ISSN 1732-9353 (suspended)
eISSN 2543-7496

Scientific Review

Engineering and Environmental Sciences

Przegląd Naukowy

Inżynieria i Kształtowanie Środowiska

Vol. 33 (1) 2024 Issue 103

Quarterly

SCIENTIFIC REVIEW
ENGINEERING AND ENVIRONMENTAL SCIENCES
Quarterly

EDITORIAL OFFICE ADDRESS
Wydział Budownictwa i Inżynierii Środowiska SGGW, ul. Nowoursynowska 159, 02-776 Warsaw, Poland
tel. (+48 22) 59 35 363, 59 35 210, 59 35 302
e-mail: srees@sggw.edu.pl
https://srees.sggw.edu.pl

Electronic version of the Scientific Review Engineering and Environmental Sciences is primary version

All papers are indexed in the data bases as follows: AGRO(Poznań), BazTech, Biblioteka Nauki, **CrossRef**, **DOAJ**, **Google Scholar**, **Index Copernicus**, INFONA, POL-Index, **SCOPUS**, SIGŻ(CBR)

Scientific Review

Engineering and Environmental Sciences

Przegląd Naukowy

Inżynieria i Kształtowanie Środowiska

Vol. 33 (1) **2024** **Issue 103**

Contents

Wydawnictwo SGGW, Warsaw 2024

Wydawnictwo SGGW wydawnictwosggw

Editorial work – Anna Dołomisiewicz, Violetta Kaska

ISSN 1732-9353 (suspended) eISSN 2543-7496

Scientific Review Engineering and Environmental Sciences (2024), 33 (1), 3–16

Sci. Rev. Eng. Env. Sci. (2023), 33 (1) https://srees.sggw.edu.pl

ISSN 1732-9353 (suspended) eISSN 2543-7496 DOI 10.22630/srees.5948

Received: 30.11.2023
Accepted: 02.01.2024

Suren G. AGHBALYAN[1]
Vazgen BAGDASARYAN[2]✉
Gayane A. VASILYAN[1]
Rafał WYCZÓŁKOWSKI[3]

[1]National Polytechnic University of Armenia, Faculty of Mining and Metallurgy, Armenia
[2]Warsaw University of Life Sciences – SGGW, Institute of Civil Engineering, Poland
[3]Czestochowa University of Technology, Department of Production Management, Poland

Investigations of stress and strain state of aluminum alloys during a hot extrusion and patterns of structure and feature formation

Keywords: casting, aluminum alloy, hot extrusion, slip, stress, matrix, pressure

Introduction

In the pursuit of new materials characterized by high strength, resistance to high temperatures, plasticity, and other valuable properties, composite materials based on metals, such as dispersion-strengthened materials, hold great potential. They can be divided into two groups based on particle size: first group where dimensions range from 0.01 μm to 0.1 μm and the content varies between 1% and 15% by volume of the material, and the second group where particle sizes exceed 1 μm, with a content of approximately 25% by volume of the material, hold great potential. Other examples of composite materials include materials composed of two metals or layered materials. However, in terms of a diverse range of valuable material properties,

✉vazgen_bagdasaryan@sggw.edu.pl

the greatest promise lies in fiber-reinforced materials, where fibers can have diameters ranging from 1 μm to several tens of micrometers, and the fiber content in the material varies from a few percent to even approximately 70% by volume.

In recent years, durable alloys and composite materials based on aluminum have been widely applied in many branches of industry, with their production made possible through casting technology (Sheng et al., 2020; Rodríguez-González, Ruiz-Navas & Gordo, 2022; Kobayashi, Funazuka, Maeda & Shiratari, 2023). However, a significant challenge lies in the fact that aluminum powder is highly oxidized, preventing it from adhering tightly to other metals, resulting in material inhomogeneity, which in turn hinders the attainment of desired mechanical properties. Additionally, aluminum powder is expensive. These issues are not present in materials produced using casting technology. On the other hand, to avoid defects in the casting process and achieve the desired fine particle structure, ensuring the desired mechanical properties, technologies utilizing high pressure, including hot extrusion, have been widely used.

For the reasons mentioned above, the research has focused on studying the state of stresses and strains, as well as changes in the structure and material properties of aluminum alloys subjected to hot extrusion. This issue appears to have significant potential for practical applications, (Johnson & Kudō, 1962; Sheng et al., 2020; Kobayashi et al., 2023).

Material and methods

To study deformation changes in metals, various empirical methods and in specific cases theoretical research methods are employed. These methods allow for the assessment of particle displacements in metals subjected to extrusion, quantitative and qualitative relationships between these displacements, as well as the evaluation of stress and strain states (Johnson & Kudō, 1962; Perlin & Raitbarg, 1975; Gun & Prudkovsky, 1979; Kenesei, Kádár, Rajkovits & Lendvai, 2004).

The research was conducted on cylindrical and semi-cylindrical samples of an aluminum alloy containing 5.7% of copper. In Table 1, the composition and physical properties of the used aluminum powder and in Table 2 of copper are presented.

TABLE 1. The composition and physical properties of the used aluminum powder

Al content [%]	Al_2O_3 content [%]	Fe content [%]	Humidity [%]	Bulk density [g·cm^{-3}]
93.8	6	0.2	0.1	0.9

TABLE 2. The composition and physical properties of the used copper powder

Cu content [%]	Fe content [%]	Pb content [%]	As content [%]	Sb content [%]	Bi content [%]	Bulk density [g·cm^{-3}]
99.5	0.02	0.05	0.005	0.01	0.30	2.5

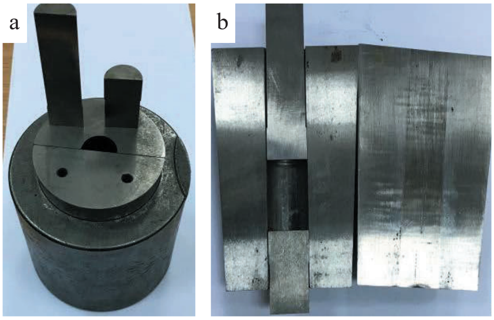


FIGURE 1. The mold for preparing samples: a – assembled; b – cross-section

Source: own elaboration.

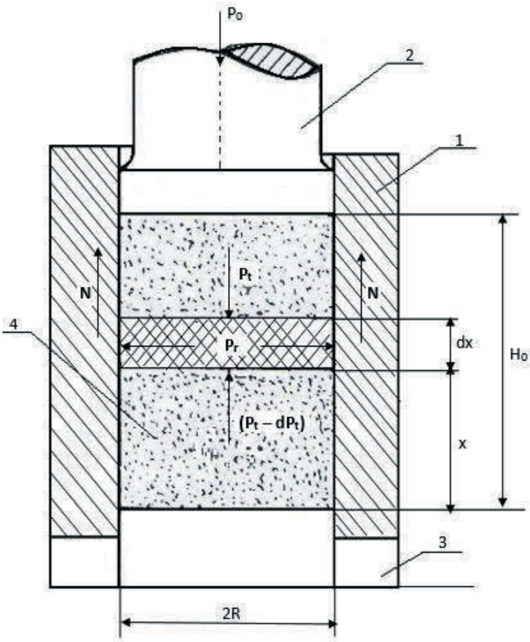


FIGURE 2. A schematic of sample creation: 1 – die, 2 – movable punch, 3 – fixed punch, 4 – metal powder

Source: own elaboration.

The samples were prepared in steel molds and subjected to a pressure of 3 t·cm^{-2} using hydraulic work (model: P474A). In Figure 1, a mold for preparing semi-cylindrical samples is presented. The height of the samples was 45 mm, and the diameter (both for cylindrical and semi--cylindrical samples) was 25 mm.

Because in the case of uniaxial compression, the distribution of compressive pressure in the sample varies along its height, it is important to know this distribution. It will allow, for example, to determine the required compression force to obtain a sample with the desired density or porosity if a known curve, such as compression pressure – density or compression pressure – porosity is available. These curves can be obtained experimentally. To determine the pressure distribution as a function of the longitudinal coordinate (x) in the sample, one considers the equilibrium of an elemental layer of the sample with a thickness dx. Figure 2 illustrates the schematic for the calculations.

Let's denote p_o, p_t as the real pressure on the top of the example and theoretical pressure calculated on the coordinate x respectively,

S as the sample surface area. Then the total force acting on the sample is P_o and P_t calculated according to the equations:

$$P_o = p_o \cdot S,\ P_t = p_t \cdot S, \tag{1}$$

Now we can write the equilibrium of an elemental layer of the sample:

$$p_t \cdot \pi \cdot R^2 = \left(p_t - dp_t\right) \cdot \pi \cdot R^2 + 2 \cdot \pi \cdot R \cdot dx \cdot N, \tag{2}$$

where, in accordance with Figure 2, R is the sample radius, dx is the thickness of the elementary layer under consideration, and N is the frictional force between the sample and the mold. We can write that $N = f \cdot \mu \cdot p_t$, where f is the coefficient of friction between the metal powder and the mold, while μ is the coefficient of lateral forces equal to $\mu = \nu/1 - \nu$, where ν the Poisson's ratio.

Substituting the above relationships into Equation (2) and simplifying it, we obtain:

$$dp_t \cdot R = 2 \cdot p_t \cdot \mu \cdot f \cdot dx, \tag{3}$$

Transforming Equation (3), we obtain the following ordinary differential equation with separated variables:

$$\frac{dp_t}{p_t} = \frac{2 \cdot \mu \cdot f \cdot dx}{R}. \tag{4}$$

By integrating both sides of Equation (4), we obtain the solution:

$$\int_{p_o}^{p_t} \frac{dp_s}{p_s} = \frac{2 \cdot \mu \cdot f}{R} \int_{H_o}^{x} dx = -\frac{2 \cdot \mu \cdot f}{R} \int_{x}^{H_o} dx, \tag{5}$$

$$\ln \frac{p_t}{p_0} = -\frac{2 \cdot \mu \cdot f}{R}\left(H_o - x\right), \tag{6}$$

or equivalently

$$\frac{p_t}{p_0} = e^{-\frac{2 \cdot \mu \cdot f}{R}\left(H_o - x\right)}. \tag{7}$$

Finally, we obtain:

$$p_t = p_o \cdot e^{-\frac{2\mu f}{R}(H_o - x)}. \tag{8}$$

Therefore, at any given moment, knowing the force acting on the upper surface of the sample and its height, the pressure distribution along the height of the sample can be determined. According to Equation (8), the pressure decreases exponentially along the height of the sample. There is a reason why lower porosity is observed in the areas closer to the press plunger. Consequently, the strength and other mechanical properties of the samples may vary in different locations.

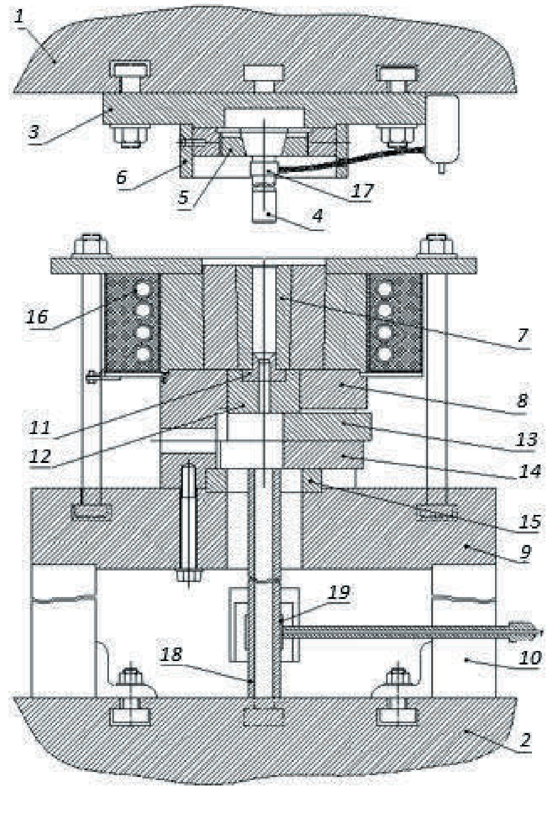


FIGURE 3. A schematic of the extruder: 1, 2 – upper and lower press plates, 3 – upper extruder plate, 4 – punch, 5 – nut, 6 – stop, 7 – container, 8 – container plate, 9 – plate, 10 – protecting pipe, 11 – die, 12 – clamp, 13 – upper wedge, 14 – lower wedge, 15 – ring, 16 – electric heater, 17 – strain gauge, 18 – force gauge, 19 – strain gauge

Source: own elaboration.

The prepared semi-cylindrical samples had a grid of coordinates applied to them (cuts and chalk markings). Subsequently, the paired samples were heated to a temperature of 550°C and then extruded through molds with diameters of 10 mm, 12 mm, and 14 mm to form cylindrical samples. Figure 3 presents a schematic of the extruder.

Figure 4 shows external forces and the resulting stresses generated during the compression of the sample (Aghbalyan & Stepanyan, 2006). All these forces and stresses change depending on boundary conditions, the speed at which the process occurs, and the physical state of the sample. Significant changes are particularly noticeable in the final phase of the process (Perlin & Raitbarg, 1975; Dunand, 2004; Dixit & Narayanan, 2013).

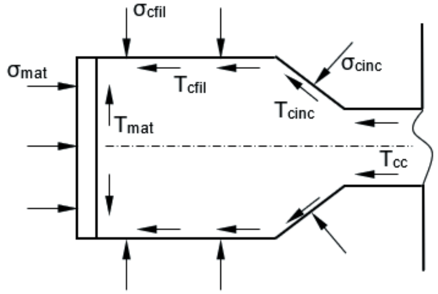


FIGURE 4. Diagram of forces and stresses acting on the sample during extrusion (in reality, the conical part has rounded walls)

Source: own elaboration.

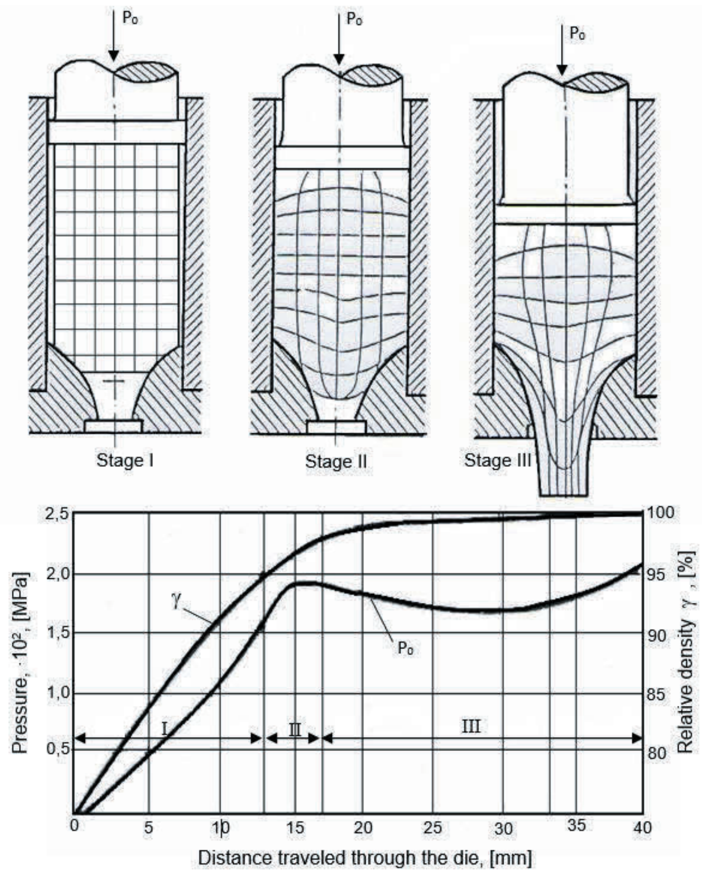


FIGURE 5. Changes in die pressure and relative density depending on the stage (I, II, III) of the hot extrusion process (P_o – the total force acting on the sample during extrusion)

Source: own elaboration.

Due to the forces acting on the sample and the deformations occurring within it, the extrusion process can be divided into three main stages (Wright, 2011):

– Stage I – the initial phase of the extrusion process, during which the sample is pressed into the container;
– Stage II – the main stage of the process characterized by a constant flow of metal;
– Stage III – the final phase, during which the constant flow of metal is disrupted, and the extrusion process concludes.

Each of these stages is characterized by a specific stress state, which can be determined primarily based on a chart depicting the relationship between the pressure in the extrusion process and the distance travelled through the die (Aghbalyan & Stepanyan, 2006). A schematic of such a chart is shown in Figure 5.

Results and discussion

As a result of the conducted research, several interesting observations were made. In Figure 6, samples with a coordinate grid applied after the hot extrusion process are shown.

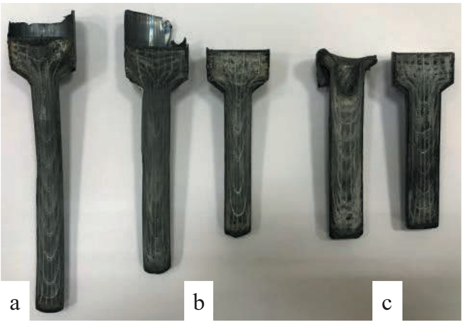


FIGURE 6. Image of samples with a coordinate grid applied after the hot extrusion process: a – for a sample with a final diameter of 10 mm; b – for samples with a final diameter of 12 mm; c – for samples with a final diameter of 14 mm

Source: own elaboration.

Based on observations from real samples, a coordinate grid scheme was created for three phases of the process. The coordinate grid and the overall deformation state diagram are presented in Figure 7. The deformation state of the metal sample subjected to hot extrusion has been determined based on a coordinate grid in each stage of the process.

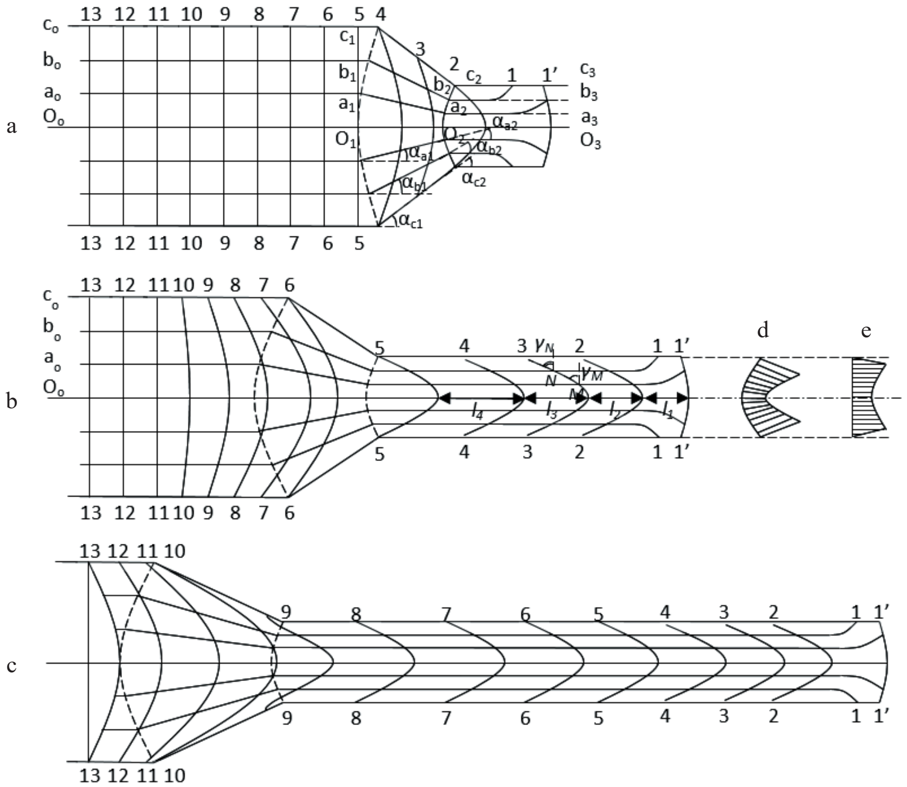


FIGURE 7. General schematic of the coordinate grid and stress state of the metal sample subjected to hot extrusion: a – in the initial phase of the process; b – in the main phase of the process; c – in the final phase of the process; d – deformation chart in the plastic deformation zone at the end of the conical part; e – deformation chart of the sample at the beginning of the conical part

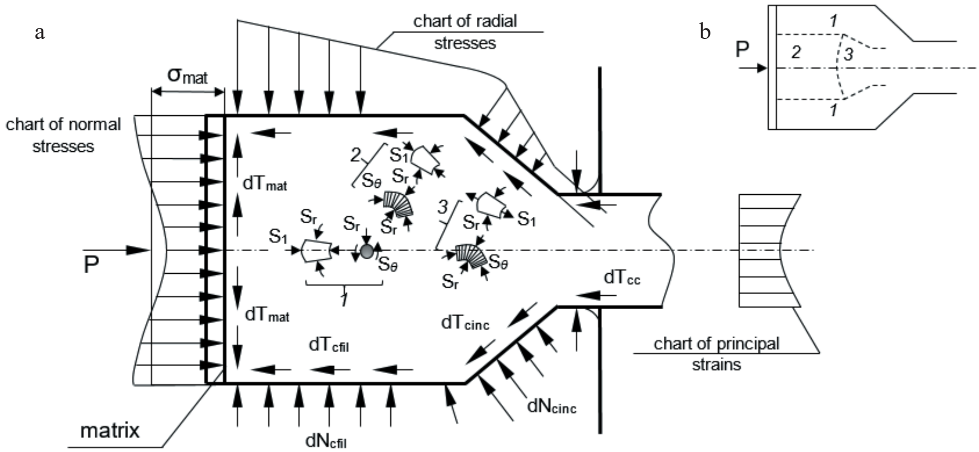
Source: own elaboration.

Based on Figure 7, the following main observations have been made:

1. After the extrusion process, all initial longitudinal lines of the coordinate grid, except for the front part of the die, remained straight, while they curved in the conical part of the die.
2. The resulting curvatures are converse to each other; for example, the angles α_{c1} and α_{c2} indicate that the deformations are not monotonic.
3. There is a noticeable relationship: $\alpha_{c1} > \alpha_{b1} > \alpha_{a1}$ and $\alpha_{c2} > \alpha_{b2} > \alpha_{a2}$, indicating that deformations decrease towards the axis of the sample (Fig. 7).
4. Longitudinal lines of the coordinate grid curved in the conical part of the die – at the beginning (points c_1, b_1, a_1) and at the end (points c_2, b_2, a_2). This indicates that in the plastic deformation zone, the front surfaces are continuous, asymmetric surfaces with convexity directed opposite to the extrusion.

5. At a certain distance from the plastic deformation zone, longitudinal lines of the coordinate grid inclined towards the axis of the sample, causing the expansion of outer layers and the narrowing of inner layers.
6. All transverse lines of the initial coordinate grid symmetrically bulged in the direction of extrusion. This indicates that inner layers moved faster than outer layers.
7. Under classical extrusion conditions, transverse lines of the coordinate grid, except for the one adjacent to the front part, took on the shape of a parabola $y^2 = p \cdot x$. The p parameter depends on the distance of the curve from the front surface of the sample.
8. Transverse lines adjacent to the front surface (in this example, one line) after the extrusion process take on the shape of a broken line 1-1'-1'-1 (Fig. 7), indicating that before deformation, the ends of the grid segments moved to the lateral surfaces, and a portion of the front surface of the sample transitioned to the lateral surface.
9. The curvature of transverse lines of the coordinate grid increases opposite to the direction of extrusion.
10. In general, in the direction opposite to the extrusion direction, there is also a slight increase in the distance between transverse lines of the grid.
11. The inclinations of transverse lines of the coordinate grid indicate that during extrusion, the concentric layers of the sample undergo not only longitudinal and transverse deformations, but also shear ones. Shear increases towards the outer layers, as evidenced by the inequality of angles $\gamma_N > \gamma_M$, describing shear at points N and M.
12. The increase in shear towards the surface indicates that the principal tensile stresses resulting from deformations and shear also increase towards the surface.
13. In the zone at the end of the conical part, the directions of principal normal stresses are different – they form an angle with the axis of the sample, which also increases towards the surface.
14. From the observed changes in the sample, it can be concluded that throughout its volume, radial and circumferential deformations are constricting deformations.

Figure 4 depicts a schematic of the forces acting on the sample subjected to extrusion (Perlin & Raitbarg, 1975; Salonine & McQueen, 2004). If we divide the sample into layers of infinitely small thickness and neglect the influence of friction with the equipment as well as friction between the layers, it will become apparent that during the extrusion of such a multilayered element, the layers will be in a state of multi-directional, uneven compression. This means that tensile deformations will be primarily passive. Under classical extrusion conditions, where the resistance of the material to deformation is nearly constant throughout the volume of the sample, the inner layers move faster than the outer layers. As a result, additional stresses are generated: tensile stresses in the outer layers (because each inner layer, moving faster than the neighboring layer, pulls it along), and compressive stresses in the inner layers (because each outer layer, moving slower than the neighboring layer, hinders it). Consequently, in the effort to equalize the stresses, two regions are formed: a region of outer layers, which is subjected to additional longitudinal tensile stresses that decrease toward the axis of the sample, and a region of inner layers, which is subjected to longitudinal compressive stresses that increase toward the axis of the sample.



P – total force acting on the die; dT_{mat}, dT_{cfil}, dT_{cinc}, dT_{cc} – elemental frictional forces of the matrix, side wall of the container, inclined container walls, and calibration surfaces respectively; dN_{cfil}, dN_{cinc} – elemental normal forces acting on the side wall of the container and inclined container walls; σ_{mat} – matrix surface stress

FIGURE 8. Diagram of forces acting on the sample, strain chart, and types of stress states during hot extrusion: a – under classical conditions; b – in the case of a sudden decrease in strength in the surface layer

Source: own elaboration.

In the zone of inner layers, additional compressive stresses, superimposing with the main compressive stresses, do not alter the stress state (Fig. 8a, Zone 1).

In the zone of outer layers (Fig. 8a, Zones 2 and 3), the additional tensile stresses, superimposing on the main compressive stresses, reduce them and change the stress state of the sample. The multi-directional uneven compression transforms into two-directional: compression in the transverse direction to the sample and tension in the longitudinal direction (Fig. 8a, Zone 3). This part is located in the final section of the conical die, where the main stresses decrease towards the narrowing of the die, while the additional stresses increase.

In the case where the strength of the material in the outer layers is significantly lower compared to the strength of the inner layers, the outer layers move faster than the inner layers. As a result, additional stresses occur: tensile stresses in the inner layers (because each outer layer, moving faster than the neighboring layer, pulls it along), and compressive stresses in the outer layers (because each inner layer, moving slower than the neighboring layer, hinders it). Thus, in the plastic deformation zone, three regions are formed:

1. A region of inner layers where additional longitudinal tensile stresses decrease towards the surface (Fig. 8b, Region 1).
2. A region of inner layers where longitudinal compressive stresses decrease towards the axis of the sample (Fig. 8b, Region 2).
3. A region of inner layers subjected to transverse compression and longitudinal tension (Fig. 8b, Region 3).

During hot extrusion, factors characterizing the internal structure of the sample include particle size and shape as well as the degree of internal non-uniformity. Metals subjected to hot extrusion are characterized by a high degree of structural non-uniformity, which results from the non-uniformity of deformations. Other factors influencing non-uniformity include changes in thermal conditions and structural changes during extrusion, followed by sample cooling (Perlin & Raitbarg, 1975; Woźnicki, Leśniak, Włoch, Leszczyńska-Madej & Wojtyna, 2016).

In general, the non-uniformity of the sample is expressed through variations in particle size in both transverse and longitudinal cross-sections. Depending on the deformation state, particle size decreases from the inner layers to the outer layers and from the front of the sample to its end. These differences are visible in Figures 9 and 10, which depict the microstructure of the sample.

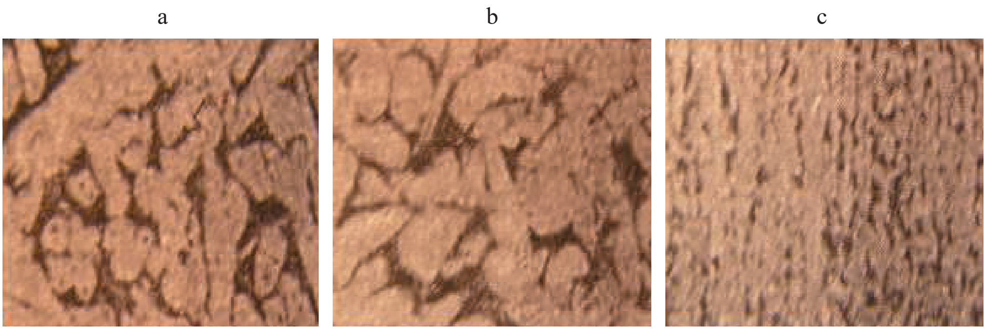


FIGURE 9. Microstructure of the internal layers of the duraluminum string (magnification ×100): a – beginning of the string; b – middle of the string; c – end of the string

Source: own elaboration.

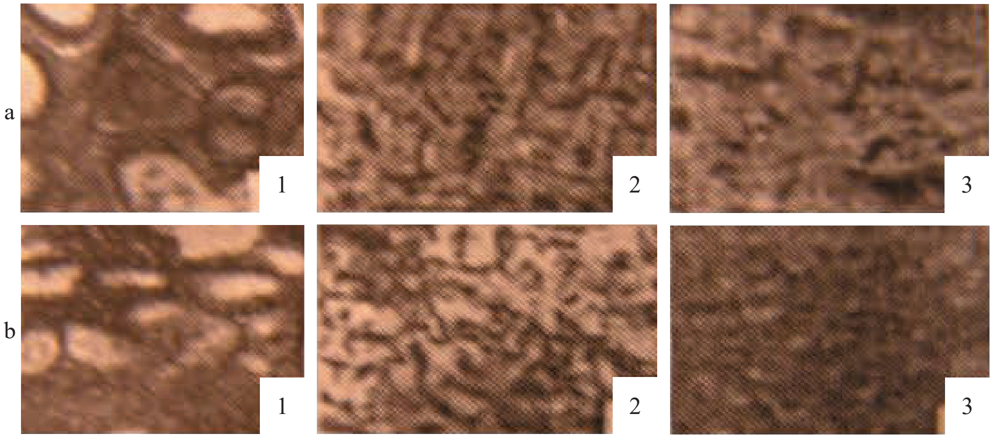


FIGURE 10. Microstructure of different sections of the duraluminum string extruded without grease (magnification ×250): a – surface of the rod; b – middle of the rod (1 – beginning of the string; 2 – middle of the string; 3 – end of the string)

Source: own elaboration.

The non-uniformity of particle dimensions increases after the sample cools down. This is because, during the heating of the sample, intense recrystallization occurs in the layers with significant deformations, resulting in the formation of an outer layer with large particles. This outer layer extends from the front of the sample towards its end. The development of such a layer and the fact that its thickness changes indicate the non-uniformity of the sample structure (Perlin & Raitbarg, 1975; Wu et al., 2020).

In the case of extrusion with the use of grease on the matrix surface or when no grease is used at all, there is either no occurrence of a layer with large particles, or it is thin.

A characteristic describing the structure of the sample is the average particle size, which depends on the conditions of the extrusion process. Increased deformation is typically accompanied by a simultaneous decrease in the average particle size, while an increase in temperature leads to a simultaneous increase in the average particle size.

In extruded samples, parallel delamination along the axis is often observed. When this effect occurs, transverse fractures exhibit a stepped surface, indicating a weakening of certain surfaces parallel to the axis of the sample. The non-uniformity of deformations and the internal structure of the sample have an impact on its mechanical properties. In the case of extrusion, the strength of the sample increases from the inner layers to the outer layer and from the front of the sample to its end. In some cases, the relative elongation follows a similar pattern but in the opposite direction.

In the case of small extrusion ratios, the mechanical properties of the sample in the inner and outer layers can differ significantly. However, when using large extrusion ratios, these differences are considerably smaller and sometimes even negligible. Many very interesting results regarding the material properties of elements subjected to hot extrusion can also be found in works (Sheng et al., 2020; Bazhenov et al., 2022; Rodríguez-González et al., 2022; Kobayashi et al., 2023).

Conclusions

It has been demonstrated that during hot extrusion, additional compressive stresses in the inner layers of the sample, superimposed on the main stresses, do not alter its stress state. The stress state remains constant and is characterized by uneven compression. Additional tensile stresses in the outer layers, superimposed on the main (compressive) stresses, reduce them and, in some cases, surpass them in magnitude, changing the stress state from multi-directional to two-directional compression: transverse compression and longitudinal tension.

It has been shown that the non-uniformity of deformations and the internal structure of the sample affect its mechanical properties. During extrusion, the strength of the material increases from the inner layers to the outer layer and from the front of the sample to its end. In the case of small extrusion ratios, the mechanical properties of the sample in the inner and outer layers can differ significantly. However, when using large extrusion ratios, these differences are considerably smaller and sometimes even negligible.

References

Aghbalyan, S. G., Stepanyan, A. M. (2006). *Tungsten-molybdenum high-speed steels and their heat treatment*. Yerevan: National Polytechnic University of Armenia (in Armenian).

Bazhenov, V., Li, A., Tavolzhanskii, S., Bazlov, A., Tabachkova, N., Koltygin, A., Komissarov, A. & Shin, K. S. (2022). Microstructure and mechanical properties of hot-extruded Mg–Zn–Ga–(Y) biodegradable alloys. *Materials*, *15* (19), 6849.

Dixit, U. S. & Narayanan, R. G. (2013). *Metal forming: technology and process modeling*. Noida: McGraw-Hill Education Private Limited.

Dunand, D. C. (2004). Processing of titanium foams. *Advanced Engineering Materials*, *6*, 369–376.

Gun, G. Y. & Prudkovsky, B. A. (1979). Automated design of extruding dies. Process automation and metal forming. *Science*, *1979*, 128–133.

Johnson, W. & Kudō, H. (1962). *The mechanics of metal extrusion*. Manchester: Manchester University Press.

Kenesei, P., Kádár, C., Rajkovits, Z. S. & Lendvai, J. (2004). The influence of cell-size distribution on the plastic deformation in metal foams. *Scripta Materialia*, *50* (2), 295–300.

Kobayashi, R., Funazuka, T., Maeda, T. & Shiratori, T. (2023). Effects of hot extrusion temperature conditions on the hardness and electrical conductivity of rapidly solidified Al-Fe alloys. *Materials*, *16* (14), 5050.

Perlin, L. & Raitbarg, L. H. (1975). *Teoriya pressovaniya metallov*. Moscow: Matallurgiya.

Rodríguez-González, P., Ruiz-Navas, E. M. & Gordo, E. (2022). Effect of heat treatment prior to direct hot-extrusion processing of Al–Cu–Li alloy. *Metals*, *12* (6), 1046.

Salonine, D. S. & McQueen, H. J. (2004). Reduction of temperature extremes at the die exit in aluminum alloys extrusion. *Materials Forum*, *28*, 1086–1091.

Sheng, X., Yang, Y., Yu, H., Wu, W., Li, K., Liu, Y., Zhao, Y. & He, G. (2020). Hot extrusion enhanced homogenization of microstructure in a spray deposition aluminum alloy. *Metals*, *10* (2), 263.

Woźnicki, A., Leśniak, D., Włoch, G., Leszczyńska-Madej, B. & Wojtyna, A. (2016). The effect of cooling rate after homogenization on the microstructure and properties of 2017 a alloy billets for extrusion with solution heat treatment on the press. *Archives of Metallurgy and Materials*, *61* (3), 1663–1670.

Wright, R. N. (2011). *Wire technology-process engineering and metallurgy*. Butterworth-Heinemann: Elsevier.

Wu, X., Xu, C., Kuan, J., Zhang, Z., Zhang, J. & Yang, W. (2020). Effects of hot extrusion temperature on mechanical and corrosion properties of Mg-Y-Zn-Zr biological magnesium alloy containing W phase and I phase. *Materials*, *13* (5), 1147.

Summary

Investigations of stress and strain state of aluminum alloys during a hot extrusion and patterns of structure and feature formation. This paper investigates the stress and strain state as well as formation processes of structure and features of aluminum alloys during the hot extrusion. It has been shown that during the hot extrusion the ring layers of an extruded element experience not only longitudinal and transverse deformations, but also a slip. The slip increases from inner layers to the surface layer. The tensile principal stresses and the sum of slip deformations also increase. It has been also demonstrated that at the exit of the pressing part the tensile principal stresses have different directions, forming an angle with extruder axis, which also increases towards the surface. In conclusion, it has been stated that the main radial and circumferential deformations act as restraining deformations.

Scientific Review Engineering and Environmental Sciences (2024), 33 (1), 17–32
Sci. Rev. Eng. Env. Sci. (2023), 33 (1) https://srees.sggw.edu.pl
ISSN 1732-9353 (suspended) eISSN 2543-7496 DOI 10.22630/srees.5247

Received: 22.08.2023
Accepted: 30.01.2024

Abdulamir Hussein QASIM✉

University of Kirkuk, College of Engineering, Kirkuk, Iraq

Sustainable debris management by linear dynamic transportation model

Keywords: disaster management, debris discharge, debris withdrawal, truck routing path, sustainable debris management

Introduction

Disasters can be defined as the failure of systems or negative results of a system. Unexpected disasters can be generated naturally (such as an earthquake, fires, etc.) or through human activities such as war, man-made pollution, nuclear explosions, etc.). Disaster waste and debris can be generated during an actual disaster, or later during inaccurate planning in response phases. Depending on their type and severity plus unsuccessful management plans, disasters will create large amounts of debris. Through understanding the actual situation of a city, it is necessary to make usable waste disposal procedures with applicable policies and scenarios before the disaster strikes the region (Hirayama, Shimaoka, Fujiwara, Okayama & Kawata, 2010; Wei, Hu & Liu, 2021).

To find a reliable method for the reconstruction of a residential area that had been destroyed, this paper starts to examine both the disaster outcomes and consequences that can result from them. It is known that the debris resulting from the demolition of buildings, whether due to natural disasters such as earthquakes or unnatural ones such as wars, is a major obstacle to reconstruction. This issue has been discussed in many kinds of literature. However, more studies are still needed, especially after the destruction caused by the recent earthquakes in Turkey, and the wars in Ukraine,

✉aqasim01@uokirkuk.edu.iq

Syria, and Yemen, for example. So far, no integrated solution has been found to meet both the financial requirements of the citizens affected by the disasters and the economic requirements of local governments.

Several studies have been conducted to establish an efficient management technique and guidelines for disaster waste. Whereas other studies conducted in recent years have demonstrated the importance of recycling technology, and suitable temporary storage sites. Lorca, Çelik, Ergun and Keskinocak (2015) presented a support tool to assist disaster management by optimizing the environmental costs and the number of recycled materials generated. By deciding on operational and readiness plans, the support tool may be useful in comprehending the repercussions at stake. Hu and Sheu (2013) proposed a linear programming model to minimize reverse logistics costs and corresponding environmental risks for post-disaster debris. On the other hand, several researchers have addressed developing cost models for predicting disaster debris and waste. Soichiro, Akiyama, Tanouchi, Egusa and Otsuka (2021) developed a dynamic hauling/transportation model (DHT model) after describing the issues that contribute to the implementation of disaster waste disposal plans. Sakaguchi, Tanouchi, Egusa and Otsuka (2018) proposed a numerical model simulating waste transportation, temporary storage areas, and final disposal site. The truck numbers in addition to capacities of temporary storage areas are located in most of the proposed models and kinds of literature.

However, although the importance of recent disaster models, the literature on off-site and on-site disaster assessment are far from common factors and complex environmental circumstances yet. Therefore, many disaster response activities may not be able to establish a long-term cooperative official guideline on how to handle the recovery stage with the assistance of the locals.

The linear dynamic transportation model (LDT) was modified after estimating the optimal time (in days) required for disaster debris disposal through the following parameters and factors; (1) debris flow volume in a certain zone area; (2) debris withdrawal volume in the selected zone; (3) debris characteristics, (4) critical route/path for vehicles. The outcome of this model will assist the local governorate with the development of a plan that more fully considers the needs of its plans through comprehend the required number and capacities of curbside pickups/truck vehicles, the ultimate capacity of temporary and secondary waste storage, optimum final disposal site and/or reusing the demolition debris resulting from the building.

Generally, the quality and quantity of debris produced vary according to the type of residence and the type of city. For example, most of the housing sector and materials used in urban cities differ from those in villages. To increase the number of application possibilities, a model specialized for a target city is demanded. In this study, we used the LDT model in Kirkuk, Iraq, as a new case study. This model can be used to compare different cities, particularly those that have been destroyed by wars, through further research. Increasing competition and disputes among countries across the world, especially in unstable political regions, raise concerns about the recovery actions needed by the government and how they would help to rebuild the damaged areas after crises and conflicts.

By implementing the LDT model, a wide range of disaster areas and damage zones will balance the finances, optimize the duration of the removal operations, manage landfill usage, and determine the number of recycled materials generated. Reusing demolition debris, instead of new raw construction materials, is an approach that this study introduced by surveying the optimum transport Path to the new construction sites. Directly and indirectly, the proposed model can support post--disaster management decisions and the challenging task of operating strategic plans for disaster awareness.

Material and methods

Needs for disaster debris model

In the absence of a comprehensive environmental monitoring system, projected population growth will continue to play an important role in producing greater amounts of solid waste and debris. Energy prices, on the one hand, and increased consumption, on the other hand, prompted researchers and institutions to seek new models and systems for converting waste into sustainable energy (Uche-Soria & Rodríguez-Monroy, 2019; Lederer, Gassner, Kleemann & Fellner, 2020). However, these models and systems would not be suitable for all types of waste, especially those resulting from the demolition of buildings such as old houses, schools, and recreational facilities. Current social, political, economic, and environmental issues are very concerning. These issues and concerns necessitate a strategic focus on catastrophe management at a certain time. With the changing fundamentals of energy prices, which are driven by flow and demand and political events, human interactions will likely continue to remain volatile (Qasim, 2021).

Although solid waste collection sites and sanitary landfills can have many health issues, the existing scenarios from the literature are only a part of an integrated management process of debris and solid waste. The challenging issue of debris management is estimating and determining a flexible approach to reuse and recycle waste, as far as the debris, particularly large material which makes it possible to conduct the disaster response recovery techniques in a short period (Chen, Kwan, Li & Chen, 2012; Chen & Zhan, 2017; Estay-Ossandon & Mena-Nieto, 2018). By integrating the mentioned parameters with the influences of post-disaster and pre-disaster factors, the LDT can help city emergency planner to identify the optimum path of debris accumulation and other management activities.

Debris model design

When determining an optimal waste treatment method, the focal point arises in determining the best method of isolating debris from the other solid waste. This is one of the fundamental problems which impact any management practice of waste disposal, dump management, site operations, and rehabilitation or recovery process in a city that tested disaster situations (Wei et al., 2021). Calculating the population density of the target city is indeed one of the key topics that can assist the major course of action (Hirayama et al., 2010). Accordingly, the optimal quantity of debris can be assumed through the building types; either they are single or multiple types, as well as notice the other structure types. Depending on how much debris is produced, it will be easier to come up with an adjustment management strategy (Qasim, 2019).

The linear debris model LDT illustrates disaster waste management; especially in regions that do not have enough information and statistical or historical data. This model will help the decision makers to evaluate a set of alternatives under catastrophic conditions, including (1) the impact of the disaster installation of unsure processes, (2) what analysis they need to design and understand patterns of transport processes, and (3) variety and amount of disaster waste in the target city. By including the selected criteria and conditions, the LDT model will facilitate efficient analysis of integrated waste management through reduction of the redundant and unnecessary processes. Moreover, LDT can be developed to cover not only the cities seen disasters but also other cities in pre-disaster situations.

Boundary condition

For disposal management during/after a catastrophic period, the damaged area will turn into an operations area of management of disposal facilities and storages step by step. To clean the area of the collected debris, or to recover the destroyed area more efficiently and smoothly, the LDT model is based on the mass transfer process and another management process before, through and after the timeline of a disaster as illustrated in Figure 1.

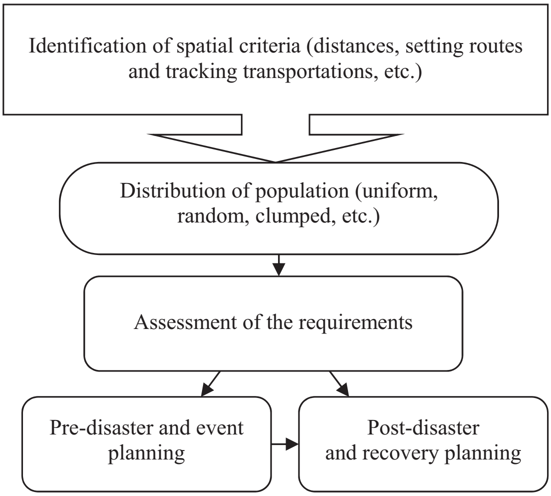


FIGURE 1. Timeline of pre-disaster and post-disaster debris management
Source: own elaboration.

The following sets of actions are designed to determine several potential scenarios as a flowchart of the debris before the disaster hits. Based on the potential disaster scenario, which will determine the forecast of debris amount, labor and equipment requirements and other management facilities such as debris treatment sites, disposal areas, and sorting and recycling processes will be identified. Any non-recyclable waste with other disaster debris generated in the target city will be transported to another collection site.

Structure of the LDT

The model of LDT is inspired by the dynamic transportation model (DHT model) with consideration of pre- and post-disaster factors, which were developed here by Soichiro et al. (2021) and Chen et al. (2012) respectively. With consideration of transport processes patterns through a variety of disaster waste in a target city, the LDT model aims to allocate debris resources among refuse collection places. Through a network of debris discharge and withdrawal paths or nodes, LDT will be able to allocate those resources reasonably.

As an example of the case study, the LDT network was established and illustrated in Kirkuk in Iraq. All the variables about debris volume in various sectors inside the city are determined during a certain period. The number of functions is expressing the relationships between the given variables and the unknown variables.

The mathematical details of the LDT model were based on quantitative environmental outcome variables:

$$D = CWV\left(\sum_{i=1}^{k} DW + \sum_{i=1}^{k} RW + \sum_{i=1}^{k} ES + \sum_{i=1}^{k} CXW\right) - \\ -\sum_{j=1}^{m} TMR - \sum_{J=1}^{M} ComW \quad \forall k, m \in n \tag{1}$$

where: D is the total volume of debris [t], CWV is the collected waste volume in each zone [m^3], DW refers to the total demolition wastes [t·m^{-3}] from buildings such as concrete, iron, ceramics, and brick, RW is the roadwork wastes result from rehabilitation process of the roads and bridges [t·m^{-3}], ES indicates to the excavation soil due to foundation excavation process [t·m^{-3}], CXW is the complex wastes which includes concrete, sand, gravel, gypsum, wood, PVC, glass, metal, plastic and carton [t·m^{-3}], TMR is the total mass of recyclable building materials minus the amount that locals could withdraw for use as construction raw materials [t], $ComW$ assigns the combustible wastes [t], I is the debris discharge path, and J refers to the debris withdrawal path of the total nodes k and m respectively, TMR and $ComW$ relate to solid wastes that can be sorted by the locals at this stage of the analysis.

No one definition of catastrophe waste management applies to all circumstances. As a result, the idea of a disaster has been employed in many different contexts, sometimes as a sign of a failure in urban planning to handle a lot of solid waste in a short length of time outside of ordinary conditions. It is relatively used sometimes as an assessment of human abilities, experiences, development, and cooperation. However, the LDT model evaluates the function of multi various variables to demonstrate a recovery process. By reviewing the relevant literature about Kirkuk, the daily waste generation in 2021 will be around 1,200 t (Mustafa, Mustafa & Mutlag, 2013). So, the concept of disaster and solid waste management has been developed by some researchers, and it's well-known that this concept can be accepted by the community.

However, the total mass of the reusable symbol TMR is developed here to simulate the raw material withdrawn from target areas. Because of the diversity of the materials among urban, suburban, or rural zones, the problem of withdrawing raw materials could become more complex and difficult. To simplify the value of TMR, we can calculate this term according to the per capita percentage of construction materials as shown in Eqs (2) and (3).

$$\sigma = C \cdot \frac{\left[\left(\frac{family}{capita}\right) \cdot HoA\right]}{DTP}, \tag{2}$$

$$TMR = \sigma \cdot HS, \tag{3}$$

where: σ refers to the per capita share of the common building material (according to commonly used materials in the target city), which will turn into debris after the demolition of the structure [$t \cdot m^{-3}$]. Coefficient C depends on the type of construction materials, such as rocks, sand, cement mortar, reinforced concrete, and gypsum, for the composite type. On the other hand, *HoA* assigns the total house or apartment number per family in the target zones and *DTP* refers to the debris transport path [m].

As the majority of the houses in Kirkuk are a composite type, the coefficient C was calculated at around 1.8 $t \cdot capita^{-1} \cdot m^{-2}$. The average housing size per person is *HS* [m^3] occupied by a person within a family which ranges between 54 and 75 m^3 of total residential space.

To indicate the effective size and volume of the debris of each node, at this stage of research, there is a need to highlight characteristic debris composition in various constructions and urban activities. The quantitative portion of each material in the temporary storage sites was set according to Eq. (4):

$$WBca = A \cdot SG \cdot ASH, \tag{4}$$

where: *WBca* is weight-based capacity allocation for each storage zone [t], A is area size [m^2], *SG* is the specific gravity [$t \cdot m^{-3}$], as shown in Tables 1 and 2. Average stacked height (*ASH*) was calculated [m] based on the temporary dumping area in selected four zones. Equation (4) and calculation method were inspired by the author from the quantitative equation which was mentioned by Soichiro et al. (2021).

TABLE 1. Weight-based capacity allocation of solid waste collection

Type of waste	Composition	Ratio [%]	Specific gravity (*SG*) [$t \cdot m^{-3}$]
Excavation soil	soil, rock, clay	7	0.64
Roadwork wastes	broken asphalt, paving stone, concrete	13	1.89
Demolition wastes	concrete, iron, ceramics, brick	69	2.23
Complex wastes	concrete, sand, gravel, gypsum, wood, PVC, glass, metal, plastic, carton	11	5.09

Source: own elaboration.

TABLE 2. Production and share rate of debris

Zone	*C*	Family	Average capita per family	House or apartment number (*HoA*)	Debris transport path (*DTP*)	*ρ*	Housing size per person (*HS*)	Total mass (*TMR*)	Combustible wastes (*ComW*)
1	1.8	74	6	64	900	2	60	95	1 870
2	1.8	48	5	42	600	1	65	79	68
3	1.8	50	5	40	580	1	68	84	270
4	1.8	45	5	38	500	1	70	86	262

Source: own elaboration.

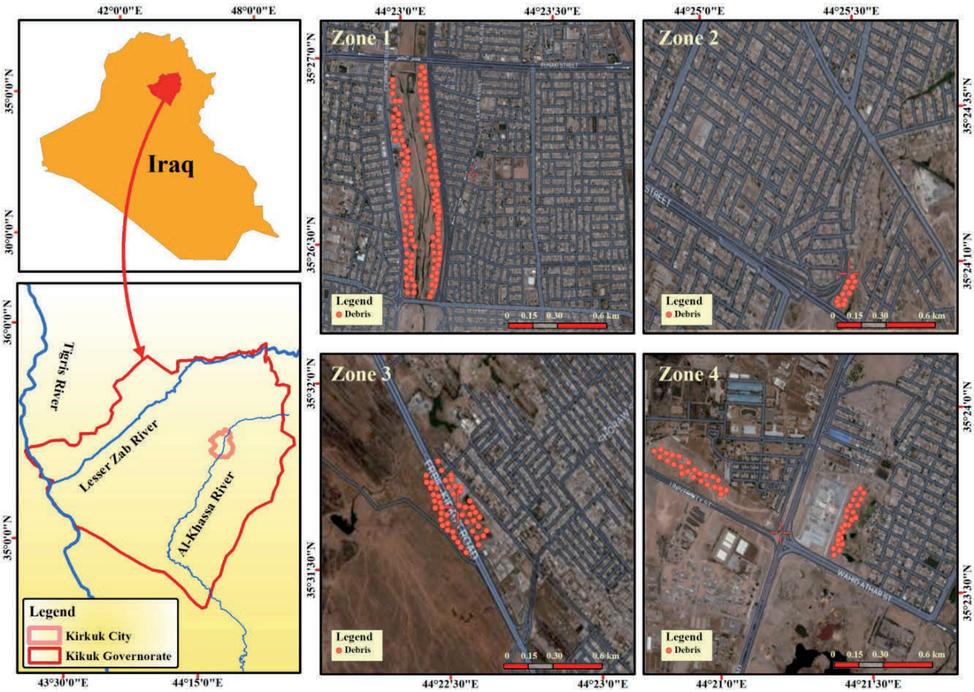


FIGURE 2. Primary debris disposal zones in Kirkuk
Source: own elaboration.

All these four temporary storages were connected to secondary temporary storages according to local conditions and solid waste type. The final disposal site will be the last collection place of all components of the secondary stations. The four selected zones are illustrated in Figure 2. These zones represent the center, east, north, and west of Kirkuk, namely, Zone 1, Zone 2, Zone 3, and Zone 4 respectively.

Model network applications

The LDT model contains 35 debris discharge (Dn) and 17 debris withdrawal (Wn) paths for the four target zones in Kirkuk. To learn more about how the paths are distributed geographically, the typical allocations of debris discharge and withdrawal paths are illustrated schematically in Figure 3.

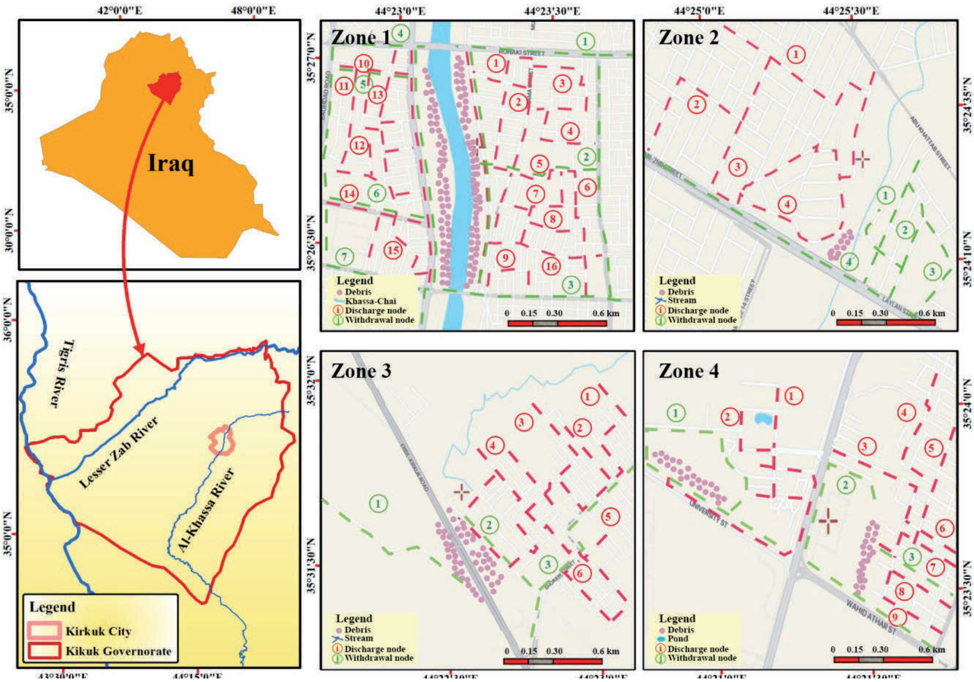


FIGURE 3. Network structure of the linear dynamic transportation model (LDT) in Kirkuk
Source: own elaboration.

To make the outcomes of LDT more reliable, the model should be calibrated and verified to a satisfactory accuracy. The model was developed by estimating several parameters as model inputs. So, the model details are significantly improved in the actual places designated for throwing debris.

The LDT model examines the relationships between projected discharge and withdrawal paths during two time periods of around four months each period. The rainy season lasts from December to April, with the dry season beginning in May and ending in November. Because of the extent of the solid waste storage location, the model can simulate the quantity of debris for the given periods, especially during the summer season.

Considering various types of housing and socioeconomic categories that live in the city, waste disposal will depend on the richness factors of the residents and the improvement of their daily income, in addition to the development programs of the city. These factors motivate people to demolish their old buildings and rebuild new ones. However, as a result of their poverty, the people who live in the surrounding zones are sometimes forced to reuse the collected debris as raw construction material in their homes. At times, debris removal and infrastructure reconstruction may depend on increased awareness of the environment by residents and decision-makers.

It is clear, the dominant portion of debris reuse effectiveness sensibly depends on the suitable collection strategy. While the efficiency of reusing the debris in the mixing zone, as when mixing with solid household waste, will be minimal. Therefore, early in the collecting process, combining municipal solid wastes with debris must be avoided. Consequently, to find out the accurate figure of how debris compositions are affecting the LDT model at each city, we need to analyze this in more detail.

Within the mentioned zones, the Statistical Package for Social Science (SPSS 22.0) was used to analyze the differences between pure debris sites and mixed debris. Thus, the indication of the total amount of debris, across different sections of other cities and zones, could hypothetically be equivalent to D in the LDT model.

The study city

The research was conducted in Kirkuk, Iraq's northeastern city. The Kirkuk holds the majority of the Kirkuk governorate's population as well as its government offices. It lies within latitude 35°21.010′–35°35.103′ N and longitude 44°18.107′–44°39.045′ E (Figs 2 and 3). The governorate of Kirkuk is divided into sub-administrative divisions known as cities or districts called Daquq, Dibis, Hawija, and Kirkuk (Qasim, 2021). Most of these cities and districts have witnessed a rise in terrorist attacks such as ISIS after 2014. The war against these terrorist groups left behind a large number of destroyed buildings and a large amount of debris.

The study area is characterized by complex topography, where small plateau ranges are the general landscape of the area. The seasonal river (Khassa-Chai) divides Kirkuk into two main parts. Despite Khassa-Chai playing a significant function for the residents of Kirkuk, the river coast became a free landfill for garbage and solid waste (Qasim, 2019). Both the local Kirkuk authority and a sizable number of people are finding it difficult to deal with this unregulated waste disposal. It is hard to isolate debris from other solid waste in the mixing zones. Most solid waste dump sites are near the debris disposal place. Furthermore, although reusing the debris as raw building materials for construction provides an excellent opportunity for residents to earn some extra money, a considerable portion of the debris ends up as useless solid waste. As a general case in Iraq, the lack of new and modern types of ways in Kirkuk is another reason for the accumulation of debris in certain places in the mentioned zones. So, in terms of disaster conditions, either due to natural disaster situations or due to terrorist attacks, Kirkuk needs to increase its efforts in debris and solid management. Thus, LDT will assist the decision-makers in Kirkuk to determine a suitable management technique as a key to the rehabilitation process during and after a disaster.

An active management approach necessitates an appropriate model to handle waste generated by disasters in a way that is both safe and environmentally sustainable. The capacity of sustainable development involves not only using proper disposal systems but also using various dynamic models. However, with ongoing political conflict and wars in Ukraine, Afghanistan, Syria, etc., in addition to the asymmetries resulting from population expansion, industrial development, and infrastructure, consequences of economic competition are highlighting the importance of sustainable debris management. To promote sustainable growth and development, transport services can create a reliable framework for human resources management and program activities.

The LDT model was applied successfully to four sites in Iraq, as shown by the figures and tables. The results obtained systematically demonstrate the model's flexibility in various circumstances and environments. However, there is a need to develop this model by increasing its application to include other unstable countries or regions, such as recent earthquake zones in Turkey and Syria, as well as war-torn countries like Ukraine, Syria, Libya, Afghanistan, etc. Thus, the proposed model presents a new research method that can be developed through prospective and new studies.

Results and discussion

In the simulation of the LDT model using the SPSS program, the observed parameters listed in this model were specified for four zones in Kirkuk. The findings show that Zone 2, Zone 3, and Zone 4 are appointed as critical lack sites for reusable materials as shown in Figure 4. Currently, due to the lack of suitable recycling facilities in the city, the large amount of waste in these zones is considered undesirable. While Zone 1 was noted as a good area for debris disposal based on the pure debris volume. Although the critical issues of collecting and/or disposing of the debris near the seasonal river Khassa-Chai, this zone offers unpolluted sites for debris.

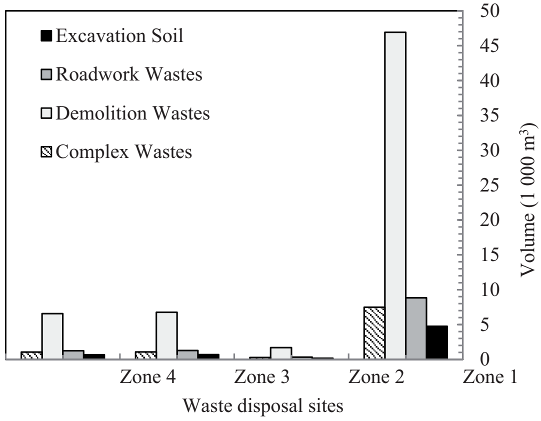


FIGURE 4. Volumetric analysis of the debris

Source: own elaboration.

After calculation of the proportions, the debris withdrawals ratio from Zone 2 was greater than other zones as shown in Table 3. Although the collected quantity of debris in Zone 1 is much greater than in Zone 2, the inefficiency of the road, as well as the truck path in Zone 1, led to a decrease in the withdrawal ratio there (Figs 5 and 6). On the other hand, the urban development process that took place in Zone 2 paid regard to the road type and networks.

TABLE 3. Withdrawal ratio of debris

Zone	Excavation soil [t]	Roadwork wastes [t]	Demolition wastes [t]	Complex wastes [t]	Total discharge [t]	Total withdrawal [t]	Withdrawal ratio [%]
1	3 189	5 923	31 436	5 012	45 560	1 965	4.31
2	115	214	1 137	182	1 648	146	8.86
3	460	854	4 531	722	6 566	354	5.39
4	447	829	4 401	701	6 378	348	5.46

Source: own elaboration.

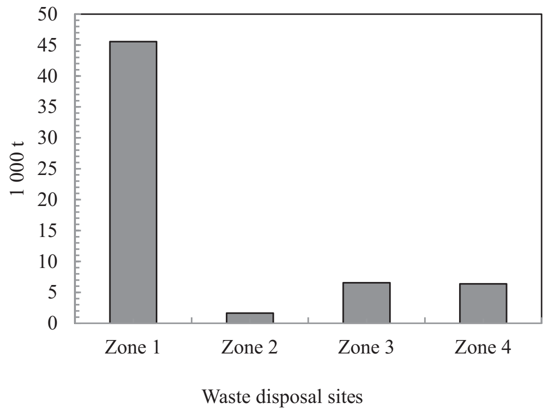


FIGURE 5. Total discharge of debris

Source: own elaboration.

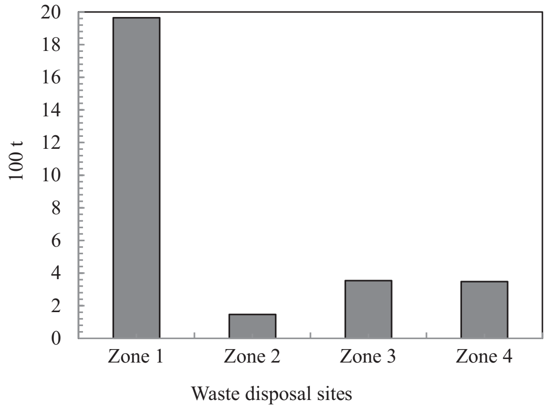


FIGURE 6. Total withdrawal of debris

Source: own elaboration.

The outcomes of the LDT show that; the transportation path of debris trucks is responsible for promoting sustainable and reliable management of debris. However, the importance of debris management does not receive the proper attention from the decision-makers yet. Thus, increasing awareness of both debris discharge and withdrawal is another significant benefit of the LDT model.

The results demonstrate that, in this application, the proposed model has several important advantages; as: LDT model was used within a controlled environment in terms of the limited variables that affect the financial status of individuals; It helped in determining the required cause and effect relationships; It helped in providing more reliable results within the target area.

On the other hand, as with all new models, certain defects have emerged during the application of the model, including; the total size of collected samples was small in comparison to the large size of the debris; Concerning the reuse of building materials, the dependence on participants' awareness and compliance was an obstacle to the increasing application of LDT model; Apart from these long-term experiences, it is difficult to maintain contact with participants; In addition, the sorting and transportation process in some regions can be costly.

Conclusions

Sustainable debris management by using the LDT model presents a rock-solid basis upon which the unexpected amount of solid waste can be effectively estimated and treated. The presence of an adequate truck path encourages the locals to reuse the debris for earning more money. As well as it provides additional assistance to implement sustainable goals during infrastructure development or disaster risk management.

The results reflect the relationship between the design of truck paths along the buildings and the ability to remove debris locally. Because of good urban design that has taken place in Zone 2, the quantity of debris reused in Zone 2 was more than in Zone 1. Although the amount of debris in Zone 1 was much greater than in Zone 2, the existence of a proper path that trucks follow in transporting the debris encouraged the locals to reuse the debris. At Zone 2, the withdrawal paths were the same number as the discharge paths. While the discharge path numbers, in the other zones were more than the withdrawal paths. So, this subject helps to conclude a kind of balance between the total amount of debris discharged and the amount withdrawn in each zone. Thus, the results are seen in the rise of debris withdrawals in Zone 1 for re-use by its individuals. Despite satisfactory results, some adverse side effects may be caused by the mixture of household waste and debris. This was one of the complex challenges associated with the safe reuse of debris.

Acknowledgments

The researcher gratefully acknowledges Kirkuk University's Civil Engineering Department – Faculty of Engineering and Environmental Research Unit – Faculty of Science for their guidance and cooperation provided during the activities of research.

References

Chen, X., Kwan, M. P., Li, Q. & Chen, J. (2012). A model for evacuation risk assessment with consideration of pre-and post-disaster factors. *Computers, environment and urban systems*, *36* (3), 207–217. https://doi.org/10.1016/j.compenvurbsys.2011.11.002

Chen, X. & Zhan, F. B. (2017). Agent-based modelling and simulation of urban evacuation: relative effectiveness of simultaneous and staged evacuation strategies. *Journal of the Operational Research Society*, *59* (1), 25–33. https://doi.org/10.1057/palgrave.jors.2602321

Estay-Ossandon, C. & Mena-Nieto, A. (2018). Modelling the driving forces of the municipal solid waste generation in touristic islands: A case study of the Balearic Islands (2000–2030). *Waste Management*, *75*, 70–81. https://doi.org/10.1016/j.wasman.2017.12.029

Hirayama, N., Shimaoka, T., Fujiwara, T., Okayama, T. & Kawata, Y. (2010). Establishment of disaster debris management based on quantitative estimation using natural hazard maps. *Waste Management and Environment*, *140*, 167–178.

Hu, Z. H. & Sheu, J. B. (2013). Post-disaster debris reverse logistics management under psychological cost minimization. *Journal of Transportation Research Part B: Methodologica*, *55*, 118–141. https://doi.org/10.1016/j.trb.2013.05.010

Lederer, J., Gassner, A., Kleemann, F. & Fellner, J. (2020). Potentials for a circular economy of mineral construction materials and demolition waste in urban areas: a case study from Vienna. *Resources, Conservation and Recycling*, *161*, 104942. https://doi.org/10.1016/j.resconrec.2020.104942

Lorca, Á., Çelik, M., Ergun, Ö. & Keskinocak, P. (2015). A decision-support tool for post-disaster debris operations. *Procedia Engineering*, *107*, 154–167. https://doi.org/10.1016/j.proeng.2015.06.069

Mustafa, S. S., Mustafa, S. S. & Mutlag, A. H. (2013). Kirkuk municipal waste to electrical energy. *International Journal of Electrical Power & Energy Systems*, *44* (1), 506–513. https://doi.org/10.1016/j.ijepes.2012.07.053

Qasim, A. H. (2019). International water model under productivity conditions: the case of the Tigris and the Euphrates. *GEOMATE Journal*, *17* (62), 29–34.

Qasim, A. H. (2021). Causes and side effects of changing water quality in Khassa-Chai river in Kirkuk, Iraq. *Scientific Review Engineering and Environmental Sciences (SREES)*, *30* (2), 271–282. https://doi.org/10.22630/PNIKS.2021.30.2.23

Sakaguchi, N., Tanouchi, H., Egusa, N. & Otsuka, Y. (2018). Development of a Disaster Waste Disposal Planning Model Aiming at Supporting of Appropriate Decision of Disaster Waste Treatment. *Journal of Japan Society of Civil Engineers, Ser. G (Environmental Research)*, *74* (5), 195–202.

Soichiro, A. S. A. I., Akiyama, T., Tanouchi, H. & Egusa, N. (2021). A numerical simulation of disaster waste disposal in Wakayama city by using DHT model. *GEOMATE Journal*, *20* (80), 23–28. https://doi.org/10.21660/2021.80.6155

Uche-Soria, M. & Rodríguez-Monroy, C. (2019). An efficient waste-to-energy model in isolated environments. Case study: La Gomera (Canary Islands). *Sustainability*, *11* (11), 3198. https://doi.org/10.3390/su11113198

Wei, L., Hu, K. & Liu, J. (2021). Quantitative analysis of the debris flow societal risk to people inside buildings at different times: a case study of Luomo village, Sichuan, southwest China. *Frontiers in Earth Science*, *8*, 627070. https://doi.org/10.3389/feart.2020.627070

Summary

Sustainable debris management by linear dynamic transportation model. Management of large debris caused by building demolition necessitates a multi-faceted approach to deal with emerging side effects. Because of emerging global challenges, such as population growth, and renovation projects, a dynamic models need to be planned and controlled. One of the key drivers of this management is determining the appropriate path for transporting waste and debris. Debris management by using the linear dynamic transportation model (LDT) is conducted to deal with the unexpected amount of debris and other solid waste. This sud-

den and unexpected large amount of solid waste might be produced by natural disasters or by man-made catastrophes either directly or indirectly. By computing several parameters in certain zones, a sensitivity analysis of each parameter is performed to obtain an optimal model for disaster debris management. Based on disaster debris volume, the model gave us an optimal explanation of the debris disposal by locals. According to the estimated parameters and conditions, significant findings appear by identifying the optimal dynamic transportation path of the debris truck. Thus, by applying the LDT model, the results showed that the efficiency/inefficiency of road types and networks clearly affect the handling of debris.

Scientific Review Engineering and Environmental Sciences (2024), 33 (1), 33–52

Sci. Rev. Eng. Env. Sci. (2023), 33 (1) https://srees.sggw.edu.pl

ISSN 1732-9353 (suspended) eISSN 2543-7496 DOI 10.22630/srees.5625

Received: 08.11.2023
Accepted: 16.01.2024

Aneta OBLOUKOVÁ
Eva VÍTKOVÁ✉
Brno University of Technology, Faculty of Civil Engineering, Czech Republic

Cost analysis of water charge rates in the Czech Republic – Case study

Keywords: water charge rates, calculation formula, water management companies, regions of the Czech Republic, vertical analysis, horizontal analysis

Introduction

The issue of water is becoming increasingly important worldwide and is currently a broadly discussed topic. A significant loss of water is a result of global warming and lack of rainfall. Groundwater depletion and drying up of wells occur in the Czech Republic yearly. It is essential to manage water resources economically, protect them systematically and prevent their pollution. It is obvious that a shortage of goods or services on the market reflects their increase in the price. The analysis of wastewater management in small and medium-sized textile (batik) enterprises in Pekalongan City was addressed by Dwidayati (2020). Information publication and data provision are another big issues for individual companies. Water information provision, which is a carrier for corporate information on water resource management, is a means of expressing corporate environmental responsibility. Therefore, He, Shen, Xu, Sun and Wang (2023) focused on companies and their water information provision.

However, the price of water in the Czech Republic is not only driven by its deficiencies due to global warming. Drinking water and water diverted through sewers

✉Eva.Vitkova1@vut.cz

is included in the list of goods with regulated prices, which is issued annually by the Ministry of Finance of the Czech Republic under the Price Act and is available in the Price Bulletin. In her article, Kejser (2016) explores the spatial heterogeneity in the public attitude towards internalizing environmental and resource costs in the price of water across the EU regions.

There are different rules for water pricing around the world. Since water is a state issue in India, for example, there are huge differences in the pricing of irrigation water from state to state. Parween, Kumari and Singh (2021) review the structures of the water pricing mechanism in different states of India and suggest a way to achieve sustainable water resources management in India. Ashoori, Dzombak and Small (2017) discuss the determination of water price and population criteria for meeting future urban water demand targets. The research concluded that water demand in Los Angeles is expected to increase by 36% between 2014 and 2025 due to climate change. It also predicted that future residential water demand in Los Angeles will be largely driven by price and population, rather than climate change and savings. The same approach applies to the pricing level of water and sewerage charge rates, which represent the provision of public service in the sense of operating water and sewer systems.

The price of water has been constantly rising according to the internet server eAGRI.cz and the Czech Statistical Office (Český statistický úřad). While in 2010 the price of drinking water was 29.10 $CZK \cdot m^{-3}$ without value added tax (VAT), in 2020 the price of drinking water rose to 41.40 $CZK \cdot m^{-3}$ without VAT. Even the year 2022 has been experiencing a considerable increase in the price of water and sewerage charge rates. Water companies attribute this to the rising costs and the need to invest in infrastructure. The coronavirus pandemic has also had an impact on water price calculations since 2020. The war in Ukraine at the beginning of 2022 has triggered a new wave of a rapid increase in the price of products and services, so the increase in prices has affected and will continue to affect all sectors of the economy in the future. This concludes that the war in Ukraine has an impact on the price of water. At present and in the given situation, it is very difficult to estimate the development of water and sewerage charge rate prices in the Czech Republic in the future. However, some mathematical models can predict their course under appropriately set calculation conditions. Oblouková (2023) deals with the issue of water and sewerage charge rate prices in the Czech Republic. The aim of the paper is to use the obtained data to predict the development of average water and sewerage charge rate prices in the Czech Republic in the years 2022–2026 using the chosen mathematical method. The most appropriate method for this research is the linear trend technique. This method has been pre-tested for relevance and suitability in previous research. However, it must be said that the chosen method does not take into account any external macroeconomic influences. It does not even consider the huge increase in inflation since February 2022, when the rapid price increases caused by the war in Ukraine began. An increase in the inflation rate is considered only at the end of the calculation, where the effect of inflation on the water and sewage charge rate prices is shown. Volf, Sušanj Čule, Žic and Zorko (2022) indicate in their studies the water quality index prediction for improvement of treatment processes in a drinking water treatment plant. Therefore, obtained models can help in the optimization of treatment processes, which depend on the quality of raw water, and overall, on the sustainability of the treatment plant.

It is often stated in the scientific literature that the Czech Republic lies in the heart of Europe as it is located approximately in the centre of Europe. The Czech Republic is not a large country. According to the Czech Statistical Office (www.czso.cz), the population as of 31 December 2022 was 10.828 million. Since 2021 there has been an increase of 3.1 million inhabitants. This increase is due to the massive immigration wave related to the armed conflict in Ukraine. Except for the decline in 2013 (by 0.4 million people), the population has grown each year in the past 10-year period (between the beginning of 2013 and the end of 2022). The Czech Republic is divided into 14 regions – the Capital City of Prague, the Central Bohemian Region, the South Bohemian Region, the Plzeň Region, the Karlovy Vary Region, the Ústí nad Labem Region, the Liberec Region, the Hradec Králové Region, the Pardubice Region, the Vysočina Region, the South Moravian Region, the Olomouc Region, the Zlín Region, and the Moravian-Silesian Region. Individual regions have different performance levels. The performance of individual regions is measured by several factors, the main ones being macroeconomic, including gross domestic product, unemployment, and others.

The eagri.cz website (www.eagri.cz) lists the price of water and sewerage charge rates nationwide and for individual regions. As water and sewerage operators transformed from state-owned companies to mostly joint stock companies in the 1990s, i.e. capital companies that aim to satisfy their shareholders through dividend payments, it is not always the case that the well-performing region has the highest water and sewerage charge rates. Therefore, Vítková, Vaňková and Oblouková (2022) focused both on the performance of the regions and the performance of the water and sewerage management companies that run them. Oblouková and Vítková (2023) in another article examined the development of water and sewerage charge rates in the Czech Republic over the last 14 years and pointed out the difference in the percentage ratio between the components of water and sewerage charge rates in the regions in comparison to the national average values. Liu, Wu, Xu and Pan (2018) examined the relationship between wastewater discharge, river water quality in the Pearl River Delta, and gross domestic product per capita. They used a logarithmic mean Divisia index (LMDI) decomposition model as well as an environmental Kuznets curve (EKC) model for their research. Maziotis, Saal, Thanassoulis and Molinos-Senante (2014) investigated changes in profit, productivity and price performance in the water and sewerage industry: an empirical application for England and Wales. This study analysed the impact of regulation on the financial performance of water and sewerage companies in England and Wales over the 1991–2008 period. In another paper, Molinos-Senante and Maziotis (2021) examined productivity growth, economies of scale and scope in the water and sewerage industry: the Chilean case, in another article. In this paper, they focused on the performance evaluation and cost driver analysis of water management companies, where they used quadratic cost functions to investigate the existence of savings in the Chilean water and sewerage industry during the 2010–2017 period.

The article deals with the issue of the price of water charge rates. The article is focused only on one aspect of the research, namely on the cost analysis of water charge rates in the Czech Republic. The sample was selected to include individual groups of company sizes, as well as cities or municipalities as public administrators. When selecting the samples, the criterion was observed that the subject should operate in a certain area, i.e. all regions of the Czech Republic should be included in the analysis.

In environmental projects, decision-making can be a complex and challenging task due to the in-built existence of compromises between environmental, socio-political, and economic factors. Jajac, Marović, Rogulj and Kilić (2019) explore a systematic approach to developing a decision support concept that includes the analysis of wastewater treatment problems, knowledge acquisition, and the identification and evaluation of criteria that bring forth an optimal solution to the location selection of wastewater treatment plants (WWTPs).

The environment is one of the most discussed topics today. It permeates all sectors of industry and services. Water management is a separate chapter. Water companies increase the price of water every year. However, each business may be affected by other factors that impact the price of water. The research described in the article aims to assess or challenge the conclusion of whether any dependence can be found between the size of the company within the representation of each price cost of water charge rates. The paper focuses specifically on water charge rates as it concerns more than 96% of the population of the Czech Republic. Sewerage charge rates have a lower percentage representation due to non-connection to the sewerage system, e.g. due to a septic tank for the house. Many variables enter the cost price of water management companies. It is known which cost components are included in the price of water, but not the average values or the percentage representation of the individual components. The result of the research is finding out which costs have the greatest influence on the price of water.

The article represents only an initial insight into the issue. The total number of samples assumed in the analysis was in the order of 100 operators in the Czech Republic, which amounts to working with 400 pieces of calculation samples, thus analysing more than 26,000 pieces of data. As the criteria for the selection of operators were set for the case study, the final number of samples was narrowed down to 14 representative operators. The article is not focused on the statistical evaluation of the analyzed data, it is an analysis of the specific costs shown in the price of water and the possible finding of connections between the amount of specific costs and the size and type of companies.

Material and methods

The Ministry of Finance is the price regulator in the water supply and sewerage sector, which regularly rearranges the price regulation settings, and the Ministry of Agriculture is the substantive regulator. The Act No. 274/2001 Coll., on water supply and sewerage for public use, as amended, regulates certain relations arising in the development, construction and operation of water supply and sewerage systems intended for public consumption in the Czech Republic. A Decree of the Ministry of Agriculture implementing Act No. 274/2001 Coll., on water supply and sewerage for public use as amended was also issued for the purpose, which is Decree No 428/2001 Coll. The primary source for determining the development of average water charge rates in the Czech Republic was the Czech Statistical Office. This is the central government body of the Czech Republic. It was established on 8 January 1969 by Act No. 2/1969 Coll., on the establishment of ministries and other central bodies of state administration. The Water supply and sewerage act as well as the price regulation use a unit price in Czech crown per cubic meter ($CZK \cdot m^{-3}$).

The case study, which deals with the cost analysis of water charge rates in the Czech Republic, focuses on the operators of water supply and sewerage systems in the Czech Republic. All data obtained from the operators in the Czech Republic was narrowed down to the time horizon of the past four years, i.e. 2018–2021, due to the complexity of the available information. All formulas and calculations were made using Microsoft Excel software.

The article first examines the issue of the price level of the water rate development in the Czech Republic as a whole and subsequently in the individual regions of the country. For each region, one representative – a water management company, which usually has only one specific supply point – was selected for the cost analysis of water charge rates, where the percentage interval of each type of cost entering the water charge price was set. Large and medium-sized companies, small and micro companies as well as specific municipalities or cities were chosen as representatives.

The following factors were observed in the selection of the sample operators:
- selection of one operator per region of the Czech Republic;
- the selection of the operator had to fulfil the condition of only one location of supply (or maximum two) within the given region of the Czech Republic;
- selection of the operator based on the above-stated points must not have interfered with other regions of the Czech Republic;
- selection of the operator was limited by the size of the company (micro, small, medium-sized, and large entities) in a 3 : 3 : 3 : 3 ratio;
- selection of the operator was limited by the representation of at least one representative of the public sector, i.e. a city or a municipality (two such representatives were considered in the case study).

To comply with the above-mentioned factors, it was necessary to work first with about 100 operators, i.e. work with 100 calculation formulas each year, i.e. a total of 400 calculation formulas (the years under consideration were 2018–2021). In total, 26,000 values were worked with to select suitable representatives of water management companies. The shortlist needed to study approximately 50 annual reports and financial statements from 2021 to establish values for number of employees, total assets, and sales to identify and select the correct size of water management companies. Following a complex and time-consuming selection process, 14 selected operators as well as representatives from each region have already been worked with and 1,736 values have been processed. Furthermore, 1,736 values were determined by vertical analysis with their assistance. These values have already been used in the analysis to establish whether there is a relationship between the percentage of costs and the size of the company in the pricing of water charge rates.

The vertical analysis is the basic methodology chosen for the cost analysis of the water charge rates of the selected Czech operators. It is one of the basic methods of financial analysis. Financial analysis is used to assess the financial performance of an entity and to evaluate the financial position of that entity. It works with data obtained from the financial statements. The outputs are primarily used for tactical and strategic investment and financing decisions and for reporting to owners, creditors, and other stakeholders. The principle of the analysis is the calculation of indicators (elementary, ratio, aggregate) that have good explanatory power concerning the economic reality under study. Elementary indicators include horizontal and vertical analysis. Vertical analysis describes the representation of individual items related to the whole. It is expressed as an absolute number or as a percentage. The second elementary indicator is the horizontal analysis, which describes the annual change of items in absolute terms or as a percentage and expresses how the items have changed compared with the previous year. In other words, it is a horizontal, i.e. a line-by-line comparison of the absolute or relative changes in the items of a given statement over time. It is a useful indicator for determining the annula percentage change (increase or decrease) in the average price of water and sewerage charge rates in the Czech Republic. The horizontal analysis is calculated according to the following relation:

$$horizontal\ analysis = \frac{indicator_t - indicator_{t-1}}{indicator_{t-1}} \cdot 100, \tag{1}$$

where *horizontal analysis* is the year-on-year change of items [%], $indicator_t$ is the price of water charge rates in the year under review [CZK·m^{-3}] and $indicator_{t-1}$ is the price of water charge rates in the previous year [CZK·m^{-3}].

A vertical analysis is suitable for the research dealt with in this paper, i.e. the cost analysis of water charge rates of selected operators in the Czech Republic. It is calculated according to the relation:

$$vertical\ analysis = \frac{indicator_i}{indicator_x + indicator_y} \cdot 100, \tag{2}$$

where *vertical analysis* is the representation of individual items related to the whole [%], $indicator_i$ is the individual cost component in the year under review [CZK], $indicator_x$ is the full own cost for calculating water charge rates in the year under review [CZK] and $indicator_y$ is the calculation profit of the year under review [CZK].

The arithmetic mean was used to determine the average values:

$$\overline{x} = \frac{1}{n} \sum_{n=1}^{n} x_i, \tag{3}$$

where $\overline{x}$ is the average values [CZK], x_i is the sum of all values to determine the average [CZK] and n is the number of values to be summed [-].

Vertical analysis is used in this paper to calculate the percentage representation of the different cost components in the pricing of water charge rates. Horizontal analysis is used to determine the evolution of water charges over the period under consideration.

The mathematical median function was also used in the Microsoft Excel calculations. The median is a value that divides a series of results into two halves. Thus, at least 50% of the values are less than or equal to the median and at least 50% of the values are greater than or equal to the median. This method is useful in calculating the representation of the main cost items in the calculation formula.

The Czech Republic is divided into three areas Bohemia, Moravia and Silesia, which are divided into 14 regions – the capital City of Prague, Central Bohemia Region, South Bohemia Region, Plzeň Region, Karlovy Vary Region, Ústí nad Labem Region, Liberec Region, Hradec Králové Region, Pardubice Region, Vysočina Region, South Moravia Region, Olomouc Region, Zlín Region, and Moravia-Silesia Region (Fig. 1).

FIGURE 1. Regions of the Czech Republic
Source: own elaboration.

Table 1 presents the price level of the average water charge rates of the Czech Republic for the 2018–2021 period. The price increases each year by an average of 4.76%. In 2019, the price increased by 3.15% compared to the previous period, while in 2021 the largest increase was 5.80%. Then, the development of average water charge rate prices in individual regions of the Czech Republic is presented.

TABLE 1. Average prices of water charge rates in the Czech Republic and in the regions of the Czech Republic without VAT in 2018–2021

Specification	2018	2019	2020	2021
	CZK·m^{-3}			
Czech Republic	38.10	39.30	41.40	43.80
Capital City of Prague Region	42.00	42.70	45.20	49.90
Central Bohemia Region	40.90	41.90	45.10	47.10
South Bohemia Region	37.00	37.60	39.20	40.50
Plzeň Region	39.20	40.80	43.60	46.10
Karlovy Vary Region	38.20	40.10	41.90	43.60
Ústí nad Labem Region	43.40	44.20	46.00	48.80
Liberec Region	44.20	44.80	45.60	48.10
Hradec Králové Region	34.90	36.10	37.40	39.00
Pardubice Region	34.10	35.90	37.40	39.40
Vysočina Region	37.10	38.50	40.10	41.80
South Moravia Region	34.50	36.10	39.40	41.60
Olomouc Region	33.10	34.10	35.80	37.10
Zlín Region	35.80	37.20	38.30	40.10
Moravia – Silesia Region	34.40	36.20	37.90	39.90

Note: 1 EUR = 25.00 CZK.

Source: own elaboration based on www.eagri.cz, www.czso.cz.

Table 2 then shows the annual percentage increase in average water charge rates in the regions of the Czech Republic over the period under review.

TABLE 2. Horizontal analysis: year-on-year change in the average water charge rates in the Czech Republic in the period of in 2018–2021

Region	2018/2019	2019/2020	2020/2021	Average
	%			
Capital city of Prague	1.67	5.85	10.40	5.97
Central Bohemia	2.44	7.64	4.43	4.84
South Bohemia	1.62	4.26	3.32	3.06
Plzeň	4.08	6.86	5.73	5.56
Karlovy Vary	4.97	4.49	4.06	4.51
Ústí nad Labem	1.84	4.07	6.09	4.00
Liberec	1.36	1.79	5.48	2.88
Hradec Králové	3.44	3.60	4.28	3.77
Pardubice	5.28	4.18	5.35	4.93
Vysočina	3.77	4.16	4.24	4.06
South Moravia	4.64	9.14	5.58	6.45
Olomouc	3.02	4.99	3.63	3.88
Zlín	3.91	2.96	4.70	3.86
Moravia – Silesia	5.23	4.70	5.28	5.07

Source: own elaboration.

Representatives of individual regions were selected for the purpose of cost analysis of water management companies in the Czech Republic. These representatives of water management companies in each region were analysed over a period of four years, i.e. 2018–2021. According to the Accounting Act No. 563/1991 Coll., it is possible to categorize entities (companies) according to their size. Following this law, companies are divided according to the number of employees, the amount of assets and the number of sales:

- micro companies;
- small companies;
- medium-size companies;
- large companies.

The size of the selected water management companies was verified on the website Justice.cz, where the Ministry of Justice of the Czech Republic presents legislation that affects the life of every citizen in the areas of civil, commercial, criminal, and procedural law. The Ministry of Justice is also the source of the regulations governing professions such as judges, lawyers, notaries, and executors. On this web portal, among other things, each company is required to publish its annual final accounts, where the information on assets and sales can be found.

The selection of operators was very difficult due to the set of variety factors mentioned above. It is typical for some operators to supply water to multiple locations, i.e. multiple regions. Another difficulty was in determining the water charge rate within the setting of the calculation formula, i.e. publishing this data for four years under review, 2018–2021. Information on the costs of individual water management companies was obtained from the eagri.cz website. According to the provisions of Section 5(3) of the Act on Water Supply and Sewerage, owners of water supply and sewerage systems are obliged to submit annually the selected data from the Property Register of Water Supply and Sewerage Systems, the so-called VÚME (*Vybrané údaje majetkové evidence*), and the selected data from the Operational Register of Water Supply and Sewerage Systems, the so-called VÚPE (*Wybrané údaje provozní evidence*), to the water authorities in whose territorial jurisdiction the property is located. It should be said that the VÚME and VÚPE data provide data beyond the scope of the CSO survey. The main reason for setting the 2018–2021 reporting period was the fact that electronically traceable data on water and sewerage charge rates have only been recorded from 2018 onwards. Table 3 shows the selected representatives/operators for each region of the Czech Republic and the distribution of the selected water supply and sewerage operators according to the size of the company – a summary of the operators and whether they are large or small companies. The assessment is based on the 2021 financial statements.

TABLE 3. Selected representatives of water supply and sewerage operators in the Czech Republic and the size of their accounting units

Region	Operator	Water supply location	Type of company
Capital city of Prague	Letiště Praha, a.s.	Letiště Praha	large account unit
Central Bohemia	Středočeské vodárny, a.s.	Kladno – Mělník	large account unit
South Bohemia	Městská Vodohospodářská s.r.o.	Třeboň	small account unit
Plzeň	Vodovody a kanalizace města Kdyně spol. s.r.o.	Kdyně	micro account unit
Karlovy Vary	CHEVAK Cheb, a.s.	Chebsko – regional price	medium-size account unit
Ústí nad Labem	Ludvíkovice municipality	Ludvíkovice	municipality
Liberec	Zásadská vodárenská společnost, s.r.o.	Zásada	micro account unit
Hradec Králové	VODA-RA s.r.o.	Radvanice	micro account unit
Pardubice	Vodovody a kanalizace Pardubice, a.s.	Pardubice	large account unit
Vysočina	Technické služby Třešť, spol. s r.o.	Třešť	small account unit
South Moravia	Vodovody a kanalizace Vyškov, a.s.	Vyškov	medium-size account unit
Olomouc	Town of Štíty	Štíty, Heroltice	city
Zlín	Slovácké vodárny a kanalizace, a.s.	Uherskohradišťsko, Uherskobrodsko – regional price	medium-size account unit
Moravia – Silesia	Vodovody a kanalizace Hlučín, s.r.o.	Hlučín	small account unit

Source: own elaboration based on www.eagri.cz, www.justice.cz.

Results and discussion

As already mentioned above, prices for water and sewerage fall among the prices that are subject to substantive regulation, i.e. prices for which the method of determination is regulated by law and other legal regulations. These are pursuant to the provisions of Sections 8 and 20 of Act No. 274/2001 Coll., on water supply and sewerage for public use, as amended, and Act No. 526/1990 Coll., on prices, as amended. The method of determination in this case means the binding procedure for calculating the price and calculating the amount of reasonable profit included in the price. Nearly 11 million people in the Czech Republic are supplied with water. The Czech Republic is made up of 14 regions, each of which differs in its performance.

Based on the above-stated data for the cost analysis of the Czech Republic's water rates included in the research, the following conclusions were drawn. It is necessary to point out that the research is still in its early stages, i.e. the input data, on the basis of which the analyzes and conclusions on this issue are compiled, can be considered as so-called pilot (initial) research, which provides insight into the given issue. All the outputs given below have been developed based on each step namely data collection, data processing and data evaluation using Microsoft Excel.

After conducting a complex analysis of the operators of the Czech Republic, finding representatives of the operators of each region of the Czech Republic so that they meet the predefined condition that the selected operator must supply water to only one location (maximum two), an analysis was carried out in which the size of the entity/company was traced to the selected operator. In this case, the selected operators were classified according to the criteria described above into large companies, medium-sized companies, small companies, micro companies, cities, and municipalities (Table 3). The next step led to the actual cost analysis of water charge rates for the identified operators in the regions of the Czech Republic. The case study was conducted to confirm or refute the assumption that there is a connection between the percentage of costs and the size of the company in the pricing of water charge rates. Vertical analyses were compiled from the calculation formulas of individual operators for the years 2018–2021, i.e. research was carried out into the representation of individual cost components entering the price of water charge rates over a four-year horizon. For illustrative purposes, one operator for which the complete calculating formula was shown and the percentage representation of each cost to all items in the water charge rates pricing formula was determined, was selected (Table 4).

Similarly to Table 4, the same process has been carried out for all 14 operators, i.e., 14 costing analyses in terms of vertical analysis have been created for the 2018–2021 period. Table 5 shows the average representation of the type of cost-material items for all 14 operators.

The most common value is a percentage representation of material ranging around 20%. For accuracy, the median was determined, which in this case was 21.58%. Thus, in this case, it can be said that the material item ranges from 16.13% to 27.19%. This methodology was applied to all the above types of cost items (Table 6).

There is no obvious connection between the size of the company and the percentage allocation of individual cost items in water charge rate determination. This means that different water management companies have different percentage representation of material, energy, wages, other direct costs, and operating costs, regardless of whether they are a micro company, a small company, a medium-sized company, a large company or a specific municipality or city. Selecting and mapping

TABLE 4. Percentage representation of individual cost items and other items constituting the price of water charge rates in Pardubice – Vodovody a kanalizace Pardubice, a.s.

No.	Cost items	2018	2019	2020	2021
		%			
1	material	22.12	21.05	20.26	22.20
1.1	- raw groundwater + surface water	7.41	6.53	5.93	5.86
1.2	- drinking water collected	8.40	8.61	9.46	11.40
1.2	- chemicals	0.29	0.28	0.32	0.47
1.4	- other material	6.02	5.62	4.56	4.48
2	energy	3.23	3.66	3.26	2.92
2.1	- electric energy	3.07	3.44	3.10	2.83
2.2	- other types of energy (gas)	0.16	0.22	0.16	0.09
3	wages	6.29	6.92	6.68	7.37
3.1	- direct wages	4.69	5.17	5.00	5.52
3.2	- other personal costs	1.59	1.74	1.68	1.85
4	other direct costs	50.75	48.28	51.90	48.05
4.1	- depreciation	23.97	22.11	20.75	18.31
4.2	- repairs	26.70	26.12	31.11	29.72
4.3	- rental of infrastructure assets	0.08	0.06	0.05	0.02
4.4	- water delivered	0.00	0.00	0.00	000
5	operating costs	4.07	3.11	3.73	3.67
5.1	- wastewater discharge charges	0.00	0.00	0.00	0.00
5.2	- other external operating costs	2.72	2.33	3.05	2.91
5.3	- other own costs	1.35	0.78	0.68	0.76
6–9	financial income and expenses, production and administrative overhead	10.46	11.49	12.03	12.90
10	full own costs	96.91	94.51	97.87	97.11
11–12	calculation profit	3.09	5.49	2.13	2.89
13	total	100.00	100.00	100.00	100.00

Source: own elaboration.

the calculation formula of each water management company to meet the set criteria was lengthy and tens of thousands of data items need to be reviewed to compare companies among themselves and to determine average values of the representation of each cost item.

However, the analysis carried out allows for the determination of an interval for each cost component, within which it is clear what percentage level the individual cost enters. The following procedure was therefore set out. For all 14 operators,

TABLE 5. Average percentage representation of material in the calculation formula for selected operators over the 2018–2021 horizon

Company size	Operator	Region	Material [%]
Large account unit	Vodovody a kanalizace Pardubice, a.s.	Pardubice	21.41
Large account unit	Středočeské vodárny, a.s.	Central Bohemia	10.40
Large account unit	Letiště Praha, a.s.	Capital City of Prague	79.51
Medium-size account unit	Slovácké vodárny a kanalizace, a. s.	Zlín	16.13
Medium-size account unit	Vodovody a kanalizace Vyškov, a.s.	South Moravia	13.83
Medium-size account unit	CHEVAK Cheb, a.s.	Karlovy Vary	9.30
Small account unit	Technické služby Třešť, spol. s.r.o.	Vysočina	22.35
Small account unit	Městská Vodohospodářská, s.r.o.	South Bohemia	82.93
Small account unit	Vodovody a kanalizace Hlučín, s.r.o.	Moravia – Silesia	27.19
Micro account unit	Vodovody a kanalizace města Kdyně spol. s r.o.	Plzeň	21.76
Micro account unit	Zásadská vodárenská společnost, s.r.o.	Liberec	19.86
Micro account unit	VODA-RA s.r.o.	Hradec Kálové	22.84
Town	Town of Štíty	Olomouc	20.11
Municipality	Ludvíkovice municipality	Ústí nad Labem	96.83

Source: own elaboration.

the main cost items have been examined and the percentage of the individual costs in relation to the total has been calculated. Subsequently, the averages of the percentage values of the individual cost items were determined from four years under review. These averages of the percentage representation of each cost item have been compared among all operators in the analysis. Subsequently, for each major cost component, their most frequent intervals for the respective operators were created. For each cost item, abnormalities were excluded from the analysis.

It can be concluded from the research carried out that it is not possible to determine an average percentage value of the main cost items that would represent companies according to their size. No similarity was found in the data. However, it was possible to determine the range of most frequently occurring values for each major cost item determined regardless of the company size. After separating the most frequently occurring values from the less frequent or exceptional ones, it can be said that in the frame of this case study:

- material ranges from 16.13% to 27.19% of the total cost and profit of the companies; it is represented by the cost of raw groundwater and surface water, potable water collected, chemicals and other materials;

TABLE 6. Average percentage representation of the main cost items in the calculation formula for selected operators over the 2018–2021 horizon, median, min and max, average of min and max

Region	Material	Energy	Wages	Other direct costs	Operating costs
			%		
Pardubice	21.41	3.27	6.81	49.75	3.65
Central Bohemia	10.40	5.57	6.45	44.02	14.47
Capital City of Prague	79.51	1.39	33.78	15.48	1.82
Zlín	16.13	5.01	30.17	25.17	7.77
South Moravia	13.83	3.61	19.17	35.83	6.98
Karlovy Vary	9.30	5.62	5.44	47.34	6.46
Vysočina	22.35	5.70	29.44	11.97	8.81
South Bohemia	82.93	4.36	3.80	8.71	4.88
Moravia – Silesia	27.19	1.53	19.91	31.83	3.04
Plzeň	21.76	10.82	33.60	6.89	12.85
Liberec	19.86	0.63	30.37	34.09	7.99
Hradec Králové	22.84	21.15	31.35	10.51	12.36
Olomouc	20.11	1.73	6.45	123.84	14.05
Ústí nad Labem	96.83	0.21	22.00	37.46	14.65
Median	21.58	3.99	20.96	32.96	7.88
Lowest interval value considered	16.13	1.39	19.17	25.17	3.04
Maximum interval value	27.19	5.70	33.78	49.75	14.65
Min and max average	21.46	3.78	27.76	38.19	9.07

Source: own elaboration.

- energy is the lowest cost item, consisting of electricity and other energy and ranging from 1.39% to 5.70%;
- wages range around 19.17–33.78%;
- other direct costs range from 25.17% to 49.75% and are represented by depreciation, repairs, rent of infrastructure assets and water supplied;
- operating costs, which include sewage discharge charge rates, external operating costs and own overheads, range from 3.04% to 14.65%.

As explained above, it was not possible to include all the data in the average, so intervals of values were set, however, it was possible to determine the average of the values taken in the interval:

- material 21.46%;
- energy 3.78%;
- wages 27.79%;

- other direct costs 38.19%;
- operating costs 9.07%.

It should therefore be noted that the median was always taken from all observed values in the case study:

- material 21.58%;
- energy 3.99%;
- wages 20.96%;
- other direct costs 32.96%;
- operating costs 7.88%.

The research described in the article aims to assess or challenge the conclusion of whether any dependence can be found between the size of the company within the representation of each price cost of water charge rates. However, the analysis shows that there is no obvious connection between the size of the company and the percentage allocation of individual cost items in water charge rate determination.

The aim of the research is also about finding out which costs have the biggest impact on the price of water. The research shows that the highest percentage representation of costs is found in other direct costs, the second highest percentage representation of costs is found in wages, the third highest representation of costs occurs in the material and the lowest intervals of percentage representation include operating costs and energy.

Conclusions

The research described in the paper aimed to determine the percentage representation of the types of costs in the unit price of water charge rates in the Czech Republic using vertical analysis, with a focus on demonstrating the dependence between the percentage representation of costs in the unit price and the size of the analyzed company. As already mentioned above, the research can be considered for the time being as pilot (initial) research, which provides an initial insight into the given issue. The conclusions drawn from the analyzes cannot be considered statistically relevant in the initial phase of the research, nor can they be understood as a representative sample for the entire Czech Republic and its individual regions. These are initial cost analyzes of the price of water, which will subsequently be confirmed by analyzing additional input data in the next step of the research. In total, 14 operators were taken into the research in the case study and had to meet certain factors in their selection, namely the selection of one operator per region of the Czech Republic.

The selection of the operator had to meet the condition of delivery to only one point of consumption (or maximum two) within the given region of the Czech Republic, the selection of the operator based on the above-stated factors must not interfere with other regions of the Czech Republic. The selection of the operator was limited by the size of the company (micro, small, medium-sized, and large companies) in the ratio 3 : 3 : 3 : 3, the selection of the operator was limited by the representation of at least one representative of the public sector, i.e. a city or a municipality (in the case study two such representatives were considered). A sample selection of 14 representatives for each region of the country was based on processing 26,000 data pieces where the years 2018–2021 were taken as the reference period, which was subsequently narrowed down to 1,736 values that entered the cost analysis. The sewerage charge rate was not addressed because not all water users are connected to the sewer system and use, for example, septic tanks.

Thus, the analysis was developed for 14 representatives/operators for which the types of costs were monitored and analysed for the 2018–2021 period. For these costs, which were narrowed down to the main item costs such as material, energy, wages, and other direct and operating costs, an interval within which the given value range was set at the very end. This interval was determined based on the minimum and maximum values that occurred in the vertical analysis. This analysis shows that there is no dependence between the size of the operator and the percentage share of the types of costs in the unit price of water charge rates. The ranges for each type of cost are shown in Table 6, where a conclusion was drawn in terms of the most frequently represented cost items from the water charge rate calculation formula, namely:

- the highest percentage representation of costs is found in other direct costs, where the largest burden is in the depreciation of infrastructure assets, whose absolute amount is in millions of CZK (1 EUR = 25.00 CZK) and ranges from 25.17% to 49.75%;
- the second highest percentage representation of costs appears in wages and salaries, ranging from 19.17% to 33.78%;
- the third highest percentage of costs occurs in materials, ranging from 16.13% to 27.19%;
- the lowest intervals of percentage representation include operating costs and energy, which range in the order of ten per cent.

It can be stated from the above-stated facts that further analysis in the ongoing research will focus on the highest percentage of costs, namely other direct costs, wages, and materials.

Acknowledgements

This paper has been worked out under the project of the specific research at the Brno University of Technology FAST-S-23-8253 – Life cycle cost analysis of construction objects.

References

Ashoori, N., Dzombak, D. A. & Small, M. J. (2017). Identifying water price and population criteria for meeting future urban water demand targets. *Journal of Hydrology*, *555*, 547–556. https://doi.org/10.1016/j.jhydrol.2017.10.047

Český statistický úřad [Czech Statistical Office] (n.d.). Retrieved from: https://www.czso.cz [accessed: 28.10.2023].

He, C., Shen, J., Xu, J., Sun, F. & Wang, B. (2023). Quality evaluation of water disclosure of Chinese papermaking enterprises based on accelerated genetic algorithm. *Scientific Reports*, *13* (1), 12225. https://doi.org/10.1038/s41598-023-39307-y

Jajac, N., Marović, I., Rogulj, K. & Kilić, J. (2019). Decision support concept to selection of wastewater treatment plant location – The case study of town of Kutina, Croatia. *Water*, *11* (4), 717. https://doi.org/10.3390/w11040717

Justice (n.d.). Retrieved from: https://www.justice.cz [accessed: 28.10.2023].

Kejser, A. (2016). European attitudes to water pricing: Internalizing environmental and resource costs. *Journal of Environmental Management*, *183* (Part 3), 453–459. https://doi.org/10.1016/j.jenvman.2016.08.074

Liu, L., Wu, T., Xu, Z. & Pan, X. (2018). The water-economy nexus and sustainable transition of the Pearl River Delta, China (1999–2015). *Sustainability (Basel, Switzerland)*, *10* (8), 2595. https://doi.org/10.3390/su10082595

Maziotis, A., Saal, D. S., Thanassoulis, E. & Molinos-Senante, M. (2015). Profit, productivity and price performance changes in the water and sewerage industry: an empirical application for England and Wales. *Clean Technologies and Environmental Policy*, *17* (4), 1005–1018. https://doi.org/10.1007/s10098-014-0852-2

Ministerstvo zemědělství [The Ministry of Agriculture of the Czech Republic] (n.d.). Retrieved fom: https://www.eagri.cz [accessed: 28.10.2023].

Molinos-Senante, M. & Maziotis, A. (2021). Productivity growth, economies of scale and scope in the water and sewerage industry: the Chilean case. *PLoS ONE*, *16* (5), e0251874. https://doi.org/10.1371/journal.pone.0251874

Oblouková, A. (2023). Predikce vývoje průměrných cen vodného a stočného České Republiky [Prediction of the development of the average prices of water and sewage in the Czech Republic]. *Mezinárodní Doktorská Konference Stavebního Inženýrství, Sborník Příspěvků* (pp. 624–630). Brno: Econ Publishing.

Oblouková, A. & Vítková, E. (2023). Analysis of water and sewerage charge rate components in the regions in comparison to the national average values. *AIP Conference Proceedings*, *2928* (1), 150018. https://doi.org/10.1063/5.0170477

Parween, F., Kumari, P. & Singh, A. (2021). Irrigation water pricing policies and water resources management. *Water Policy*, *23* (1), 130–141. https://doi.org/10.2166/wp.2020.147

Vítková, E., Vaňková, L. & Oblouková, A. (2022). Comparison of the price level of the water and sewerage charge rates and macroeconomic indicators in the Czech Republic. *Scientific Review Engineering and Environmental Sciences*, *31* (2), 113–123. https://doi.org/10.22630/srees.2958

Volf, G., Sušanj Čule, I., Žic, E. & Zorko, S. (2022). Water Quality Index prediction for improvement of treatment processes on drinking water treatment plant. *Sustainability (Basel, Switzerland)*, *14* (18), 11481. https://doi.org/10.3390/su141811481

Vyhláška č. 428/2001 Sb., kterou se provádí zákon č. 274/2001 Sb., o vodovodech a kanalizacích pro veřejnou potřebu a o změně některých zákonů (zákon o vodovodech a kanalizacích) [Decree No. 428/2001 Coll., implementing Act No. 274/2001 Coll., on water supply and sewerage for public use and on amendments to certain acts (Act on water supply and sewerage)].

Vyhláška č. 450/2009 Sb., kterou se provádí zákon č. 526/1990 Sb., o cenách, ve znění pozdějších předpisů [Decree No. 450/2009 Coll., implementing Act No. 526/1990 Coll., on prices, as amended].

Zaenuri & Dwidayati, N. (2020). The analysis of wastewater management at small and medium-size textile (batik) enterprises in Pekalongan city. *Journal of Physics: Conference Series*, *1567* (4), 42057. https://doi:10.1088/1742-6596/1567/4/042057

Zákon č. 274/2001 Sb., o vodovodech a kanalizacích pro veřejnou potřebu a o změně některých zákonů (zákon o vodovodech a kanalizacích), ve znění pozdějších předpisů [Act No. 274/2001 Coll., on Water Supply and Sewerage for Public Use and on Amendments to Certain Acts (Act on Water Supply and Sewerage), as amended].

Zákon č. 526/1990 Sb., o cenách, ve znění pozdějších předpisů [Act No. 526/1990 Coll., on prices, as amended].

Zákon č. 563/1991 Sb., o účetnictví [Act No. 563/1991 Coll., on Accounting].

Summary

Cost analysis of water charge rates in the Czech Republic – Case study. The article was conceived as an initial insight into the issues related to the representation of individual type costs (e.g. material, labor costs, property depreciation, etc.) in the price of water in the Czech Republic. The aim of the article was to point out the possibility of dependence between the size of the company operating the infrastructural property of water supply and sewerage and the representation of individual costs in the water price within the framework of the case study. As a sample that formed the outputs of the case study, 14 companies were

taken, which were selected according to the unified regions of the Czech Republic. Both basic mathematical methods and elementary methods used in financial analysis were used in the analysis. Within the scope of the case study, it can be stated that there is no dependence between the size of the companies and the representation of costs. Among the largest costs from the point of view of financial representation are other direct costs, where the costs of depreciation, property repairs, rental property, as well as wage costs and material costs are mainly represented. Insignificant costs include, for example, energy costs, which are only represented in the range of 1.39–5.70% of the total costs. Therefore, in order for the results included in the case study to be considered statistically relevant, it is necessary to expand the sample and confirm or refute the initial findings published in this article.

Scientific Review Engineering and Environmental Sciences (2024), 33 (1), 53–68

Sci. Rev. Eng. Env. Sci. (2023), 33 (1) https://srees.sggw.edu.pl

ISSN 1732-9353 (suspended) eISSN 2543-7496 DOI 10.22630/srees.5619

Received: 02.11.2023
Accepted: 05.01.2024

Michał TRZECIAK✉

Daria SIKORSKA

Institute of Environmental Engineering, Warsaw University of Life Science – SGGW, Poland

Application of UAV and ground measurements for urban vegetation cooling benefit assessment, Wilanów Palace case study

Keywords: ecosystem services, NDVI, urban vegetation, urban heat island, climate change, UAV, cultural heritage sites

Introduction

Cultural heritage sites (CHS), particularly those in the suburbs of large European metropolises, play a unique and multifaceted role in urban landscapes. Examples such as the Wilanów Palace in Warsaw, once royal residences, have undergone intense urbanization in recent decades. These sites, absorbed into the urban fabric, have transformed into enclosed urban parks surrounded by high-rise buildings, offering not only a glimpse into shared cultural history, but also acting as crucial green spaces in urban environments (Tengberg et al., 2012). They contribute to local climate regulation, air quality improvement,

✉m.trzeciak99@gmail.com

and biodiversity conservation, yet the specific role of vegetation in these settings has not been thoroughly examined, marking a notable gap in urban ecology research, particularly in the context of CHS management (Wang et al., 2020).

Despite their significance, CHS like the Wilanów Palace are increasingly threatened by urban development, leading to ecological challenges such as habitat degradation and landscape fragmentation (Tengberg et al., 2012). This highlights an urgent need for effective strategies to evaluate and enhance the ecological functions and ecosystem services of CHS within complex urban ecosystems (Tengberg et al., 2012). This emphasizes an urgent need for tailored strategies to evaluate and enhance the ecological functions and ecosystem services of CHS within complex urban ecosystems (Andersson et al., 2021).

Focusing on ecosystem services, especially climate regulation, urban parks and green spaces within CHS have been documented to mitigate urban heat island effects (Bowler, Buyung-Ali, Knight & Pullin, 2010; Peng et al., 2021; Sikorski et al., 2021). Previous studies have utilized microclimatic approaches, employing temperature sensors to compare vegetated and non-vegetated areas, revealing the mitigation of urban heat island effects (Heidt & Neef, 2008; Trojanowska, 2022). However, advancements in remote sensing technologies have enabled more sophisticated multi-methodological assessments. Tools like the normalized difference vegetation index (NDVI) offer quantitative and qualitative measures of vegetation which can be correlated with land surface temperature (LST) to assess temperature distribution within urban green spaces and evaluate cooling benefits of urban green spaces (Weng, Lu & Schubring, 2004; Pettorelli et al., 2005).

In this study we focus on the following research questions:

1. What are the comparative cooling effects of different vegetation classes (dense high, sparse high, dense low, sparse low vegetation) within the CHS context?
2. How does the integration of remote sensing and ground-based measurements enhance the understanding of ecosystem services, particularly cooling benefits, provided by urban green spaces in CHS like Wilanów Palace?
3. How do the findings on temperature variability across different vegetation types contribute to the development of sustainable urban planning and conservation strategies for CHS?

The overarching goal of this study is to deepen our understanding of the cooling effects of urban greenery in CHS and provide information on sustainable urban planning and conservation strategies.

Material and methods

Study area

The study focuses on the Wilanów Palace, a recognized CHS in the suburban precincts of Warsaw, Poland's capital, with a population nearing 2 million and an area of 517.2 km^2 (Statistics Poland, 2022). This site, with its origins in the 17th century, holds significant historical value. Integral to this study is the extensive green space surrounding the palace, which is not only a key aesthetic component, but also a vital ecological resource. These vegetated areas contribute substantially to the ecosystem services of the locale, with their value previously assessed to exceed EUR 500,000 annually in ecological benefits (Zawojska, Szkop, Czajkowski & Żylicz, 2016). The focus of our investigation is to understand the role these green spaces play in local climate regulation, specifically their contribution to cooling within the urban ecological context of the Wilanów Palace. The palace's green spaces consist of a variety of vegetation, including tall trees, manicured gardens, and dense shrubbery, which collectively enhance the local climate. The site offers a unique combination of meticulously maintained gardens and near-wild forested areas, making it an ideal location for assessing the cooling influence of different types of vegetation within a CHS.

Given the palace's location adjacent to Wilanów Lake, the study also considers the potential synergistic effects of water bodies and vegetation in climate regulation (Setiawan & Gawryszewska, 2023). The palace's verdant landscapes serve as a natural oasis within the urban fabric of Warsaw, offering a compelling case study for the assessment of cooling benefits through a combination of remote sensing technologies and ground-based measurements.

Cooling benefit assessment

Our study investigating the cooling benefits at the Wilanów Palace integrated various data collection techniques to evaluate the impact of vegetation on local climate regulation. The methodology encompassed in-situ ground measurements of topsoil temperature (TST), complemented by UAV-derived thermal imagery. Additionally, we employed LAI-ground measurements, LiDAR, and NDVI for a detailed classification of vegetation, which was subsequently correlated with temperature data (Fig. 1).

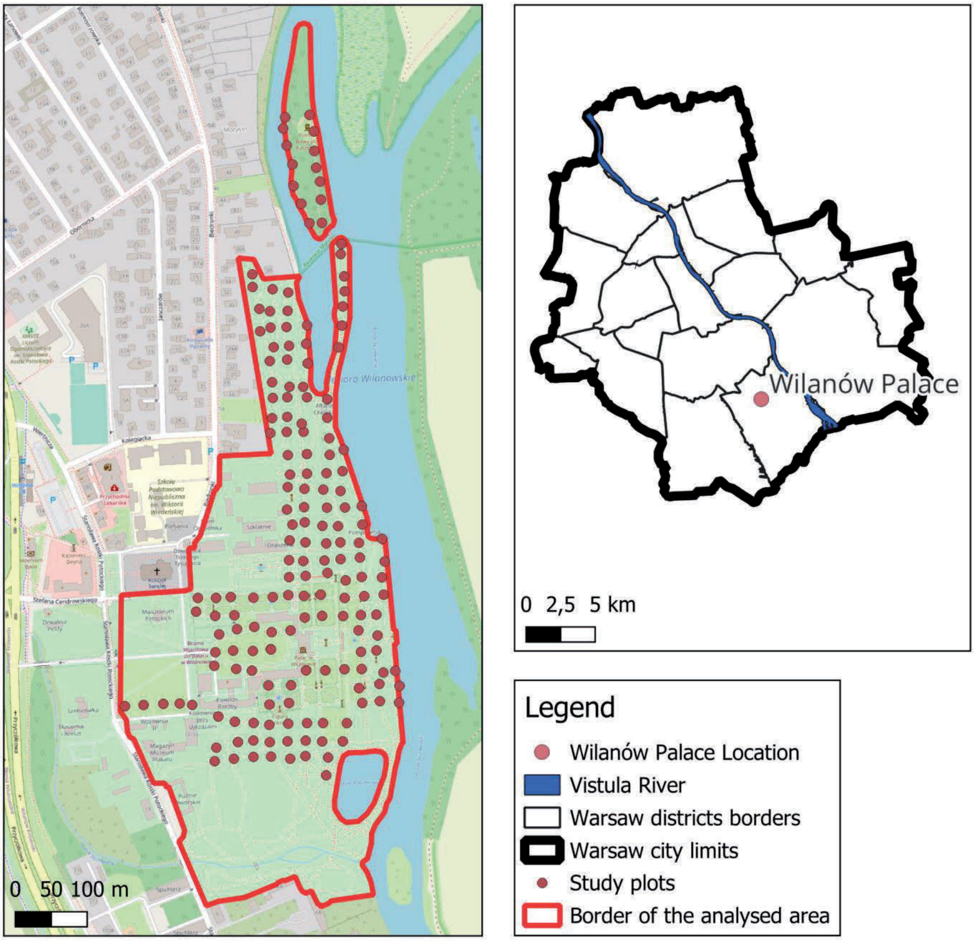


FIGURE 1. Study area within Wilanów Palace, with dots representing the specific locations of detailed ground temperature measurements

Source: Points, overlay area – own elaboration; background – OpenStreetMap contributors (2023) Planet dump [data file from 2023]. Retrieved from https://planet.openstreetmap.org.

All measurements were conducted on the hottest summer day, characterized by a maximum temperature exceeding 30°C over a period of one week. The data recorded, both ground-based and aerial, provided precise, localized insights into temperature variations and vegetation characteristics (Fig. 1).

Field measurements

The Wilanów Palace grounds were methodically segmented into a 25 × 25 m^2 grid, specifically designed to represent the site's diverse vegetation. In study plot, we recorded the topsoil temperature (TST) at the center, making adjustments for any physical obstructions. This approach included all vegetation types, extending under tree canopies where feasible, to comprehensively assess the cooling effects of various vegetation. The top 0–10 cm soil layer was analyzed using a TDR HH2 Moisture Meter by Delta-T Devices. Additionally, in each plot, the leaf area index (LAI) was measured as an indicator of leaf volume (Nagendra, 2002) using the SunScan Canopy Analysis System, also by Delta-T Devices. For assessing the temperature of impermeable surfaces, a FLIR C3 thermal camera was employed. The data collection, conducted on August 2, 2023, resulted in a total of 170 temperature and vegetation samples. Precise geographical coordinates for each data point were recorded using a handheld GPS device with an accuracy of 0.5 m (Fig. 1).

Land surface temperature using UAV

To acquire a comprehensive temperature gradient within the study area, we employed UAV thermal imagery captured by a DJI Mavic 3T drone to acquire surface temperature data within our study area. The UAV mission was conducted concurrently with ground sample collection. Thermal images were processed using IRimage-UAV software (Irujo, 2022) integrated within ImageJ software (Schneider, Rasband & Eliceiri, 2012) to generate raster tiles in degrees Celsius (Schneider et al., 2012). Subsequently, we utilized OpenDroneMap software to further process the thermal imagery. This software was utilized to mosaic raster tiles, allowing us to acquire an accurate orthophotomap that presents a LST.

Vegetation classification

For the vegetation classification in our study, we initially accessed light detection and ranging (LiDAR) data from a publicly available repository managed by the Head Office of Geodesy and Cartography (Główny Urząd Geodezji i Kartografii [GUGiK] 2022). Utilizing the R programming language (R Core Team, 2021) and the lidR package (Roussel et al., 2020), we processed the raw LiDAR data (laz files format) to produce two key models: a digital terrain model (DTM) and a digital surface model (DSM). The DTM was developed using a spatial interpolation algorithm based on a k-nearest neighbor approach (k-NN) with inverse-distance weighting, while the DSM was constructed via Delaunay triangulation (DT) of first returns, employing linear interpolation within each triangle. By subtracting the DTM from the DSM, we obtained a model representing the height of all objects in the study area.

Concurrently, we acquired high-resolution (0.25 × 0.25, June 2022) aerial color-infrared (CIR) photographs (GUGiK, 2022) and converted them into normalized difference vegetation index (NDVI) images using the R package terra (Hijmans et al., 2023) following Crippen's method (Crippen, 1990). The NDVI images were then resampled to a spatial resolution of 1 × 1 m. Applying a thresholding approach, pixels with NDVI values below 0.2 were classified as non-vegetated areas (Hashim, Abd Latif & Adnan, 2019).

The next phase involved merging the height model and the processed NDVI images. Here, raster values equal to or greater than 2 m were categorized as high vegetation, and those below 2 m as low vegetation. This process resulted in a final raster product categorizing the study area into three distinct classes: non-vegetated areas, low vegetation, and high vegetation.

Further refinement was achieved by incorporating field-collected LAI data, allowing us to subdivide the low and high vegetation categories into sparse and dense types. Consequently, our research area was delineated into five classes: non-vegetated areas, sparse low vegetation, dense low vegetation, sparse high vegetation, and dense high vegetation (Turner et al., 2015).

Finally, we correlated each ground sample point with its corresponding LST, NDVI, and vegetation class values, creating a comprehensive dataset for our analysis.

Statistical analysis

In our statistical analysis, the primary objective was to elucidate the interaction between the vegetation structure and temperature variations. Recognizing the heterogeneity in our dataset, with some segments deviating from normal distribution, we adopted a dual approach. For data adhering to a normal distribution, parametric tests were utilized, while non-parametric tests were applied to the non-normally distributed segments. Specifically, to explore the relationship between NDVI values and both ground and surface temperatures, Spearman's correlation tests were conducted. Additionally, our analysis aimed to identify significant temperature differences across various vegetation classes. This involved the use of Dunn's multiple comparison test for non-normally distributed data and the Tukey's test for data following a normal distribution. These tests were instrumental in discerning the distinct thermal characteristics attributable to different vegetation types.

All statistical procedures were meticulously carried out using R software, version 3.4.4, ensuring rigorous data analysis and reliable results.

Results

Temperature variability across vegetation types

Our findings indicate a clear relationship between vegetation type and temperature. In areas characterized by the most dense and highest vegetation, the lowest temperatures were recorded, corroborating the cooling benefits of such vegetation (Fig. 2). Conversely, the highest temperatures were observed on sandy and concrete paths, as well as on grass lawns. This temperature gradient, attributable to the cooling effects of vegetation, is vividly illustrated in the thermal imagery captured along one of the paths in Wilanów Park (Fig. 3).

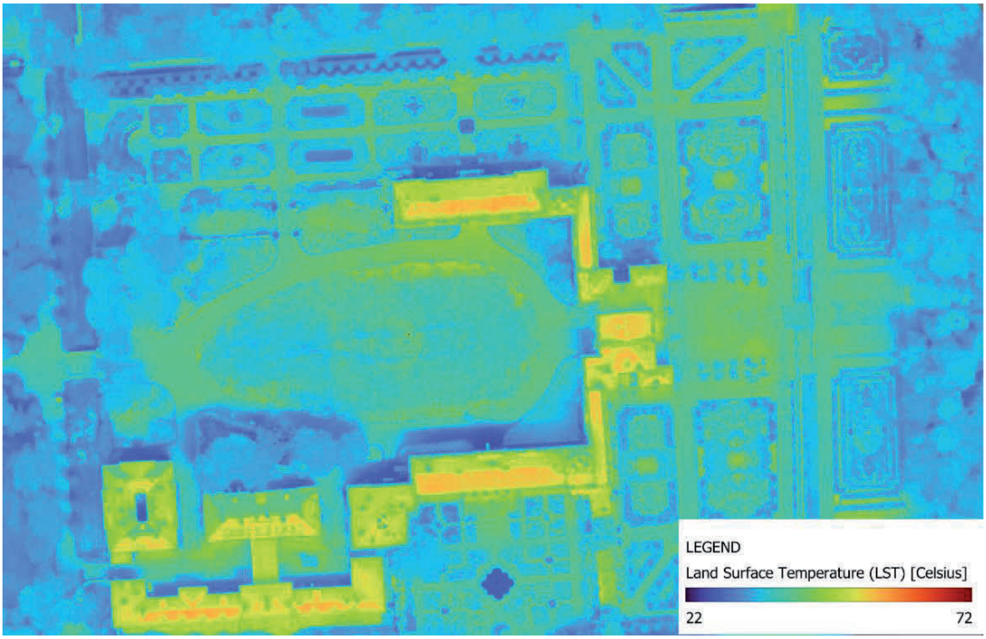


FIGURE 2. Fragment of LST map acquired from the drone aerial photography
Source: own elaboration.

FIGURE 3. Infrared (right), and RGB (left) picture, taken down one of the paths in the Wilanów Park
Source: own elaboration.

Land and topsoil surface temperatures

The UAV-based LST measurements ranged from 26.8° to 47.5°C, aligning with the ground-based topsoil temperature (TST) measurements, which varied between 19.5° and 29.2°C (Fig. 2). Notably, the highest recorded LST reached 72°C at some locations (metal rooftops of buildings), which warrants further investigation for its potential implications on microclimates within the CHS (Baryła, Karczmarczyk, Bus & Witkowska-Dobrev, 2019).

Temperature characteristics among vegetation classes

Statistical analysis revealed significant differences in temperature characteristics among various vegetation classes (Table 1). Specifically, areas with higher LAI values exhibited lower LST and TST compared to areas with lower LAI, substantiating the cooling benefits of denser vegetation.

Statistical analysis ANOVA with post-hoc Tukey, HSD for TST distinguished four types of vegetation: dense high vegetation, sparse high vegetation, dense low vegetation and sparse low vegetation. Analysis with Dunn's test showed that all these groups have significantly different from each other.

The Spearman's correlation test indicated significant negative correlations between both TST and NDVI (–0.57) and between LST and NDVI (–0.41), (Table 2). These findings suggest that increased vegetation, as indicated by higher NDVI values, is associated with lower temperatures.

TABLE 1. Comparison of vegetation characteristics and temperature (TST and LST) for distinguished vegetation classes, significant differences represented by homogenous groups (a–e) at $p < 0.05$

Vegetation class	Mean LAI	LAI *SD*	Mean TST	Homogenous groups – Tukey's $p < 0.05$	TST SD	Mean LST (UAV)	Homogenous groups – Dunn's $p < 0.05$	LST *SD*
Dense high vegetation	4.50	3.13	27.74	c	1.46	33.32	e	1.34
Sparse high vegetation	3.52	3.11	28.54	bc	1.99	33.61	d	2.69
Dense low vegetation	4.35	2.86	28.14	bc	1.93	33.37	c	3.95
Sparse low vegetation	2.88	2.56	29.12	b	1.39	34.16	b	3.23
Without vegetation	NA	NA	30.51	a	1.58	36.89	a	4.13

Source: own elaboration.

TABLE 2. Correlations matrix between temperature characteristics and normalized difference vegetation index ($p < 0.05$) Spearman's correlation

Parameter	Mean	*SD*	1
Normalized difference vegetation index NDVI	0.29	–0.19	–
Topsoil temperature (field measurements)	28.97	2.00	–0.57
Land surface temperature (UAV)	34.58	3.57	–0.41

Source: own elaboration.

It is worth noting that the lowest TST values were particularly observed in areas with dense vegetation, both low and high (Fig. 4). Furthermore, both LST and TST were significantly lower in vegetated areas compared to non-vegetated areas (Figs 4 and 5).

The correlation coefficient between TST and NDVI is –0.57; and between LST and NDVI is –0.41; and the *p*-value of both tests is below 0.05, which means that both of those correlations are significant (Table 2).

Additionally, we must also note that TST values were lowest especially in areas with dense vegetation, for both low, and high vegetation (Fig. 4). Moreover, it is worth mentioning that both LST, and TST were significantly lower in vegetated areas in comparison to areas without vegetation (Figs 4 and 5).

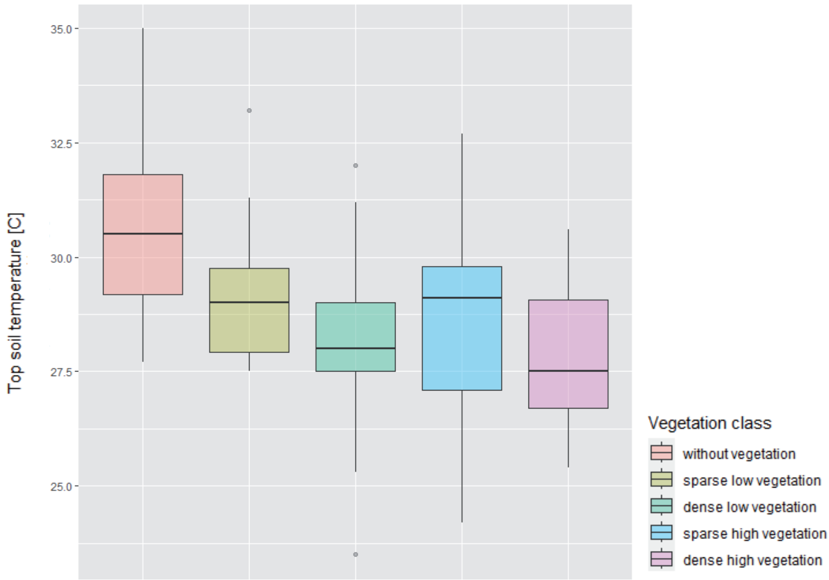


FIGURE 4. Boxplot illustrating topsoil temperature (TST) across different vegetation types
Source: own elaboration.

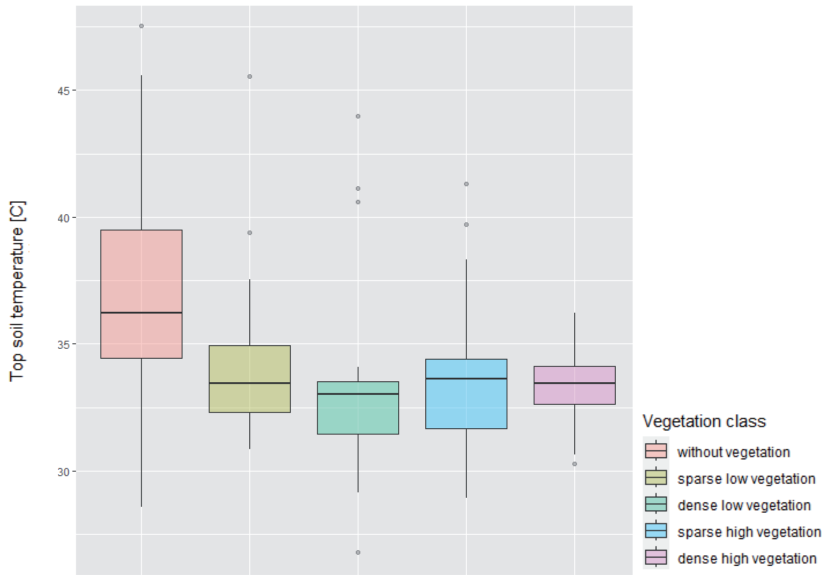


FIGURE 5. Boxplot illustrating land surface temperature (LST) derived from UAV equipped with thermal camera
Source: own elaboration.

Discussion

Our study, leveraging remote sensing techniques for urban ecosystem service assessment (García-Pardo, Moreno-Rangel, Domínguez-Amarillo & García-Chávez, 2022), highlights the significant impact of vegetation on both land surface temperature (LST) and topsoil temperature (TST) within the contrasting environment of a city park. This approach, necessitating high-resolution spatial data (Chen, Xu, Yang, Wu & Zhu, 2020; Voogt & Oke, 2023), is exemplified in our case study of the Wilanów Palace's urban park with its dense tree cover and open areas. Our findings align with existing research (Oke, 1987; Chen, Zhao, Li & Yin, 2006), emphasizing urban parks, particularly older ones, as 'cool islands' significantly influencing their surroundings. For instance, park vegetation can noticeably cool air temperatures up to 520 m away, with reductions of 1.6°C at 130 m and 0.9°C at 280 m (Aram, Solgi, Garcia & Mosavi, 2020).

We confirmed the general principle that parks broadly impact air and ground temperatures, a crucial aspect for CHS like the Wilanów Palace. However, this also raises questions about the extent to which TST, easily and inexpensively measured through remote sensing, can substitute for LST. While a close relationship between LST and TST is often assumed, this premise, particularly at a micro-scale, requires further investigation. LST is more directly linked to human comfort, health, and mortality (Shahidan, Shariff, Jones, Saalleh & Abdullah, 2010; Masselot et al., 2023), whereas TST plays a significant role in drought monitoring for plants, energy transfer modeling in ecosystems, and global carbon cycle assessment (Xu, Qu, Hao, Zhu & Gutenberg, 2020). High-resolution applications often rely on the fusion of satellite-measured TST data and validation based on vegetation parameters, primarily the normalized difference vegetation index (NDVI), (Xu et al., 2020).

Field measurements indicated significant temperature variations between lawns and wooded areas (Figs 4 and 5), with more pronounced differences than between types of greenery. The influence of shading was a key factor in these differences. The results demonstrate that both lawns and tree canopies can effectively cool surfaces, thus aiding in reducing urban heat islands during heatwaves. However, lawns have a minimal impact on human comfort, whereas tree shade can provide effective local cooling (Armson, Stringer & Ennos, 2012). Urban climate models show that the average TST in urban areas decreases by only about 3.06°C, highlighting the significant role of shading (Schwaab et al., 2021).

It is important to note that in the context of LST, we found dense high vegetation to be statistically warmer than a low one (Fig. 5). One possible explanation of this phenomenon is LST being measured from air, canopies of tall trees have an increased exposure to sunlight what vegetation is contributing to higher temperatures.

Our study at the Wilanów Palace revealed a crucial disparity between LST, recorded via UAV, predominantly reflecting canopy temperature, and TST, more indicative of the temperature experienced by park users. While the lowest LST across vegetation categories ranged from 35° to 45°C, TST measurements were significantly lower, with dense low vegetation recording a minimum of 19°, and tall sparse vegetation around 22°. The median TST values were notably lowest for tall dense vegetation.

This disparity suggests that UAV-derived LST, while insightful for canopy temperatures, may not fully capture the thermal experience at the ground level. The lower TST in areas with dense, tall vegetation indicates more substantial cooling benefits, enhancing thermal comfort for park visitors. This distinction is vital for urban planners and conservationists in managing green spaces within CHS, aiming to maximize ecological benefits and human comfort. Furthermore, the study underscores the role of trees in limiting solar radiation penetration and lowering surface and air temperatures in shaded areas (Roy, Byrne & Pickering, 2012; Schwaab et al., 2021). The temporal-spatial variability of evapotranspirative cooling, changing with the leafiness of deciduous trees (Song & Wang, 2015), and the potential adverse effects of urban trees on air pollution dispersion (Morani, Nowak, Hurabayashi & Calfapietra, 2011; Li & Wang, 2018) are also significant considerations.

Conclusion and future directions

The study's implications are multifaceted. It reinforces the vital role of green spaces in CHS, especially amid urban expansion, and informs adaptive landscape management strategies for historical sites like Wilanów Palace (Akkurt et al., 2020). Our methodologies offer a replicable framework for similar studies in other CHS, contributing to the broader discourse on microclimate regulation within these unique environments. However, future research should expand this study's scope by including additional environmental variables such as humidity, wind speed, and solar radiation, and considering temporal variations. By deepening our understanding of these dynamics, we can make informed decisions that optimally preserve both the historical and ecological value of CHS landscapes.

In conclusion, vegetation significantly impacts both LST and TST, with the potential to reduce temperature gradients by over 5°C in CHS. Such comprehensive studies will enable informed decisions for preserving the historical and ecological values of CHS landscapes.

Acknowledgements

Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.

This research was funded by the National Science Centre (Poland) grant 2020/39/B/HS4/03240.

References

Akkurt, G. G., Aste, N., Borderon, J., Buda, A., Calzolari, M., Chung, D., Constanzo, V., Del Pero, C., Evola, G., Huerto-Cardenas, H. E., Lo Faro, A., Lucchi, E., Marletta, L., Nocera, F., Pracchi, V. & Turhan, C. (2020). Dynamic thermal and hygrometric simulation of historical buildings: Critical factors and possible solutions. *Renewable and Sustainable Energy Reviews*, *118*, 109509. https://doi.10.1016/j.rser.2019.109509

Andersson, E., Haase, D., Anderson, P., Cortinovis, C., Goodness, J., Kendal, D., Lausch, A., McPhearson, T., Sikorska, D. & Wellmann, T. (2021). What are the traits of a social-ecological system: Towards a framework in support of urban sustainability. *npj Urban Sustainability*, *1* (1), 14. https://doi.10.1038/s42949-020-00008-4

Aram, F., Solgi, E., Garcia, E. H. & Mosavi, A. (2020). Urban heat resilience at the time of global warming: evaluating the impact of the urban parks on outdoor thermal comfort. *Environmental Sciences Europe*, *32*, 1–15. https://doi.10.1186/s12302-020-00393-8

Armson, D., Stringer, P. & Ennos, A. R. (2012). The effect of tree shade and grass on surface and globe temperatures in an urban area. *Urban Forestry & Urban Greening*, *11* (3), 245–255. https://doi.10.1016/j.ufug.2012.05.002

Baryła, A., Karczmarczyk, A., Bus, A. & Witkowska-Dobrev, J. (2019). Surface temperature analysis of conventional roof and different use forms of the green roof. *Scientific Review Engineering and Environmental Sciences (SREES)*, *28* (4), 632–640. https://doi.10.22630/PNIKS.2019.28.4.57

Bowler, D. E., Buyung-Ali, L., Knight, T. M. & Pullin, A. S. (2010). Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning*, *97* (3), 147–155. https://doi.10.1016/j.landurbplan.2010.05.006

Chen, X., Xu, Y., Yang, J., Wu, Z. & Zhu, H. (2020). Remote sensing of urban thermal environments within local climate zones: A case study of two high-density subtropical Chinese cities. *Urban Climate*, *31*, 100568. https://doi.10.1016/j.uclim.2019.100568

Chen, X. L., Zhao, H. M., Li, P. X. & Yin, Z. Y. (2006). Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote Sensing of Environment*, *104* (2), 133–146. https://doi.10.1016/j.rse.2005.11.016

Crippen, R. E. (1990). Calculating the vegetation index faster. *Remote Sensing of Environment*, *34* (1), 71–73. https://doi.10.1016/0034-4257(90)90085-Z

García-Pardo, K. A., Moreno-Rangel, D., Domínguez-Amarillo, S. & García-Chávez, J. R. (2022). Remote sensing for the assessment of ecosystem services provided by urban vegetation: A review of the methods applied. *Urban Forestry & Urban Greening*, *74*, 127636. https://doi.10.1016/j.ufug.2022.127636

Hashim, H., Abd Latif, Z. & Adnan, N. A. (2019). Urban vegetation classification with NDVI threshold value method with very high resolution (VHR) Pleiades imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 237–240. https://doi.10.5194/isprs-archives-XLII-4-W16-237-2019

Head Office of Geodesy and Cartography [GUGiK] (2022). Data available free of charge for download from www.geoportal.gov.pl [accessed: 31.10.2023].

Heidt, V. & Neef, M. (2008). Benefits of urban green space for improving urban climate. In *Ecology, planning, and management of urban forests: International perspectives* (pp. 84–96). New York, NY: Springer.

Hijmans, R. J., Bivand, R., Forner, K., Ooms, J., Pebesma, E. & Sumner, M. D. (2022). *Package 'terra'*. Maintainer: Vienna, Austria.

Irujo, G. P. (2022). IRimage: open source software for processing images from infrared thermal cameras. *PeerJ Computer Science*, *8*, e977. https://doi.10.7717/peerj-cs.977

Li, Q. & Wang, Z. H. (2018). Large-eddy simulation of the impact of urban trees on momentum and heat fluxes. *Agricultural and Forest Meteorology*, *255*, 44–56. https://doi:10.1016/j.agrformet.2017.07.011

Masselot, P., Mistry, M., Vanoli, J., Schneider, R., Iungman, T., Garcia-Leon, D., Ciscar, J. C., Feyen, L., Orru, H., Urban, A., Breitner, S., Huber, V., Schneider, A., Samoli, E., Stafoggia, M. de'Donato, F., Rao, S., Armstrong, B., Nieuwenhuijsen, M., …, Aunan, K. (2023). Excess mortality attributed to heat and cold: a health impact assessment study in 854 cities in Europe. *The Lancet Planetary Health*, *7* (4), e271–e281. https://doi.10.1016/S2542-5196(23)00023-2

Morani, A., Nowak, D. J., Hirabayashi, S. & Calfapietra, C. (2011). How to select the best tree planting locations to enhance air pollution removal in the MillionTreesNYC initiative. *Environmental Pollution*, *159* (5), 1040–1047. https://doi.10.1016/j.envpol.2010.11.022

Nagendra, H. (2002). Opposite trends in response for the Shannon and Simpson indices of landscape diversity. *Applied Geography*, *22* (2), 175–186. https://doi.10.1016/S0143-6228(02)00002-4

Oke, T. R. (1987). *Boundary layer climates*. London: Routledge.

OpenStreetMap (2023). *Background*. Retrieved from: https://planet.openstreetmap.org [accessed: 31.10.2023].

Peng, J., Dan, Y., Qiao, R., Liu, Y., Dong, J. & Wu, J. (2021). How to quantify the cooling effect of urban parks? Linking maximum and accumulation perspectives. *Remote Sensing of Environment*, *252*, 112135. https://doi.10.1016/j.rse.2020.112135

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J. & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, *20* (9), 503–510. https://doi.10.1016/j.tree.2005.05.011

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Roussel, J. R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R., Meador, A. S., Bourdon, J. F., Boissieu, F. de & Achim, A. (2020). lidR: an R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment*, *251*, 112061. https://doi.10.1016/j.rse.2020.112061

Roy, S., Byrne, J. & Pickering, C. (2012). A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. *Urban Forestry & Urban Greening*, *11* (4), 351–363. https://doi.10.1016/j.ufug.2012.06.006

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9* (7), 671–675. https://doi.10.1038/nmeth.2089

Schwaab, J., Meier, R., Mussetti, G., Seneviratne, S., Bürgi, C. & Davin, E. L. (2021). The role of urban trees in reducing land surface temperatures in European cities. *Nature Communications*, *12* (1), 6763. https://doi.10.1038/s41467-021-26768-w

Setiawan, W. & Gawryszewska, B. (2023). Urban garden as a water reservoir in an urban area – a literature review. *Scientific Review Engineering and Environmental Sciences (SREES)*, *32* (3), 221–237. https://doi.10.22630/srees.4960

Shahidan, M. F., Shariff, M. K., Jones, P., Salleh, E. & Abdullah, A. M. (2010). A comparison of *Mesua ferrea* L. and *Hura crepitans* L. for shade creation and radiation modification in improving thermal comfort. *Landscape and Urban Planning*, *97* (3), 168–181. https://doi.10.1016/j.landurbplan.2010.05.008

Sikorski, P., Gawryszewska, B., Sikorska, D., Chormański, J., Schwerk, A., Jojczyk, A., Ciężkoski, W., Archiciński, P. Łepkowski, M., Dymitryszyn, I., Przybysz, A., Wińska-Krysiak, M., Zajdel, B., Matusiak, J. & Łaszkiewicz, E. (2021). The value of doing nothing – How informal green spaces can provide comparable ecosystem services to cultivated urban parks. *Ecosystem Services*, *50*, 101339. https://doi.10.1016/j.ecoser.2021.101339

Song, J. & Wang, Z. H. (2015). Impacts of mesic and xeric urban vegetation on outdoor thermal comfort and microclimate in Phoenix, AZ. *Building and Environment*, *94*, 558–568. https://doi.10.1016/j.buildenv.2015.10.016

Statistics Poland (2022). *Area and population in the territorial profile in 2022.* Warsaw: Statistics Poland. Retrieved from: https://stat.gov.pl/en/topics/population/population/area-and-population-in-the-territorial-profile-in-2022,4,16.html [accessed: 31.10.2023].

Tengberg, A., Fredholm, S., Eliasson, I., Knez, I., Saltzman, K. & Wetterberg, O. (2012). Cultural ecosystem services provided by landscapes: Assessment of heritage values and identity. *Ecosystem Services*, *2*, 14-26. https://doi.10.1016/j.ecoser.2012.07.006

Trojanowska, M. (2022). Climate change mitigation and preservation of the cultural heritage – a story of the Municipal Park in Rumia, Poland. *Land*, *11* (1), 65. https://doi.10.3390/land11010065

Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A. K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koch, L. P., Leimgruber, P., Taubenboeck, H., Wegmann, M., Wikelski, M. & Woodcock, C. (2015). Free and open-access satellite data are key to biodiversity conservation. *Biological Conservation*, *182*, 173–176. https://doi.10.1016/j.biocon.2014.11.048

Voogt, J. A. & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment*, *86* (3), 370–384. https://doi.10.1016/S0034-4257(03)00079-8

Wang, Z., Yang, Z., Shi, H., Han, F., Liu, Q., Qi, J. & Lu, Y. (2020). Ecosystem health assessment of world natural heritage sites based on remote sensing and field sampling verification: Bayanbulak as case study. *Sustainability*, *12* (7), 2610. https://doi.10.3390/su12072610

Weng, Q., Lu, D. & Schubring, J. (2004). Estimation of land surface temperature – vegetation abundance relationship for urban heat island studies. *Remote sensing of Environment*, *89* (4), 467–483. https://doi.10.1016/j.rse.2003.11.005

Xu, C., Qu, J. J., Hao, X., Zhu, Z. & Gutenberg, L. (2020). Surface soil temperature seasonal variation estimation in a forested area using combined satellite observations and in-situ measurements. *International Journal of Applied Earth Observation and Geoinformation*, *91*, 102156. https://doi.10.1016/j.jag.2020.102156

Zawojska, E., Szkop, Z., Czajkowski, M. & Żylicz, T. (2016). Economic valuation of ecosystem services provided by the Wilanów Park. A benefit transfer study. *Economics and Environment*, *59* (4), 11–11. Retrieved from: https://www.ekonomiaisrodowisko.pl/journal/article/view/207 [accessed: 31.10.2023].

Summary

Application of UAV and ground measurements for cooling benefit assessment, the Wilanów Palace case study. This research at the Wilanów Palace, Warsaw, assesses urban greenery's cooling impacts in a cultural heritage site using remote sensing and on-site measurements, highlighting vegetation's importance in urban climate control. The study combines soil temperature data, UAV thermal imagery, leaf area index (LAI), LiDAR, and NDVI analyses. Findings demonstrate a strong link between vegetation density and temperature: UAV land surface temperature (LST) ranged from 26.8° to 47.5°C, peaking at 72°C, while ground-based temperatures were between 19.5° and 29.2°C, lowest in dense vegetation areas. The statistical analysis confirmed significant temperature differences across vegetation types, with higher LAI areas showing lower temperatures. These results validate the cooling effect of dense vegetation, emphasizing green spaces' significance in urban climate regulation within cultural heritage sites. The study informs sustainable urban design and conservation, underlining the critical role of vegetation in improving urban microclimates.

Scientific Review Engineering and Environmental Sciences (2024), 33 (1), 69–94

Sci. Rev. Eng. Env. Sci. (2023), 33 (1) https://srees.sggw.edu.pl

ISSN 1732-9353 (suspended) eISSN 2543-7496 DOI 10.22630/srees.5258

Received: 02.09.2023
Accepted: 19.01.2024

Sachin Dadu KHANDEKAR[1]✉

Dinesh Shrikrishna ASWAR[2]

Pandurang Digamber SABALE[3]

Varsha Sachin KHANDEKAR[4]

Mohankumar Namdeorao BAJAD[5]

Shivakumar KHAPLE[6]

[1, 2, 5]Savitribai Phule Pune University, Sinhgad College of Engineering, Maharashtra, India
[3]Deccan College Deemed University, Senior Professor and Head, Maharashtra, India
[4]Analytics-Analyst, Cummins Global Services & Analytics India Office Campus, Pune (Maharashtra), India
[6]JSPM Narhe Technical Campus, Narhe, SPPU, Maharashtra, India

Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station)

Keywords: wavelet transform, artificial neural network, multiple linear regression, streamflow, Daubechies wavelet, time series

Introduction

For water authorities to effectively manage water reserves for different water users like hydropower generation, agricultural, domestic, flood management, etc., it is necessary to forecast river flow hours, days, months, or possibly longer in advance. Large volumes of dynamic, non-linear, and noisy data can be handled effectively using soft computing approaches, especially when the underlying physical relations are not understood.

✉khandekarsd@gmail.com

Khandekar, S. D., Aswar, D. S., Sabale, P. D., Khandekar, V. S., Bajad, M. N., Khaple, S. (2024). Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station). *Sci. Rev. Eng. Env. Sci.*, *33* (1), 69–94. DOI 10.22630/srees.5258

For rainfall-runoff modeling and runoff forecasting, neural network models are successfully applied and can be found in the literature (Dawson & Wilby, 1998; Jain & Chalisgaonkar, 2000; Raghuwanshi, Singh & Reddy, 2006; Tayfur & Singh, 2006; Besaw, Rizzo, Bierman & Hackett, 2010; Rezaeian Zadeh, Amin, Khalili & Singh, 2010). The ASCE Task Committee (2000) reviews hydrologic applications of artificial neural network (ANN). Solaimani (2009) utilized ANN for modeling the rainfall runoff relationship in a catchment area located in a semiarid region of Iran by adopting feed forward back propagation for the rainfall forecasting with various algorithms with performance of multi-layer perceptions. The monthly stream flow of Jarahi watershed was analyzed to calibrate the given models. The monthly hydrometric and climatic were ranged from 1969 to 2000. The results extracted from the comparative study indicated that ANN was more appropriate and efficient to predict the river runoff than the classic regression model. Using the feed-forward back propagation neural network (FFBPNN) and the cascade forward back propagation neural network (CFBPNN) models, Mohseni and Muskula (2023) created rainfall--runoff-based models in the Yerli sub-catchment of the upper Tapi basin using data from 36 years from 1981 to 2016 and concluded that the developed ANN model was capable of predicting runoff quite accurately.

Wavelet transform (WT) has gained popularity recently as a helpful method for examining trends, periodicities, and variations in time series. Similar to the Fourier transform and short-time Fourier transform, a WT is a powerful mathematical tool for signal processing that can analyze both stationary and non-stationary data and produce time and frequency information at a higher resolution that is not possible with the former. The WT offers a multiresolution analysis; that is, at low scales (high frequency), it provides poor frequency resolution and better time resolution; at high scales (low frequency), it provides poor frequency resolution and higher time resolution. Such information is significant in real practice for all time-series signals. The WT breaks down a non-stationary time series into a specific number of stationary time series. The WT is then used to integrate other single prediction techniques to increase prediction accuracy. Another trustworthy hybrid model for application in time series forecasting issues is the wavelet-ANN (WANN). To estimate runoff discharge for the Ligvanchai watershed in Tabriz, Iran, Nourani, Komasi and Mano (2009) explored the rainfall-runoff modeling utilizing the wavelet-ANN technique. The time series were decomposed up to four levels using Haar, Daubechies (db2), Symlet (sym3), and Coiflet (coif1) wavelets. The model's outcomes demonstrated the great quality of the Haar wavelet when compared to the others. Kisi (2009b) proposed neuro-wavelet model for forecasting daily intermittent stream flow. Forecasting accuracy of neuro-wavelet model was better than single ANN model. To predict the flow of the Malbrabha river in India, Nayak, Venkatesh, Krishna and Jain (2013) developed a wavelet neural network (WNN) hybrid model employing the db5 wavelet for one time step ahead forecasting. According to the results, the WANN model outperformed the ANN model. Nourani, Baghanam, Adamowski and Gebremichael (2013) utilized feed forward neural networks (FFNN) to model the rainfall-runoff process on a daily and multi-step (two days, three days and four days) ahead time scale. Authors used db4 and Haar wavelets to remove noise from runoff time series and discovered that the performance of the FFNN was improved by applying WT to the raw runoff data. Also, it was concluded that db4 wavelet exhibits superior results as compared to Haar wavelet. Shafaei and Kisi (2017) compared the WANN model with ANN and SVM for prediction of short-term daily river flow of Ajichai river of Iran. A db5 mother wavelet was used to decompose the raw data and concluded that the WANN model performed better than ANN and SVM. Wang et al. (2022) developed hybrid models by combining the wavelet theory with five diverse types of machine learning models such as support vector machine (SVM)-radial basis function, SVM-polynomial, decision tree, gradient boosting, random forest, and long- or short-term memory. The db4 mother wavelet was used in the study. A comparison revealed that hybrid models exhibit better estimates than the stand-alone ones. To estimate monthly stream flows in Amasya, Turkey, Katipoğlu (2023a) combined a discrete wavelet transform and a feedforward backpropagation neural network. The author decomposed various meteorological data using different wavelets such as Haar, Daubechies 2, Daubechies 4, Discrete Meyer, Coiflet 3, Coiflet 5, Symlet 3, and Symlet 5, and found that the results of Coiflet 5 was superior. Katipoğlu (2023b) developed wavelet-ANFIS (W-ANFIS) model using db10 mother wavelet for predicting monthly Bitlis river flows in Turkey and concluded that the W-ANFIS model proved successful. Other studies on runoff prediction using wavelet-ANN are found in literature (Rao & Krishna, 2009; Adamowski & Sun, 2010; Linh et al., 2021; Kumar, Kumar, Kumar, Elbeltagi & Kuriqi, 2022). Nourani, Baghanam, Adamowski and Kisi (2014), Khandekar and Deka (2016) review the application of WT in hydrology. Application of WT in drought prediction, groundwater prediction, evaporation prediction can be found in the literature (Djerbouai & Souag-Gamane, 2016; Patil & Deka, 2017; Araghi, Adamowski & Martinez, 2020; Katipoğlu, 2023c; Katipoğlu, 2023d; Katipoğlu, 2023e).

In time series forecasting issues, the wavelet-MLR (WMLR) is another trustworthy hybrid model. Kisi (2010) combined discrete wavelet transform and linear regression (WR) for short-term stream flow forecasting of two stations in Turkey. In comparison to ANN and ARMA models, WR models are found to be more superior. Kisi (2011) proposed a wavelet regression (WR) model for daily river stage forecasting of two stations on the Schuylkill river in Philadelphia. The result of a WR model was superior to the ANN models. Zhang, Zhang and Sing (2018) developed four models – MLR, ANN, wavelet coupled with MLR (W-MLR) and ANN (W-ANN) – for stream flow forecasting at four stations in the East River basin in China. All models showed similar performance in forecasting stream flow one-day ahead, while W-MLR and W-ANN performed better in five-day ahead forecasting. Other studies on runoff prediction using wavelet-MLR is found in literature (Kisi, 2009a; Budu, 2013; Shoaib et al., 2018; Khazaee Poul, Shourian & Ebrahimi, 2019).

Most previous investigations employed a selected mother wavelet type. Also, more research is required to fully understand the potential of Daubechies wavelets of different orders in studying hydrologic time series behavior. So, in this study, it is suggested to create hybrid models by coupling wavelet transform with ANN and MLR, with the following objectives:

1. To compare the results of all hybrid models (WANN, WMLR) developed using db1, db2, db3, db8 and db10 Daubechies wavelets for multiple lead times (two days, four days, seven days, 14 days). Also, to compare results of hybrid models with standalone ANN and MLR models.
2. To investigate the effect of Daubechies wavelets db1, db2, db3, db8, and db10 on forecasting accuracy.
3. To investigate how decomposition level affects model effectiveness.

Study area and data collection

With an average discharge at its mouth of 19,830 cumec, Brahmaputra is the fourth-largest river in the world (Goswami, 1985). Brahmaputra river Pancharatna station is selected for the study. Ten-year (Jan 1990 – Dec 1999) daily flow data were collected from Water Resources Department, Assam, India. The catchment area up to Pancharatna station is 532,000 km^2. The seasonal monsoon rhythm and the Himalayan snow's freeze-thaw cycle influence the river's hydrologic regime. During the flood season, there are noticeable large changes in discharge over a short period of time. A maximum differential of roughly 17,000 m$^3 \cdot$s^{-1} in 24 h (June 7–8, 1990) and 24,000 m$^3 \cdot$s^{-1} in 48 h (June 7–9, 1990) was observed in rising limb (Sarma, 2005). The location of Pancharatna station is shown in Figure 1.

Figure 2 displays the observed time series discharge at Pancharatna station. Figure 2 demonstrates that the discharge is remarkably non-stationary, particularly during monsoon season (from June to September, each year). The contribution of flow from snow melting, during February to April, causes the discharge in the rising limb to fluctuate as well.

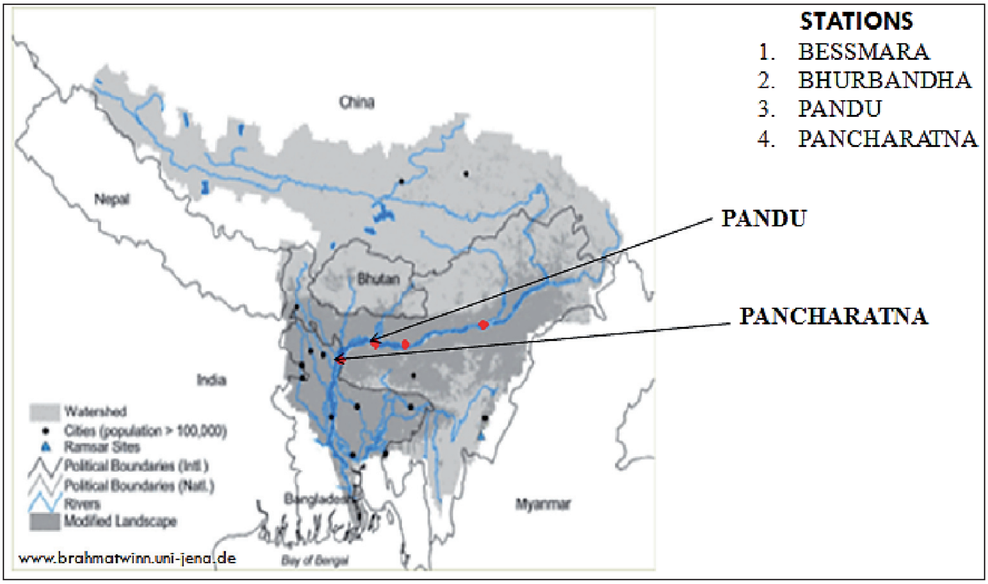


FIGURE 1. Location of the gauging sites
Source: Khandekar (2014).

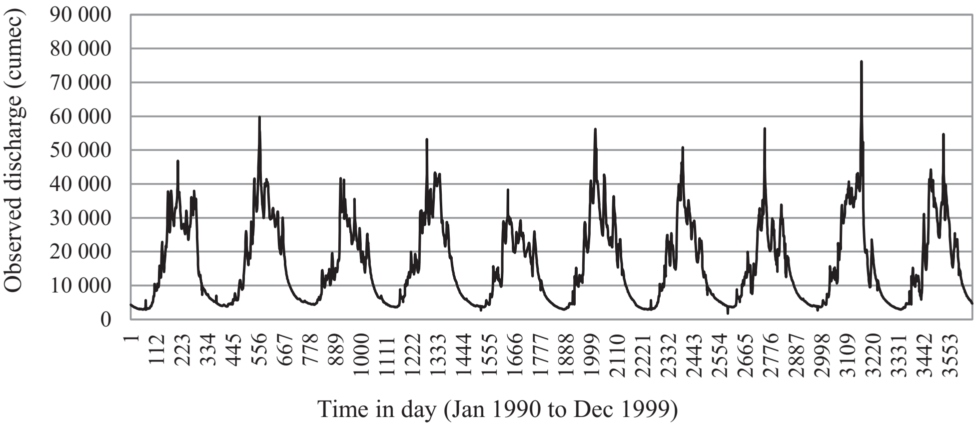


FIGURE 2. Observed flow series
Source: Khandekar (2014).

Table 1 displays the statistical properties of flow data and demonstrates substantial variability. The terms Q_{mean}, Q_{max}, Q_{min}, S_d, and C_x in the table stand for the mean, maximum, minimum, standard deviation, and skewness, respectively.

Khandekar, S. D., Aswar, D. S., Sabale, P. D., Khandekar, V. S., Bajad, M. N., Khaple, S. (2024). Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station). *Sci. Rev. Eng. Env. Sci.*, *33* (1), 69–94. DOI 10.22630/srees.5258

TABLE 1. Statistical properties of flow data

Statistical parameter	Training	Testing	All
Q_{mean} [m^3·s^{-1}]	16 159	16 236	16 161
Q_{max} [m^3·s^{-1}]	59 832	76 236	76 236
Q_{min} [m^3·s^{-1}]	2 628	1 723	1 723
S_d [m^3·s^{-1}]	11 783	12 388	11 965
C_x	0.726	0.968	0.809

Source: Khandekar (2014).

The discharge throughout the study period exhibits significant fluctuations, as shown in Table 1 (minimum = 1,723 cumec, the highest = 76,236 cumec). The standard deviation is determined to be 11,965 cumec, showing a significant dispersion of values from the mean. Additionally, the observed flows show significant positive coefficient of skewness (C_x = 0.809), indicating that the data has a more scattered distribution about the mean.

Methodology

Wavelet transform (WT)

Wavelet theory is discussed thoroughly in works by Mallat (1998), and Labat, Ababou and Mangin (2000). Wavelet transform of a raw signal has the ability of providing time-frequency, using a range of window sizes. It divides the input signal into wavelets (small waves), which are scaled and shifted versions of the original wavelet, called mother wavelet. There are two forms of wavelet transformation: the continuous wavelet transform (*CWT*) and the discrete wavelet transform (*DWT*). The *CWT* of a signal $x(t)$ is described:

$$CWT\left(a,b\right)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}x\left(t\right)\cdot\psi^{*}\left(\frac{t-b}{a}\right)\cdot \mathrm{d}t. \tag{1}$$

In Eq. (1), the transformed signal is a function of two variables, *a* and *b*, which represent the scale and translation factor of the function $\psi(t)$, respectively; * corresponds to complex conjugate (Mallat, 1998).

The $\psi(t)$ is the transforming function, called as mother wavelet, is mathematically defined as:

$$\int_{-\infty}^{\infty}\psi\left(t\right)\mathrm{d}t=0. \tag{2}$$

From a data set of length (L), *CWT* produces L^2 coefficients. Therefore, duplicate data is trapped within the coefficients, which may or may not be a desired quality (Nourani et al., 2009; Rajaee, Nourani, Zounemat-Kermani & Kisi, 2011). The analysis will be significantly more precise and efficient, resulting in *N* transform coefficients, if scales and positions are chosen based on the powers of two. The resulting transform is called DWT, and has the form:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_o^m}} \psi\left(\frac{t - nb_o a_o^m}{a_o^m}\right), \tag{3}$$

where the wavelet translation and dilation are each controlled by an integer, m, and the other by n; b_o which must be larger than 0, is the location parameter; a_o is a specified fixed dilation step greater than 1. The most common and simplest choice for parameters a_o and b_o are 2 and 1 (time steps), respectively. This power of two logarithmic scaling of the translations and dilations is known as the dyadic grid arrangement (Mallat, 1989). For parameters a_o and b_o, 2 and 1 (time steps), respectively, are the most typical options. According to Mallat (1989), this power-of-two logarithmic scaling of the translations and dilations is referred to as the dyadic grid layout.

Two sets of functions, known as high-pass (wavelet function) and low-pass (scaling function) filters, are operated by DWT. The original time series undergo processing via high-pass and low-pass filters (as shown in Fig. 3), followed by a down-sampled process that discards every second data point (Deka & Prahlada, 2012). The detailed (D_1, D_2, ..., D_n) and approximation (A_1, A_2, ..., A_n) coefficients, which represent the low frequency and high frequency components of the original signal, respectively, are derived after the signal is passed through high pass and low pass filters. There will be a total of $n + 1$ coefficients at every given n^{th} decomposition level, where there will be one series of approximation coefficients at the n^{th} level (A_n) and n series of detailed coefficients (D_1, D_2, ..., D_n). The sum of $A_n + D_1 + D_2 + ... + D_n$ is equal to the original signal.

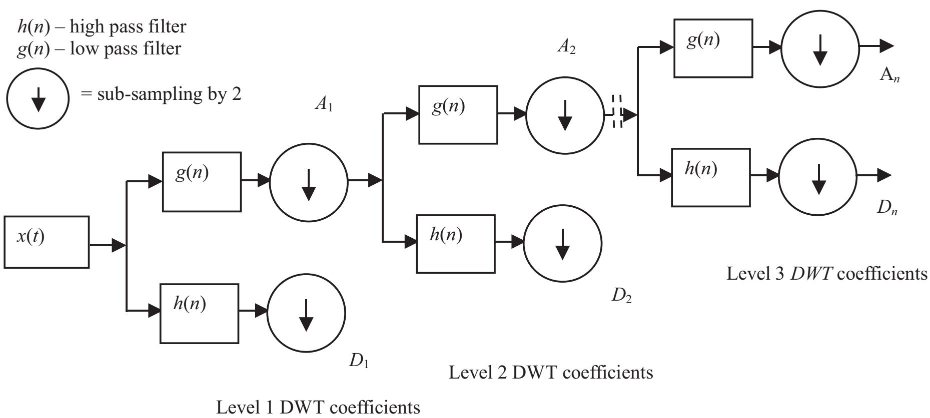


FIGURE 3. Wavelet decomposition tree
Source: Khandekar (2014).

Artificial Neural Network (ANN)

Neural networks are interconnected groups of artificial neurons that may be utilized as a computational model for information processing based on the connectionist approach to computing. An ANN may be thought of mathematically as a universal approximator with the capacity to learn from instances without the need of explicit physics.

Three-layer-feedforward artificial neural networks are most frequently utilized in hydrologic time series modeling (Jain & Chalisgaonkar, 2000; Raghuwanshi, Singh & Reddy, 2006; Tayfur & Singh, 2006; Bajirao, Kumar, Kumar, Elbeltagi & Kuriqi, 2021; Katipoğlu, 2023a). In the current study, the Lavenberg–Marquardt (LM) learning function and tangent sigmoid as the transfer function were employed in the feedforward ANN. The LM technique was used to train the ANN since it is more effective and faster than the traditional gradient descent technique (Kisi, 2009b; Mohseni & Muskula, 2023). In three-layer-feedforward ANN, information flow takes place from the input to the output side. Weights regulate the strength of a signal that passes across connections between neurons, altering the information that is transferred between them. While not connected to nodes in the same layer, nodes in one layer are connected to those in the next one. As a result, a node's output in a layer depends exclusively on the inputs and weights it receives from the previous layers. Every node multiplies each input by its weight, adds up the result, and then runs the total through a transfer function to get the desired outcome. Typically, this transfer function is represented as a sigmoid function, which is an S-shaped curve that increases continuously. The "S" upper and lower limb attenuation keeps the raw sums smoothly within predetermined bounds. The transfer function additionally adds a nonlinearity that improves the network's capacity to represent complicated functions (Jain & Chalisgaonkar, 2000). The sigmoid function is monotonically growing, continuous, and differentiable everywhere.

Multiple linear regression (MLR)

The MLR is one of the statistical methods which represents a mathematical equation expressing the relation between two or more explanatory variable and a response variable. The MLR attempts to model the link between two or more explanatory variables and a response variable by fitting a linear equation to the observed data (Sharma, Isik, Srivastava & Kalin, 2013). The relationship between the dependent variable y and the independent variable x is described by the MLR equation as follows:

$$y = A_0 + A_1 x_1 + A_2 x_2 + \cdots + A_n x_n \tag{4}$$

where: x_n is the value of n^{th} predictor, A_0 is the regression constant, A_i is the coefficient of n^{th} predictor.

Model development

Input selection in model

For deciding the number of input parameters, no fixed rule is available. The values of the future time step (Q_{t+n}) in any time-series forecasting are inevitably reliant on the antecedent values Q_t, Q_{t-1}, Q_{t-2}, …, Q_{t-j} (Table 2). But because the value of j (lag) is not known beforehand, it is difficult to determine how many lags in the past will lead to higher efficiency. In hydrologic time series forecasting, determining j is crucial since it can help to reduce information loss and the omission of critical input variables that could interfere with training.

TABLE 2. Correlation coefficients for the flow series

Output* (Q_{t+n})	Input (Q_{t-j})*									
	Q_t	Q_{t-1}	Q_{t-2}	Q_{t-3}	Q_{t-4}	Q_{t-5}	Q_{t-6}	Q_{t-7}	Q_{t-8}	Q_{t-9}
Q_{t+2}	0.983	0.971	0.957	0.942	0.929	0.915	0.903	0.891	0.880	0.871
Q_{t+4}	0.957	0.942	0.928	0.915	0.902	0.891	0.880	0.870	0.862	0.854
Q_{t+7}	0.915	0.902	0.891	0.880	0.870	0.862	0.854	0.846	0.839	0.832
Q_{t+14}	0.846	0.839	0.832	0.825	0.818	0.812	0.805	0.799	0.792	0.784

*n – lead time; j – lag.

Source: Khandekar (2014).

Numerous researchers (Sudheer, Gosain & Ramsastri, 2002; Budu, 2013; Nayak et al., 2013; Khazaee Poul et. al., 2019; Sun, Niu & Sivakumar, 2019; Katipoğlu, 2023a, 2023b, 2023c) have used the technique based on the statistical features such as cross-correlation, autocorrelation and partial correlation of the data series in order to identify a distinct input vector. In the current research, based on the autocorrelation coefficient between the relevant variables (shown in Table 2), the input vectors to the models are chosen. To forecast the value of discharge, based on the following various input combinations were taken into account (Khandekar, 2014):

1. Q_t.
2. Q_t, $Q_{(t-1)}$.
3. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$.

Khandekar, S. D., Aswar, D. S., Sabale, P. D., Khandekar, V. S., Bajad, M. N., Khaple, S. (2024). Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station). *Sci. Rev. Eng. Env. Sci.*, *33* (1), 69–94. DOI 10.22630/srees.5258

4. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$.
5. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$, $Q_{(t-4)}$.
6. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$, $Q_{(t-4)}$, $Q_{(t-5)}$.

ANN and MLR model development

At the first stage, to forecast discharge, ANN and MLR models without data preprocessing were created. In the present study it was decided to predict discharge for multiple lead times such as two days, four days, seven days and 14 days because the Brahmaputra river is one of the major rivers in India which carries heavy flood during monsoon season, so it is necessary to design flood warning system well in advance. First 70% data was used to train the model and last 30% data was used for testing the model. Initially, a three-layer feedforward backpropagation ANN were applied in the study. For each lead time, an optimal input combination is obtained by providing each of one to six input combination obtained through autocorrelation as input to ANN and by varying number of neurons in the hidden layer from two to 10 through trial and error (Budu, 2013; Moosavi, Vafakhah, Shirmohammadi & Behnia, 2013). The optimal input combination that gave the minimum root mean square error (*RMSE*) during testing period for each lead time is shown in Table 3. The output layer has only one neuron which is the discharge value for the given lead time. From viewpoint of comparing all models, same input combination is employed for the MLR and all hybrid models. The time series data normalized between zero and one (Nourani et al., 2009) by dividing the discharge value by the maximum one.

TABLE 3. Optimal input combination

Lead time [day]	Input parameter*	Output parameter** $Q_{(t+n)}$
2	Q_t, $Q_{(t-1)}$	$Q_{(t+2)}$
4	Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+4)}$
7	Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+7)}$
14	Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+14)}$

*Q_t – current discharge value; $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$ – one-, two- and three-time step past discharge values; **n – lead time.

Source: Khandekar (2014).

Development of hybrid models (WANN and WMLR)

After developing the ANN and MLR models, hybrid models were developed. First of all, using DWT, for developing hybrid models, the normalized input data were decomposed into approximation and detail coefficients. Since all hydrological data are observed at discrete time intervals, all hybrid models used DWT to process time series data in the form of approximations and details at various levels so that gross and small features of a signal can be separated (Deka & Prahlada, 2012). Then, to obtain the output at a predetermined lead time, the approximation and detail coefficient were fed as input to ANN and MLR. Without decomposition, the output signals were preserved as normalized original series. The complete flow chart of model development is shown in Figure 4.

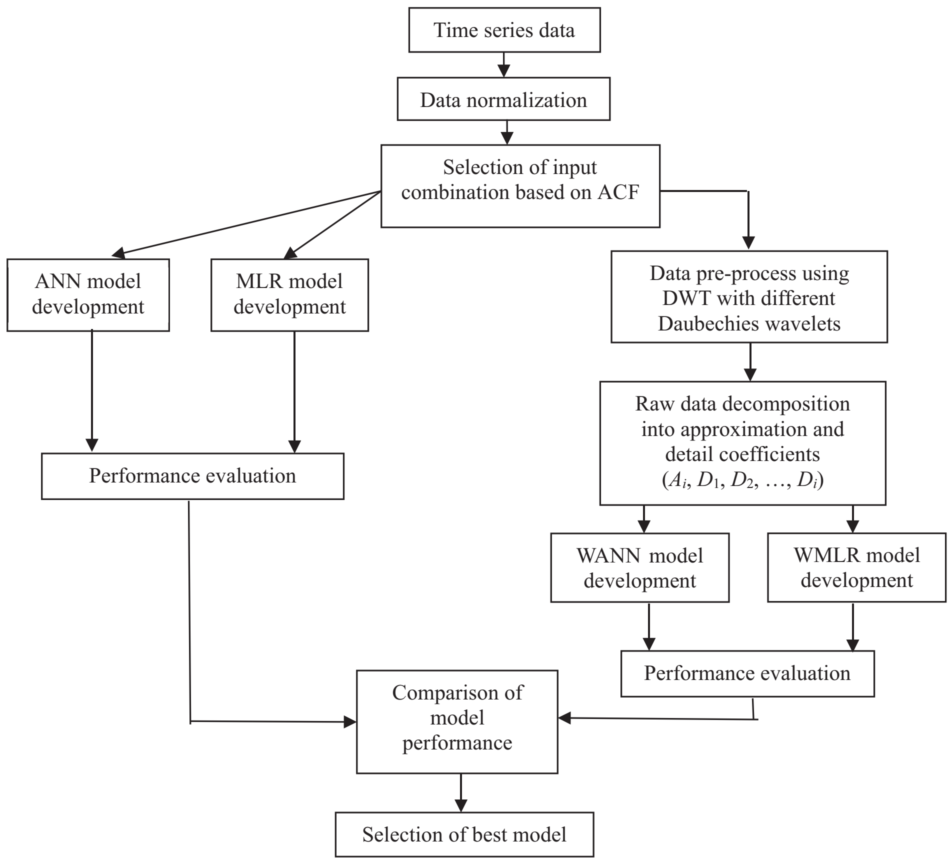


FIGURE 4. Flow chart of model development
Source: own elaboration.

Selection of mother wavelets and decomposition level

The mother wavelet that will be used depends on the data to be analyzed (Nayak et al., 2013). As per authors' knowledge, there is no comparative study on prediction of hydrological parameter using group of Daubechies wavelets. So, in this research,

Khandekar, S. D., Aswar, D. S., Sabale, P. D., Khandekar, V. S., Bajad, M. N., Khaple, S. (2024). Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station). *Sci. Rev. Eng. Env. Sci.*, *33* (1), 69–94. DOI 10.22630/srees.5258

an irregular wavelet, Daubechies (db) of orders 1 (db1), 2 (db2), 3 (db3), 8 (db8), and 10 (db10), illustrated in Figure 5, was chosen to deal with very irregular signal form. Daubechies wavelets of order N (db$N$) are all asymmetric, orthogonal, and biorthogonal. They are compactly supported wavelets with extremal phase and highest number of vanishing moments for a given support width (Misiti, Misiti, Oppenheim & Poggi, 2010). Daubechies wavelet has a support width of $2N-1$. Except for db1 (haar), no Daubechies wavelet has an explicit expression. Wavelets with compact support or a narrow window function are appropriate for a local analysis of the signal.

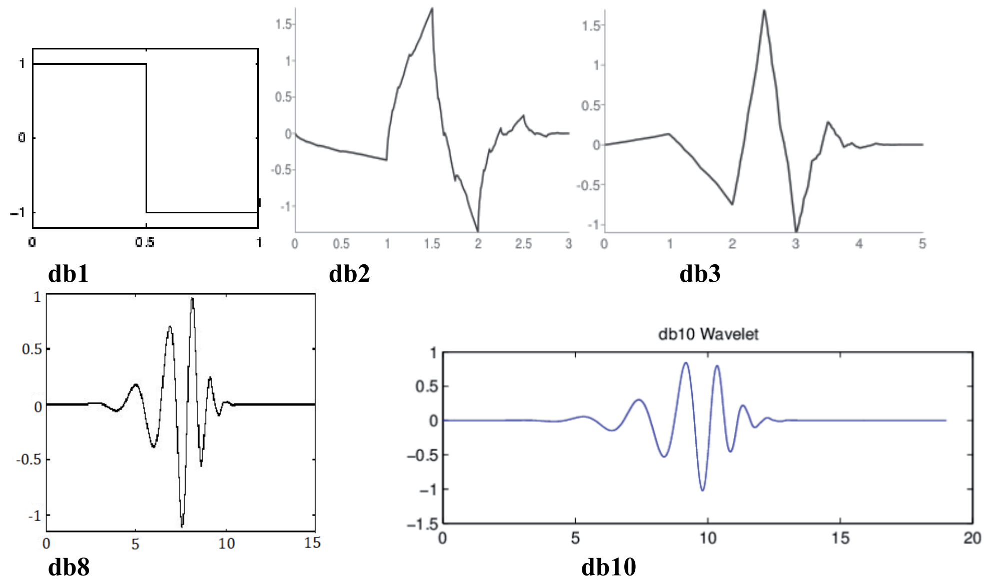


FIGURE 5. Daubechies wavelets

Source: Khandekar (2014).

The optimal decomposition level was obtained using the formula $l = \text{int}\,[\log(L)]$, (Nourani et al., 2014; Khazaee et al., 2019; Tarate et al., 2021; Katipoğlu, 2023a, 2023b, 2023c), where l is the decomposition level, L represents the number of time series data, *int* represents the integer part function, and log represents logarithm with base 10. In the present study, L is 3,650 so l is approximately four. However, in our study, we decomposed raw signal up to fifth level. At any l^{th} decomposition level, DWT produces l detailed coefficients and one approximation coefficient at l^{th} level.

Performance criteria

The following evaluation measures were used to compare model performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{L}(Q_{\mathrm{obs}} - Q_{\mathrm{com}})^2}{L}}, \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{L}(Q_{\mathrm{obs}} - Q_{\mathrm{com}})^2}{\sum_{i=1}^{L}(Q_{\mathrm{obs}} - \overline{Q}_{\mathrm{obs}})^2}, \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{L}\left|Q_{\mathrm{obs}} - Q_{\mathrm{com}}\right|, \tag{7}$$

$$B = \frac{\sum_{i=1}^{L} Q_{\mathrm{comp}}}{\sum_{i=1}^{L} Q_{\mathrm{obs}}}, \tag{8}$$

$$SI = \frac{RMSE}{\sum_{i=1}^{L} \overline{Q}_{\mathrm{obs}}}, \tag{9}$$

where: $RMSE$, R^2, MAE, B, SI, L, Q_{obs}, Q_{com}, and $\overline{Q}_{\mathrm{obs}}$ are root-mean squared error ($RMSE$), determination coefficient (R^2), mean absolute error (MAE), bias (B), scatter index (SI), number of observations (L), observed data (Q_{obs}), computed values (Q_{com}), mean of observed data ($\overline{Q}_{\mathrm{obs}}$), respectively. A R^2 of 0.9 or more is considered very satisfactory, 0.8 to 0.9 represents a fairly good model, and less than 0.8 is considered unsatisfactory (Dawson & Wilby, 2001).

Results and discussion

ANN and MLR model results

The results of all models are presented in Tables 4–7. It can be seen from Tables 4–7 that in comparison with MLR models, the ANN model has performed better for all lead times, except for the fourteen-day one. It is observed that the effectiveness of both models decreases while increasing lead time. This could be due to considerable fluctuations in the data around mean values, such as a high standard

Khandekar, S. D., Aswar, D. S., Sabale, P. D., Khandekar, V. S., Bajad, M. N., Khaple, S. (2024). Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station). *Sci. Rev. Eng. Env. Sci.*, *33* (1), 69–94. DOI 10.22630/srees.5258

deviation. Tables 4–7 also indicates the optimal ANN structure (e.g. for a two-day lead time, 2-8-1 means two neurons in the input layer, eight neurons in the hidden layer, and one neuron in the output layer).

TABLE 4. Values of statistical parameters for two-day lead time

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	*MAE*	*BIAS*	*SI*	*RMSE*	R^2	*MAE*	*BIAS*	*SI*	–
ANN	1 764.66	0.977	1 044.28	1.002	0.109	2 463.33	0.960	1 401.83	1.014	0.152	2-8-1
MLR	1 828.26	0.976	1 058.12	1.000	0.113	2 535.96	0.958	1 293.19	1.000	0.156	–
WANN-db1*l*5	1 199.43	0.989	757.47	1.008	0.074	1 795.99	0.979	1 144.51	1.018	0.110	12-2-1
WMLR-db1*l*5	1 228.25	0.989	680.19	1.000	0.076	1 758.64	0.980	913.41	1.000	0.109	–
WANN-db2*l*4	960.30	0.993	540.14	1.001	0.059	1 415.36	0.987	796.43	1.007	0.087	10-2-1
WMLR-db2*l*5	1 093.66	0.991	595.74	1.000	0.068	1 478.74	0.986	769.43	0.999	0.091	–
WANN-db3*l*4	813.84	0.995	469.56	1.001	0.050	1 202.33	0.990	698.70	1.007	0.074	10-3-1
WMLR-db3*l*5	909.81	0.994	510.23	1.000	0.056	1 311.38	0.989	629.76	0.999	0.081	–
WANN-db8*l*5	553.07	0.998	368.93	1.000	0.034	1 054.10	0.992	543.92	1.000	0.065	12-2-1
WMLR-db8*l*5	526.83	0.998	296.78	1.000	0.033	775.11	0.996	406.20	0.999	0.048	–
WANN-db10*l*5	481.40	0.998	298.51	0.999	0.030	933.06	0.994	436.49	0.999	0.057	12-2-1
WMLR-db10*l*5	471.99	0.998	261.79	1.000	0.029	751.87	0.996	369.69	1.000	0.046	–

Note: *MAE* and *RMSE* are $m^3 \cdot s^{-1}$ unit (Results of only optimum decomposition level are shown).

Source: own elaboration.

TABLE 5. Values of statistical parameters for four-day lead time

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	*MAE*	*BIAS*	*SI*	*RMSE*	R^2	*MAE*	*BIAS*	*SI*	
ANN	3 133.09	0.929	2 067.41	1.005	0.194	3 890.35	0.901	2 566.07	1.031	0.239	4-8-1
MLR	3 136.83	0.929	1 988.84	1.000	0.194	3 932.06	0.899	2 274.72	1.000	0.243	–
WANN-db1*l*5	1 918.22	0.973	1 229.16	1.008	0.118	2 590.84	0.956	1 717.58	1.018	0.159	24-3-1
WMLR-db1*l*5	2 088.67	0.968	1 268.59	1.000	0.129	2 909.44	0.945	1 618.86	1.000	0.179	–
WANN-db2*l*5	1 431.04	0.985	868.66	1.001	0.088	1 986.38	0.974	1 137.20	1.009	0.122	24-2-1
WMLR-db2*l*4	1 922.49	0.973	1 160.03	1.000	0.119	2 342.13	0.964	1 354.14	0.999	0.144	–
WANN-db3*l*5	1 280.06	0.988	790.25	1.001	0.079	1 683.59	0.982	1 034.97	1.004	0.104	24-2-1
WMLR-db3*l*5	1 528.75	0.983	924.38	1.000	0.094	2 067.67	0.972	1 084.63	0.999	0.127	–
WANN-db8*l*4	962.64	0.993	601.76	1.001	0.059	1 465.35	0.986	774.11	1.000	0.090	20-2-1
WMLR-db8*l*5	940.03	0.994	556.62	1.000	0.058	1 213.32	0.990	664.47	0.999	0.075	–
WANN-db10*l*5	822.67	0.995	527.85	1.003	0.051	1 477.12	0.986	755.19	1.004	0.091	24-2-1
WMLR-db10*l*5	785.37	0.995	460.22	1.000	0.048	1 174.80	0.991	612.69	1.000	0.072	–

Note: *MAE* and *RMSE* are in $m^3 \cdot s^{-1}$ unit (Results of only optimum decomposition level are shown).

Source: own elaboration.

TABLE 6. Values of statistical parameters for seven-day lead time

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	*MAE*	*BIAS*	*SI*	*RMSE*	R^2	*MAE*	*BIAS*	*SI*	
ANN	4 308.72	0.866	2 801.05	1.002	0.266	5 185.33	0.825	3 441.99	1.032	0.318	4-3-1
MLR	4 490.92	0.854	2 986.84	1.000	0.277	5 376.87	0.811	3 429.75	0.999	0.331	–
WANN-db1*l*5	2 830.19	0.942	1 961.30	1.004	0.175	4 131.69	0.889	2 942.38	1.016	0.254	24-3-1
WMLR-db1*l*5	3 122.62	0.929	1 942.73	1.000	0.193	4 123.65	0.889	2 415.44	0.999	0.254	–
WANNdb2*l*5	2 038.56	0.970	1 408.02	1.004	0.126	2 894.02	0.945	1 927.13	1.024	0.178	24-2-1
WMLR-db2*l*5	2 623.55	0.950	1 617.63	1.000	0.162	3 193.40	0.933	1 937.76	0.999	0.196	–
WANN-db3*l*5	1 683.52	0.979	1 151.42	1.003	0.104	2 525.32	0.958	1 639.74	1.013	0.155	24-3-1
WMLR-db3*l*5	2 129.55	0.967	1 336.34	1.000	0.132	2 987.57	0.942	1 640.06	0.999	0.184	–
WANN-db8*l*5	1 313.63	0.987	871.91	0.999	0.081	2 038.33	0.973	1 233.97	0.997	0.125	24-2-1
WMLR-db8*l*5	1 318.74	0.987	837.08	1.000	0.082	1 702.96	0.981	1 042.41	0.999	0.105	–
WANN-db10*l*5	1 187.51	0.989	792.06	0.999	0.073	1 929.58	0.975	1 060.01	1.005	0.119	24-2-1
WMLR-db10*l*5	1 167.65	0.990	733.49	1.000	0.072	1 585.02	0.984	881.76	1.000	0.097	–

Note: *MAE* and *RMSE* are in $m^3 \cdot s^{-1}$ unit (Results of only optimum decomposition level are shown).

Source: own elaboration.

TABLE 7. Values of statistical parameters for 14-day lead time

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	*MAE*	*BIAS*	*SI*	*RMSE*	R^2	*MAE*	*BIAS*	*SI*	
ANN	5 704.97	0.765	4 035.67	1.014	0.353	7 084.21	0.673	5 116.37	1.068	0.433	4-2-1
MLR	4 584.37	0.848	3 204.37	1.000	0.283	5 415.77	0.809	3 630.94	0.999	0.333	–
WANN-db1*l*5	4 173.09	0.874	2 765.27	1.007	0.258	4 913.55	0.843	3 425.92	1.036	0.300	24-4-1
WMLR-db1*l*5	4 761.19	0.836	3 215.26	1.000	0.294	6 050.48	0.762	3 886.88	0.999	0.371	–
WANN-db2*l*5	3 071.96	0.932	2 062.78	1.004	0.190	4 716.30	0.855	3 017.38	1.038	0.288	24-3-1
WMLR-db2*l*5	4 037.44	0.882	2 571.67	1.000	0.249	4 876.84	0.845	3 109.41	0.999	0.299	–
WANN-db3*l*5	2 420.95	0.958	1 737.07	1.014	0.149	3 848.56	0.904	2 663.88	1.031	0.235	24-2-1
WMLR-db3*l*5	3 172.17	0.927	2 083.46	1.000	0.196	4 770.31	0.852	2 694.26	0.999	0.292	—
WANN-db8*l*5	1 922.74	0.973	1 353.08	1.012	0.119	3 746.15	0.908	2 190.53	1.020	0.229	24-2-1
WMLR-db8*l*5	1 985.33	0.971	1 325.19	1.000	0.123	2 612.09	0.955	1 731.46	0.999	0.160	–
WANN-db10*l*5	2 225.74	0.964	1 600.63	0.995	0.138	3 539.60	0.918	2 287.83	0.986	0.217	24-3-1
WMLR-db10*l*5	1 670.31	0.979	1 137.67	1.000	0.103	2 196.46	0.968	1 435.57	1.000	0.134	–

Note: *MAE* and *RMSE* are in $m^3 \cdot s^{-1}$ unit (Results of only optimum decomposition level are shown).

Source: own elaboration.

WANN and WMLR model results

The normalized observed data was decomposed using Daubechies wavelets of order 1 (db1), 2 (db2), 3(db3), 8 (db8) and 10 (db10) up to fifth level decomposition, which were fed as input to ANN and MLR, making the models as WANN(db*Nl*i) and WMLR(db*Nl*i), respectively, where *N* is the order of Daubechies wavelet and *i* is decomposition level. The performances of these hybrid models only for best decomposition level (low *RMSE*) are presented in Tables 4–7. Results of Tables 4–7 show that wavelet-based hybrid models perform significantly better than the standalone model.

Effect of Daubechies wavelet order on model efficiency

TABLE 8. Percent improvement in *RMSE* with increase in wavelet order from db1 to db10

Lead time [day]	Improvement in *RMSE* [%]
2	57.25
4	54.66
7	61.56
14	55.30

Source: own elaboration.

After comparing the results of hybrid models with respect to wavelet order, it was observed that all models' efficiency is increasing with wavelet order, highest being at 10th order. For example, for four-day lead time (during testing period), from the Table 4 it was found that for WANN model the value of R^2 increased from 0.956 (WANN-db1*l*5) to 0.986 (WANN-db10*l*5) and for WMLR model the increase is from 0.945 (WMLR-db1*l*5) to 0.991 (WMLR-db10*l*5). Similar trend was observed for all lead times. Table 8 shows percent improvement in *RMSE* values with increase in Daubechies wavelet order from db1 to db10. The average percent improvement in *RMSE* is found to be 57.19%. Figure 6 shows a sample plot of effect of Daubechies wavelet order on *RMSE* for four-day lead time.

In general, for all lead times, from the analysis it was found that for lower order wavelets (db1, db2, db3), WANN model performance was better as compared to the WMLR model, while for higher order wavelets (db8, db10), WMLR was found to be superior as compared to WANN. A careful study of Tables 4–7 reveals that the forecasting ability of each hybrid model improves with increasing wavelet order. According to the observed time series flow data (Fig. 2), it appears to occur in a high frequency during the monsoon season (June to September, i.e. four months out of the year) and in a low frequency during the other non-monsoon eight months. Wider-support wavelets can capture low frequencies, while wavelets with a smaller support can capture high frequencies. The support width of a Daubechies wavelet of order *N* is equal to $2N - 1$ (Misiti

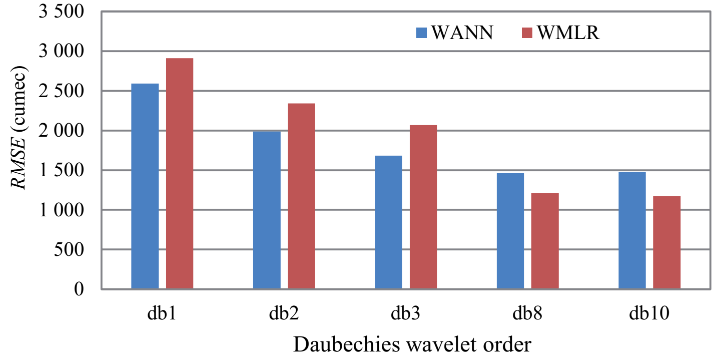


FIGURE 6. Effect of Daubechies wavelet order on *RMSE* (four-day lead time)

Source: own elaboration.

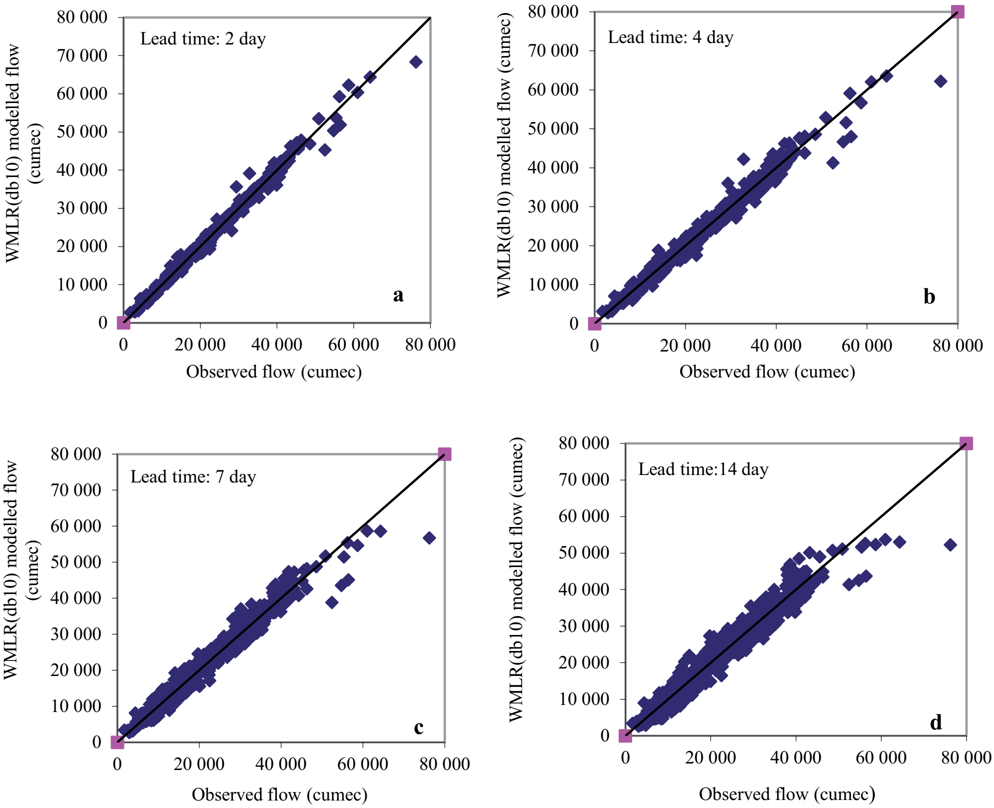


FIGURE 7. Scatter plot for lead time during different testing period

Source: own elaboration.

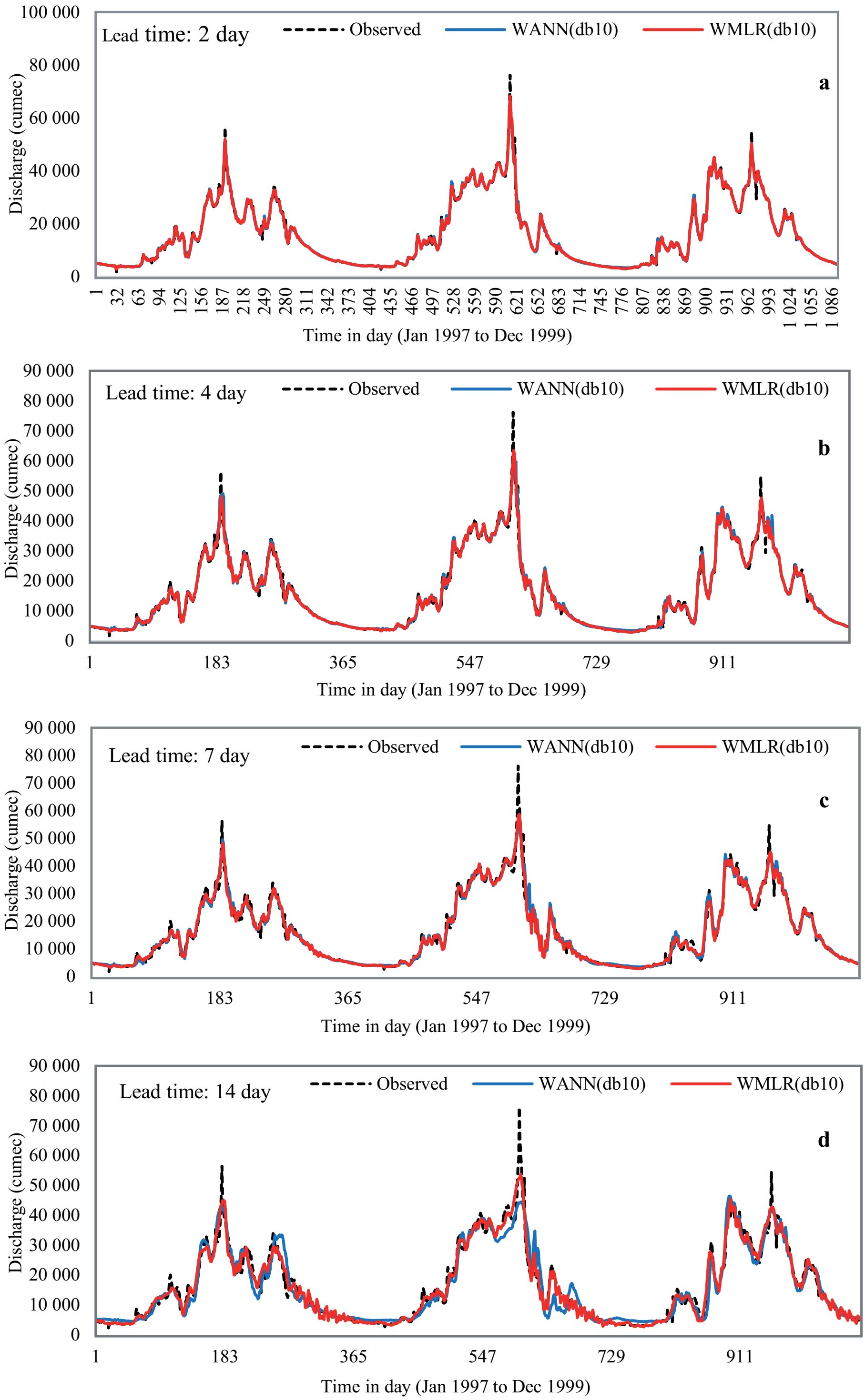


FIGURE 8. Time series plot for lead time during different testing period influence of decomposition level on model performance

Source: own elaboration.

et al., 2010), hence the support widths of db1, db2, db3, db8, and db10, wavelets are one, three, five, 15, and 19, respectively. In brief, compared to db1, db2, db3, and db8 wavelet-based forecast models, the db10 wavelet has a reasonable support and good time-frequency localization properties, which together allow the model to capture both the underlying trend and the short-term variations in the time series. This conclusion is consistent with the findings of (Nourani et al., 2013; Katipoğlu, 2023b, 2023e), who demonstrated that higher order mother wavelet (db4 and db10) offered substantially better results than lower order Haar (db1) wavelet. Scatter plots for all lead times for best model [WMLR(db10)] are shown in Figure 7. The scatter plots showed that the majority of the points were quite close to the 45° line, with only a few having greater magnitudes of observed flow on the lower side, indicating model underestimate (BIAS slightly below 1.0). WMLR(db10) models performance was very satisfactory ($R^2 > 0.9$) (Dawson & Wilby, 2001) for all lead times. Time series plots are shown in Figure 8 for all lead times.

As mentioned earlier, the optimum decomposition level was determined using the formula l = int [log (L)]. In our study, we decomposed raw signal up to fifth level. For the best model WMLR(db10), the effect of decomposition level on R^2 is presented in Table 9. A careful examination of Table 9 demonstrates that model efficiency enhances with decomposition level. The initial resolution level of the original time series captures the high frequency components, and as the decomposition level (scale) is increased, the signal becomes smoother and more stationary. As a result, prediction errors were not affected by scale. Figure 9 shows effect of decomposition level on R^2 for WMLR(db10) model for all lead times.

TABLE 9. Effect of decomposition level on determination coefficient (R^2) for WMLR-db10 model (testing period)

Two-day lead time		Four-day lead time		Seven-day lead time		14-day lead time	
Model type	R^2	Model type	R^2	Model type	R^2	Model type	R^2
WMLR-db10/*l*1	0.976	WMLR-db10/*l*1	0.914	WMLR-db10/*l*1	0.824	WMLR-db10/*l*1	0.656
WMLR-db10/*l*2	0.995	WMLR-db10/*l*2	0.966	WMLR-db10/*l*2	0.861	WMLR-db10/*l*2	0.684
WMLR-db10/*l*3	0.996	WMLR-db10/*l*3	0.990	WMLR-db10/*l*3	0.967	WMLR-db10/*l*3	0.760
WMLR-db10/*l*4	0.996	WMLR-db10/*l*4	0.991	WMLR-db10/*l*4	0.983	WMLR-db10/*l*4	0.953
WMLR-db10/*l*5	0.996	WMLR-db10/*l*5	0.991	WMLR-db10/*l*5	0.984	WMLR-db10/*l*5	0.968

Source: own elaboration.

Khandekar, S. D., Aswar, D. S., Sabale, P. D., Khandekar, V. S., Bajad, M. N., Khaple, S. (2024). Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station). *Sci. Rev. Eng. Env. Sci.*, *33* (1), 69–94. DOI 10.22630/srees.5258

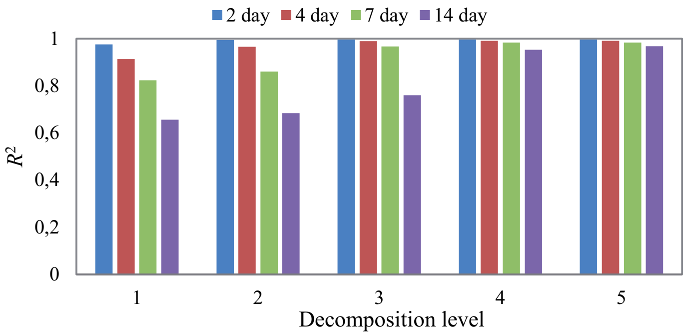


FIGURE 9. Effect of decomposition level on determination coefficient (R^2) for WMLR(db10) model

Source: own elaboration.

Analysis of results for monsoon season during testing period

Finally, because the Brahmaputra river carries substantial flood during the monsoon season (June to September), an attempt was made to assess the model's accuracy during the monsoon period (for three years in testing from 1997 to 1999) for

TABLE 10. Values of statistical parameters for WMLR(db10) model for monsoon season (June to September) in testing period

Year	*RMSE*	R^2	*MAE*	*BIAS*	*S.I.*
		Two-day lead time			
1997	890.55	0.995	596.966	1.000	0.034
1998	1 448.88	0.995	767.802	0.999	0.042
1999	1 070.20	0.996	629.899	0.999	0.034
		Four-day lead time			
1997	1 572.65	0.984	1 072.41	1.000	0.059
1998	2 188.39	0.989	1 153.71	0.999	0.063
1999	1514.68	0.992	968.89	0.998	0.048
		Seven-day lead time			
1997	2 243.79	0.967	1 667.90	0.998	0.084
1998	2 971.27	0.980	1 677.60	1.001	0.086
1999	1 797.46	0.989	1 209.73	0.999	0.058
		14-day lead time			
1997	3 035.86	0.938	2 292.43	1.002	0.115
1998	3 722.21	0.967	2 295.27	1.006	0.111
1999	2 588.93	0.976	1 972.25	0.998	0.084

Note: *RMSE* and *MAE* are in cumec unit.

Source: own elaboration.

the WMLR(db10) model. The statistical parameter values for the WMLR(db10) model during testing period of monsoon season are shown in Table 10. As shown in Table 10, the WMLR(db10) model demonstrated good performance during monsoon season producing extremely satisfactory results for all lead times, despite substantial non-stationarity. Figure 10 compares flow series between observed, WMLR(db10) and WANN(db10) modelled flow for monsoon season for two-day lead time. From Figure 10 it is clear that the WMLR(db10) model has captured well almost all peaks, except the highest.

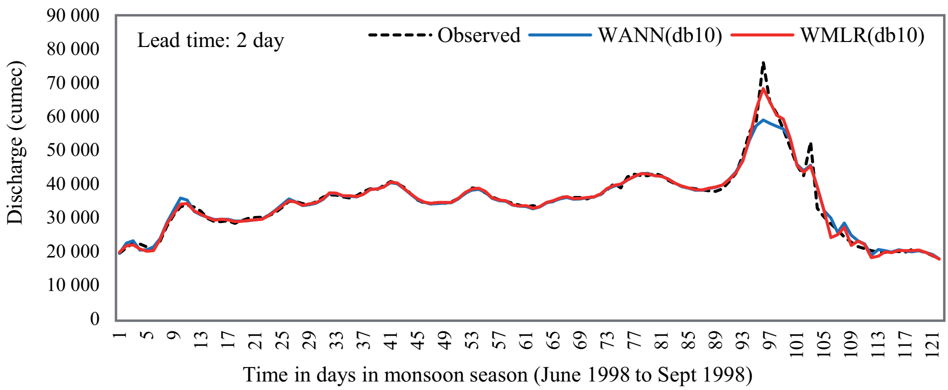


FIGURE 10. Flow series comparison between observed, WMLR(db10) and WANN(db10) modelled flow for monsoon season

Source: own elaboration.

Conclusions

In this study, hybrid models were created by coupling wavelet transform with ANN and MLR to predict the Brahmaputra River flow at Pancharatna station for two-, four-, seven-, and 14-day lead time using 10-year daily flow data. Daubechies wavelets db1, db2, db3, db8, and db10 were used to decompose the observed raw flow data up to fifth level, which were used as input to ANN and MLR. After comparing the results of WANN and WMLR models with respect to wavelet order, it was found that both hybrid models' efficiency is increasing with wavelet order, highest being at 10^{th} order. Comparing the hybrid models' outcomes, it was discovered for all lead times that the WMLR-db10 model had produced more reliable and superior results than the WANN model. The average percent improvement in *RMSE* with increase in wavelet order from one to 10 was found to be 57.19%. Also, it was concluded that with increase in the decomposition level the model's efficiency was increased. Finally, it was concluded that the wavelet transform as a preprocessing tool has proved to be best for mapping the relation between input and output as compared to a single model. To evaluate the forecasting effectiveness of suggested wavelet coupled hybrid models, other hydrological time series variables, such as rainfall, temperature, and evapotranspiration, can be employed as model inputs. To investigate more accuracy, the work can be expanded with additional wavelet types. Study can further be applied at other stations with highly non-stationary data.

References

Adamowski, J. & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, *390*, 85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033

Araghi, A., Adamowski, J. & Martinez, C. J. (2020) Comparison of wavelet-based hybrid models for the estimation of daily reference evapotranspiration in different climates. *Journal of Water and Climate Change*, *11*, 39–53. https://doi.org/10.2166/wcc.2018.113

ASCE Task Committee on Application of Artificial Neural Network in Hydrology (2000). Artificial neural networks in hydrology-II: hydrologic applications. *Journal of Hydrologic Engineering*, *5* (2), 124–137.

Bajirao, T. S., Kumar, P., Kumar, M., Elbeltagi, A. & Kuriqi, A. (2021). Potential of hybrid wavelet-coupled data-driven-based algorithms for daily runoff prediction in complex river basins. *Theoretical and Applied Climatology*, *145* (3–4), 1207–1231. https://doi.org/10.1007/s00704-021-03681-2

Besaw, L. E., Rizzo, D. M., Bierman, P. R. & Hackett, W. R. (2010). Advances in ungauged stream flow prediction using artificial neural networks. *Journal of Hydrology*, *386* (1–4), 27–37. https://doi.org/10.1016/j.jhydrol.2010.02.037

Budu, K. (2013). Comparison of wavelet-based ANN and regression models for reservoir inflow forecasting. *Journal of Hydrologic Engineering*, *19* (7), 1385–1400. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000892

Dawson, C. W. & Wilby, R. (1998). An artificial neural network approach to rainfall runoff modeling. *Hydrological Sciences Journal*, *43* (1), 47–66. https://doi.org/10.1080/02626669809492102

Dawson, C. W. & Wilby, R. L. (2001). Hydrological modelling using artificial neural network. *Progress in Physical Geography: Earth and Environment*, *25* (1), 80–108. https://doi.org/10.1177/030913330102500104

Deka, P. C. & Prahlada, R. (2012). Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. *Ocean Engineering*, *43*, 32–42. https://doi.org/10.1016/j.oceaneng.2012.01.017

Djerbouai, S. & Souag-Gamane, D. (2016). Drought forecasting using neural networks, wavelet neural networks, and stochastic models: case of the Algerois Basin in North Algeria. *Water Resources Management*, *30*, 2445–2464. https://doi.org/10.1007/s11269-016-1298-6

Goswami, D. C. (1985). Brahmaputra River, Assam, India: physiography, basin denudation, and channel aggradation. *Water Resources Research*, *21* (7), 959–978. https://doi.org/10.1029/WR021i007p00959

Jain, S. K. & Chalisgaonkar, D. (2000). Setting up stage-discharge relations using ANN. *Journal of Hydrologic Engineering*, *5* (4), 428–433. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:4(428)

Katipoğlu, O. M. (2023a). Monthly streamflow prediction in Amasya, Türkiye, using an integrated approach of a feedforward backpropagation neural network and discrete wavelet transform. *Modeling Earth Systems and Environment*, *9* (2), 2463–2475. https://doi.org/10.1007/s40808-022-01629-7

Katipoğlu, O. M. (2023b). Evaluation of the success of the hybrid wavelet-based ANFIS approach in the estimation of monthly stream flows of the Bitlis River, Turkey. *Water Supply*, *23* (2), 836–850. https://doi.org/10.2166/ws.2023.024

Katipoğlu, O. M. (2023c). Evaporation prediction with wavelet-based hyperparameter optimized K-nearest neighbors and extreme gradient boosting algorithms in a semi-arid environment. *Environmental Processes*, *10* (4), 50. https://doi.org/10.1007/s40710-023-00669-0

Katipoğlu, O. M. (2023d). Combining discrete wavelet decomposition with soft computing techniques to predict monthly evapotranspiration in semi-arid Hakkâri province, Türkiye. *Environmental Science and Pollution Research*, *30* (15), 44043–44066. https://doi.org/10.1007/s11356-023-25369-y

Katipoğlu, O. M. (2023e). Prediction of streamflow drought index for short-term hydrological drought in the semi-arid Yesilirmak Basin using Wavelet transform and artificial intelligence techniques. *Sustainability*, *15* (2), 1109. https://doi.org/10.3390/su15021109

Khandekar, S. D. (2014). *Hybrid wavelet transform-neural network approach for short term and long term time series flow forecasting* (PhD thesis). National Institute of Technology Karnataka, Mangalore.

Khandekar, S. D. & Deka, P. C. (2016). Applications of wavelet transform technique in hydrology – a brief review. In A. Sarma, V. Singh, S. Kartha & R. Bhattacharjya (Eds), *Urban Hydrology, Watershed Management and Socio-Economic Aspects. Water Science and Technology Library*. Vol. 73. Cham: Springer. https://doi.org/10.1007/978-3-319-40195-9_19

Khazaee Poul, A., Shourian, M. & Ebrahimi, H. (2019). A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. *Water Resources Management*, *33*, 2907–2923. https://doi.org/10.1007/s11269-019-02273-0

Kişi, Ö. (2009a). Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting. *Hydrological Processes*, *23* (25), 3583–3597. https://doi.org/10.1002/hyp.7461

Kişi, Ö. (2009b). Neural networks and wavelet conjunction model for intermittent stream flow forecasting. *Journal of Hydrologic Engineering*, *14* (8), 773–782. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000053

Kişi, Ö. (2010). Wavelet regression model for short-term stream flow forecasting. *Journal of Hydrology*, *389* (3–4), 344–353. https://doi.org/10.1016/j.jhydrol.2010.06.013

Kişi, Ö. (2011). Wavelet regression model as an alternative to neural networks for river stage forecasting. *Water Resource Management*, *25*, 579–600. https://doi.org/10.1007/s11269-010-9715-8

Kumar, M., Kumar, P., Kumar, A., Elbeltagi, A. & Kuriqi, A. (2022). Modeling stage–discharge–-sediment using support vector machine and artificial neural network coupled with wavelet transform. *Applied Water Science*, *12* (5), 87. https://doi.org/10.1007/s13201-022-01621-7

Labat, D., Ababou, R. & Mangin, A. (2000). Rainfall–runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. *Journal of Hydrology*, *238* (3–4), 149–178. https://doi.org/10.1016/S0022-1694(00)00322-X

Linh, N. T. T., Ruigar, H., Golian, S., Bawoke, G. T., Gupta, V., Rahman, K. U., Sankaran, A. & Pham, Q. B. (2021). Flood prediction based on climatic signals using wavelet neural network. *Acta Geophysica*, *69* (4), 1413–1426. https://doi.org/10.1007/s11600-021-00620-7

Mallat, S. G. (1989). Multiresolution approximations and wavelet orthonormal bases of L^2 (R). *Transactions of the American Mathematical Society*, *315* (1), 69–87.

Mallat, S. G. (1998). *A wavelet tour of signal processing*. San Diego: Academic Press.

Misiti, M., Misiti, Y., Oppenheim, G. & Poggi, J. (2010). *Wavelet toolbox: For use with MATLAB*. The MathWorks, Natick.

Mohseni, U. & Muskula, S. B. (2023). Rainfall-runoff modeling using artificial neural network – a case study of purna sub-catchment of Upper Tapi Basin, India. *Environmental Sciences Proceedings*, *25* (1), 1–8. https://doi.org/10.3390/ECWS-7-14232

Moosavi, V., Vafakhah, M., Shirmohammadi, B. & Behnia, N. (2013). A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. *Water Resources Management*, *27*, 1301–1321. https://doi.org/10.1007/s11269-012-0239-2

Nayak, P., Venkatesh, B., Krishna, B. & Jain, S. K. (2013). Rainfall-runoff modeling using conceptual, data driven, and wavelet based computing approach. *Journal of Hydrology*, *493*, 57–67. https://doi.org/10.1016/j.jhydrol.2013.04.016

Nourani, V., Baghanam, A. H., Adamowski, J. & Gebremichael, M. (2013). Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. *Journal of Hydrology*, *476*, 228–243. https://doi.org/10.1016/j.jhydrol.2012.10.054

Nourani, V., Baghanam, A. H., Adamowski, J. & Kisi, O. (2014). Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. *Journal of Hydrology*, *514*, 358–377. https://doi.org/10.1016/j.jhydrol.2014.03.057

Nourani, V., Komasi, M. & Mano, A. A. (2009). Multivariate ANN-Wavelet Approach for Rainfall–Runoff Modeling. *Water Resources Management*, *23*, 2877–2894. https://doi.org/10.1007/s11269-009-9414-5

Patil, A. P. & Deka, P. C. (2017). Performance evaluation of hybrid Wavelet-ANN and Wavelet-ANFIS models for estimating evapotranspiration in arid regions of India. *Neural Computing and Applications*, *28*, 275–285. https://doi.org/10.1007/s00521-015-2055-0

Raghuwanshi, N. S., Singh, R. & Reddy, L. S. (2006). Runoff and sediment yield modeling using artificial neural networks: Upper Siwane River, India. *Journal of Hydrologic Engineering*, *11* (1), 71–79. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:1(71)

Rajaee, T., Nourani, V., Zounemat-Kermani, M. & Kisi, O. (2011). River suspended sediment load prediction: application of ANN and wavelet conjunction model. *Journal of Hydrologic Engineering*, *16* (8), 613–627. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000347

Rao, Y. S. & Krishna, B. (2009). Modelling hydrological time series data using wavelet neural network analysis. *IAHS publication*, *333*, 101–111.

Rezaeian Zadeh, M., Amin, S., Khalili, D. & Singh, V. P. (2010). Daily outflow prediction by multi layer perceptron with logistic sigmoid and tangent sigmoid activation functions. *Water Resources Management*, *24*, 2673–2688. https://doi.org/10.1007/s11269-009-9573-4

Sarma, J. N. (2005). Fluvial process and morphology of the Brahmaputra River in Assam, India. *Geomorphology*, *70* (3–4), 226–256. https://doi.org/10.1016/j.geomorph.2005.02.007

Shafaei, M. & Kisi, O. (2017). Predicting river daily flow using wavelet-artificial neural networks based on regression analyses in comparison with artificial neural networks and support vector machine models. *Neural Computing and Applications*, *28*, 15–28. https://doi.org/10.1007/s00521-016-2293-9

Sharma, S., Isik, S., Srivastava, P. & Kalin, L. (2013). Deriving spatially distributed precipitation data using the artificial neural network and multilinear regression models. *Journal of Hydrologic Engineering*, *18* (2), 194–205. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000617

Shoaib, M., Shamseldin, A. Y., Khan, S., Khan, M. M., Khan, Z. M., Sultan, T. & Melville, B. W. (2018). A comparative study of various hybrid wavelet feedforward neural network models for runoff forecasting. *Water Resources Management*, *32*, 83–103. https://doi.org/10.1007/s11269-017-1796-1

Solaimani, K. (2009). Rainfall-runoff prediction based on artificial neural network (a case study: Jarahi watershed). *American-Eurasian Journal of Agricultural & Environmental Sciences*, *5* (6), 856–865.

Sudheer, K. P., Gosain, A. K. & Ramasastri, K. S. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological Processes*, *16* (6), 1325–1330. https://doi.org/10.1002/hyp.554

Sun, Y., Niu, J. & Sivakumar, B. (2019). A comparative study of models for short-term streamflow forecasting with emphasis on wavelet-based approach. *Stochastic Environmental Research and Risk Assessment*, *33*, 1875–1891. https://doi.org/10.1007/s00477-019-01734-7

Tayfur, G. & Singh, V. P. (2006). ANN and fuzzy logic models for simulating event-based rainfall-runoff. *Journal of Hydraulic Engineering*, *132* (12), 1321–1330. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:12(1321)

Wang, K., Band, S. S., Ameri, R., Biyari, M., Hai, T., Hsu, C. C., Hadjouni, M., Elmannai, H., Chau, K. W. & Mosavi, A. (2022). Performance improvement of machine learning models via wavelet theory in estimating monthly river streamflow. *Engineering Applications of Computational Fluid Mechanics*, *16* (1), 1833–1848. https://doi.org/10.1080/19942060.2022.2119281

Zhang, Z., Zhang, Q. & Sing, V. P. (2018). Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. *Hydrological Sciences Journal*, *63* (7), 1091–1111. https://doi.org/10.1080/02626667.2018.1469756

Khandekar, S. D., Aswar, D. S., Sabale, P. D., Khandekar, V. S., Bajad, M. N., Khaple, S. (2024). Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station). *Sci. Rev. Eng. Env. Sci.*, *33* (1), 69–94. DOI 10.22630/srees.5258

Summary

Hybrid wavelet transform – MLR and ANN models for river flow prediction: Case study of Brahmaputra river (Pancharatna station). In this research, discrete wavelet transform (DWT) is combined with MLR and ANN to develop WMLR and WANN hybrid models, respectively, for the Brahmaputra river (Pancharatna station) flow forecasting. Daily flow data for the period of 10 year were decomposed (up to fifth level) into detailed and approximation coefficients (using Daubechies wavelets db1, db2, db3, db8 and db10) which were fed as input to MLR and ANN to get the predicted discharge values two days, four days, seven days and 14 days ahead. For all lead times, the WMLR-db10 model was found to be superior as compared to WANN-db1, WANN-db2, WANN-db3, WANN-db8, WMLR-db1, WMLR-db2, WMLR-db3, WMLR-db8 and single MLR and ANN models. During testing period, the values of determination coefficient (R^2) and *RMSE* for WMLR-db10 model for two-, four-, seven- and 14-day lead time were found to be, respectively, 0.996 (751.87 $m^3 \cdot s^{-1}$), 0.991 (1,174.80 $m^3 \cdot s^{-1}$), 0.984 (1,585.02 $m^3 \cdot s^{-1}$), and 0.968 (2,196.46 $m^3 \cdot s^{-1}$). Also, it was observed that for lower order wavelets (db1, db2, db3) WANN's performance was better, and for higher order wavelets (db8, db10) WMLR's performance was better. Correspondingly, it was observed that all hybrid models' efficiency increased with increase in the decomposition level.

Scientific Review Engineering and Environmental Sciences (2024), 33 (1), 95–111

Sci. Rev. Eng. Env. Sci. (2023), 33 (1) https://srees.sggw.edu.pl

ISSN 1732-9353 (suspended) eISSN 2543-7496 DOI 10.22630/srees.4894

Received: 02.05.2023
Accepted: 12.07.2023

SITENDER[1]✉

Deepak Kumar VERMA[2]

Baldev SETIA[3]

[1, 2]U.I.E.T., Maharshi Dayanand University, Department of Civil Engineering, India
[3]National Institute of Technology Kurukshetra, India

Support vector regression tree model for the embankment breaching analysis based on the Chamoli tragedy in Uttarakhand

Keywords: embankment breaching, multi-objective data, catastrophic collapses, rock-ice avalanche, Chamoli tragedy

Introduction

The degradation of earth fill or rock fill components by water flow, either over or through levee bodies or the surrounding foundation, is the primary source of wall breaches in hills and bridges. Water flow through dam bodies causes pipe or seepage failure, whereas water flow over dam bodies generates flash floods in reservoirs and accompanying exterior erosion. Overtopping, piping/seepage, slips, and other difficulties have all been identified as important causes of embankment dam failure, according to a prior investigation on the dam collapse data (Li et al., 2021). Whatever the nature of the accident factors, multiple field investigations and theoretical studies have demonstrated that the final failure mechanisms of hydraulic structures are either superior power or pipeline failure. Landslide barriers, as opposed to hydraulic constructions, are constructed quickly and are made up of a diverse mass of significant sector or rockfill parts (Donghui et al., 2016). Furthermore, in catastrophe dams,

✉sitender0123456@gmail.com

there are no drain zones to maintain pore pressure and no impermeable or filter sections to stop drainage flow.

Landslide waterfalls, on the other hand, typically have a large width to height ratio (representing their hydraulic gradient) and softer upstream and downstream slopes, which may reduce the likelihood of pipe and slope collapse. Upstream to downstream slope percentages of earth-filled dams are normally 1 : 1 to 1 : 2; upstream and downstream slope ratios of clay core rock-fill dams are typically 3 : 1 to 5 : 1 and 1 : 3 to 1 : 2, respectively; and cementitious face rockfill dams are typically 1 : 1.50 to 1 : 1.13 (Samela et al., 2015). The upstream and downstream slope ratio is influenced by the cohesion and friction angles of landslide dam bodies, which typically range from 1 to 3 and have higher to higher percentages than embankment dams. Overtopping failure in landslide dams is thus more common than pipeline failure (Aksoy, Kirca, Burgan & Kellecioglu, 2016).

Landslide dams, like embankment dams, can break fatally due to overtopping or pipeline failure. In this context, it is critical to emphasize that technical solutions are frequently used to reduce the danger of dam breaches once landslide dams have been formed (Khanduri, 2019). Diverting input water upstream, utilizing pumps or syphons for drainage, constructing drainage tunnels through abutments, and constructing sewage spillways through dams are all popular engineering risk reduction strategies (Xu et al., 2009). Spillway excavation, which has been shown to be an effective disaster mitigation strategy, is most commonly used to reduce the peak breach flow of large dammed lakes. As a result, while studying the dynamics causing landslide dam failures, engineering precautions should be considered (Gariano & Guzzettai, 2016). Physical model testing is the primary technical tool for evaluating the mechanisms involved in slope and landslide river flooding. The next sections (Khanduri, 2020) detail the government's breach processes for man-made and natural dams made of earthfill and rock-fill materials, where the collapse was triggered by which factor or pipe line.

The majority of people, particularly in India, are aware of the destruction in Kedarnath and the Idukki valley. This is partly due to the extensive media coverage it received as a result of the unfortunate deaths (Rautela, 2013). This tragedy, however, ravaged the whole Higher Himalayan area of Uttarakhand, from the Kali river valley in the east to the Yamuna river valley in the west, with 5 of the state's 13 districts – namely, Rudraprayag, Chamoli (Fig. 1), Uttarkashi, Bageshwar, and Pithoragarh – being affected the most.

On 7 February 2021, at around 10.45 AM, a part of the Nanda Devi Glacier in Uttarakhand's Chamoli area, near Joshimath, erupted, resulting in a deadly glacial avalanche with an embankment. It is geologically adjacent to the Main Central Thrust (MCT) in the Indian Himalaya, a seismically active and extremely vulnerable tectonic zone (Khanduri, 2017). A glacier fracture near the Raini hamlet has inun-

FIGURE 1. Chamoli tragedy affected region
Source: https://economictimes.indiatimes.com.

dated the Tapovan area of Joshimath in Uttarakhand's Chamoli district. As a result, the flash flood in the Dhauliganga, Rishi Ganga, and Alaknanda rivers is reminiscent to the 2013 Kedarnath catastrophe, which was caused by a Himalayan tsunami. The most recent tragedy resulted in multiple deaths and property damages, as well as full devastation to the Rishiganga and NTPC hydroelectric generating stations in Dhauliganga and Rishiganga. A significant portion of the enormous flood destroyed other infrastructure, such as bridges and roadways. Under the guise of Operation Jeevan, the NDRF, SDRF, and ITBP continue to carry out rescue and relief activities, saving numerous lives (Khanduri, 2021).

Background of the event

A large landslide happened on 7 February 2021, in the remote Chamoli district of northern India's Uttarakhand state. The avalanche originated on the Ronti Glacier, which is located at 79.732° E longitude and 30.3750° N latitude. Because the catastrophic flood was discovered in the middle of the day in the Rishi Ganga river, a tributary of the Ganges, it was first described to as a 'glacial' burst or glacial lake outburst flood disaster (Rana et al., 2021). As a result, the occurrence was named a rock ice avalanche. However, satellite pictures indicated that the source location is mostly made up of ice-covered bedrock, with no lake or lake burst involved. According to seismic waveform data gathered by the Indian Institute of Remote Sensing (IIRS) at Mandal and Nainital stations, the occurrence occurred about 4:52 UT (10:22 AM local time), as see in Figure 2. The vast majority of those killed or missing in the ensuing flood worked on the Tapovan-Vishnugad and Rishi Ganga hydropower projects, both of which were seriously damaged by the tragedy (Valdiya, 2014).

FIGURE 2. Outburst of avalanche of ice rock
Source: https://english.jagran.com.

Individual families of 4–5 people own land ranging from 0.3 to 0.8 ha in this area. The area's primary bio-resource is forest, which includes subtropical pine, Himalayan wet temperate forest, oak, and mixed forest types. Human and animal population growth is harming this forest cover. The presence of farming on steeper slopes (> 60°), as seen in Devpuri and Patal Kharak, indicates that population pressure has compelled people to employ unusual slopes (Rautela & Pande, 2006).

Geographical and administrative setup of Uttarakhand

Uttarakhand, India, is located in the country's northwest and became the country's 27th state in 2000, after splitting from the state of Uttar Pradesh. As seen in Figure 3, Uttarakhand shares its northern border with Tibet, its eastern border with Nepal, its southern border with Uttar Pradesh, and its western border with Himachal Pradesh. Garhwal and Kumaun are the 2 divisions of the state, each with 13 districts. Pauri Garhwal, Tehri Garhwal, Chamoli, Haridwar, Dehradun, Uttarkashi, and

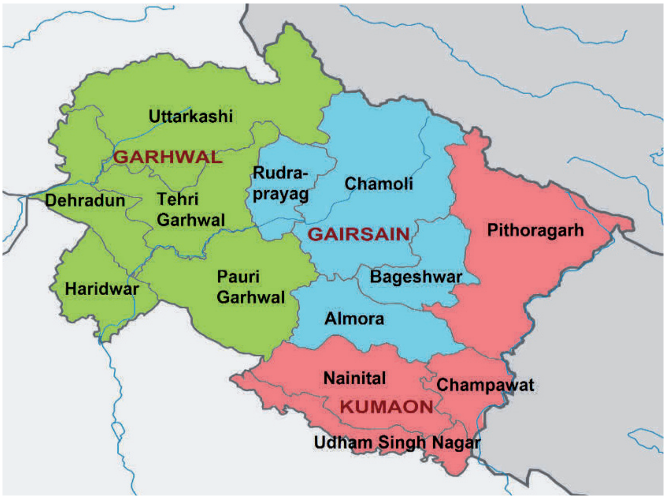


FIGURE 3. Uttaraghand landmark
Source: www.shutterstock.com.

Rudraprayag are part of the Garhwal division, whereas Almora, Bageshwar, Champawat, Nainital, Udham Singh Nagar, and Pithoragarh are part of the Kumaun division (Indian Standard [IS], 1983).

Due to its position in the middle range of the Himalayan Mountains, the state comprises 53,483 km^2 and 86.07% of its total area is classed as hilly. The lowest point is 200 m in the south and the highest point is 8,000 m in the north. Aside from political divisions, the state is split geographically as follows:

– Upper hills include Uttarkashi, Chamoli, Rudraprayag, Pithoragarh, and Bageshwar.
– Middle hills – Tehri-Garhwal, Garhwal, Almora, and Champawat, as well as the hill areas of Nainital and Dehradun's Chakrata tehsil.
– Foothills – the rest of Dehradun, Haridwar, Udham Singh Nagar, and Nainital.

Socio-economic condition

Uttarakhand has a population of 10,086,292 people and observed a population growth of 18.81% between 2001 and 2011, somewhat higher than the national average of 17.64%. Despite the state's expanding population, the average population density is 189 people per km^2, which is less than half of the national average of 382 people per km^2. This is due to a scarcity of habitable land, since forest occupies 61.91% of the territory and glacial terrain covers 13.73% (Fig. 4).

Sitender, Verma, D. K., Setia, B. (2024). Support vector regression tree model for the embankment breaching analysis based on the Chamoli tragedy in Uttarakhand. *Sci. Rev. Eng. Env. Sci.*, *33* (1), 95–111. DOI 10.22630/srees.4894

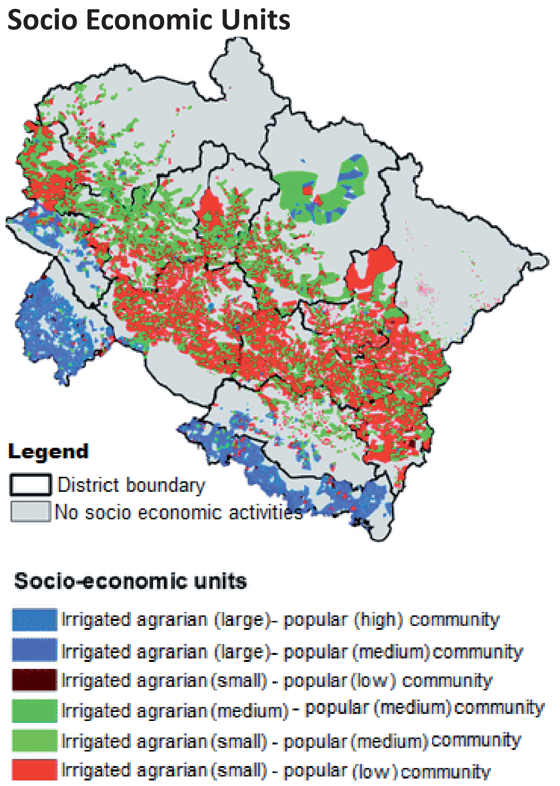


FIGURE 4. Socio-economic representation of Uttarakhand
Source: Kumar and Bhattacharjya (2020).

Because Uttarakhand is home to India's main pilgrimage circuit as well as other tourist spots, the state has a substantial floating population. As a result, tourism-related industries such as hotels and transportation are growing, and in 2012, 27 million people visited the Uttarakhand Festival of Tourism in 2013. While tourism is growing, farming, forestry, horticulture, and livestock production continue to be the key sources of revenue in Uttarakhand. Despite the difficult terrain, agriculture accounts for 21.72% of total land surface and employs 7,585% of the population (Kayal, 2010). Farmers mostly cultivate grains, and agriculture contributes for 22.41% of the state's GDP. Uttarakhand is witnessing a huge population expansion as a result of the state's demographic qualities, as well as an inflow of pilgrims and visitors. People are becoming increasingly reliant on agriculture for a living, and the tourist sector is growing.

Geological condition

While hydrological catastrophes are the most common, Uttarakhand is located in a tectonically sensitive location and has been hit by a number of seismic events in the past (Fig. 5). Between 1803 and 1999, there were 36 earthquakes with Richter magnitudes greater than 5.0, with 12 earthquakes with magnitudes larger than 6.0 occurring solely in the 21st century (Heim & Gansser, 1939). The most recent and important earthquakes, both of which caused substantial damage, were the Uttarkashi earthquake in 1991 and the Chamoli earthquake in 1999.

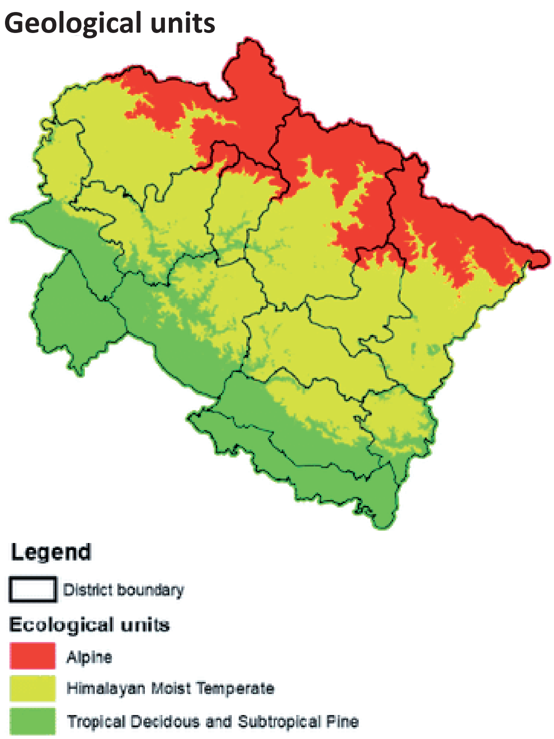


FIGURE 5. Geological condition of Uttarakhand

Source: Socio-Ecological Systems (SESs) – Identification and Spatial Mapping in the Central Himalaya.

The Uttarkashi earthquake had a magnitude of 6.6, and home collapses resulted in the loss of life. Roads and bridges were also severely destroyed, causing rescue efforts to be delayed. The Chamoli earthquake in 1999 had a Richter scale magnitude of 6.8. The earthquake damaged over 90% of the dwellings in the Chamoli district, with the majority of them being non-engineered buildings that stood after the earthquake (Valdiya, 1980). This earthquake also damaged school buildings since the roofs were composed of slate or galvanized corrugated iron and the walls were not linked to the roof, allowing them to collapse easily.

Parameters causing embankment breaching

Effect of slope

Higher embanking walls, which are always unstable, were necessary for terracing steeper slopes. The residents relocate to upslope locations that are generally considered unsafe for human settlement by any measure. This is mostly due to the lack of effective alternatives, which has led residents to neglect the terrain's fragility. Recent development efforts, especially those involving watershed management, have increased terrain deterioration.

Changes in slope gradient are a common source of landslides and other types of mass waste. A change in slope gradient alters the internal tension of the rock or soil mass, whereas a change in shear stress alters the distribution of equilibrium conditions (Fig. 6).

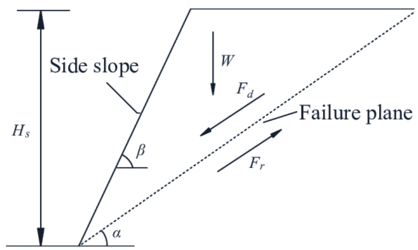


FIGURE 6. Failure plane of slope
Source: own elaboration.

Human-caused changes in slope height, on the other hand, increase shear stress and the creation of fractures in weak zones, which can become saturated with water and break, as witnessed in the Devpuri region.

Water pressure due to blockages

Water tanks and check dams were built along the channel routes above the Patal Kharak and Suri Kharak. During the disaster, debris and trees gathered on these structures, creating a larger natural barrier that hinders the flow of the canals. All of the dams burst within minutes due to increasing water pressure, resulting in a deadly flash flood downstream. Road construction is a major source of harm in the Patal Kharak and Devpuri neighborhoods. A massive quantity of debris has accumulated on the steep slopes as a result of road building, impeding streams during cloudburst occurrences and producing natural dams.

Improper drainage

According to reports, the road sides lacked adequate drainage systems. The Khansar Gad valley lacks a meteorological observatory. The sole observatory in the vicinity, Gairsain, is 15.5 km away. The Gairsain observatory reported 16 mm of rainfall on 12 July 2007, indicating that this is a rather brief occurrence. It had been pouring non-stop for over 6 h before to this incident. bedding planes or joints. The aperture of the rough joints enlarges owing to dilatation when the slide begins (Pol et al., 2022). As a result, the sliding plane works as a natural water channel. Because of poor surface drainage, infiltration rates are significant in the overburden area, creating increased pore-water pressure and restarting the slope.

Hydrostatic stress in old landslide areas

As a result of the poor hydrological conditions, secondary structural vulnerabilities in the host rocks, debris, and pre-existing slip surfaces in previous landslide zones are also reactivated. Rain and melt water penetrate rock joints, causing hydrostatic tension. Rain raises soil pore water pressure, which reduces shear resistance (Grämiger et al., 2020). The bulk of the surface layers in this area are loose, unconsolidated soils connected by Pleistocene solifluction lobes. Because clay predominates in such debris cones and terraces, they become impermeable when water penetration exceeds a critical limit as pore water pressure increases. Water rushes up the hill, bringing with it tones of dirt, rock, and other debris.

The difficulty is exacerbated by the height of the embanking wall and the outward slope of the terraces. Because of the tremendous bulk of the soil and the regular rains, these terraces have become unstable. Lithology is one of the key causes of slope instability. In the study region, quartzites produce steep slopes because they are unweathered, hard, massive, and resistant to weathering, whereas phyllites form depressions and valleys because they are prone to weathering and erosion. Older alluvium is often exceedingly compacted and durable, whereas slide debris is often loose with low shear and erosion resistance. In geology, phyllite is classified as R3 rock and is part of the moderately strong rock group The R4 grade jointed quartzite and meta-basics are classified as strong rock. Massive Kalu Kharak quartzite has an R6 classification and is considered an extremely strong rock.

Agricultural practices

Because agriculture is the primary source of income for the slope's inhabitants, there is an urgent need to create alternative agricultural practices or encourage non-land-based economic activity in this region. Landslides have been documented in a few locations as a result of a change in slope gradient induced by terrace and construction excavation. We are all aware that the mountain agricultural system cannot be seen just through the lens of field farming; significant contributions from the sectors of animal husbandry and conservation are also necessary. In this context, we also urge that suitable grazing sites be chosen in eco-fragile zones, as grazing has been shown in multiple studies to worsen soil erosion and landslide occurrences.

Cloudbursts

Cloudbursts are thunderstorms or convectional precipitation caused by the fast condensation of locally formed cumulonimbus clouds. Such events are widespread in the Higher Himalayan valley slopes, where agriculture terraces can be seen. The majority of landslides are restricted to agricultural grounds since most of the area's residents live above the destabilized landslide debris and solifluction lobes, where cultivable fields are easily prepared. Villagers are converting steep mountain slopes into geologically and structurally unsound agricultural terraces. The current study discovered that structural, lithological, and geomorphological variables all have a major impact on landslides during cloudburst occurrences. Landslides were especially prevalent in this area during the monsoon rains, as water is a significant inducer of land sliding. Land use practices and settlements have considerably contributed to the loss of lives and property in terms of man-made impacts.

Road construction

Road development has had a crucial influence in the devastating landslides. The roads in the area were poorly designed and built, resulting in slope instability and massive debris generation. Instead of maintaining the side drain, road construction personnel dumped debris and big chunks of quartzite on the gullies and steep slopes. As a result, comprehensive slope stability evaluation is critical, rubbish should be disposed of in appropriate areas, and correct drainage systems are essential since water seepage causes stability issues. Long-term solutions should involve appropriate drainage, such as culverts and hillside catch water drains. To prevent the negative effects of future cloudbursts, it is necessary to identify and analyze vulnerable areas in the region, as well as establish strategies and programs for execution.

Parameters regression analysis with support vector regression model

Using multivariate linear regression, a linear connection is formed between the response variable and one or more explanatory factors (MLR). The response variable is sometimes referred to as the 'dependent variable', whilst the descriptive variables are referred to as the 'independent variables'. Linear regression is defined as

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + an \cdot xn + n = 1, 2, \ldots, m,$$

where y is the dependent variable, an is the model parameter, xn is the independent variable, and m is the random error component.

The MNLR methodology is a statistical method for fitting appropriate non-linear functions to collected data to maximize the determination coefficient. The connection is represented as a non-linear function between independent and dependent variables that may be used to calculate model parameters. The MNLR models can handle a variety of functions, including exponential, logarithmic, and power functions, with the power function being most suited to the dam-break scenario.

The regression tree (RT) method is then used to split the feature space and to choose the most significant features and discard the rest. Furthermore, we construct the support vector regression (SVR) model using the RT methodology's key variables as well as the RT method's prediction outputs, which are used as additional input information in the SVR input vector space. The inclusion of RT output as an input feature increases the complexity of the feature space and improves class separability. The proposed hybrid RT-SVR model's robustness is based on the selection of important features and the incorporation of RT model regression results into the SVR model. The suggested hybrid RT-SVR paradigm's informal work-flow. To train and build a decision tree model that estimates important attributes, a regression tree approach is utilized. The RT technique prediction result is employed as an extra feature in the SVR model's input feature space. To create an appropriate SVR model, the needed RT input variables are then retrieved, along with an extra input variable (RT output). The SVR approach is employed with a Gaussian kernel function for non-linear regression challenges, and regression results are reported. This technique is a two-step pipeline strategy that starts with RT feature selection and then uses SVR to increase its performance by using all of the RT algorithm's outputs in an additional analysis. Our suggested model may be utilized to identify important characteristics that will achieve a certain aim in order to address a water quality problem, as well as to estimate boiler output pH using critical causal process factors. The system was designed to address the issue of water quality, but it may also be used to manage other complicated linear systems.

The root mean square error (*RMSE*) is a statistic used to calculate the absolute difference between observed and simulated data. This statistical metric has a value between 0 and +, with lower values suggesting better simulation results. The Nash–Sutcliffe efficiency (*NSE*) (Shrestha et al., 2023) is the normal state of the least-squares error function, expressing the ratio of residual variance to data variance for evaluating the effectiveness of existing theories. The *NSE* value ranges from –1 to +1, with +1 being the best. As a result, if this value exceeds 0.5, the model performs a simulation properly. The coefficient of determination (R^2) is a model accuracy metric that shows the relationship between observed and predicted values. This index also indicates a portion of the observed value range that may be justified by simulated values. The *RMSE*, *NSE*, and R^2 indices are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(Q_{s,i} - Q_{o,i}\right)^2},$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{s,i} - Q_{o,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{o,i} - \overline{Q_o}\right)^2},$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(Q_{s,i} - \overline{Q_s}\right)\left(Q_{o,i} - \overline{Q_o}\right)\right]^2}{\sum_{i=1}^{n}\left(Q_{s,i} - \overline{Q_s}\right)^2 \sum_{i=1}^{n}\left(Q_{o,i} - \overline{Q_o}\right)^2},$$

where $Q_{s,i}$ is the simulated weight, $Q_{o,i}$ is the observed weight, s is the average of parameter weights, o is the average of parameter observed weights, I is the number of steps, and n is the total number of parameters in the above relationships.

Sensitivity analysis of parameters

In general, the soil erodibility coefficient (k_d) and critical shear stress can influence landslide dam breaches (c). Furthermore, for a large dammed lake, spillway excavation is recognized as an effective disaster-reduction approach. The impacts of k_d, c, and spillway excavation on the landslide dam breach were investigated using a parameter sensitivity analysis.

The soil erodibility coefficient (k_d) was multiplied by 0.5, 1.0, and 2.0 to test the model's sensitivity, while all other parameters remained fixed (Fig. 7). The sensitivity analysis findings for k_d are displayed (Fig. 8). Breach hydrographs with varied k_ds values were created because breach flow discharge is critical for disaster consequences.

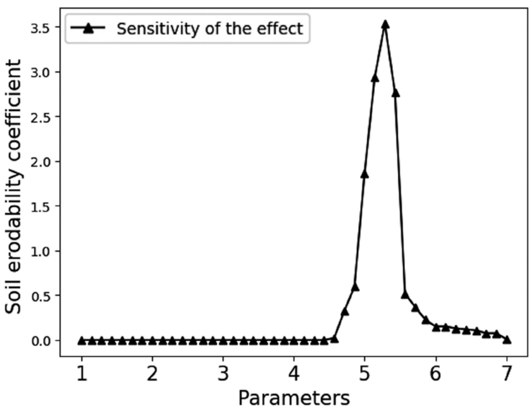


FIGURE 7. Soil erodability coefficient

Source: own elaboration.

Because the height of the landslide dam had a direct influence on the storage capacity of the dammed lake, the output parameters of peak breach flow and breach breadth rose significantly for the dam without a spillway.

However, the eventual breach size increased somewhat with increasing k_d due to the rising soil erodibility coefficient with depth and the quick attenuation of hydrodynamic conditions in the case of a big peak breach

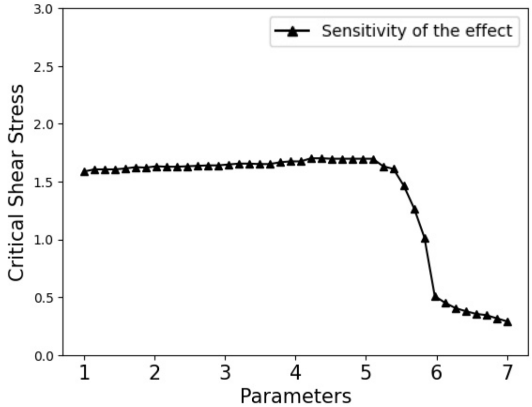


FIGURE 8. Critical shear stress

Source: own elaboration.

flow (Fig. 9). According to the sensitivity analysis results, peak breach flow and peak discharge time were more susceptible to the erodibility coefficient, but breach size was less sensitive.

Furthermore, the water overtopped the landslide dam crest for the dam without a spillway, and the time to peak discharge happened, according to the findings of numerical calculations. The dam's released water storage was approximately 784 million m^3, whereas the dam with a spillway was 615 million m^3. Furthermore, because the dam lacked a spillway, the dam breach method took less time. As a result, spillway excavation may lower the peak breach flow while concurrently increasing the elapsed time of the dam breach, which is an effective disaster mitigation strategy. As a result, for a dammed lake with a substantial reservoir capacity, spillway excavation should be the initial technical intervention.

Peak breach flow was also the most sensitive of all output parameters, fluctuating by more than 50% for different soil erodibility coefficients. As a result, selecting k_d precisely is crucial for numerical simulation of the landslide dam breach process.

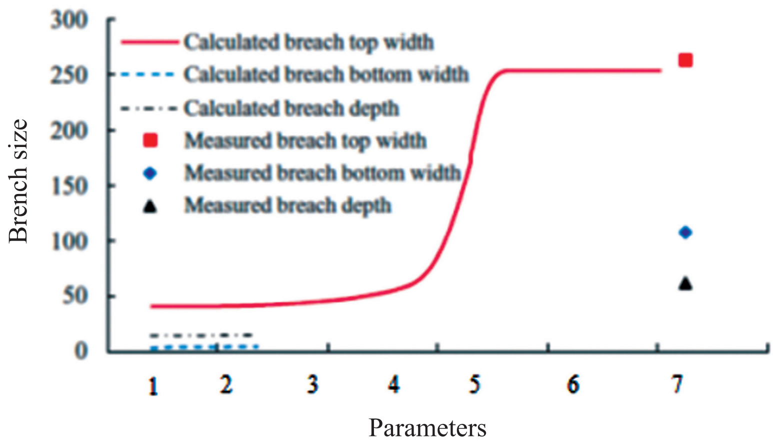


FIGURE 9. Brench size

Source: own elaboration.

The sensitivity analysis results showed that the soil erodibility coefficient had a significant influence on the landslide dam breach process, but the soil critical shear stress had a minor impact. Furthermore, early spillway excavation in dammed lakes with large reservoir capacity can considerably lower the peak flow generated by a landslide dam failure. However, given the growing rate of the water level rise and construction capacity, the spillway's construction circumstances and sectional design should be evaluated holistically.

Conclusion and discussion

Cloudbursts and other comparable occurrences have already wreaked havoc in the Uttarakhand area, where population density and construction activity have skyrocketed. Every year, floods inundate 4–5 villages in Uttarakhand, causing devastating destruction to life and property, yet we are still lacking a well-developed plan for dealing with such calamities. Landslides threaten over 300 communities in the Garhwal Himalaya, yet little is being done for preparation. Based on this research, the following recommendations have been offered for dealing with similar crises:

– The actual origin of a breach is unknown, and experts are still unable to predict the likelihood of a cloudburst occurring. The current threshold value for this occurrence is unknown. As a result, scientists will need a regional climatic-hydrological atlas in the future to establish the association between water activities and landslide activity.
– Despite the lack of awareness of the region's geographical peculiarities, several development activities are ongoing. These procedures should only be implemented after thorough expert examinations. Several regions have been identified as having extremely high hazards, necessitating the development of risk management methods for these areas.
– Some spots inside villages were discovered to be particularly safe via study, and such regions should be recognized in a village-level disaster management plan that should be considered for any future development plans.
– The most damage was seen in regions where homeowners built houses near seasonal or permanent streams. Depending on the site conditions, structures should have a minimum edge distance of 4 to 5 m. Houses should be constructed on non-erodible stratum and have an impermeable drain floor. On steep terrain, RCC-framed constructions with large plinth beams are safe. In this hilly location, a study of drainage systems, whether permanent or seasonal, is critical to lowering mortality in Devpuri-like situations.

– Overhanging boulders and overburden (thickness > 1 m) from steep slopes should be removed to reduce potential future danger.
– Planting trees native to the region with site species may increase slope stability, visual appeal, and the area's micro-ecosystem.
– Given the vulnerability of the region, the government and non-governmental organizations should work together to create a volunteer, community-based watch, monitoring, and alert system. This will be beneficial not only during a crisis, but it will also instill required self-confidence (and hence self-reliance), which is a fundamental goal of any effective recovery program. Despite major R&D operations and knowledge establishment at many institutions and universities, the transfer of expertise to user agencies remains a hurdle. User agencies may occasionally request that the institutions address specific problems. However, there is no framework in place to transmit expertise from R&D institutes to implementing agencies or to refresh knowledge. Capacity-building programs for this sort of research and information dissemination to user organizations are crucial.

References

Aksoy, H., Kirca, V. S. O., Burgan, H. I. & Kellecioglu, D. (2016). Hydrological and hydraulic models for determination of flood-prone and flood inundation areas. *The 7th International Water Resources Management Conference of ICWRS*, *373*, 137–141.

Donghui, S., Liu, S., Ding, Y., Guo, W., Xu, B., Xu, J. & Jiang, Z. (2016). Characterizing the May 2015 Karayaylak Glacier surge in the eastern Pamir Plateau using remote sensing. *Journal of Glaciology*, *62* (235), 944–953.

Gariano, S. L. & Guzzettai, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, *162*, 227–252.

Grämiger, L. M., Moore, J. R., Gischig, V. S., Loew, S., Funk, M. & Limpach, P. (2020). Hydromechanical rock slope damage during Late Pleistocene and Holocene glacial cycles in an Alpine valley. *Journal of Geophysical Research: Earth Surface*, *125* (8), e2019JF005494.

Heim, A. & Gansser, A. (1939). Central Himalaya: geological observations of the Swiss expedition 1936. *Memoir Society Helvetica Science Nature*, *73*, 1–245.

Indian Standard [IS] (1983). *Criteria for earthquake resistant design of structures. Part 1.* New Delhi: Bureau of Indian Standards.

Kayal, J. R. (2010). Himalayan tectonic model and the great earthquakes: an appraisal; Geomatics. *Natural Hazards and Risk*, *1* (1), 51–67.

Khanduri, S. (2017). Disaster hit Pithoragarh District of Uttarakhand Himalaya: causes and implications. *Journal of Geography and Natural Disasters*, *7* (3), 2–5.

Khanduri, S. (2019). Natural hazards in the townships of Nainital, Uttarakhand in India. *International Journal of Engineering Applied Sciences and Technology*, *3* (12), 42–49.

Khanduri, S. (2020). Cloudbursts over Indian sub-continent of Uttarakhand Himalaya: A traditional habitation input from Bansoli, District-Chamoli, India. *International Journal of Earth Sciences Knowledge and Applications*, *2* (2), 48–63.

Khanduri, S. (2021). Formation and failure of natural dams in Uttarakhand Himalaya: an observation from Lwarkha, Chamba Tahsil of Tehri Garhwal District, India. *International Journal of Earth Sciences Knowledge and Applications*, *3* (1), 12–22.

Kumar, D. & Bhattacharjya, R. K. (2020). Study of Integrated Social Vulnerability Index SoVIint of Hilly Region of Uttarakhand, India. *Environmental and Climate Technologies*, *24* (1), 105–122. https://doi.org/10.2478/rtuect-2020-0007

Li, Z., Zhou, F., Han, X., Chen, J., Li, Y., Zhai, S., Han, M. & Bao, Y. (2021). Numerical simulation and analysis of a geological disaster chain in the Peilong valley, SE Tibetan Plateau. *Bulletin of Engineering Geology and the Environment*, *80*, 3405–3422.

Pol, J. C. (2022). *Time-dependent development of Backward Erosion Piping* (doctoral thesis). Delft University of Technology, Delft.

Rana, N., Sharma, S., Sundriyal, Y., Kaushik, S., Pradhan, S., Tiwari, G., ... & Juyal, N. (2021). A preliminary assessment of the 7th February 2021 flashflood in lower Dhauli Ganga valley, Central Himalaya, India. *Journal of Earth System Science*, *130*, 1–10.

Rautela, P. (2013). Lessons learnt from the deluge of Kedarnath, Uttarakhand, India. *Asian Journal of Environment and Disaster Management*, *5* (2), 167–175.

Rautela, P. & Pande, R. K. (2006). Non-monsoonal landslides in Uttaranchal Himalaya (India): implications upon disaster mitigation strategy. *Disaster Prevention and Management*, *15* (3), 448–460.

Samela, C., Manfreda, S., Paola, F. D., Giugni, M., Sole, A. & Fiorentino, M. (2016). DEM-based approaches for the delineation of flood-prone areas in an ungauged basin in Africa. *Journal of Hydrologic Engineering*, *21* (2), 06015010.

Shrestha, F., Steiner, J. F., Shrestha, R., Dhungel, Y., Joshi, S. P., Inglis, S. ... & Zhang, T. (2023). A comprehensive and version-controlled database of glacial lake outburst floods in High Mountain Asia. *Earth System Science Data*, *15* (9), 3941–3961.

Valdiya, K. S. (1980). *Geology of the Kumaun Lesser Himalaya*. Dehradun, India: Wadia Institute of Himalayan Geology.

Valdiya, K. S. (1989). Trans-Himadri intracrustal fault and basement upwarps south of Indus Tsangpo Suture Zone. In L. L. Malinconico & R. J. Lillie (Eds), *Tectonics of Western Himalaya.* Washington, DC: Geological Society of America.

Valdiya, K. S. (2014). Damming rivers in the tectonically resurgent Uttarakhand Himalaya. *Current Science*, *106* (12), 1658–1668.

Xu, J., Grumbine, R. E., Shrestha, A., Eriksson, M., Yang, X., Wang, Y. & Wilkes, A. (2009). The melting Himalayas: cascading effects of climate change on water, biodiversity, and livelihoods. *Conservation Biology*, *23* (3), 520–530.

Summary

Support vector regression tree model for the embankment breaching analysis based on the Chamoli tragedy in Uttarakhand. This study used the analysis to provide considerable support of historical distortion in the Himalayan Chamoli tragedy of 2021. According to multi-objective data and survey results, a precursor event occurred in 2016, and a linear fracture grew at joint planes, suggesting that the 2021 rock ice avalanche will fail retrogressively. To analyze breaching, this study considers seven distinct criteria such as slope, water pressure, and faulty drainage, hydrostatic stress, agricultural operations, cloudbursts, and road building. Based on these characteristics, the support vector regression (SVR) model is utilized to analyze the sensitivity of the link between these parameters. The application of support vector regression analysis on the Chamoli instance confirmed our conclusion that embankment breaching causes glacier retreat and other consequences in increasing sensitivity to the characteristics of fractured rock masses in tectonically active mountain belts. Recent advances in environmental monitoring and geological monitoring systems can be used with the proposed SVR model to provide further information on the location and time of the impending catastrophic collapses in high hill regions.

Instruction for Authors

The journal publishes in English languages, peer-reviewed original research, critical reviews and short communications, which have not been and will not be published elsewhere in substantially the same form. Author of an article is required to transfer the copyright to the journal publisher, however authors retain significant rights to use and share their own published papers. The published papers are available under the terms of the principles of Open Access Creative Commons CC BY-NC license. The submitting author must agree to pay the publication charge (see Charges).

The author of submitted materials (e.g. text, figures, tables etc.) is obligated to restricts the publishing rights. All contributors who do not meet the criteria for authorship should be listed in an Acknowledgements section of the manuscript. Authors should, therefore, add a statement on the type of assistance, if any, received from the sponsor or the sponsor's representative and include the names of any person who provided technical help, writing assistance, editorial support or any type of participation in writing the manuscript.

Uniform requirements for manuscripts

The manuscript should be submitted by the Open Journal System (OJS) at https://srees.sggw.edu.pl/about/submissions. All figures and tables should be placed near their reference in the main text and additionally sent in a form of data files (e.g. Excel, Visio, Adobe Illustrator, Adobe Photoshop, CorelDRAW). Figures are printed in black and white on paper version of the journal (color printing is combined with an additional fee calculated on a case-by-case basis), while on the website are published in color.

The size of the manuscript should be limited up to 10 pages including overview, summary, references and figures (the manuscript more than 13 pages is unacceptable); Please set the text format in single column with paragraphs (A4 paper format), all margins to 25 mm, use the font Times New Roman, typeface 12 points and line spacing one and half.

The submitted manuscript should include the following parts:

- name and SURNAME of the author(s) – up to 5 authors
- affiliation of the author(s), ORCID Id (optional)
- title of the work
- key words
- abstract (about 500 characters)
- text of the paper divided into: Introduction, Material and Methods, Results and Discussion, Conclusions, References and Summary
- references in APA style are listed fully in alphabetical order according to the last name of the first author and not numbered; please find the details below
- post and mailing address of the corresponding author:

  Author's address:
  Name, SURNAME
  Affiliation
  Street, number, postal code, City
  Country
  e-mail: adress@domain

- Plagiarism statement (https://srees.sggw.edu.pl/copyright)

Reference formatting

In the Scientific Review Engineering and Environmental Sciences the APA 6th edition style is used.

Detailed information

More information can be found: https://srees.sggw.edu.pl