@gmail.com)
Cracow University of Technology

Andrzej Kułak (kulak@oa.uj.edu.pl)
Electronics and Telecommunications, Institute of Electronics, Faculty of Computer Science, AGH University of Science and Technology

Agata Kołodziejczyk (maria.kolodziejczyk@esa.int)
European Space Agency, Advanced Concepts Team, Noordwijk, The Netherlands

Matt Harasymczuk (matt@harasymczuk.pl)
Warsaw University, European Space Agency, Advanced Concepts Team, Noordwijk, The Netherlands (joanna.kuzmaa@gmail.com)

Aleksandra Ławrynowicz (olalawrynowicz@gmail.com)
Pomorski Park Naukowo-Technologiczny Gdynia

Joanna Kuźma (joanna.kuzmaa@gmail.com)
Faculty of Chemistry, Wrocław University of Science and Technology

Tomasz Brol (t.brol@ssh.gliwice.pl)

The design and the performance of stratospheric mission in the search for the Schumann resonances

Projekt i realizacja misji stratosferycznej w poszukiwaniu rezonansów Schumanna

Abstract

The technical details of a balloon stratospheric mission that is aimed at measuring the Schumann resonances are described. The gondola is designed specifically for the measuring of faint effects of ELF (Extremely Low Frequency electromagnetic waves) phenomena. The prototype met the design requirements. The ELF measuring system worked properly for entire mission; however, the level of signal amplification that was chosen taking into account ground-level measurements was too high. Movement of the gondola in the Earth magnetic field induced the signal in the antenna that saturated the measuring system. This effect will be taken into account in the planning of future missions. A large telemetry dataset was gathered during the experiment and is currently under processing. The payload consists also of biological material as well as electronic equipment that was tested under extreme conditions.

Keywords: The Schumann resonances, ELF (Extremely Low Frequencies), balloon mission

Streszczenie

W artykule opisano szczegóły techniczne misji balonowej, mającej na celu pomiar rezonansów Schumanna. Gondola została zaprojektowana do pomiaru słabych efektów ELF (fale elektromagnetyczne o skrajnie niskiej częstotliwości). Prototyp spełnił założenia projektowe. Układ do pomiaru ELF działał poprawnie przez całą misję, jednak poziom wzmocnienia sygnału, który był dobrany zgodnie z pomiarami naziemnymi, okazał się za duży. Ruch gondoli w ziemskim polu magnetycznym indukował w antenie sygnał, który wprowadzał układ pomiarowy w stan nasycenia. Ten efekt zostanie wzięty pod uwagę w planowaniu przyszłych misji. Podczas eksperymentu zebrano duży zestaw danych telemetrycznych, który jest obecnie analizowany. Ładunek zwierał również materiał biologiczny oraz sprzęt elektroniczny, który był testowany w warunkach ekstremalnych.

Słowa kluczowe: rezonanse Schumanna, ELF, misje balonowe

1. Introduction

ELF are the waves that are usually connected with natural atmospheric phenomena. They are defined to have frequency from 3 Hz to 3 kHz. Their observation can be affected by industrial activity that disturb the measurements – the most common is 50 Hz radiation of electric network and radio transmissions. Therefore, the measuring device has to be designed in the way that it can discard these strong disturbances.

One of the most important ELF phenomena are the Schumann resonances, which were predicted and measured in the 1950s (see [2]) and the references therein. They occur as the Earth-ionosphere waveguide is constantly powered by the electromagnetic waves from lightning. They interfere giving a characteristic spectrum of amplifications at frequencies 7.83 Hz and higher harmonics 14.3 Hz, 20.8 Hz etc. [1]. The frequency location of the resonances is connected with the parameters of the atmosphere; therefore, they can be used for measuring of their properties on Earth and other planets, e.g., in future missions on Mars [3, 4]. There are also some indications that the low frequency electromagnetic fields influence biological systems; however, there is no general description of this phenomenon – see review in [5].

Stratospheric balloon missions are the most versatile, in the sense of cost-results optimisation, way of performing measurements in the environment closely connected to the cosmic space or in the higher layers of the atmosphere. As the ascending and descending phases are slower than in rocket carriers the gathered datasets can be large. It is therefore natural to use this platform to measure the resonances, as was suggested in [3] in the case of Mars. The first initial results of measuring the resonance were reported in [6], therefore, it suggests that the proper design of the balloon ELF mission is of great importance in atmospheric research and future cosmic exploration.

The paper is organised as follows. In the next section the overview of the electronic antenna system and then the design of the gondola will be presented. Next, the brief description of the mission will be outlined and conclusions from the first iteration will be presented.

2. Antenna system

The Schumann resonance has two components – the electric one which is vertical and the magnetic one. The experiment was aimed at measurement of the first case.

The construction of antenna as a standard dipole [7] is unsuitable for balloon experiments due to the long wavelengths. The most appropriate choice, as the space and mass of package is constrained by the avionic law, is the short dipole active antenna of length of a few centimetres comparing to the wavelength of hundreds of kilometres of ELF waves. In the field of ELF wave it behaves like the electromotive source with negligible resistance and inductance. Therefore, it has to be connected with amplifier with large input impedance and small capacity. The most optimal length for balloon missions is 20 cm, which, based on the ground level measurements of the Schumann resonances, would generate output on the level of 90 μV, which results from the standard theory of short dipole of given length and estimated value of electric field on the ground level.

The scheme of the amplifier system is a small modification of the design from [8] called ELA 1 with passive antenna. For summary see also [9]. This design was used to observe ELFs on the ground [10, 11, 12, 13]. It was supplied with a Chebyshev filter reducing aliasing. The output was connected to the ADC described in the next section.

In the ground tests the dominating signal 50 Hz of electric power was visible, that showed the antenna system worked. In this design the induction of signal in the antenna by movement in Earth magnetic field was not taken into account as the effect depends the parameters of flight and wind.

3. Data acquisition system

The system of data acquisition consists of two computers for backup purposes:

1. RaspberryPi 3 B, 4-channel 12bit ADC converter – ADS1015, GPS and IMU (Inertial Measurement Unit) GY-801;
2. Arduino Due, 16 bit ADC converter – ADS1115 and GPS.

The systems were charged by the TP-LINK TL-PB10400 power bank with 10 400 mAh capacity. The power bank has two ports and one of them charged the RaspberryPi and Arduino computers and the second one the antenna system. There was also a second battery (Colorovo PowerBox 6800 mAh) connected to the YI Action 2 camera, which was also the device that was tested against low temperatures and extreme stratospheric conditions. The 10 400 mAh power supply was too large for a 2 h mission (as it was tested before the mission); however it was used in order to prevent the effect of low temperature on the capacity of chemical power sources. The data were saved on fast input-output transaction SD cards. The data frame format used in the first system was as follows:

```
GPS: DATA
GPS: DATA
[ACCELERATION x,y,z] [GYROSCOPE x,y,z] [MAGNETIC FIELD x,y,z]
[
'TIME', ADC,
'TIME', ADC,
...
]
...
END
```

where the first two lines are GPS data, then the data from IMU. The next part is the data from the timer and the corresponding ADC readout of 3000 samples and finally END marker. The average sampling ratio was 300 Hz, which is sufficient to detect the Schumann resonances. The clock was synchronized with GPS at startup of the system before the start.

The data frame of the second system has the following format:

GPS
ADC
...
GPS
ADC
...

where GPS stands for the data from GPS and ADC denotes the data from ADC. The Arduino was not connected to the RTC (Real Time Clock), therefore it saved the ADC data until the GPS interrupted, which ended the frame – the number of ADC readouts depends on the frequency of GPS interruptions. Its average sampling rate was 600-700 Hz – the spread results from interrupt-driven design.

4. Gondola

The gondola was designed to meet standards of aviation law. Its mass was 1.69 kg. The whole gondola was made from pieces of XPS (Extruded Polyester) glued together. The outer layer was covered by aluminium foil in order to prevent electrostatic discharges which could disrupt ELF measurements. The sharp end of the antenna was protected by a piece of XPS. Figure 1 presents its cross section along the centre.

The bottom isolated compartment was occupied by the antenna system. The upper part was occupied by the acquisition systems, GPS and battery. In addition APRS (Automatic Packet Reporting System) which allowed us to localize the balloon online was also present. In the top part there was also a place for a camera, which was placed for tests in stratospheric conditions – there was a hole for the lens; see Figure 2.

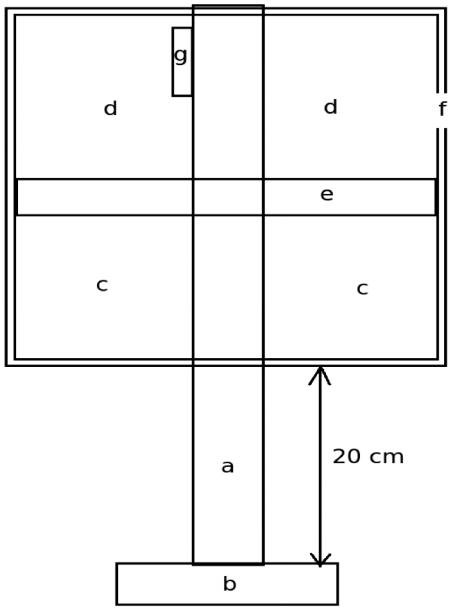


Fig. 1. Schematic sketch of gondola (no real sizes): a – antenna (plastic pipe covered by metallic foil); b – protection of the antenna sharp end (XPS material); c – bottom compartment (for antenna system); d – top compartment (for acquisition system, telemetry, battery and camera); e – partition wall from XPS material; f – hole for camera; g – Inertial Measurement Unit

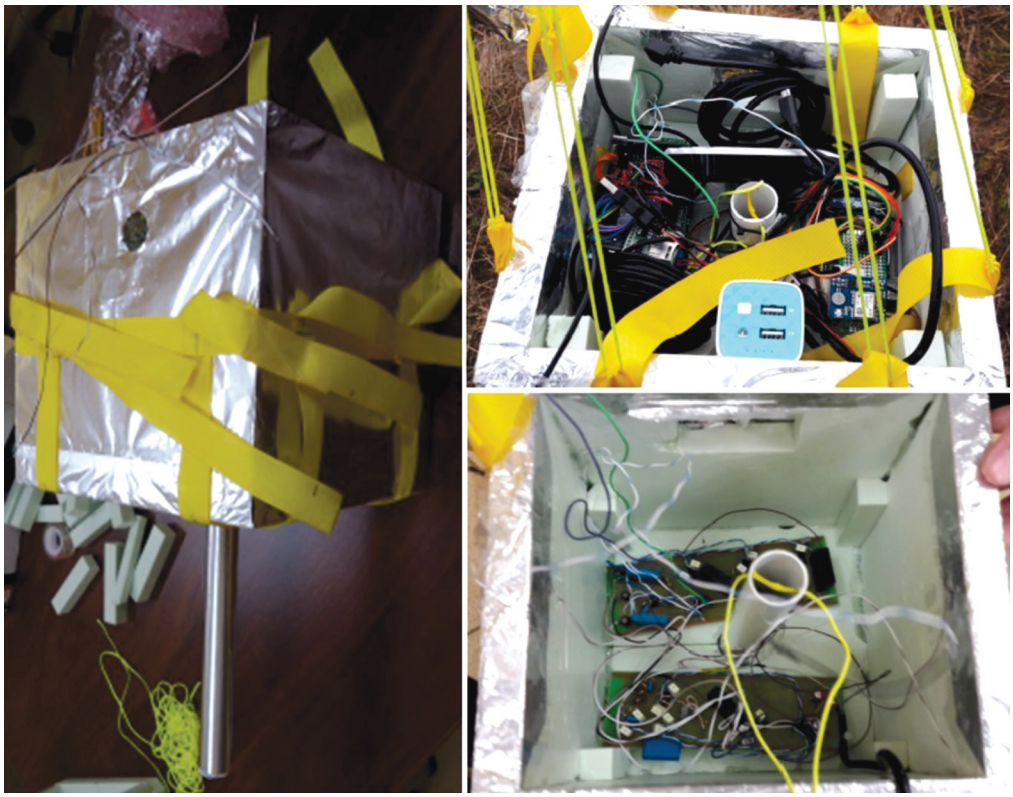


Fig. 2. Capsule view – side view, top compartment, bottom compartment

The additional payload of biological samples was attached to the side walls of the gondola as well as to the long string and hang below the gondola.

The gondola was attached to the parachute by nylon strings and the parachute was connected with a balloon in such a way that in case it popped it automatically opens during the fall.

For the experiment abHwoyee HY-1200 balloon model was selected as it is sufficient to reach 30 km with a payload of 2 kg when filled with hydrogen gas.

In the next section the mission will be described.

5. Mission

The mission started on 27 November 2016 at 9:18 am of CET, i.e. GMT (+1 h) when the balloon was released near Gliwice, Poland, see Figure 3. The decision of the start was preceded by a simulation of the trajectory using wind predictions at [14]. The prediction for the balloon trajectory shortly after the start of the balloon is presented in Figure 4 and the path of flight from GPS data is presented in Figure 5.

Fig. 3. Start of the balloon

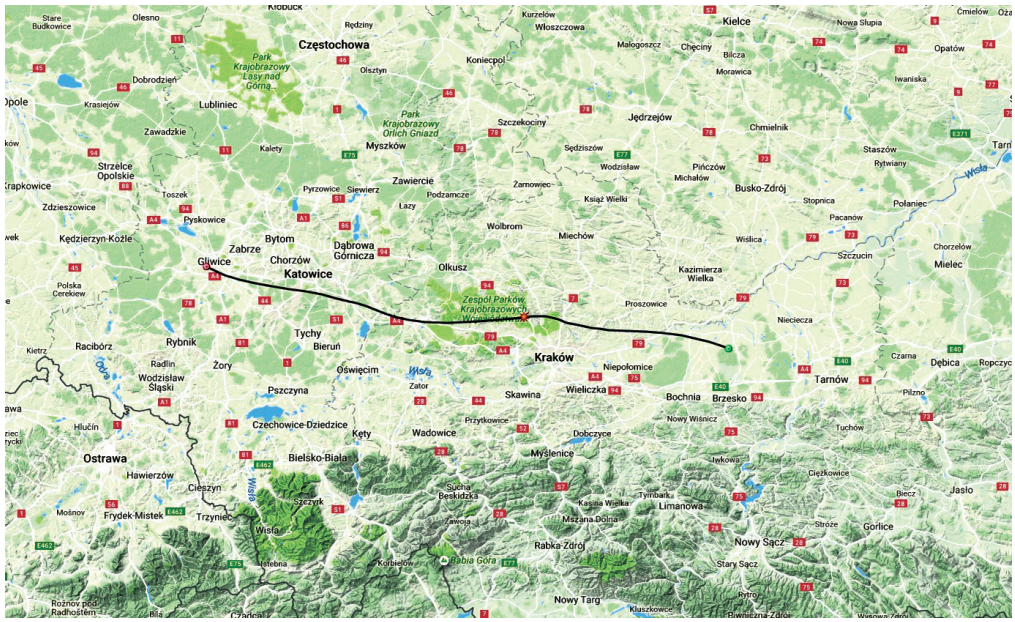


Fig. 4. Trajectory of the balloon predicted according to the wind simulation on 24th of November from [14]

The comparison of Figure 5 with the simulation made on 24 of November on Figure 4 indicates that the simulation quite well agrees with the real path.

In addition the height profile of the path of the balloon from APRS data are presented in Figure 6. The height plot for the data from GPS is presented in Figure 7. It can be seen that the balloon had constant vertical speed during ascending (line); however, after it popped its vertical velocity was large (almost horizontal part of the trajectory) until it started to decelerate when the parachute was slowly opening. A more detailed analysis of the flight can describe the dynamics of the atmosphere and the gondola-balloon system and this will be presented in a separate paper. The balloon popped at an attitude of 30 km above sea level and during the descent the parachute opened as can be seen in Figure 8.

The flight lasted 2 hours. The gondola travelled 135 km from the starting to the landing point and the lowest temperature it was exposed to was −55°C.

Additional information from the mission, including photos, are available on [18].

The preliminary results of the experiment will be described in the next section.

6. Preliminary results

Preliminary analysis of the data from the Schumann resonances measurement system indicated that the system was saturated for the entire mission as is presented in Figure 9.

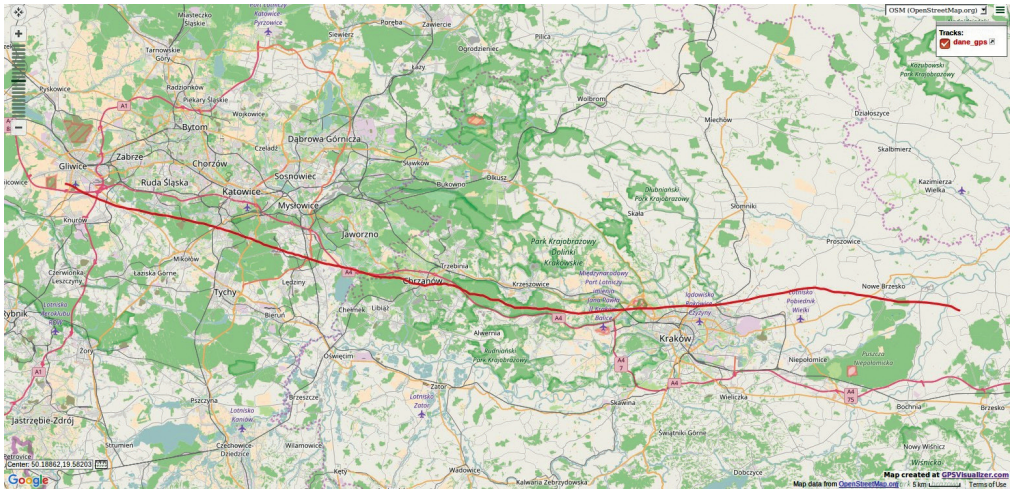


Fig. 5. Trajectory of the balloon generated from GPS data from [16]

This indicates that the level of amplification was too high. Therefore the redesign of this part of the system is required before the next mission. The excitement was caused by the movement in the Earth's magnetic field as it was tested after the mission in ground tests.

The analysis of data from the GPS and accelerometer indicates that it is possible to make a dynamical model of the gondola-balloon system and atmospheric dynamics. The analysis deserves another publication which is currently under development.

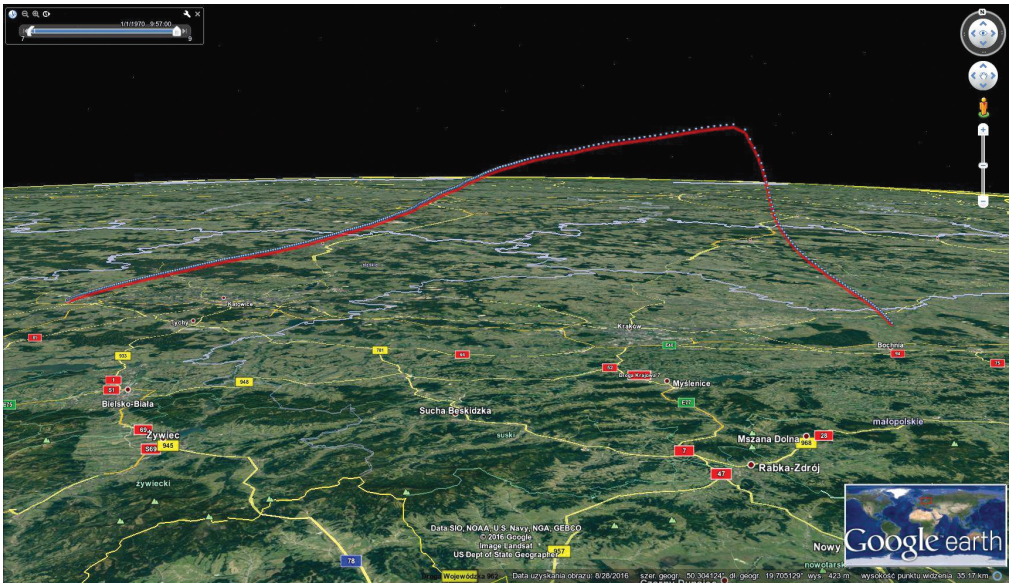


Fig. 6. The path of the balloon from APRS data visualized in Google Earth service [15]

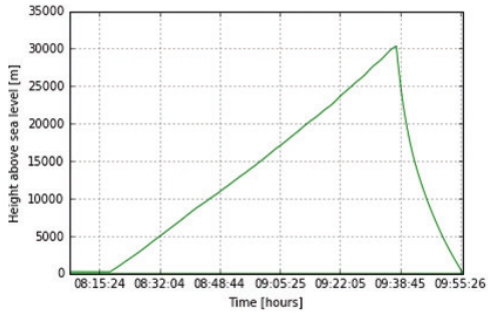


Fig. 7. Height profile for GPS data. Time is given with the respect to GMT

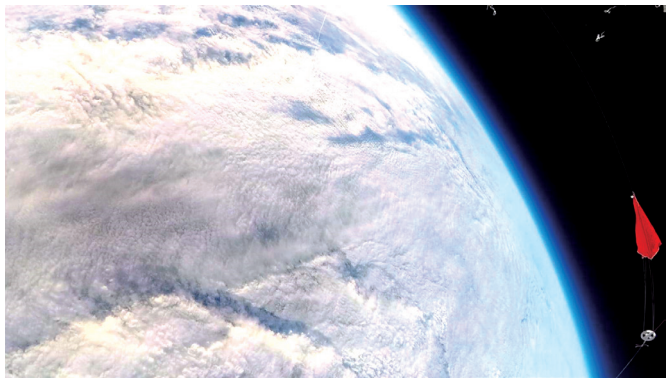


Fig. 8. Balloon popping. The opening of the parachute (red) and the remnants of the balloon are visible. The large cloud cover and the atmosphere layer that gradually passes to the space can be noticed

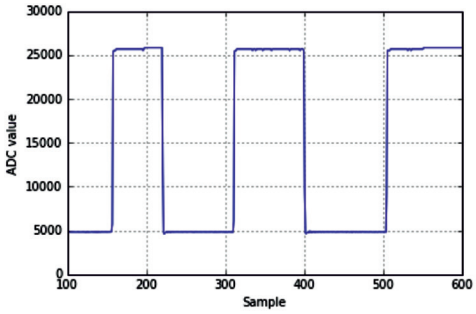


Fig. 9. Saturation of the ELF system. The system oscillates between low and high state

The analysis of the biological part of the experiment is currently in progress; however, no living bacteria and fungi were detected.

The camera was tested under stratospheric conditions. It worked after the landing and the film from the mission is available at [17].

7. Conclusions

The experiment described in the paper was intended as a proof of concept for stratospheric ELF missions. Although no Schumann resonance was registered the analysis of the results enables the system to be redesigned for future missions. Additional data including telemetry and biological samples were gathered and the equipment was tested under extreme conditions. The concepts used in this experiment can be adapted to future similar balloon missions on Mars.

The experiment was funded by the Faculty of Physics, Mathematics and Computer Science of Tadeusz Kościuszko Cracow University of Technology. We would like to thank Modular Analog Research Station M.A.R.S., Astronomia Nova organisations and YI.pl company for support. Last but not least, we would like to thank the Hackerspace group in retrieving the gondola after landing.

References

[1] MacGorman D.R., Rust W.D., *The electrical nature of storms*, Oxford University Press, 1998.

[2] Schumann W.O., Knig H., *Uber die Beobactung von Atmospherics bei geringsten Frequenzen*, Naturwissenschaften, 41 (8), 1954, 183-184, DOI:10.1007/BF00638174.

[3] Molina-Cuberos G.J., Morente J.A., Besser B.P., Porti J., Lichtenegger H., Schwingenschuh K., Salinas A., Margineda J., *Schumann resonances as a tool to study the lower ionospheric structure of Mars*, Radio Science, 41, 1, 2006, DOI: 10.1029/2004RS003187.

[4] Kozakiewicz J., Kulak A., Kubisz J., Zietara K., *Extremely Low Frequency Electromagnetic Investigation on Mars, Earth, Moon, and Planets*, 118, 2, 2016, 103-115.

[5] Kirschvink J.L., *Comment on "Constraints on biological effects of weak extremely-low-frequency electromagnetic fields"*, Physical Review A, 46, 4, 1992.
[6] Benda R. et al., *Measurements of atmospheric electrical parameters and ELF electromagnetic emissions during a meteorological balloon flight*, Geophysical Research Abstracts, Vol. 18, EGU2016-7725, EGU General Assembly, 2016.
[7] Balanis C.A., *Modern Antenna Handbook*, Wiley-Interscience, 2008.
[8] Kulak A., Knapik J., Kubisz J., *Przenośna stacja do obserwacji widm dynamicznych pól elektromagnetycznych w zakresie ULF*, Proceedings of the VII National Symposium of Radio Science (URSI), 1996.
[9] Kulak A., Kubisz J., Klucjasz S., Michalec A., Mlynarczyk J., Nieckarz Z., Ostrowski M., Zieba S., *Extremely low frequency electromagnetic field measurements at the Hylaty station and methodology of signal analysis*, Radio Science, 49, 6, 2014, 361-370, DOI: 10.1002/2014RS005400.
[10] Kulak A., Kubisz J., Michalec A. et al., *Solar variations in extremely low frequency propagation parameters: I. A two-dimensional telegraph equation (TDTE) model of ELF propagation and fundamental parameters of Schumann resonances*, Journal of Geophysical Research, 108(A7), 1270, 2003, DOI:10.1029/2002JA009304.
[11] Kulak A., Mlynarczyk J., Zieba S. et al., *Studies of ELF propagation in the spherical shell cavity using a field decomposition method based on asymmetry of Schumann resonance curves*, Journal of Geophysical Research-Space Physics, 111, A10, A10304, 2006, DOI: 10.1029/2005JA011429.
[12] Nieckarz Z., Kulak A., Zieba S. et al., *Comparison of global storm activity rate calculated from Schumann resonance background components to electric field intensity E-OZ*, Atmospheric Research, 91, 2-4, 2009, 184-187, DOI: 10.1016/j.atmosres.2008.06.006.
[13] Nieckarz Z., Zieba S., Kulak A. et al., *Study of the Periodicities of Lightning Activity in Three Main Thunderstorm Centers Based on Schumann Resonance Measurements*, Monthly Weather Review, 137, 12, 2009, 4401-4409, DOI: 10.1175/2009MWR2920.1.
[14] http://predict.habhub.org/ (access: 2.10.2017).
[15] https://www.google.com/earth/ (access: 2.10.2017).
[16] http://www.gpsvisualizer.com/ (access: 2.10.2017).
[17] Movie from the mission: https://www.youtube.com/watch?v=3Cyj75oyoII (access: 2.10.2017).
[18] Additional material from the mission are available at http://fizyk.ifpk.pk.edu.pl/~rkycia/groups/elfHunters.html (access: 2.10.2017).