الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Centre Universitaire AbdelhafidBoussouf -Mila

Institut des Sciences et de la Technologie

Département de Génie Civil et Hydraulique

N°Ref :……

Graduation thesis prepared to obtaine

The MASTER degree

Speciality :Urban Hydraulics

Predictive Modeling of Soil Erosion Using GIS and Artificial Intelligence. Tafna watershed case.

Presented by:

☞ FERKHI Zineddine

In front of a jury composed of :

Mr.	**KEMOUKH Sami**	**Chairman**
Mrs	**ALLIA Zineb**	**Examiner**
Mr.	**MOUSSOUNI Abderzak**	**supervisor**
Mr	**ZEGHMAR Amer**	**Co-supervisor**

Academic year: 2024/2025

Dedication

*First of all, I thank **Allah** for giving me the courage and willpower to finish my work.*

I dedicate this work with all my love and gratitude:

*To **my parents**, who gave me life and instilled in me the values of hard work, patience, and perseverance. Thank you for your unconditional love, your silent sacrifices, and your unwavering support throughout my academic journey.*

*To **my brothers and sisters**, for always being there with encouragement, prayers, and kind words when I needed them most.*

*To **my teachers**, who guided me with knowledge and passion, and to everyone who contributed to my education this work is a tribute to your dedication.*

To my colleagues in all the years in this university

To my friends who been with all the time by my side either I was right or wrong.

*And to **all those who believed in me**, near or far this achievement is also yours.*

Acknowledgements

I would like to express my deepest gratitude to all those who supported and guided me throughout the completion of this thesis.

First and foremost, my sincere thanks go to ***Dr. Moussouni****, my academic advisor, for his guidance, patience, and constant availability. Their expertise, insightful advice, and constructive feedback greatly contributed to the quality of his work. It has been an honor to work under your supervision.*

I also extend my thanks to ***Dr. Zeghmar Amer,*** *for their valuable observations and recommendations, which helped shape and improve this research.*

My appreciation goes to the entire ***faculty and staff of the urban hydraulics*** *for providing a supportive academic environment and the resources needed for this study.*

A heartfelt thanks to my ***classmates, colleagues, and friends****, whose support, collaboration, and encouragement made this journey both intellectually enriching and personally fulfilling.*

Lastly, I am deeply grateful to my ***family****, who has always stood by me with love, patience, and strength. Without their support, this work would not have been possible.*

To all of you ***thank you from the bottom of my heart****.*

Content table

Figures list:

Tables list

الملخص
تُعد التعرية المائية من التحديات الرئيسية التي تؤدي إلى انخفاض إنتاجية الأراضي الزراعية الواقعة في منابع الأحواض المائية بسبب فقدان التربة والأسمدة والعناصر الغذائية والمادة العضوية، مما يقلل في الوقت نفسه من عمر المنشآت الهيدروليكية في المصب مثل السدود، بسبب تراكم الرواسب بسرعة في خزاناتها. وتحدث التعرية المائية نتيجة لعوامل فيزيائية مناخية وأنثروبوجينية، مثل شدة الأمطار وتضاريس الأرض واستعمالات الأراضي والغطاء النباتي وخصائص التربة، حيث تتضافر هذه العوامل جميعها بشكل خاص في حوض تافنة الواقع في شمال غرب الجزائر لتسبب وتشجع على التعرية المائية.

تهدف هذه الدراسة إلى تقدير فقدان التربة بسبب التعرية المائية وتقييم المناطق المعرضة لهذا التغير في حوض تافنة. وقد تم استخدام المعادلة العالمية المعدلة لفقدان التربة (RUSLE) لتقدير الفقد في التربة والمناطق المتأثرة ضمن بيئة نظم المعلومات الجغرافية .(GIS) تأخذ تطبيقات نموذج RUSLE بعين الاعتبار خمسة عوامل رئيسية وهي: معامل التعرية المطرية؛ العامل الطبوغرافي (طول وانحدار المنحدر)؛ معامل قابلية التربة للتعرية؛ عامل الغطاء النباتي/استعمالات الأراضي/الغطاء؛ وممارسات مكافحة التعرية. وقد سمحت عملية تقييم هذه العوامل ودمجها مكانياً في نظم المعلومات الجغرافية بمحاكاة الفقدان السنوي المحتمل للتربة وتحديد المناطق ذات الأولوية للإدارة.

تشير النتائج إلى أن فقدان التربة في حوض تافنة يتراوح بين 2.92 و5.87 طن/هكتار/سنة حسب الانحدار واستعمالات الأراضي ومعامل التعرية المطرية، مما يدل على وجود تباين في قابلية التعرية. وبينما تبقى الغالبية العظمى من الحوض تحت العتبة المقبولة للتعرية (أقل من 7.0 طن/هكتار)، تسجل بعض المناطق المحلية فقداناً أعلى بكثير؛ حيث تعرف العديد من المناطق معدلات فقدان للتربة تتجاوز 45 طن/هكتار/سنة في نقاط حرجة محلية على منحدرات شديدة وضعيفة الغطاء النباتي داخل الحوض. وتسهم هذه المناطق غير المعالجة بشكل غير متناسب في إنتاج الرواسب داخل الحوض، وبالتالي تُعد مناطق ذات أولوية في الحفاظ على التربة والإدارة.

كما تم إجراء نمذجة لحركة الرواسب باستخدام الشبكات العصبية الاصطناعية – (ANN) تحديداً إصدارات الانتشار الخلفي الأمامي(FFBP) ، والانتشار الخلفي المتسلسل الأمامي(CFBP) ، والمتعدد الطبقات (MLP) في بيئة MATLAB – وذلك بالاعتماد على القياسات الفعلية لتركيز الرواسب العالقة ومعدلات الجريان. وقد اعتُبرت نتائج أداء النماذج في التنبؤ بإنتاج الرواسب مُرضية، كما تم التحقق منها باستخدام معامل الارتباط، ومعامل ناش–سوتكليف(NSE) ، (R)معامل الارتباط، مما يبرز فائدتها في توقع إنتاج الرواسب.

الكلمات المفتاحية :التعرية المائية، نقل الرواسب، الشبكات العصبية الاصطناعية(ANN) ، حوض تافنة، نظم المعلومات الجغرافية(GIS) ، .RUSLE

Abstract

Water erosion still presents a significant challenge that leads to the negative change of productivity of arable land located sources of watersheds through loss of soil, fertilizers, nutrients, and organic matter, which simultaneously reduces the longevity of hydraulic structures downstream, such as dams, due to quickly accumulating sediments in their reservoirs. Water erosion is triggered by factors that include a combination of physico-climatic and anthropogenic agents, such as rainfall intensities, terrain morphology, land use, vegetative cover, and soil characteristics, and in the Tafna watershed in northwestern Algeria, all of these factors are particularly effective in the triggering and encouragement of water erosion.

The aim of this research is to estimate soil loss by water erosion and to assess areas susceptible to this change in the Tafna watershed. The Revised Universal Soil Loss Equation (RUSLE) was used to estimate soil loss and areas affected, within a Geographic Information System (GIS). The application of the RUSLE model considers five main factors, which are Rainfall Erosivity; Topographical factor (Slope, length, slope steepness); Soil Erodibility; Cover Management/Land cover/ Vegetation; and Anti-Erosion Practices. Their assessment and spatial integration in GIS allowed the simulation of potential annual soil loss and identification of areas with priority issues for management.

The results indicate that soil loss in the Tafna watershed is as low as 2.92 t/ha/year and as high as 5.87 t/ha/year based upon slope, land use, and rainfall erosivity, indicating a variability of erosion potential. While, also the vast majority of the watershed continues to be below an acceptable erosion threshold (less than 7.0 t/ha), local areas experience considerably higher soil loss; many areas experience soil loss rates over 45 t/ha/year in local hotspots in steep poorly vegetated sub-basins in the watershed catchment. These unresolved critical zones contribute an inordinate share of the sediment yield in the watershed and thus, are indeed priority areas for soil conservation and management.

Also, modeling of sediment transport was undertaken using artificial neural networks (ANN) - specifically the Multi Layer Perceptron (MLP) versions in MATLAB - based upon measured suspended sediment concentrations and flow rates. The performance of the models to predict sediment yield were deemed satisfactory, as validated through determination of correlation constant and Nash–Sutcliffe Efficiency (NSE),The correlation coefficient(R) highlighting their utility in forecasting sediment yield.

Keywords: Soil erosion, Sediment transport, Artificial Neural Networks (ANN), Tafna watershed, GIS, RUSLE.

Résumé

L'érosion hydrique reste un défi important qui entraîne une diminution de la productivité des terres arables situées en amont des bassins versants par la perte de sol, d'engrais, de nutriments et de matière organique, réduisant simultanément la longévité des ouvrages hydrauliques en aval, tels que les barrages, en raison de l'accumulation rapide de sédiments dans leurs réservoirs. L'érosion hydrique est déclenchée par une combinaison de facteurs physico-climatiques et anthropiques, tels que l'intensité des précipitations, la morphologie du terrain, l'utilisation des terres, la couverture végétale et les caractéristiques du sol. Dans le bassin versant de la Tafna, au nord-ouest de l'Algérie, tous ces facteurs contribuent particulièrement au déclenchement et à l'intensification de l'érosion hydrique.

L'objectif de cette recherche est d'estimer la perte de sol due à l'érosion hydrique et d'évaluer les zones sensibles à ce phénomène dans le bassin versant de la Tafna. L'équation universelle révisée de perte de sol (RUSLE) a été utilisée pour estimer la perte de sol et les zones affectées, dans un système d'information géographique (SIG). L'application du modèle RUSLE prend en compte cinq facteurs principaux : l'érosivité des précipitations ; le facteur topographique (longueur et pente de la pente) ; l'érodibilité du sol ; la couverture végétale/l'utilisation des terres/la végétation ; et les pratiques anti-érosives. Leur évaluation et leur intégration spatiale dans le SIG ont permis la simulation de la perte de sol annuelle potentielle et l'identification des zones prioritaires à gérer.

Les résultats indiquent que la perte de sol dans le bassin versant de la Tafna varie entre 2,92 t/ha/an et 5,87 t/ha/an selon la pente, l'utilisation des terres et l'érosivité des précipitations, montrant une variabilité du potentiel d'érosion. Alors que la grande majorité du bassin versant reste en dessous du seuil acceptable d'érosion (moins de 7,0 t/ha), certaines zones locales enregistrent des pertes de sol nettement plus élevées ; de nombreuses zones subissent des pertes de sol supérieures à 45 t/ha/an dans des points chauds locaux situés dans des sous-bassins escarpés et faiblement végétalisés du bassin versant. Ces zones critiques non traitées contribuent de manière disproportionnée à la production de sédiments dans le bassin et doivent donc être considérées comme prioritaires pour la conservation des sols et la gestion.

En outre, la modélisation du transport de sédiments a été réalisée à l'aide de réseaux de neurones artificiels (RNA) – en particulier la version Multi Layer Perceptron (MLP) sous MATLAB – basées sur les concentrations mesurées de sédiments en suspension et les débits. Les performances des modèles pour prédire la production de sédiments ont été jugées satisfaisantes, comme validées par la détermination du coefficient de corrélation, de l'efficacité de Nash–Sutcliffe (NSE) et le coefficient de correlation (R), soulignant leur utilité pour la prévision de la production de sédiments.

Mots-clés : Érosion des sols, Transport de sédiments, Réseaux de neurones artificiels (RNA), Bassin versant de la Tafna, SIG, RUSLE.

General introduction

Water erosion of soils is one of the most serious and extensive forms of land degradation on Earth, with downstream consequences on agricultural efficiency, water resources, and ecosystem sustainability. In semi-arid regions coupled with climate variability, such as the Mediterranean basin, increasing environmental pressures arising from deforestation, unsustainable land-use practices and increasing anthropogenic pressures have amplified rates of soil loss. The case of Algeria, more precisely the Tafna watershed located in the northwestern portion of the country, demonstrates the vulnerability of such territories where surface erosion, sedimentation and environmental degradation directly threaten agricultural production processes.

Understanding and quantifying the processes that generate soil erosion is a requisite consideration for ecosystem and environmental management as well as for sustainable land-use planning, water resource management, and the agricultural development process. Traditional empirical models, such as the Universal Soil Loss Equation (USLE) and more recently the Revised Universal Soil Loss Equation (RUSLE), have long been used as tools for estimating long-term averages of annual soil loss due to sheet and rill erosion. In combination with Geographic Information Systems (GIS) and remote sensing information (satellite-imagery, elevation etc), RUSLE has provided researchers a spatially explicit approach to understanding erosion risk and to identifying which areas may be more threatened.

While erosion is non-linear and complex in nature, which include climatic, topographic, edaphic, and anthropogenic variables - traditional methods for measuring soil erosion failure don't capture such inter-relational dependence and process interactions with sufficient precision, especially in contrasting and spatially complex environments. Artificial intelligence (AI) and machine-learning (ML) methods have joined the ranks in the last few years as plausible alternatives or complements to these traditional models. ML algorithms (eg. Random Forests, Support Vector Machines, Artificial Neural Networks etc) are capable of learning from large amounts of data to find patterns or relationships that the linear models have neglected, and thus can improve prediction of erosion susceptibility and sediment yield.

This graduation project proposes an integrated framework that combines the strengths of both empirical and data-driven approaches to model and analyze soil erosion risk in the Tafna watershed, a region characterized by steep slopes, variable rainfall patterns, and increasing anthropogenic disturbance. The study leverages the RUSLE model within a GIS environment

to estimate soil loss spatially, and integrates it with machine learning techniques trained on biophysical and environmental data to improve predictive accuracy.

The objectives of this research are multiple:

- To assess the spatial distribution and intensity of soil erosion within the Tafna basin using RUSLE-GIS methodology.
- To evaluate the added value of machine learning models in erosion susceptibility mapping.
- To compare hydrological model outputs with empirical estimates for validation and cross-analysis.
- To identify priority areas for conservation and propose informed land management strategies.

This work contributes to bridging the gap between theoretical modeling and practical land management in semi-arid regions, offering a replicable methodology for erosion analysis under current and future environmental pressures. By demonstrating the synergy between conventional and AI-driven tools, it seeks to enhance decision-making in watershed protection and natural resource sustainability.

CHAPTER I:
BIBLIOGRAPHIC SYNTHESIS

I.1. Introduction

Soil is a vital component of terrestrial ecosystems, providing essential nutrients, moisture, and habitat for diverse biological communities. However, soil erosion poses a significant threat to land productivity, water quality, and ecological balance, emerging as one of the most pressing environmental challenges today. Erosion, driven by hydraulic, wind, and freeze-thaw forces, leads to detrimental effects such as sedimentation in waterways and depletion of land resources (Bach et al., 2020; Zhang et al., 2024).

Traditional methods for studying soil erosion are often labor-intensive and geographically limited. In contrast, remote sensing technology has revolutionized the field by enabling rapid data collection and spatial analysis, thereby enhancing our understanding of erosion processes (Wu et al., 2024). This technology encompasses a variety of methods, including satellite imaging and GIS analysis, which allow for comprehensive assessments of erosion characteristics and dynamics across large areas (Gong et al., 2022).

This study reviews existing literature on remote sensing applications in soil erosion research, highlighting key methodologies, influential factors, and the role of advanced models.

I.2. Soil Erosion

Soil erosion is a fundamental geomorphological process that reshapes terrestrial landscapes and plays a critical role in the degradation of soil stability and fertility across diverse ecological zones. Driven by natural forces such as water and wind, soil erosion begins when external agents detach and transport soil particles, leading to surface degradation, reduced agricultural productivity, and altered landforms. This process contributes significantly to land degradation and the long-term loss of arable land. Research has shown that soil erosion can lead to an estimated 0.3% annual reduction in global agricultural productivity, posing a direct threat to food security (Sartori et al., 2019; Borrelli et al., 2020).

I.2.1. Types and Agents of Soil Erosion

I.2.1.1. Water Erosion

Water erosion is the most common form of accelerated erosion. It results from two main energy sources: the impact of falling raindrops (vertical force) and surface runoff (horizontal force). Raindrop impact detaches soil particles, while surface runoff transports them downslope (Telkar et al., 2015). Several types of water erosion are recognized (Figure I.1):

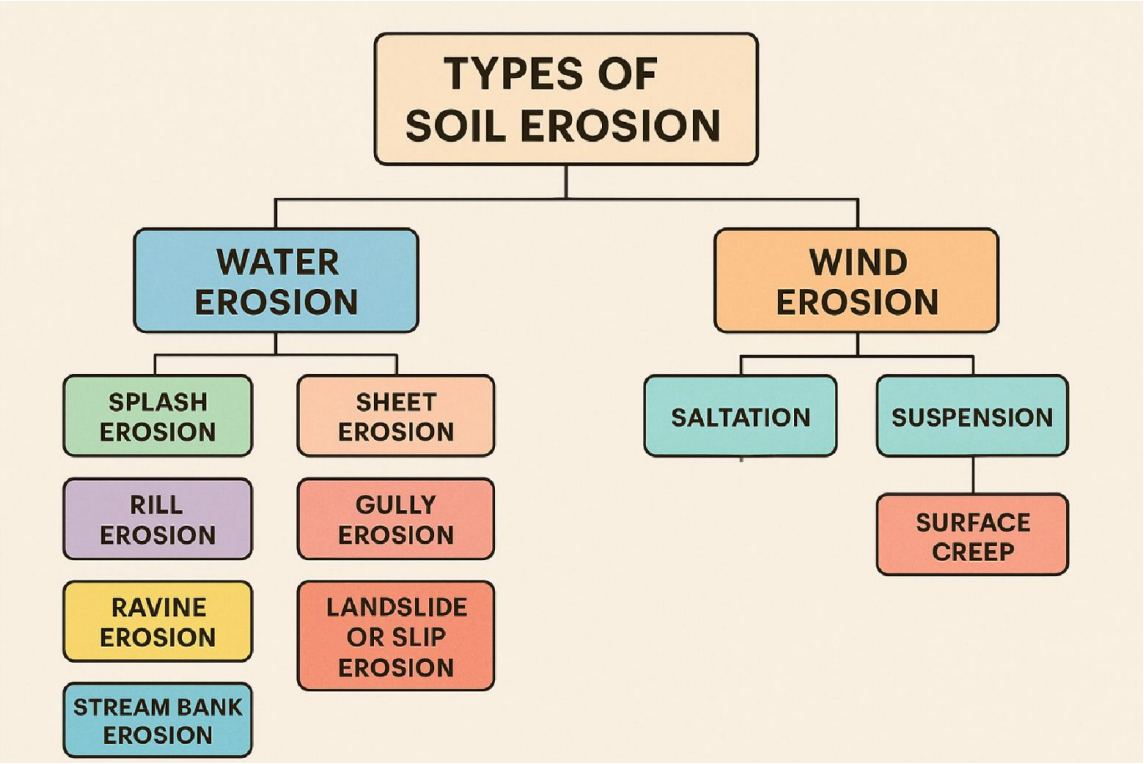


Figure I.1: Classification of Soil Erosion Types Based

I.2.1.2. Splash Erosion

Splash erosion is the initial stage of the soil erosion process, where individual raindrops strike the soil surface with enough force to dislodge particles. These dislodged particles are then propelled a short distance from their original location, leading to surface sealing and reduced infiltration. Although this form of erosion may appear minor, it plays a critical role in initiating more severe erosion types by loosening soil for subsequent transport by runoff (Figure I.2).

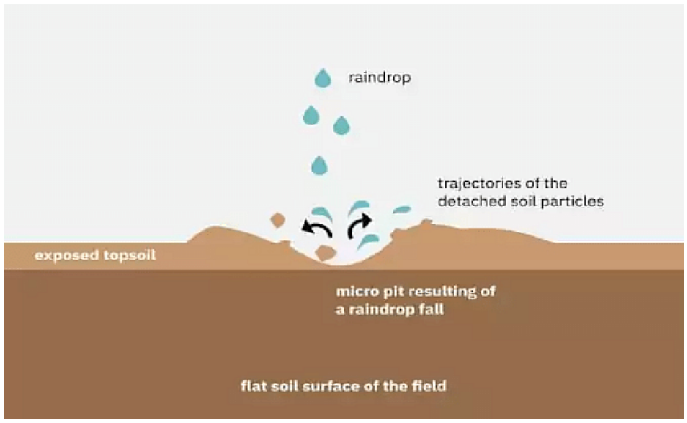


Figure I.2: Illustration of Splash Erosion Caused by Raindrop Impact

I.2.1.3. Sheet Erosion

Sheet erosion involves the uniform removal of a thin layer of topsoil over large surface areas during rainfall events. This type of erosion often goes unnoticed because it does not leave visible scars on the land, yet it cumulatively results in substantial loss of fertile soil. Due to its subtle and widespread nature, it is sometimes referred to as the "silent thief" or the "death of the farmer." It frequently follows splash erosion and often precedes more severe forms such as rill or gully erosion. The progression and classification of these erosion types are depicted in Figure I.3.

Figure I.3. Visual Representation of Sheet Erosion

I.2.1.4. Rill Erosion

Rill erosion occurs when surface runoff begins to concentrate, forming small but clearly defined channels known as rills. These channels indicate a transition from sheet erosion to more localized erosion patterns. While rills can typically be removed through standard tillage practices, neglecting them allows the erosion process to intensify, potentially leading to gully formation. The formation and progression of rill erosion are illustrated in Figure I.4.

Figure I.4. Formation of Rill Erosion Due to Concentrated Surface Runoff

I.2.1.5. Gully Erosion

Gully erosion represents a more advanced and visibly destructive form of soil erosion. It occurs when concentrated surface runoff carves deep, wide channels referred to as gullies

into the landscape. These features can severely fragment the land, making it difficult or impossible to cultivate. Gully erosion often results from the unchecked progression of rill erosion and can expand rapidly during intense rainfall events. The typical appearance and impact of gully formation are depicted in Figure I.5.

Figure I.5. Development of Gully Erosion from Concentrated Runoff in Degraded Landscapes

I.2.1.6. Ravine Erosion

Ravine erosion is the prolonged and advanced stage of gully erosion, typically observed in deep alluvial soils. Over time, continuous runoff enlarges gullies into extensive, deep, and wide chasms called ravines that permanently scar the landscape. These formations significantly reduce land usability and agricultural productivity. The severe degradation and characteristic form of ravines are illustrated in Figure I.6.

Figure I.6. Formation of Ravine Erosion as an Advanced Stage of Gully Development

I.2.1.7. Landslide or Slip Erosion

Landslide or slip erosion occurs on steep slopes exceeding 20%, especially in mountainous or hilly terrains. This form of erosion is typically triggered by heavy rainfall, seismic activity, or the saturation of soil layers, which reduces cohesion and causes large masses of soil and rock

to slide downslope. Such events can lead to sudden and severe land degradation, infrastructure damage, and loss of vegetation. The dynamics and visual characteristics of this erosion type are illustrated in Figure I.7.

Figure I.7. Landslide (Slip) Erosion Triggered by Rainfall or Slope Instability in Steep Terrain

I.2.1.8. Stream Bank Erosion

Stream bank erosion occurs primarily during high-flow conditions in rivers and streams, when increased water velocity and volume can cause channels to shift or meander. As the water undercuts and erodes the banks, large sections of soil become detached and are carried downstream, contributing to sedimentation and the degradation of adjacent land. This process can significantly impact agricultural areas, infrastructure, and ecosystems along riverbanks. The process and effects of stream bank erosion are illustrated in Figure I.8.

Figure I.8. Stream Bank Erosion During High-Flow Events

I.2.2. Wind Erosion

Wind erosion is a major form of soil degradation, particularly in arid and semi-arid regions with minimal vegetation and loose, dry soil surfaces. Strong winds mobilize soil particles especially during dust storms leading to the removal of fertile topsoil, exposure of nutrient-

poor subsoil, and the formation of sand dunes. This process significantly reduces land productivity. Wind erosion occurs primarily through saltation, suspension, and surface creep, as illustrated in Figure I.9.

Figure I.9. Primary Processes of Wind Erosion in Dryland Environments

I.2.2.1. Saltation

The primary mechanism, involving the bouncing movement of soil particles (0.1–0.5 mm in diameter) caused by wind pressure and inter-particle collision. Saltation accounts for 50–75% of wind-driven soil erosion.

I.2.2.2. Suspension

Fine soil particles (<0.1 mm) are lifted and suspended in the air for long distances. These are eventually deposited when gravity overcomes the uplifting wind force.

I.2.2.3. Surface Creep

Larger soil particles (typically >0.5 mm) are rolled or dragged along the ground surface due to wind pressure and collision with saltating particles.

I.2.3. Factors That Enhance Soil Degradation

Soil erosion occurs when the land surface is left bare and exposed to the forces of wind and raindrops. The severity of erosion depends on the intensity of rainfall and wind activity across unprotected land(Gandhi krishiVishwavidyalaya, n.d.; Pal et al., 2025). According to Pimentel and Burgess (2013), several key factors contribute to soil degradation:

I.2.3.1. Rainfall and Surface Runoff

Rainfall initiates the breakdown of soil aggregates and disperses soil particles. Runoff primarily transports lighter materials such as organic matter, silt, and fine sand. However, during intense rainfall events, even larger and denser soil particles may be displaced, accelerating soil loss.

I.2.3.2. Agricultural Activities

Farming practices can significantly alter soil structure and reduce organic matter content, making soils more vulnerable to erosion. In particular, land tillage disrupts soil cohesion, weakening its structure and increasing erosion susceptibility. Conversely, less intensive agricultural practices tend to exert a reduced impact on soil degradation.

I.2.3.3. Vegetative Cover

Plants and ground cover play a critical role in protecting soil from erosion by anchoring soil particles and reducing the force of rainfall impact. Areas lacking sufficient natural vegetation are more prone to erosion, particularly during heavy rains or dry, windy conditions.

I.2.3.4. Wind

Wind contributes to soil degradation, particularly in arid and semi-arid regions where soils are dry and loosely structured. Light winds may have minimal impact, but strong winds can transport lighter soil particles especially sandy soils over long distances, leading to significant erosion.

I.2.3.5. Slope of the Land

Topography also influences soil erosion. Steep slopes accelerate water flow, increasing the volume and speed of runoff, which in turn intensifies soil displacement (Chen et al., 2011; Nenadović et al., 2013). Sloped lands are particularly at risk when vegetation is sparse or removed.

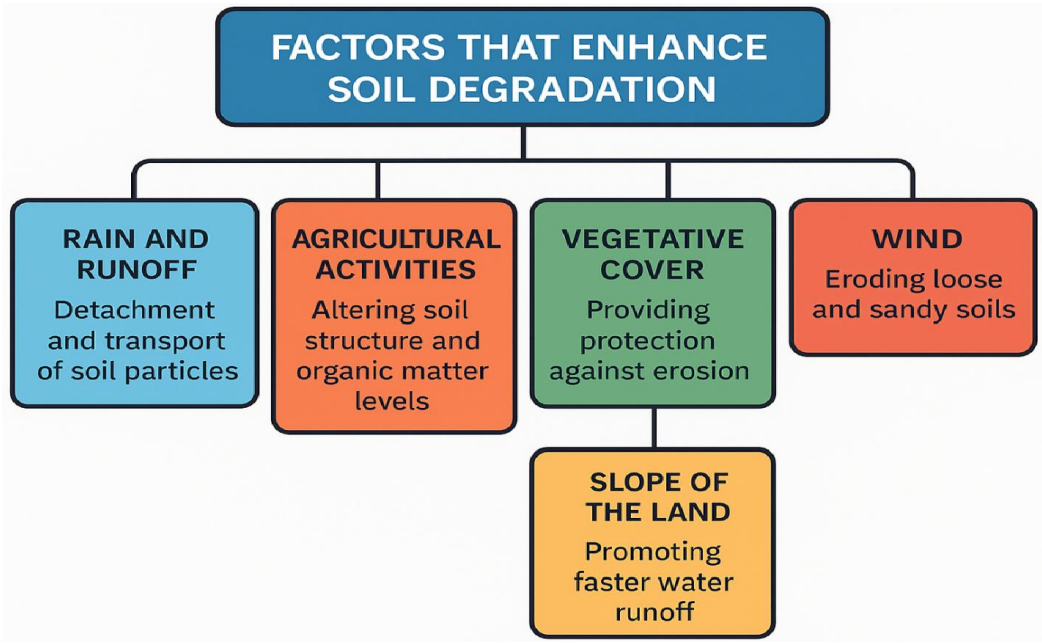


Figure I.10. Primary Processes of Wind Erosion

I.3. Remote Sensing Data Sources Available for Soil Erosion Research

Soil erosion is a complex and dynamic process influenced by a combination of natural and anthropogenic factors, including climate, topography, soil characteristics, vegetation cover, geology, and land use practices (Zhang et al., 2024). Effective monitoring and modeling of soil erosion require detailed, multi-scale data to capture both spatial variability and temporal dynamics.

Remote sensing has emerged as a powerful tool for observing and analyzing these factors due to its multi-temporal, multi-spectral, and multi-resolution capabilities. Satellite imagery facilitates the extraction of critical erosion-related variables such as vegetation indices, land cover changes, rainfall distribution, and terrain parameters (e.g., slope and aspect) (Gong et al., 2022). The availability of long-term satellite datasets enables researchers to assess historical erosion trends and predict future scenarios over large and inaccessible areas, overcoming many limitations of traditional field-based methods (Seutloali et al., 2018).

Today, Earth observation missions such as Landsat, Sentinel, MODIS, and others provide an extensive archive of freely accessible imagery, making satellite remote sensing the most widely adopted approach in soil erosion studies. The simplicity of data acquisition and preprocessing further supports its integration into erosion modeling frameworks. Figure 1 and Table 1 summarize the most commonly used satellite data sources, classified by spatial resolution and application type in soil erosion research.

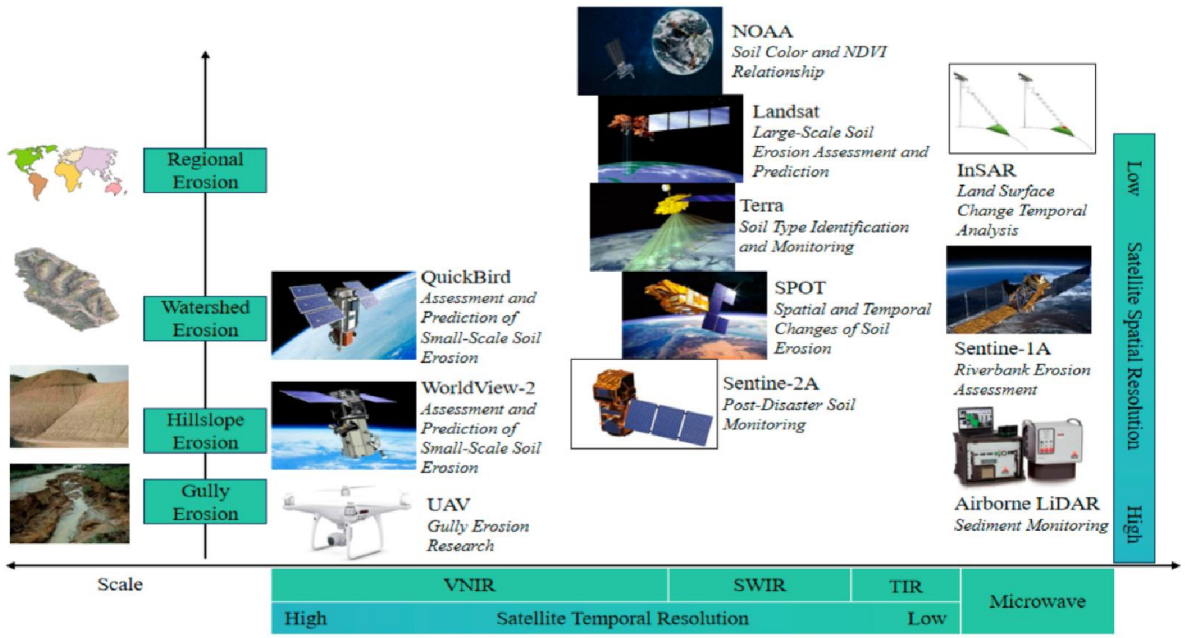


Figure I.12. Common remote sensing data sources and their applications in soil erosion (Wang et all., 2024).

Table I.1. Basic information on common remote sensing data sources used in soil erosion (Wang et all., 2024).

Data Type	Satellite Sensor	Spatial Resolution (m)	Temporal Resolution (day)	Spectral Domains Involved	Application Cases
Low Spatial Resolution	NOAA AVHRR	1100	6	VNIR, SWIR, TIR	Soil erosion monitoring, climate
Medium Spatial Resolution	Landsat MSS	80	18	VNIR	LULC change, erosion modeling
	Landsat TM	30/120	16	VNIR, SWIR, TIR	
	Landsat ETM+	15/30/60	16	VNIR, SWIR, TIR	
	Landsat OLI, TIRS	15/30/100	16	VNIR, SWIR, TIR	
	SPOT HRV	10/20	26	VNIR	
	SPOT HRVIR	10/20	26	VNIR, SWIR	
	CBERS-01/02	20/78/156/258	3/26	VNIR, SWIR, TIR	
	Sentinel-2A MSI	10/20/60	5	VNIR, SWIR	
	Terra ASTER	15/30	16	VNIR, SWIR, TIR	
	GF-2	0.81/3.24	5	VNIR	Urban analysis, precision agriculture

	IKONOS	0.82/3.2	3	VNIR	
	QuickBird	0.61/2.44	1–6	VNIR	
	WorldView-2	0.46/1.85	1.1/3.7	VNIR	
	ZY-1-02C	2.36/5/10	3–5	VNIR	
	CBERS	2.36/20/258	3	VNIR	
High Spatial Resolution	PAN, IRS, MUX, WFI	5/10/20/40/ 73/80	3/26	VNIR, SWIR, TIR	
Radar Data	Sentinel-1 C-SAR	5–40	12	C-Band	Flood mapping, terrain structure
Hyperspectral Data	AISA	Sub-meter	--	VNIR	Mineralogy, vegetation studies

I.5. Soil Erosion Models

A review of global soil erosion modelling practices reveals the use of 435 distinct models and model variants, though overlaps in naming suggest that some are functionally equivalent. Among these, models in the (R)USLE familyincluding USLE, RUSLE, USLE-SDR, RUSLE-SDR, and SEDDare by far the most widely applied, accounting for over 1200 applications, or around 41% of total usage. This figure would increase further if related empirical models such as WaTEM/SEDEM, EPIC, SWAT, and USPED were grouped under the (R)USLE umbrella.

Other commonly used process-based models include WEPP (7.4%), LISEM (1.9%), and EROSION-3D (1%), while regional models like PESERA (0.8%) and EUROSEM (0.6%) are less frequently used. Among the empirical alternatives to (R)USLE, SWAT (6%), WaTEM/SEDEM (4.6%), and (R)MMF (2%) are prominent(Borrelli et al., 2021).

Trends over time, broken into four-year intervals, show a consistent increase in the use of (R)USLE, SWAT, and WaTEM/SEDEM, with moderate growth for models like WEPP, AGNPS, MMF, Erosion-3D, and LISEM. In contrast, EUROSEM usage has declined. This evolution in model preference and deployment over time is illustrated in Figure I.13, which presents a 2D digital infographic summarizing the prevalence, development, and spatial resolution of soil erosion modelling across global applications.

In terms of spatial resolution, more than 56% of studies did not report grid size. Among those that did, very high-resolution outputs (≤5 m) represented 7.2%, high-resolution outputs (>5–

25 m) accounted for 11.9%, medium resolution (>25–150 m) for 19.8%, and moderate resolution (>150–300 m) for 1.6%. A small portion (~3%) used coarse resolutions between 330 and 60,000 meters. There is a clear trend toward finer spatial resolutions in more recent applications, particularly at watershed and hillslope scales. However, a consistent link between model type and grid size is absent, except in large-scale studies where empirical (R)USLE and (R)WEQ models dominate.

Regarding model validation, it was performed in about 58% of the reviewed studies. The most common method involved comparing model results to measured sediment yield (26%), followed by comparisons with field-measured erosion rates (18%), outputs from other models (10%), and expert judgment (3%). This indicates a growing commitment to validation and model reliability, especially in newer studies.

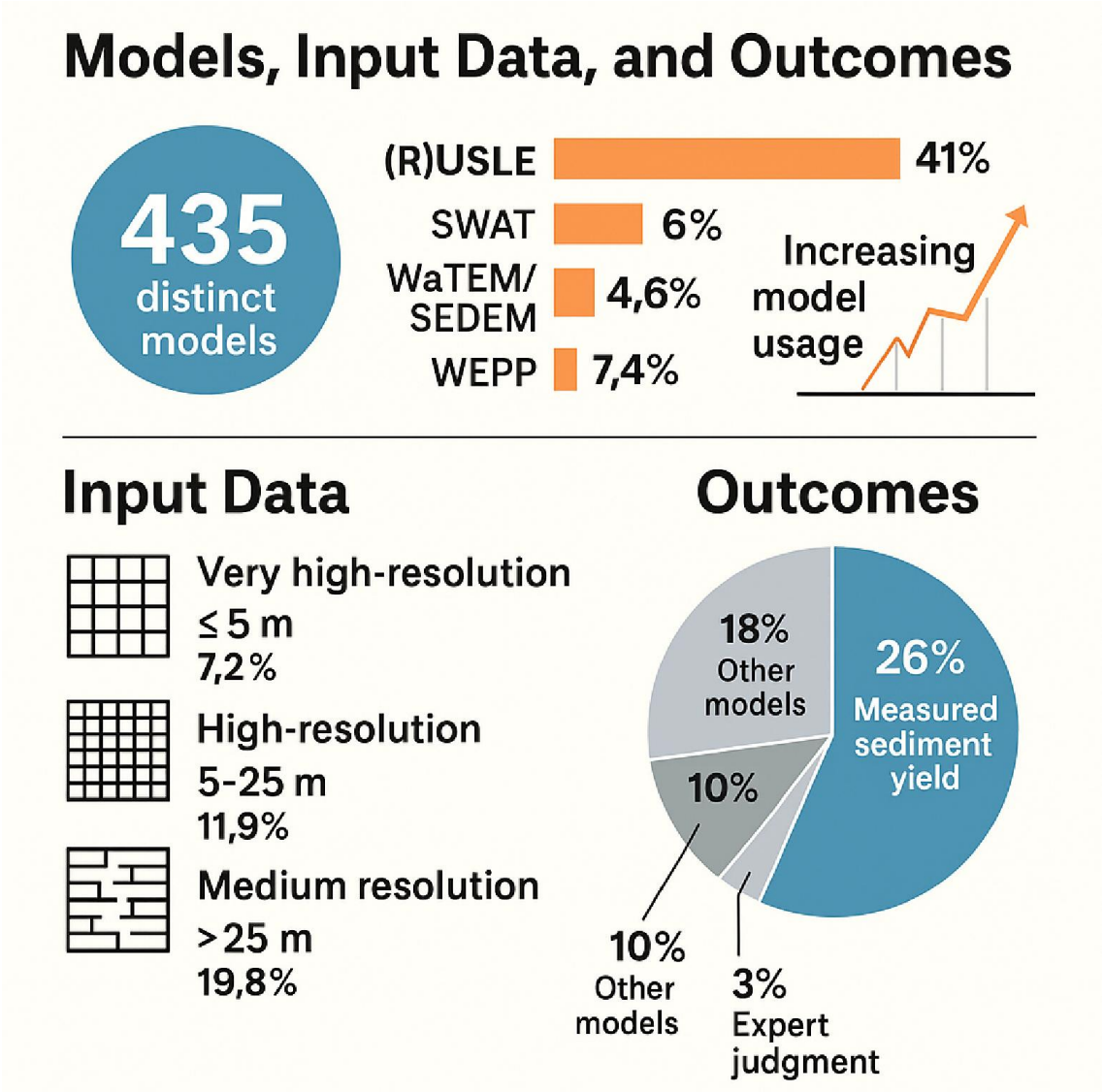


Figure I.13. Global Trends and Applications of Soil Erosion Models

I.4.1. The Revised Universal Soil Loss Equation (RUSLE)model

The Revised Universal Soil Loss Equation (RUSLE) is widely applied within GIS frameworks to estimate annual soil erosion across varied landscapes, including tropical, agricultural, and mountainous watersheds, as well as regions experiencing dynamic land-use change(Ganasri and Ramesh, 2016; Ghosal and Das Bhattacharya, 2020; Jahun et al., 2015).

RUSLE integrates five erosion-related factors, drawn from three key data sources:

Climatic and topographic data – provide inputs for calculating the rainfall-runoff erosivity factor (R) and the topographic factor (LS), which combines slope length (L) and slope steepness (S) derived from digital elevation models (DEMs).

Crop or land cover database – used to determine the cover and management factor (C).

Soil data – informs the soil erodibility factor (K) through texture and organic matter properties.

The general method involves estimating each of the RUSLE factors (R, K, LS, C, P) using diverse data sources such as satellite imagery, climate records, topographic maps, and soil surveys. The topography of the landscape, especially slope and elevation change, plays a critical role in defining the LS factor, which governs how terrain influences erosion potential.

The model calculates average annual soil loss using the formula:

$$A=R\times K\times L\times S\times C\times P$$

Where:

A = Predicted annual soil loss (t/ha/year)

R = Rainfall-runoff erosivity factor

K = Soil erodibility factor

L = Slope length factor

S = Slope steepness factor

C = Cover-management factor

P = Support practice factor

I.4.2. The Modified Universal Soil Loss Equation (MUSLE)

The Modified Universal Soil Loss Equation (MUSLE) was developed as an event-based enhancement to the RUSLE model, allowing for more accurate estimation of soil erosion at shorter temporal scales such as daily or per storm event(CHESIRE, 2022; Mohammadi et al., 2025; Reda et al., 2024). Unlike RUSLE, which uses rainfall erosivity (R) to estimate long-term average erosion, MUSLE replaces the R factor with runoff volume (Q) and peak runoff rate (qp)key hydrological parameters that directly influence sediment transport during rainfall events. This modification enables MUSLE to be integrated into hydrological models such as

SWAT (Soil and Water Assessment Tool), making it suitable for continuous simulations under varying land use and climate conditions. The MUSLE formula is expressed as:

$A=11.8\times(Q\times q_p)^{0.56}\times K\times C\times P\times LS$

Where:

- A = Sediment yield per event (tons)
- Q = Runoff volume (mm)
- q_p = Peak runoff rate (m^3/s)
- K, C, P, and LS = Soil erodibility, cover-management, conservation practices, and topographic factors, respectively

I.4.3. The Soil and Water Assessment Tool (SWAT)

The Soil and Water Assessment Tool (SWAT) is a semi-distributed, process-based watershed model developed to predict the long-term impacts of land management practices on water, sediment, and agricultural chemical yields(Lei et al., 2024; Zhao et al., 2024). It operates at the watershed scale and divides the area into multiple sub-basins and Hydrologic Response Units (HRUs), which represent unique combinations of land use, soil type, and slope class(Muauz et al., 2024; Prakash et al., 2024).

SWAT simulates the water balance using the equation:

$SWt=SW_0+\sum (R_{day}-Q_{surf}-E_a-W_{seep}-Q_{gw})$

Where:

- SW_t is the final soil water content,
- SW_0 is the initial soil water content,
- R_{day} is daily precipitation,
- Q_{surf} is surface runoff,
- E_a is evapotranspiration,
- W_{seep} is percolation,
- $Q_{gw}$is return flow.

I.4.4. The Water Erosion Prediction Project (WEPP)

The Water Erosion Prediction Project (WEPP) model is a physically-based, continuous simulation model designed to predict soil erosion and sediment transport on hillslopes and small watersheds(Bhat et al., n.d.; MAHRUKH et al., 2025; Ugwu et al., 2024). Unlike empirical models such as RUSLE or MUSLE, WEPP simulates fundamental hydrological and erosion processes using climate, soil, topography, and land management data.

WEPP calculates soil erosion by simulating the following processes:

- Infiltration and surface runoff generation,
- Soil detachment by raindrop impact and overland flow,
- Sediment transport and deposition across the slope or channel.

The core equation for sediment transport capacity in WEPP is:

$Tc=a\times q\times S^{b}$

Where:

- Tc = sediment transport capacity,
- q = flow discharge per unit width,
- S = slope gradient,
- a,b = empirical coefficients based on soil and flow characteristics.

I.5. Conclusions

This comprehensive review underscores the complexity of soil erosion and the evolving role of remote sensing and modeling techniques in understanding and mitigating its impacts. Soil erosion, driven primarily by water and wind, continues to degrade landscapes, reduce agricultural productivity, and disrupt ecosystems—particularly in regions with poor land management or insufficient vegetation cover.

Remote sensing technologies, such as satellite imagery and GIS, have transformed erosion research by enabling large-scale, multi-temporal monitoring of erosion dynamics. Coupled with digital elevation models, land use data, and climatic records, these tools offer unprecedented insights into spatial patterns and temporal changes in erosion risk.

Modeling tools like RUSLE, MUSLE, SWAT, and WEPP each contribute distinct advantages. RUSLE remains the most widely used due to its simplicity and empirical strength, while MUSLE introduces hydrologic responsiveness suitable for event-based analysis. SWAT enables holistic watershed-scale assessments, integrating land management and hydrological interactions, and WEPP provides a detailed, process-based simulation of erosion mechanisms. However, a significant limitation persists in the lack of field validation across many modeling applications, which can undermine result reliability.

Moving forward, integrating high-resolution remote sensing, validated process-based models, and localized field data is essential for developing robust, actionable soil conservation strategies. Improved model calibration and validation, especially in data-scarce, erosion-prone regions, are critical for guiding sustainable land management and informing global efforts to combat soil degradation.

CHAPTER II:
PRESENTATION OF THE STUDY AREA

II.1.Introduction

A watershed is a land unit defined by natural boundaries where all surface water drains toward a single outlet (Anctil et al., 2012). This chapter focuses on the Tafna watershed, located in the extreme northwest of Algeria. The goal is to present a detailed description of its main morphometric features. The approach is descriptive and based on spatial data analysis. Morphometric analysis is a reliable tool for watershed evaluation and supports effective planning and integrated water resource management (Pingale et al., 2012). A clear understanding of these physical characteristics is necessary to design sustainable strategies for soil conservation, erosion control, and optimal water use across the Tafna basin.

II.2.Geographic overview

Oued Tafna (Latitude: 35° 05' N; Longitude: 1° 20' W) rises from the mountainous regions of Tlemcen in northwestern Algeria. It flows over a course of about 170 kilometers before emptying into the Mediterranean Sea near Rachgoun. The Tafna watershed, located at the extreme west of Algeria. It is bounded by 1° and 2° west longitude and34°5' at 35°3' north latitude. Covers an area of 7700 km2,less than one third of its surface area is located on Moroccan territory. However, 5340 km2 is on the Algerian territory (Aboura, 2006; ABH, 2006). The basin is delimited by Tlemcen Mountains, mainly composed of mountains in the south (800 to 1400 m of altitude). This orographic structure, which is dominated to the north by the Taras Mountains (1081 m) of narrow width, constitutes an important barrier against precipitation (Meddi *et al.*,2013). The hydrographic network of the Tafna River is composed of two main wadis, the East Isser and the Tafna. It lies about 500 kilometers west of Algiers.

The Tafna watershed is part of the Tell Atlas system. It is bounded:

- **To the south**: by the Tlemcen Mountains, forming a rugged natural barrier with steep slopes and marked altitudinal contrasts
- **To the east**: by the Oued Isser sub-watershed, transitioning toward the central Tell
- **To the north**: by the coastal strip and hills that separate the basin from the Mediterranean Sea
- **To the west**: by the Moroccan border and the lower catchments of the Moulouya River system

The basin features a varied topography ranging from high-altitude zones in the southern mountains to intermediate hills and low coastal plains. This topographic diversity strongly affects rainfall distribution, hydrological response, and soil types across the basin, resulting in high spatial variability in erosion dynamics.

Oued Tafna is the main watercourse draining this region. It collects surface runoff from seasonal and irregular rainfall, making it a vital water source. Snowmelt from higher altitudes contributes significantly to stream flow in wet years. The oued plays a key role in local ecosystems, agriculture, and rural livelihoods.

The watershed holds economic importance in the region. Agriculture, especially cereal production, olive cultivation, and orchards, relies on water availability from Tafna and its tributaries. The forested zones in the upper basin support biodiversity and provide ecological balance.

However, the Tafna watershed faces several environmental challenges, notably soil erosion. Steep slopes, degraded vegetation cover, and unsustainable land use practices accelerate surface runoff and sediment loss, increasing pressure on natural and agricultural systems.

The geographical location and precise limits of the watershed are shown in Figure II.1.

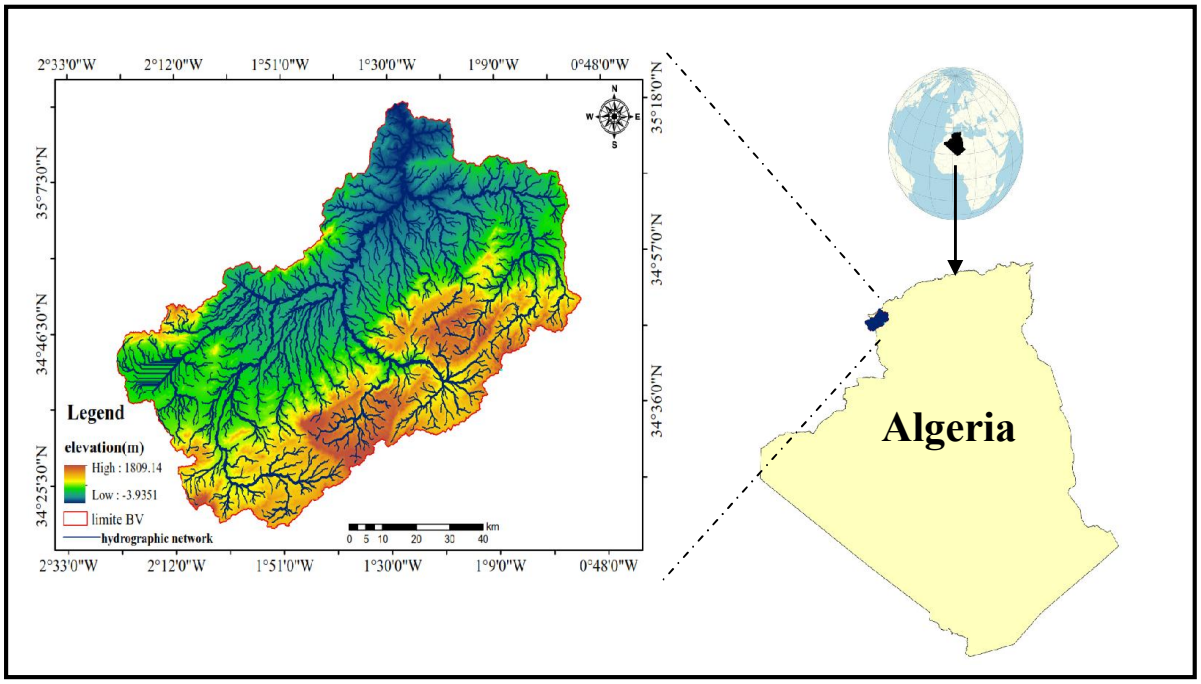


Figure II.1. Geographic location and boundaries of the Tafna watershed

II.3.Topographical Characteristics of the Study Area

Using a Geographic Information System (GIS) with remote sensing data allowed identification of the topographic profile of the Tafna watershed. A Digital Elevation Model (DEM) of the SRTM type (Shuttle Radar Topography Mission) with a spatial resolution of 30 meters was downloaded and integrated into the ArcGIS software interface. After various operations and data processing within the GIS environment, it was possible to calculate several relevant topographic and morphometric parameters, which we present below:

II.4. Altimetry

The DEM demonstrates that the Tafna watershed exhibits significant elevation variation, ranging from approximately **-3.93 m** (likely coastal or river mouth depression zones) to about **1809.14 m** in the southern highlands. This elevation gradient defines the basin's geomorphology, flow dynamics, and erosion potential as its shown in the figure bellow (Figure II.2.)

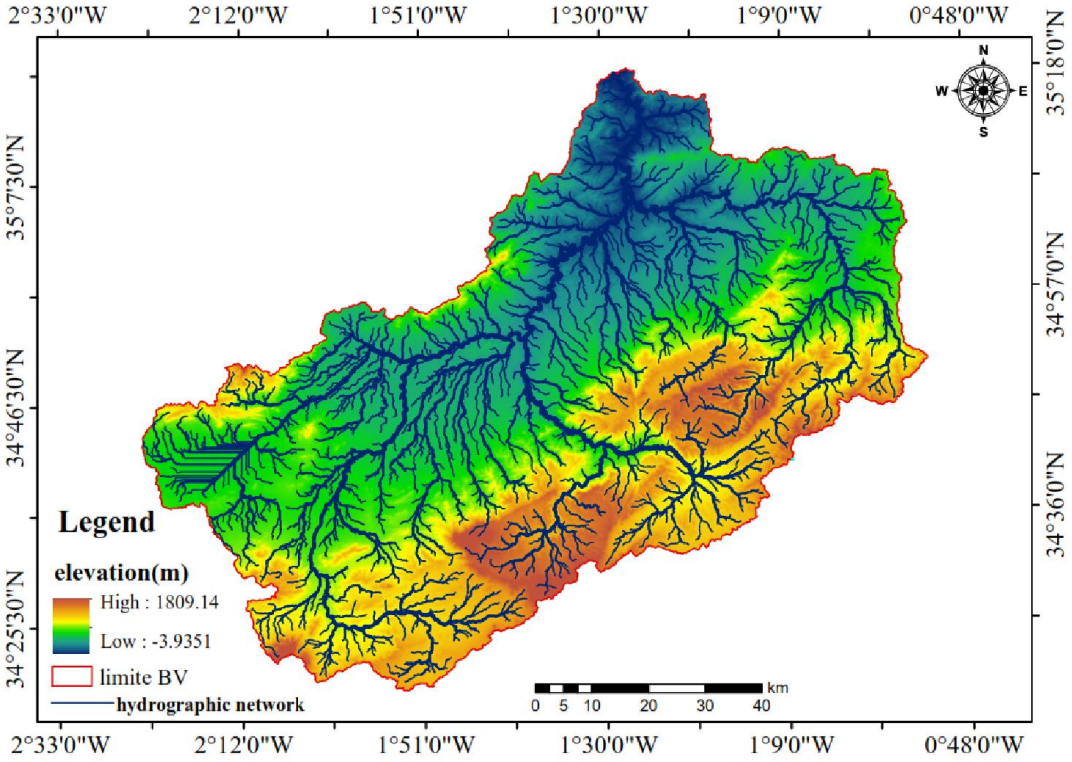


Figure II.2. Tafna watershed MNT

II.5. Altitude classes

In order to conduct a more comprehensive analysis of the altimetric characteristics of the Tafna watershed, the elevation values were classified into ten categories with an average interval of 200 meters. Furthermore, the surface areas corresponding to each class were extracted using GIS techniques, allowing for a more detailed interpretation of the spatial distribution of elevation across the watershed domain (Figure II.3).

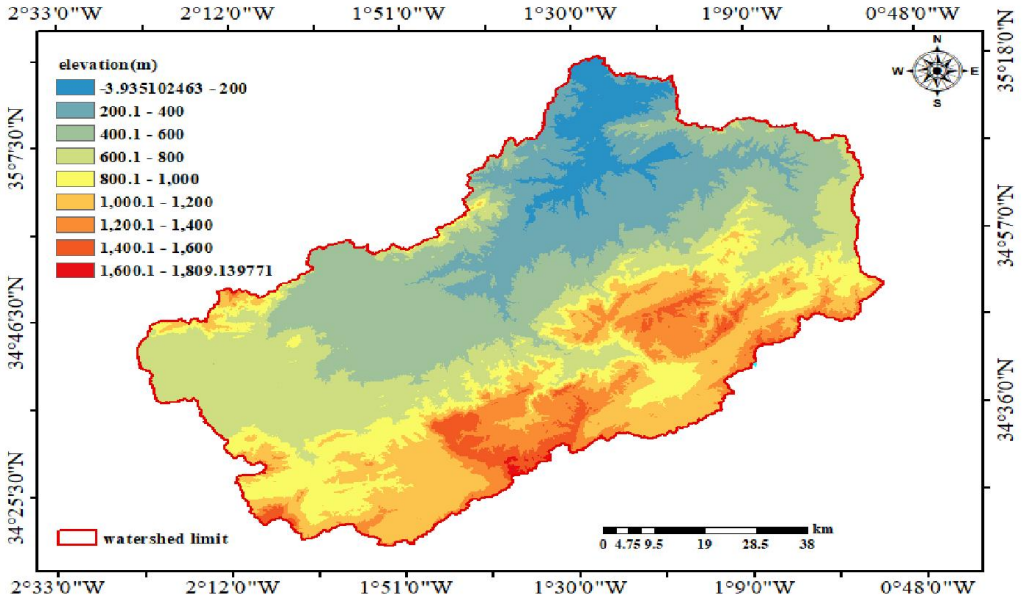


Figure II.3. Elevation Classes within the Tafna Watershed

(Figure II.3) reveals a distinct altitudinal distribution, with a clear dominance of elevation classes ranging from 200 to 1000 meters. These elevation ranges, represented in light blue,

green, and yellow tones, occupy approximately 60 to 70% of the total watershed area. They are primarily concentrated in the northern and central regions, extending along a northwest southeast axis.

The southeastern and southwestern regions of the watershed indicate areas of highest elevation (1400 -1809 m) represented by orange-red to deep red coloring, and they represent the extreme rugged and steepest regions in the watershed that may have higher continued runoff generation potential while having a risk for water erosion. This elevation class is about 15 -20% total functional area.

Conversely, the lowest elevation class (below -4m to 200m) is in northern and north east part of the watershed which were shown with dark blue coloring, which recognizes these areas represent the down stream zones or potential outlets for the water catchment areas that have a convergence or accumulation of water. The flat land forms and topography in this area would suggest that these areas will likely enter contribute critical seasonal habitat for retention and deposition processes.

II.6. Hypsometric Curve Interpretation

The hypsometric curve provides a synthetic view of the Tafna watershed's topographic structure, and it represents the distribution of area of the watershed. The area is depicted as a function of elevation. This curve is a reflection of the potential dynamic state of equilibrium of the basin, and it can be a very powerful means of comparison—not only of different watersheds, but also of different sections of the watershed itself, such as sub-basins.

In the Tafna watershed, the hypsometric curve aids in the assessment of average precipitation distribution, while also giving us a good understanding of the hydrologic and hydraulic behavior of the watershed and its drainage network, which is very useful for interpreting hydrological/ hydraulic behavior and basin variability through time. Furthermore, the curve can be used to estimate the morphological age of the watershed and the extent of erosion development, especially helpful in planning integrated erosion control programs.

The curve gives a general idea of the altimetric profile of the Tafna watershed and it is also an effective preliminary means of assessing the relief's susceptibility to various types of erosion (e.g. hydraulic or wind). As a visual and quantitative reading of the curve, it allows us to better specify the spatial dynamics involved in the evolution of the watershed.

Figure x8 plots the cumulative percentages of the area between successive altitude classes on the x-axis and mean elevations of these classes on the y-axis. The graphical view allows a fairly good reading of the altitudinal distribution of the Tafna watershed, for example, the curve shows that the area of the watershed found mainly in the mid elevation range (200-1000m) of the curve where the terrain is typified by moderate slopes and balanced runoff-infiltration processes. In contrast, the higher elevation classes—located primarily in the southern sectors of the watershed, which exhibit steep gradients—have the potential to generate rapid surface runoff, providing further erosive energy to the flood peaks associated

with significant storm events.

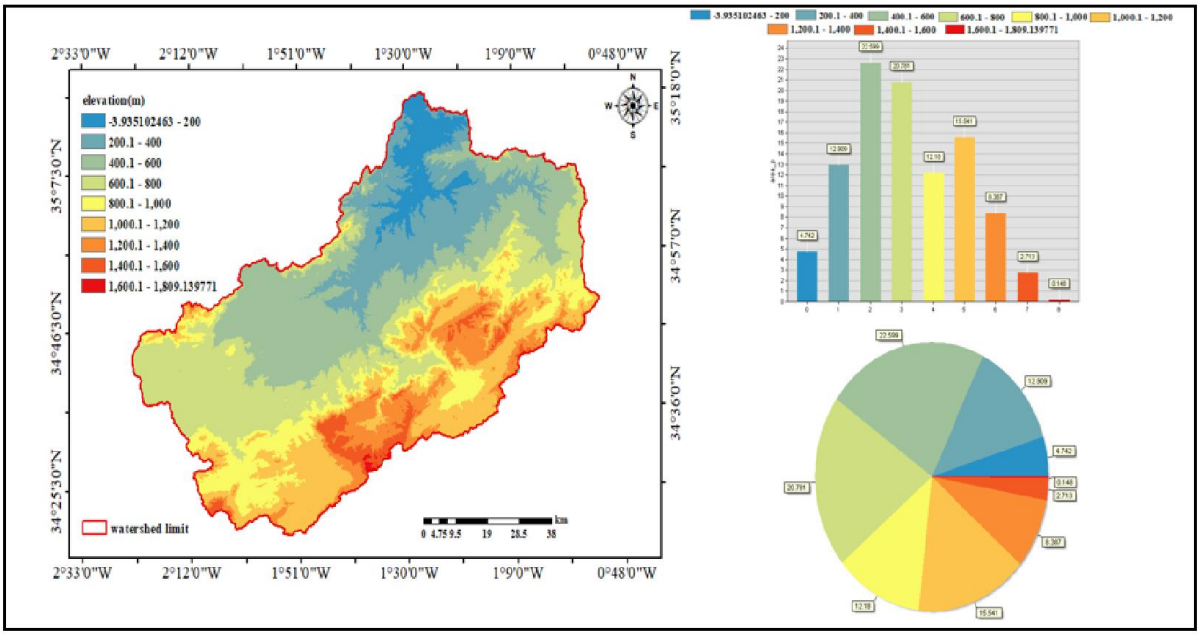


Figure II.4.Altitude classes of the Tafna watershed

The distribution of partial areas according to the watershed altitudes is shown in Table II.1.

Table II.1. Distribution of the partial surface area according to the altitude of the watershed

Elevation class (m)	Class centre	perimeter	Area	Area km	Area %	Cumulative area %
-3.935-200	98.032	535400.5	368281892.15	368.2819	4.742138	4.742138
200-400	300	1453508.83	1002501368.23	1002.501	12.90859	17.65073
400-600	500	1702751.03	1755059311.03	1755.059	22.59881	40.24954
600-800	700	1714245.59	1613847536.83	1613.848	20.78052	61.03006
800-1000	900	1919073.43	945917523.37	945.9175	12.18	73.21006
1000-1200	1100	2077818.52	1206976444.56	1206.976	15.54149	88.75155
1200-1400	1300	1499938.1	651364130.97	651.3641	8.387214	97.13876
1400-1600	1500	549339.17	210729882.7	210.7299	2.713439	99.8522
1600-1809	170.45	64499.53	11478612.79	11.47861	0.147803	100

II.7. Spatial Distribution Analysis of Slopes

The calculation of slopes and the analysis of the spatial distribution of gradients are fundamental for understanding the morphogenic potential of watersheds (Durand Delga, 1969). It is primarily the magnitude of slopes and their locations within the watershed that determine the erosive dynamics (Bouhali, 2016).

By deriving the previously used Digital Elevation Model (DEM) (Figure II.2) based on planimetric distances, it was possible to create a slope map for the boundaries of the Tafna watershed. The slopes, calculated using GIS, are expressed as percentages, yielding values ranging from 0% to 40%. The most dominant slopes, ranging from 0% to 5%, cover 45.408% of the total area, followed by 24.528% for the 5.1–10% class. The 10.1–15% slope class is the least dominant, representing 15.085% of the area. However, the average slope of the watershed is estimated at 15%.

To facilitate the interpretation of results and provide a clear representation of the spatial distribution of slopes within the studied watershed, it was prudent to group all values into defined classes (Figure II.5).

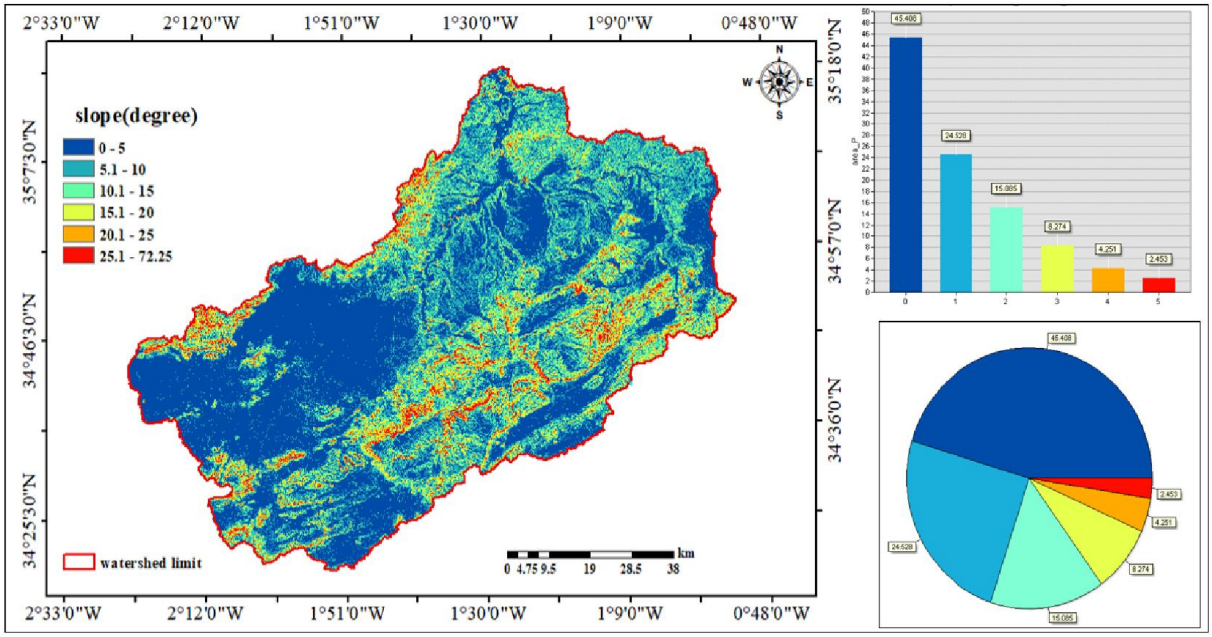


Figure II.5.slope classes and area % covered by every class in the Tafna watershed

Total percentages sum to 100%, which aligns perfectly. Now, scaling to 7,700 km²:

- 0–5°: 45.408% of 7,700 km² = 3,496.416 km²
- 5.1–10°: 24.528% of 7,700 km² = 1,888.656 km²
- 10.1–15°: 15.085% of 7,700 km² = 1,161.545 km²
- 15.1–20°: 8.274% of 7,700 km² = 637.098 km²
- 20.1–25°: 4.251% of 7,700 km² = 327.327 km²
- 25.1–72.25°: 2.453% of 7,700 km² = 188.881 km²

Total: 3,496.416 + 1,888.656 + 1,161.545 + 637.098 + 327.327 + 188.881 = 7,700 km²

II.8. Sediment Transport Implications

With this revised distribution, the Tafna watershed's sediment dynamics shift notably:

- 0–5° (3,496.416 km²): Avg. slope S≈0.087 rad. This 45.4% flat area is a major sediment sink, with low bed shear stress promoting deposition of suspended load (silt, clay). Over 3,496 km², expect significant aggradation in channels or floodplains during high flows.
- 5.1–10° (1,888.656 km²): Avg. S≈0.174 rad. This 24.5% zone transitions to bed load transport (sand, small gravel) during peak flows, acting as a key sediment conveyor.

- 10.1–15° (1,161.545 km²): Avg. S≈0.262 rad. This 15.1% area boosts erosion, potentially mobilizing coarser bed load and contributing to gully formation.

- 15.1–20° (637.098 km²): Avg. S≈0.349rad. At 8.3%, this steepens the game, increasing erosion risk and sediment supply.
- 20.1–25° (327.327 km²): Avg. S≈0.436 rad. This 4.3% zone is a significant erosion source, likely yielding gravel or cobbles.
- 25.1–72.25° (188.881 km²): Avg. S≈1.26rad. The 2.5% steepest areas are prime erosion hotspots, driving high-energy sediment transport during rains.

The increased proportion of steeper slopes (22.1% >10°) compared to earlier estimates suggests a more dynamic sediment system, with the 188.881 km² >25° zones as critical sources. Seasonal Mediterranean rains could flush substantial sediment from these areas, with the 3,496 km² flat zone trapping it downstream.

II.9. Overall Slope Index (Ig)

To calculate the Overall Slope Index (Ig) , we use the formula:

$$Ig = \frac{H_{95\%} - H_{5\%}}{Lb}$$

where:

- $H_{95\%}$is the elevation at which 95% of the watershed area lies below,
- $H_{5\%}$is the elevation at which 5% of the watershed area lies below,
- (Lb) is the length of the equivalent rectangle in kilometers,
- (Ig) is expressed in meters per kilometer (m/km).

$\boldsymbol{H_{5\%}}$**:** This is the elevation where the cumulative area reaches 5%. The cumulative area increases from 4.742138% at 200 m (end of the -3.935–200 m class) to 17.65073% at 400 m (end of the 200–400 m class). Since 5% lies between 4.742138% and 17.65073%, we interpolate within the 200–400 m class.

The area percentage of the 200–400 m class is 12.90859% (17.65073% - 4.742138%). Assuming a linear distribution within the class:

$$\text{Proportion within class} = \frac{5 - 4.743}{12.9085} \approx 0.01997$$

The elevation range is 400 m - 200 m = 200 m, so:

$$\boldsymbol{H_{5\%} = 200 + 0.01997 \times 200 \approx 204\ m}$$

$H_{95\%}$: This is the elevation where the cumulative area reaches 95%. The cumulative area increases from 88.75155% at 1200 m (end of the 1000–1200 m class) to 97.13876% at 1400 m (end of the 1200–1400 m class). Since 95% lies between 88.75155% and 97.13876%, we interpolate within the 1200–1400 m class.

The area percentage of the 1200–1400 m class is 8.387214% (97.13876% - 88.75155%):

$$\text{Proportion within class} = \frac{95 - 88.75}{8{,}387} \approx 0.745$$

The elevation range is 1400 m - 1200 m = 200 m, so:

$$H_{95\%} = 1200 + 0.745 \times 200 \approx 1349\ m$$

The elevation difference is:

$$D = H_{95\%} - H_{5\%} = 1349 - 204 = 1145\ m$$

The length of the equivalent rectangle (Lb)

$$Lb \approx 180\ km$$

(Ig):

$$Ig = \frac{D}{Lb} = \frac{1145}{180} \approx 6.36\ m/km$$

II.10. Soil types

Soil classes in Tafna Watershed is show in figure II.6.

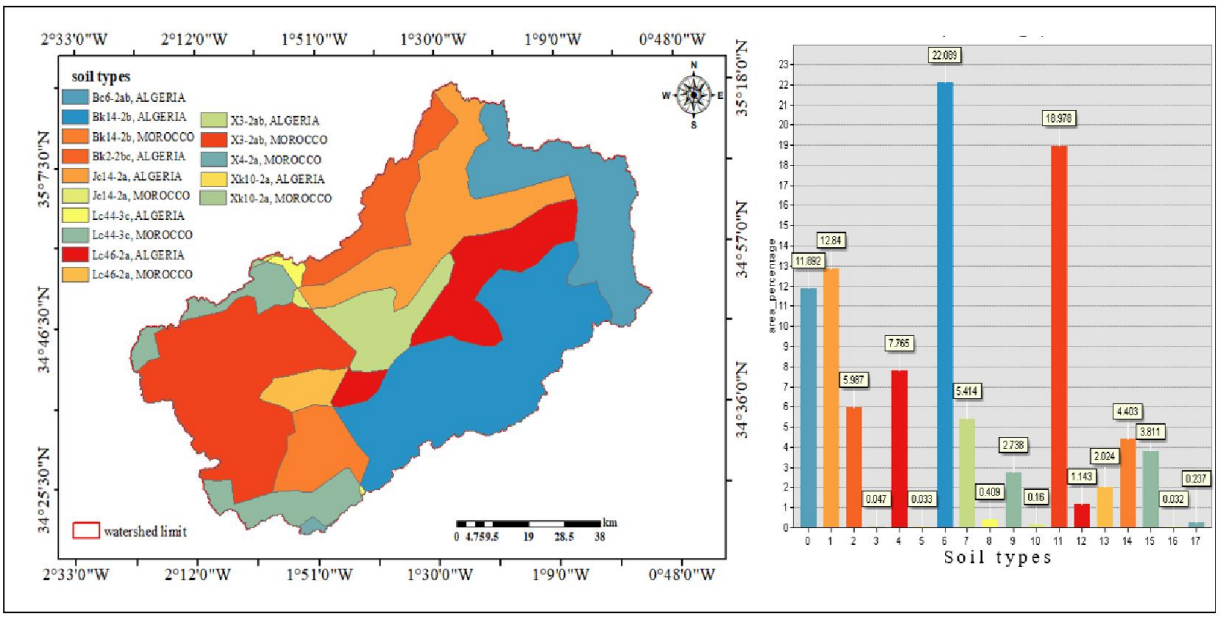


Figure II.6. Soil classes in Tafna Watershed

The Tafna watershed, located in northwest Algeria, exhibits a diverse range of soil types, as depicted in the provided soil type map and associated bar chart. These soil types are critical for understanding the watershed's hydrological behavior, erosion potential, and agricultural suitability. The map uses color-coded regions to represent different soil classifications, with the bar chart providing the percentage area coverage for each soil type. Below is a description of the soil types identified in the region, based on the legend and data.

II.11. Soil Types and Distribution

II.11.1. Be6-2b, Algeria

Percentage Area is 22.089%. This soil type occupies the largest area within the watershed. It is likely a well-drained, moderately fertile soil, possibly a calcic or cambic type, common in semi-arid Mediterranean climates. Its extensive coverage suggests it plays a significant role in the watershed's water retention and erosion dynamics.

II.11.2. Bk14-2b, Algeria

Percentage Area is 18.978%. Representing the second most common soil, this type may be a variant of calcic or gypsic soils with a moderate organic content. These soils are typically found in semi-arid regions and can support limited agriculture, though they may be prone to crusting and erosion under intense rainfall.

II.11.3. Bk14-2b, Morocco

Percentage Area is 13.842%. Similar to the Algerian Bk14-2b but with regional variations, this soil type likely reflects a calcic or gypsic profile influenced by Moroccan soil-forming processes. Its presence indicates cross-border consistency in soil characteristics, covering a notable portion of the watershed.

II.11.4. X3-2ab, Morocco

Percentage Area is 11.892%. This soil type could be a lithic or shallow soil, often found on steeper slopes or rocky terrains. Its moderate coverage suggests it contributes to runoff and sediment transport, particularly in the watershed's hilly areas.

II.11.5. X3-2ab, Algeria

Percentage Area is 7.785%. A variant of the Moroccan X3-2ab, this soil likely shares similar shallow or rocky characteristics but is adapted to Algerian conditions. It occupies a smaller but significant area, influencing local erosion patterns.

II.11.6. Jc14-2a, Algeria

Percentage Area is 5.414%. This soil type may be a vertisol or clay-rich soil, known for its shrink-swell properties. Its limited distribution suggests it occurs in specific low-lying or flat areas, impacting water infiltration and stability.

II.11.7. Lc44-3c, Algeria

Percentage Area is 4.403%. Likely a leptosol or shallow skeletal soil, this type is typical of rocky or mountainous regions. Its small coverage indicates it is confined to higher elevations or steep slopes, contributing to rapid runoff.

II.11.8. Lc46-2a, Algeria

Percentage Area is 3.811%. Another leptosol variant, this soil may have slightly deeper profiles than Lc44-3c but remains rocky and poorly developed. It adds to the watershed's erosion potential in steeper zones.

II.11.9. Lc46-2a, Morocco

Percentage Area is 2.738%. A Moroccan counterpart to the Algerian Lc46-2a, this soil type reinforces the presence of shallow, rocky soils along the watershed's boundaries, influencing sediment yield.

II.11.10. Xh10-2a, Algeria

Percentage Area is 2.024%. This soil could be a xerosol or aridic soil, characterized by low organic matter and high carbonate content. Its minor coverage suggests localized occurrence in dry, flat areas.

II.11.12. Xh10-2a, Morocco

Percentage Area is 1.143%. A variant of the Algerian Xh10-2a, this soil type is similarly aridic and occupies a small area, likely in the watershed's drier margins.

II.11.13. Jc14-2a, Morocco

Percentage Area is 0.409%. A minor occurrence of vertisol-like soil, this type mirrors the Algerian Jc14-2a but is less prevalent, suggesting localized clay-rich deposits.

II.11.14. Lc44-3c, Morocco (Light Green)

Percentage Area is 0.047%. A trace amount of leptosol, this soil type is negligible in area but indicates the presence of rocky outcrops in specific Moroccan sections.

II.11.15. Spatial Distribution

The map shows a heterogeneous distribution, with Be6-2b (Algeria) dominating the central and western parts, while Bk14-2b (Algeria and Morocco) and X3-2ab (Algeria and Morocco) are interspersed across the watershed.

Steeper, rocky soils (e.g., Lc44-3c, Lc46-2a) are concentrated in the northern and eastern edges, aligning with the higher slope classes (>10%) previously analyzed.

Flat to gently sloping areas (0–5%, 45.408% of the watershed) are likely dominated by Be6-2b and Bk14-2b, facilitating deposition, while steeper zones (>25%, 2.453%) host more erodible soils like Lc46-2a.

II.12. Conclusion

The Tafna watershed, nestled in northwestern Algeria, showcases a striking mix of geomorphological and hydrological traits that deeply affect water resource management and efforts to curb erosion. Its altimetric profile stretches from low-lying areas near sea level up to peaks soaring past 1800 meters, shaping a diverse range of slope dynamics and runoff patterns across the basin. The mid-elevation belt, spanning 200 to 1000 meters, forms the heart of the watershed, acting as a buffer where infiltration and surface flow strike a balance. Meanwhile, the rugged southern mountains stand out as key zones for runoff generation and erosion, driving significant sediment movement.

A hypsometric breakdown reveals the basin’s mature stage, shedding light on its ongoing erosional evolution and sediment transport processes. Steep slopes, covering more than 22% of the watershed, play an outsized role in sediment production, especially during the intense storms typical of the Mediterranean climate. This insight is backed by slope distribution data and gradient analyses derived from digital elevation models, which pinpoint certain sub-regions as particularly vulnerable to erosion triggered by water flow.

The variety in soil types further shapes how the watershed responds to water. Dominant soils like Be6-2b and Bk14-2b, found in flatter regions, help retain water and trap sediment, while

the shallow, rocky soils such as Lc44-3c and Lc46-2a—in steeper areas heighten the risk of erosion. This interplay between terrain, slope angles, and soil makeup highlights the intricate nature of sediment movement and water spread across the basin.

The uneven relief and hydrologic conditions make the Tafna watershed especially responsive to shifts in climate and land use. As a result, any solid management strategy needs to weave together this detailed morphometric and spatial knowledge to guide targeted actions. Priorities should include erosion control in the steep, delicate zones, sediment management in flat deposition areas, and land use planning that fits the natural landscape.

This chapter lays a solid groundwork on the Tafna watershed's physical makeup, providing essential insights for integrated water resource management (IWRM) moving forward.

CHAPTER III:

MAPPING AND MODELING OF SOIL WATER EROSION OF TAFNA WATERSHED

CHAPTER III: MAPPING AND MODELING OF SOIL WATER EROSION OF TAFNA WATERSHED

III.1. Introduction

Soil erosion by water is more than just a scientific concern it's a real and growing threat to the environment, agriculture, and water systems, especially in fragile regions like the Tafna watershed in northwestern Algeria. Here, the combination of intense seasonal rains, steep terrain, and often unsustainable land practices makes the land highly vulnerable. Erosion strips away fertile topsoil, clogs rivers and dams with sediment, and reduces the land's long term productivity (Mazour&Roose, 2002; Meddi&Toumi, 2015). To manage these risks effectively, we need tools that help us see where erosion is happening and how severe it is.

That's where the Revised Universal Soil Loss Equation, or RUSLE, comes in. This model, developed by Renard et al. (1991, 1997), lets us estimate how much soil is being lost over time by combining several key factors like rainfall, soil type, slope, vegetation, and land use practices. More importantly, RUSLE works well with modern geospatial technologies such as GIS and remote sensing (Parveen & Kumar, 2012; Kateb et al., 2020), which makes it especially valuable in large and complex landscapes.

In this chapter, we apply the RUSLE model to the Tafna watershed using a combination of ArcGIS and Google Earth Engine. We've pulled in multiple data sources from rainfall estimates using CHIRPS, to terrain data from digital elevation models, to vegetation indices derived from satellite images. This integrated approach allows us not only to map erosion risk across the basin, but also to identify priority areas where action is most urgently needed (Hamdan et al., 2021; Sahli et al., 2019). Our aim is to build a clear, spatial understanding of how soil erosion operates here and how it can be managed more sustainably.

III.2. RUSLE model

The RUSLE model represents an evolution of the original USLE, incorporating improvements in the estimation of topographic (LS), cover-management (C), and support practice (P) factors, as well as updated rainfall erosivity (R) and soil erodibility (K) data (Renard et al., 1991; McCool et al., 1987, 1989). Unlike its predecessor, RUSLE accounts for temporal variability and spatial heterogeneity by integrating GIS and remote sensing tools**,** allowing for detailed erosion risk mapping across large and diverse watersheds (Parveen & Kumar, 2012; Sahli et al., 2019).

Widely applied across the globe and particularly adapted for the Mediterranean and North African contexts (Mazour&Roose, 2002; Meddi&Toumi, 2015), RUSLE is a practical and flexible tool that supports land conservation planning, sediment yield analysis, and the design of soil erosion mitigation strategies in both data-rich and data-scarce environments (Kateb et al., 2020; Hamdan et al., 2021).

III.2.1.RUSLE Structure and Factors

The RUSLE model estimates soil loss using the equation:

$A = R \times K \times LS \times C \times P$

Where:

- A: predicted average annual soil loss (tons per hectare per year)

- R: rainfall-runoff erosivity factor, derived from rainfall intensity and volume (Renard, 1997)
- K: soil erodibility factor, which expresses the rate at which soil particles detach and move based on soil texture, structure, permeability, and organic content
- LS: slope length and steepness factor; Moore and Burch (1986), McCool et al. (1987, 1989) refined its calculation to better reflect real topography
- C: cover-management factor; represents the protective effects of vegetation and land use (Mazour and Roose, 2002; Roose, 1994)
- P: support practice factor; includes practices like strip cropping, contour farming, or terracing that reduce runoff velocity and soil transport

Each factor can be mapped using spatial datasets and incorporated into GIS to produce erosion risk maps.

III.2.2. Key Applications and Use Cases

- **Erosion Risk Zoning**
  Sahli et al. (2019) applied RUSLE in the Soummam watershed (Algeria) using GIS to map areas of high erosion potential. This allowed spatial prioritization for conservation measures.
- **Dam Siltation Assessment**
  Ranzi et al. (2012) used RUSLE to evaluate suspended sediment yield in Vietnam's Lo River. Their results highlighted the role of land use and upstream reservoir operation in sediment delivery.
- **Agricultural Planning**
  Parveen and Kumar (2012) used a combined RUSLE-GIS approach to assess soil loss risk in Jharkhand, India, helping identify areas where agricultural land needed immediate soil conservation intervention.
- **Watershed Management**
  Kateb et al. (2020) modeled sediment transport in the Beni Haroun dam catchment (Algeria) using RUSLE integrated with the SWAT model to analyze land use impacts under two scenarios.

III.2.3. RUSLE Strengths

- Empirical and straightforward
- Requires moderate input data
- Validated across many climates and landscapes
- Easily integrated with GIS and remote sensing tools
- Can be used for future scenario analysis when combined with land use or climate change models (Somura et al., 2009; Sahli et al., 2019)

III.2.4. Limitations and Solutions

Despite its utility, RUSLE does not simulate:

- Gully or rill erosion
- Sediment deposition or routing

- Event-based erosion dynamics

These limitations can be addressed by coupling RUSLE with:

Hydrological models such as HEC-HMS (used in runoff simulation studies like Ismael et al., 2017), machine learning methods (Hasanuzzaman et al., 2022; Jimeno-Sáez et al., 2022) to enhance prediction accuracy, physically-based models like SWAT, which simulate sediment transport over time and space (Neitsch et al., 2011).

III.2.5. Relevance in Current Research

With climate variability intensifying rainfall patterns and land use rapidly changing due to urbanization and agriculture, tools like RUSLE are critical. They help quantify the impact of these changes on soil erosion and sediment yield, informing long-term land management strategies.

III.3. Rustle model factors

III.3.1.Rainfall Erosivity (R-Factor)

Rainfall erosivity (R-factor) quantifies the impact of raindrop energy and intensity on soil detachment and runoff generation. It is expressed in MJ·mm/(ha·h·year), and plays a critical role in RUSLE by representing the climatic aggressiveness of a region (Wischmeier & Smith, 1978; Renard et al., 1997). In Mediterranean and semi-arid environments like the Tafna watershed, highly variable seasonal rainfall intensities can amplify soil loss risk (Meddi & Toumi, 2015; Jayawardena et al., 2017). The R-factor was derived from long-term CHIRPS precipitation datasets, interpolated and processed in a GIS environment. Spatial analysis of R is essential to identify erosion-prone zones and support hydrological modeling efforts such as HEC-HMS (Hamdan et al., 2021).

Figure III.1 quantifies the erosive power of rainfall over the Tafna watershed using CHIRPS daily precipitation data

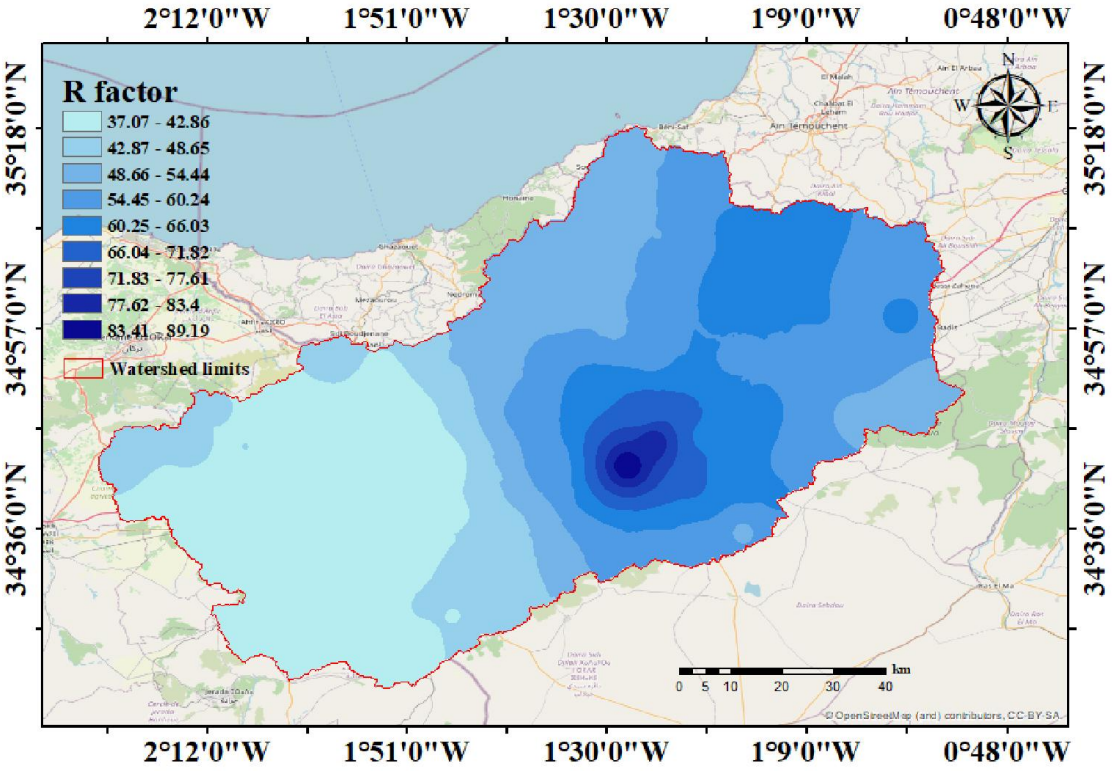


Figure III.1: Spatial Distribution of Rainfall Erosivity (R Factor) in the Tafna Watershed Based on CHIRPS Data.

The map shows the spatial distribution of rainfall erosivity (R-Factor) in the Tafna watershed, with values ranging from 37.07 to 89.19 MJ·mm/(ha·h·yr). The R-Factor is a critical input in RUSLE, as it reflects the kinetic energy of rainfall events, which controls their potential to detach soil particles (Renard et al., 1991; Wischmeier & Smith, 1978).

III.3.1.1. Spatial Distribution and Interpretation

Central-eastern region (dark blue): These areas display the highest R-factor values (>83 MJ·mm/(ha·h·yr)), indicating intense rainfall events and high erosive potential. This coincides with areas likely affected by convective storms or orographic rainfall (Meddi & Toumi, 2015).

Southwest and western parts (light blue): Here, values drop below 45 MJ·mm/(ha·h·yr), suggesting lower rainfall intensities and reduced erosivity. This aligns with more semi-arid zones and rain-shadowed regions in the Tafna basin.

This spatial variability in erosivity strongly correlates with topographic and climatic gradients, as noted by Mazour & Roose (2002) for NW Algeria.

III.3.1.2. Hydrological and Erosion Implications

High R-Factor zones are likely to contribute significantly to soil loss, especially where combined with steep slopes and degraded vegetation (as per Sahli et al., 2019; Ranzi et al., 2012).

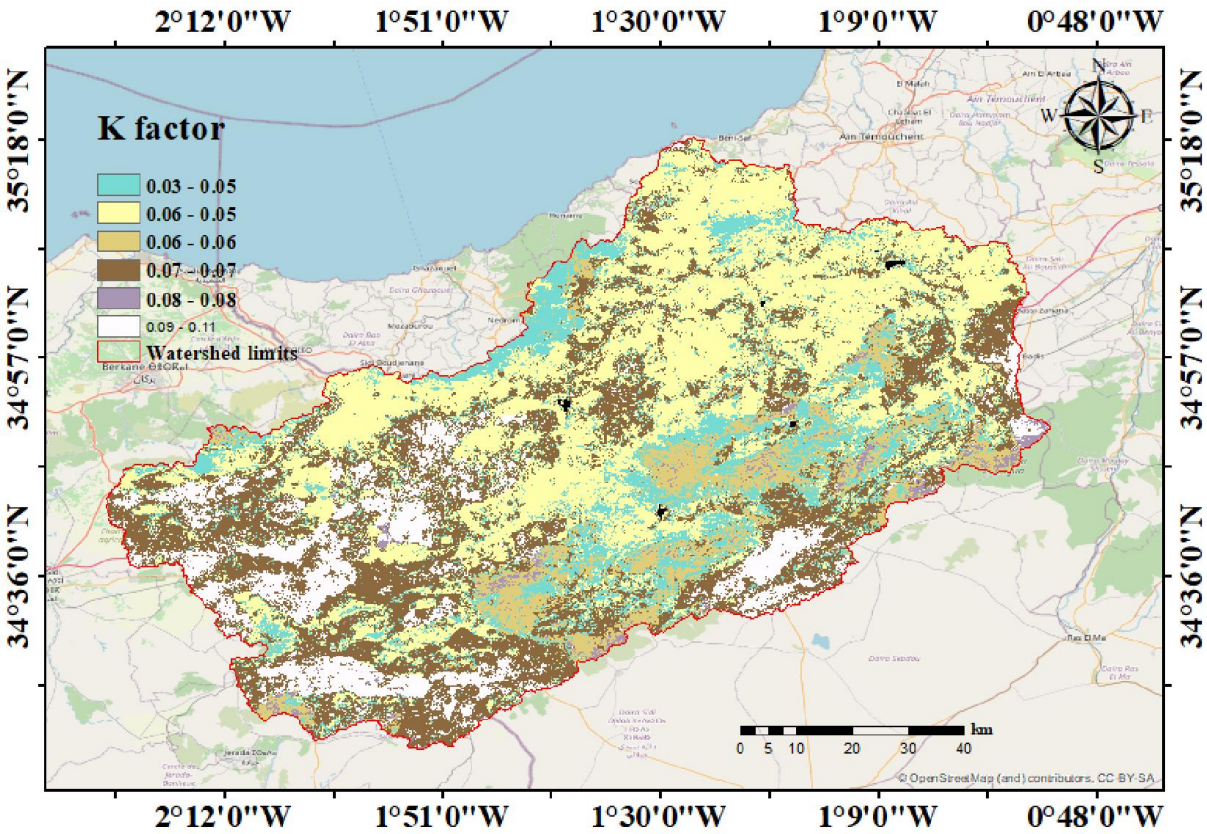
In modeling efforts, these zones would be the priority areas for conservation (e.g., reforestation, contour farming).

Integrating such spatial data into HEC-HMS or RUSLE in ArcGIS improves hydrological prediction accuracy (Hamdan et al. 2021; Ismael et al. 2017), particularly for runoff and sediment yield estimation.

III.3.2.Soil Erodibility (K-Factor)

The K-factor represents the susceptibility of soil particles to detachment and transport by rainfall and surface runoff. Influenced by soil texture, organic matter, structure, and permeability, it is expressed in t·ha·h/(ha·MJ·mm) (Renard et al., 1997; Wischmeier & Smith, 1978). In the Tafna watershed, K was computed using harmonized soil data and reflects the geopedological variability of the basin. Regions with fine silty textures or low organic content present higher erodibility (Toy et al., 2002; Mazour & Roose, 2002), while sandy or well-aggregated soils exhibit lower values. Accurate mapping of K is crucial for identifying soil types that are inherently vulnerable to erosion, as shown in regional studies (Kateb et al., 2020; Sahli et al., 2019).

FigureII.2. Represent the P-factor (support practice factor) in the RUSLE model quantifies the effectiveness of soil conservation measures in reducing the rate of water erosion.



FigureII.2. Soil Erodibility (K Factor) Map of the Tafna Watershed Derived from Pedological and Textural Data.

The map reveals spatial variability in soil erodibility, reflecting the influence of soil texture, structure, permeability, and organic matter content. In general:

Higher K values are observed in zones dominated by silt-loam or fine-textured soils, indicating a greater susceptibility to erosion under rainfall impact. These zones are particularly vulnerable if they coincide with steep slopes or intense land use.

Lower K values correspond to coarser soils (e.g., sandy or gravelly textures) or soils with better structure and organic content, which exhibit greater resistance to particle detachment and surface runoff.

This spatial distribution of the K factor highlights critical areas for soil conservation interventions, especially where high erodibility overlaps with erosive rainfall and steep topography.

Values range from 0 (maximum erosion control) to 1 (no control), with intermediate values indicating partial effectiveness of conservation techniques.

The P-factor map of the Tafna watershed is divided into seven classes, indicating varied conservation levels.

Table III.1. Classification of Soil Erodibility (K-Factor) in the Tafna Watershed

Class (P)	Conservation Level	Color	Spatial Observation
0	Ideal (maximum control)	Light Yellow	Rare, minimal patches
0.01–0.3	High control	Light Green	Found in isolated zones, mostly in central and southern areas
0.31–0.4	Moderate control	Green	Moderate presence in hilly regions
0.41–0.6	Low control	Dark Green	Dominant across the watershed, especially central and western parts
0.61–0.9	Very low control	Navy Blue	Found across mid-slopes and lower terrain
0.91–1	No control (critical)	Purple	Concentrated in urban edges and along bare slopes

These spatial patterns suggest:

- Better conservation efforts in southern and central hilly zones due to terracing or vegetative barriers.
- Poor conservation (P > 0.6) across flatter or heavily farmed areas, likely lacking in erosion control infrastructure.
- Urban and peri-urban zones show P = 1, meaning no conservation support at all matching*Kateb et al. (2020)* and *Sahli et al. (2019)* in similar Algerian basins.

- Low P values (<0.3) align with areas practicing contour plowing, vegetative strips, or retaining vegetation (Mazour & Roose, 2002).
- High P values (>0.6) may reflect:
    - Lack of soil conservation policies.
    - Expanding agriculture on steep slopes.
    - Urban development and road construction (*Marouf (2012)* and *Ouzerbane (2019))*.

III.3.3.Slope Length and Steepness (LS-Factor)

The LS-factor integrates the influence of topographyon erosion processes by accounting for both slope length (L) and slope steepness (S). It reflects the accumulation of runoff along a slope and the increasing energy with gradient (Moore & Burch, 1986; McCool et al., 1987, 1989). The factor was derived from a high-resolution Digital Elevation Model (DEM) using GIS-based hydrological tools. Within the Tafna watershed, steep slopes in mountainous sub-basins present elevated LS values, correlating with high sediment detachment potential. This terrain-driven influence on soil loss is particularly pronounced in poorly vegetated or deforested hill zones, in line with findings from Sahli et al. (2019) and Marouf & Remini (2011).

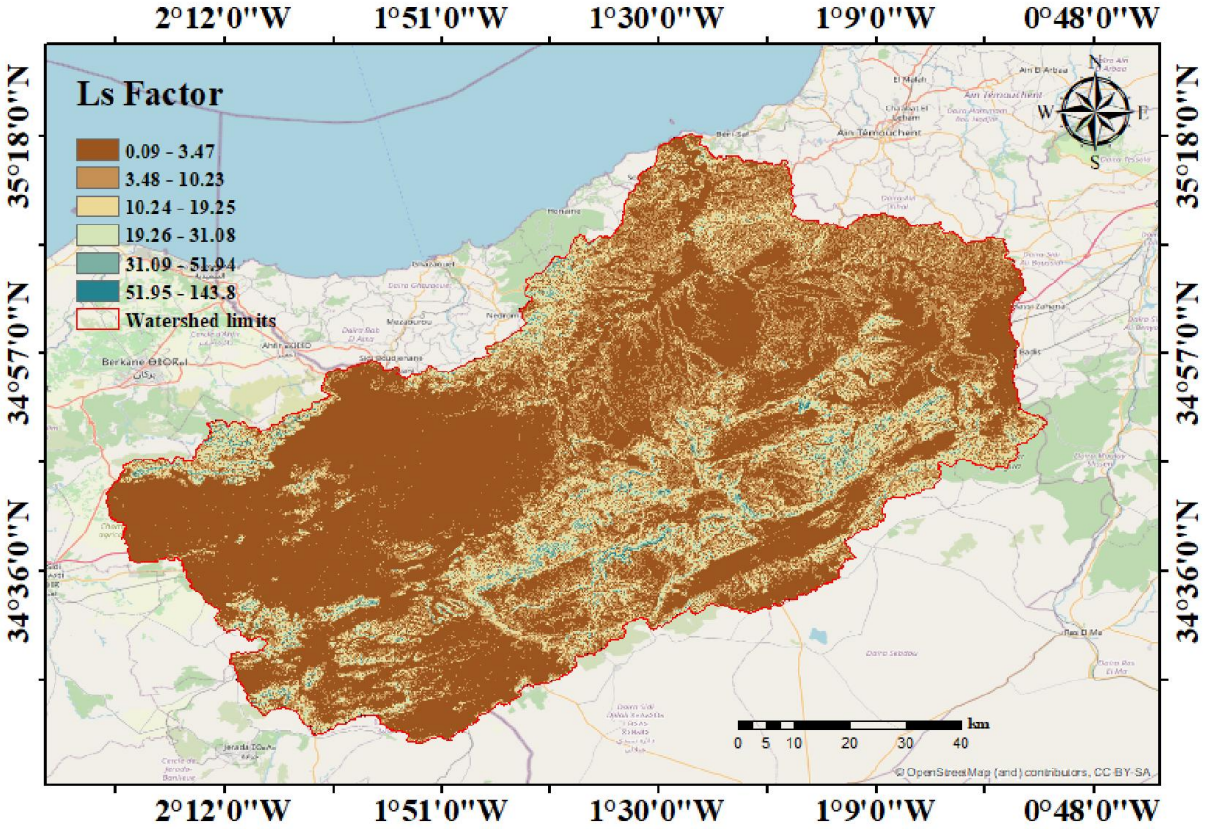


Figure III.3.1. Spatial Distribution of LS Factor in the Tafna Watershed Derived from DEM Analysis.

The LS-factor map captures the spatial variability of slope length and steepness within the Tafna watershed.It reflects topographic control on erosion potential as per the RUSLE model framework (Renard et al., 1997).

III.3.3.2. Topographic Patterns

- **Low LS values (0.09–3.47)**:
  - Dominant in flatter areas, especially in the northern and southeastern basin zones.
  - Indicates low slope gradients or short flow paths.

  - Representslowerosionsusceptibility.
- **Moderate LS values (3.48–19.25)**:
  - Covers the majority of the watershed.
  - Common in undulating terrains.
  - Suggests moderate overland flow concentration and slope.
- **High to very high LS values (≥31.09)**:
  - Concentrated along deeply incised valleys, drainage lines, escarpments, and converging slopes.
  - Especially evident in the south-central and west-central parts.
  - High erosion potential if not counteracted by vegetation or land management practices.

- These high-LS areas likely coincide with zones of higher runoff velocity (relevant for HEC-HMS flow routing).
- Strong LS values may contribute to concentrated sediment transport, as described in Ranzi et al. (2012) and Marouf & Remini (2011).
- Overlap with bare soil or agricultural land will increase RUSLE A-values dramatically (check against C and P layers).

III.3.3.3. Area Distribution by LS Classes

Estimate spatial impact using class ranges shown on the map:

Table III.2. Area Distribution by LS Factor Classes in the Tafna Watershed.

LS Class (Range)	Interpretation	Estimated Area Coverage (Visual)
0.09 – 3.47	Flat zones, low erosion risk	Moderate to high
3.48 – 10.23	Gentle slopes, stable unless bare	Very high
10.24 – 19.25	Rolling hills, moderate erosion potential	High
19.26 – 31.08	Steeper slopes, early rill formation risk	Medium
31.09 – 51.94	Strong slopes, concentrated flow paths	Low
51.95 – 143.8	Critical steep channels/erosion hotspots	Very low

- Majority of the watershed falls between 3.48–19.25.
- Less than 10% of the watershed has LS values above 31, but these zones contribute disproportionately to erosion and sediment production.

III.3.4. Cover Management (C-Factor)

The cover management factor (C-factor) represents the effect of vegetative cover and land **use** on soil erosion. It accounts for the protective role of crops, forests, and residues in reducing raindrop impact and runoff velocity (Renard et al., 1997; Wischmeier & Smith, 1978). C was estimated using NDVI-derived classifications from Landsat imagery (e.g.,

Jasinski, 1990; Sader & Winne, 1992), allowing for dynamic assessment of vegetation cover. Within the Tafna watershed, agricultural zones and degraded lands exhibit higher C values, indicating minimal protection, whereas forested and natural vegetative areas present lower C, signifying better soil conservation. This factor is vital in land-use planning and supports sustainable agricultural strategies (Mazour & Roose, 2002) (Marouf, 2012).

The figureIII.4 shows how land cover affects erosion potential in the Tafna watershed.

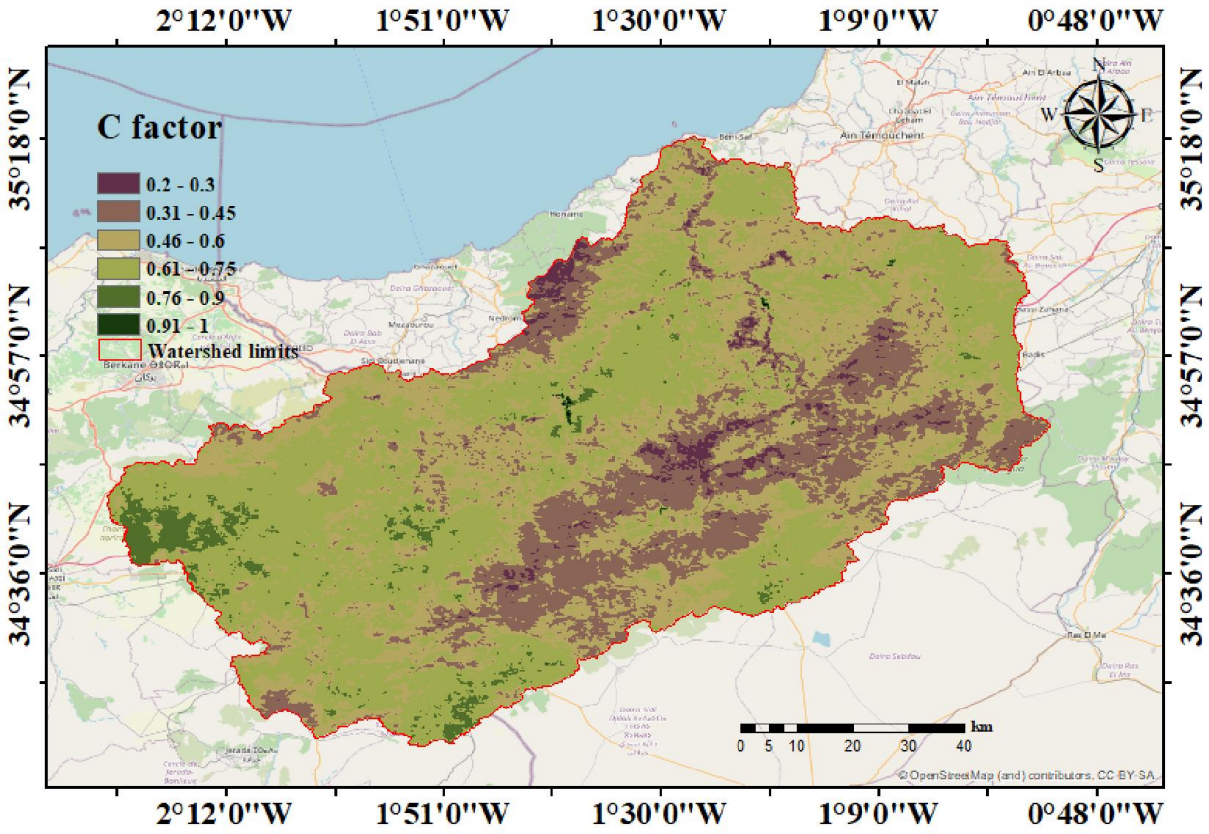


Figure III.4. Spatial Distribution of Cover Management (C Factor) in the Tafna Watershed Based on Land Use Classification.

The C-factor reflects the effect of land cover and land management on soil erosion, alues range from 0 (full protection) to 1 (no protection).Based on NDVI or land use classification (Jasinski, 1990; Sader & Winne, 1992).

III.3.4.1. Spatial Distribution of C-Factor Classes

Table III.3. Spatial Distribution of C-Factor Classes in the Tafna Watershed

C-Factor Class	Cover Type (Indicative)	Visual Dominance (Map)
0.2 – 0.3	Dense forest, wetlands	Sparse (smalldark green)
0.31 – 0.45	Shrubs, semi-natural land	Low
0.46 – 0.6	Mixed vegetation, light crops	Moderate
0.61 – 0.75	Sparsevegetation, dryland ag.	High
0.76 – 0.9	Baresoil, overgrazed land	Moderate to high

0.91 – 1	Urban zones, bare rock, dry ag	Moderate

III.3.4.2. Interpretation of Coverage

- Majority of watershed has medium to high C values (0.46–0.9).
  - Indicates vulnerable land use types (fallow, degraded rangelands).
- Very few areas with dense vegetation (C < 0.3) remain.
  - Found mainly in scattered forested highlands (north-central zones).
- High-C areas (>0.75) visible in:
  - Southernslopebelts
  - Central zonesnearerosion-pronecatchments

This spatial structure aligns with patterns in Sahli et al. (2019) for Soummam Basin and Marouf (2012) for Beni Haroun.

III.3.4.3. Interaction with LS Factor (Synergistic Risk Zones)

- Overlay zones of **C > 0.6** and **LS > 31** = critical erosion hotspots
- These areas contribute maximum values of RUSLE's **A = R × K × LS × C × P**
- They should be prioritized for soil conservation:
  - Afforestation
  - Stone lines
  - Agroforestry buffers

III.3.4.4. Spatial Contribution Summary

Table III.4 shows the spatial contribution of C-factor zones to erosion in the Tafna watershed.

Table III.4. Spatial Contribution of C-Factor Zones to Erosion in the Tafna Watershed

Zone Type	C Range	Estimated Area Share	Impact
Protected zones	0.2 – 0.45	Low	Minimal erosion contribution
Vulnerable agri zones	0.46 – 0.75	Moderate to high	Major source of sediment transport
Critical erosion zones	>0.75	Moderate	High erosion risk, runoff producer

III.3.5.Support Practice (P-Factor)

The P-factor accounts for conservation practices that reduce erosion by modifying the flow direction or velocity of surface runoff, such as contour farming, terracing, and strip cropping (Renard et al., 1997). It ranges from 0 (maximum protection) to 1 (no support practices). In data-scarce regions like the Tafna basin, P was estimated based on land use and slope class combinations, a method validated by Sahli et al. (2019) and Parveen & Kumar (2012). Areas with steep slopes and minimal conservation practices display high P values, indicating urgent need for intervention. Conversely, regions with vegetative buffers or engineered control

structures show lower values. P-factor mapping enhances decision-making for erosion control investments.

Figure III.5depicts spatial variation in conservation practices affecting soil erosion potential.

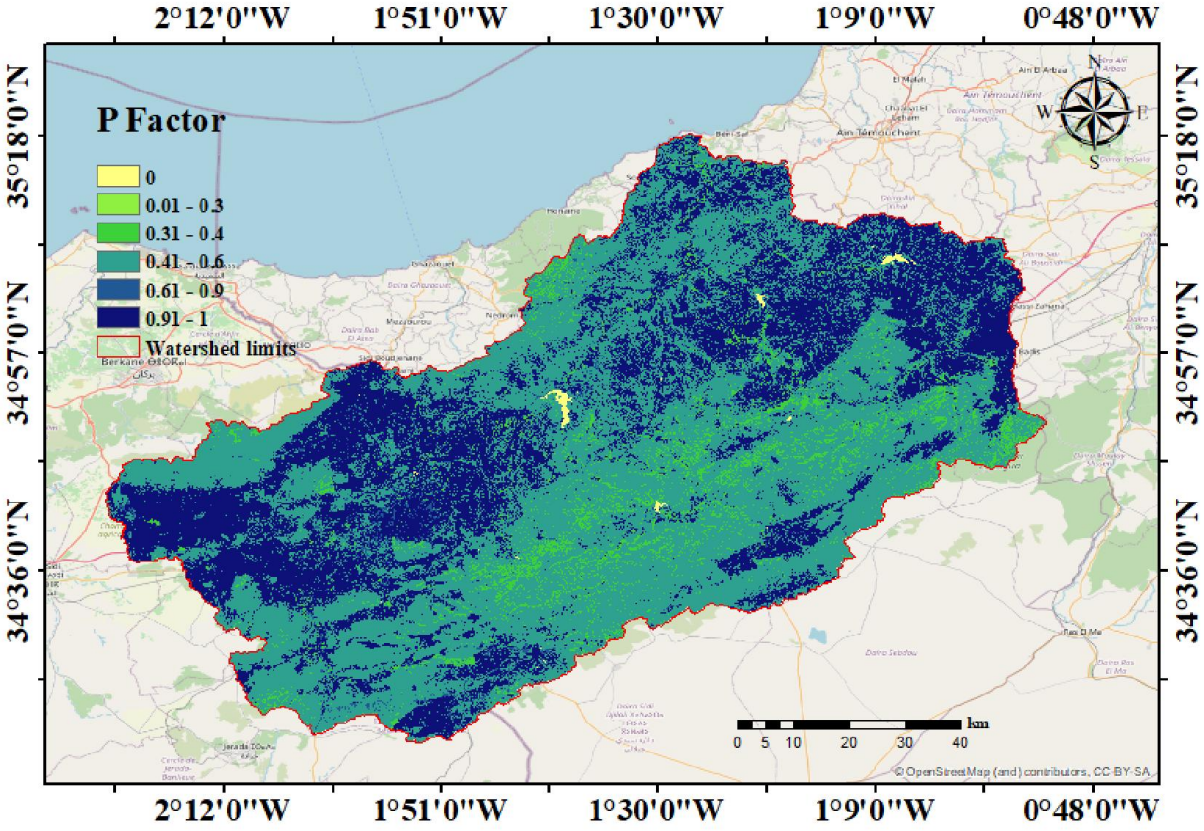


Figure III.5. Support Practice Factor (P-Factor) Map of the Tafna Watershed.

The **P-Factor** (support practice factor) reflects the degree to which soil conservation measures (like contouring, terracing, or strip cropping) are implemented to reduce runoff and prevent erosion. It is a dimensionless factor ranging from 0 (maximum protection) to 1 (no protection), and it plays a significant role in modifying predicted soil loss in the RUSLE model (Wischmeier & Smith, 1978; Renard et al., 1997).

The P-Factor values in the Tafna watershed show a wide spatial variability:

Table III.5. Classification of Support Practice (P-Factor) in the Tafna Watershed

P-Factor Class	Description	Color Range	Observations
0.00–0.30	Strong Conservation	Light Green	Found in central and western zones; indicates effective practices (e.g., contouring).
0.31–0.60	Moderate Conservation	Green to Cyan	Scattered throughout, possibly mixed land uses (fallow, rainfed farming).
0.61–0.85	Poor Conservation	Blue	Found on steep slopes and transitional zones.
0.86–1.00	No Conservation /	Red	Concentrated in the northwestern and

	Degraded		southeastern sub-basins.

This spatial pattern mirrors topographic and land use conditions. Regions with high P-values (red zones) likely correspond to steep slopes, low vegetation cover, and unmanaged agriculture, consistent with Sahli et al. (2019) and Mazour & Roose (2002).

III.3.5.1. Land Management and Conservation Implications

- **Zones with P > 0.85** are priority zones for intervention through terracing, reforestation, or controlled grazing.
- The combination of high P, R, and LS factors significantly increases the risk of erosion and sediment export.
- GIS-integrated P-Factor layers help identify where conservation farming techniques should be implemented (Parveen & Kumar, 2012).

P-Factor values were derived from land use/land cover classification and slope, based on the method by Sahli et al. (2019). Remote sensing and DEM processing in ArcGIS enabled a spatially distributed assignment of conservation levels.

III.3.6. Annual Soil Loss (A)

The final RUSLE output (A)represents the mean annual soil lossin tons per hectare per year (t/ha/year). It results from the combined effect of all factors: climate, soil, topography, land cover, and human practices. Mapping A spatially enables the identification of erosion hotspots, guiding prioritization of soil conservation interventions (Renard et al., 1997; Moriasi et al., 2007). In the Tafna watershed, annual soil loss was computed using GIS-based multiplication of all factor layers. Results reveal severe erosion risks in southern and central sub-watersheds, correlating with high LS, K, and P zones. These outputs serve as critical input for hydrological sediment yield modeling (e.g., HEC-HMS or SWAT) and watershed restoration planning (Hamdan et al., 2021; Sahli et al., 2019).

Figure III.6 shows spatial distribution of predicted soil erosion rates (t/ha/yr) based on RUSLE modeling.

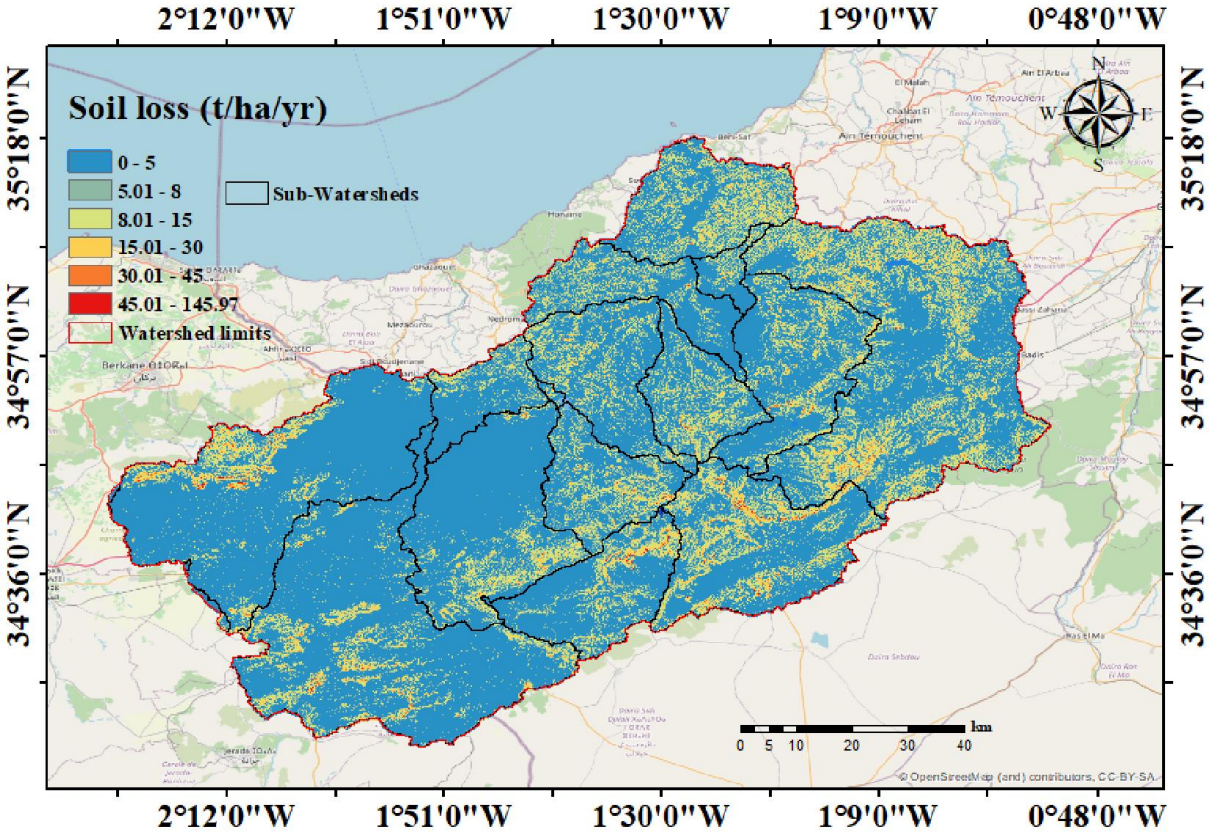


Figure III.6. Estimated Annual Soil Loss (A) Map Using RUSLE in the Tafna Watershed.

This map illustrates the average **annual soil loss (A)** in t/ha/year across the Tafna watershed, calculated using the RUSLE model. The values range from less than 5 to approximately 146 t/ha/year, offering a comprehensive view of erosion intensity influenced by rainfall erosivity, topography, soil erodibility, vegetation cover, and conservation practices (Renard et al., 1997).

- **Low erosion zones (0–5 t/ha/yr, blue)**: These dominate most of the basin, particularly in flat, forested, or well-managed agricultural areas. These regions represent stable environments with effective land cover and limited erosive force.
- **Moderate erosion zones (5–15 t/ha/yr, green to yellow)**: These occur mainly in the central sub-watersheds, likely corresponding to transitional zones with moderate slopes, mixed cropping, and patchy vegetation.
- **Severe erosion zones (15–45+ t/ha/yr, orange to red)**: These are concentrated in the southern, central, and some eastern sub-basins, indicating steep slopes, high R-values, poor vegetation cover, and limited soil conservation. These coincide with sediment-prone areas identified in past studies (Mokhtari, 2017; Marouf, 2012).

This spatial heterogeneity confirms the findings of Sahli et al. (2019) and Kateb et al. (2020) in similar Algerian watersheds where erosion correlates with terrain, rainfall, and land use.

III.3.6.1. Hydrological and Management Implications

- High erosion risk zones should be priority targets for intervention, particularly where they overlap with headwaters and agricultural lands.
- Conservation techniques such as reforestation, check dams, and slope stabilization should be concentrated in high-loss zones (Ranzi et al., 2012; Hamdan et al., 2021).
- This soil loss map can serve as a critical input for sediment yield modeling in HEC-HMS or SWAT for watershed-scale planning.

III.3.6.2. MethodologicalConsiderations

- The final erosion map was generated by multiplying all RUSLE factors (R, K, LS, C, and P), with factor layers derived from remote sensing, CHIRPS data, DEMs, and classified land use.
- GIS integration enables dynamic updates to the model by incorporating climate and land use changes (Parveen & Kumar, 2012).

III.3.7. Average Soil Loss Map in the Tafna Watershed

The average soil loss map offers a spatially aggregated view of erosion severity across the sub-watersheds within the Tafna basin. Rather than pixel-level detail, this approach summarizes erosion risk by hydrological unit, providing an effective way to prioritize conservation actions and manage sediment sources at a watershed scale. In the context of RUSLE (Revised Universal Soil Loss Equation), the soil loss value A (t/ha/year) represents the potential long-term average annual erosion, as influenced by the interaction of all contributing factors (R, K, LS, C, and P).

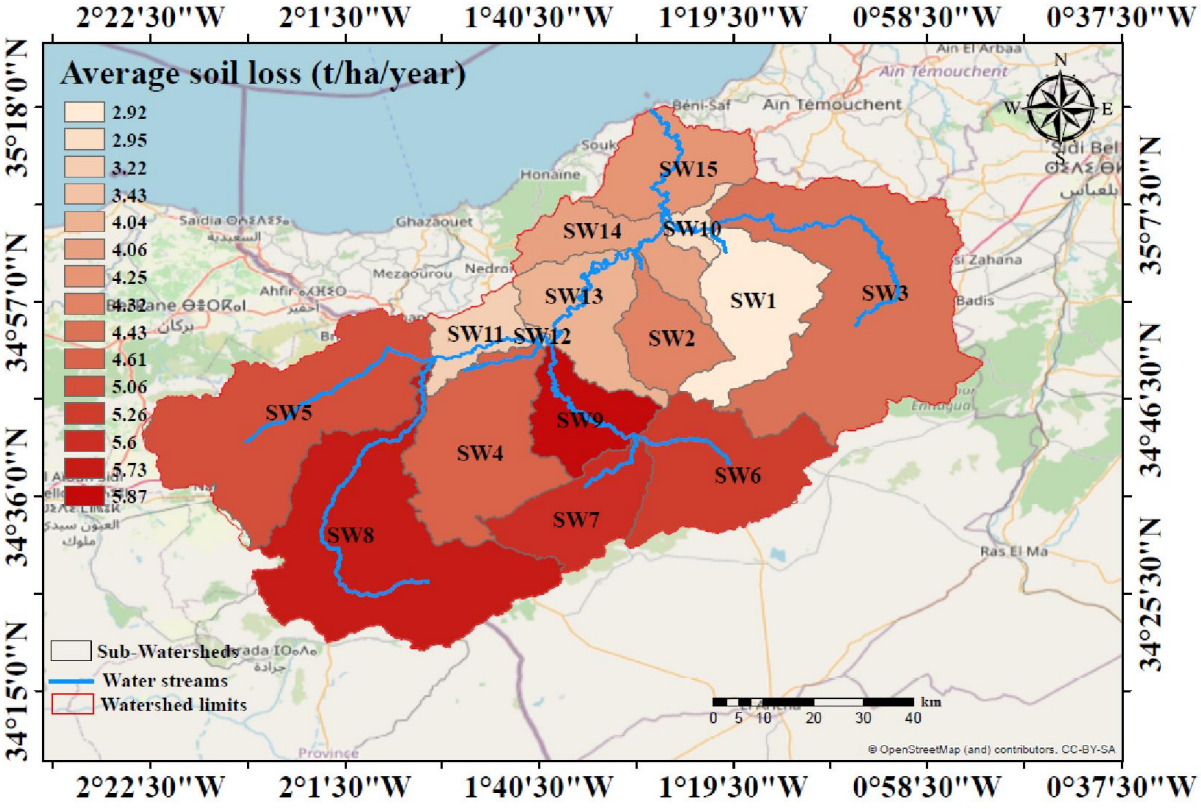


Figure III.7. Average Annual Soil Loss per Sub-Watershed in the Tafna Basin (t/ha/year)

The map represents the **mean annual soil loss (t/ha/year)** calculated at the **sub-watershed level** using the RUSLE model. Average values range from **2.92 to 5.87 t/ha/year**, with each sub-watershed delineated and color-coded according to its computed erosion rate.

This aggregated visualization provides a strategic overview of erosion severity across administrative or hydrological units, supporting watershed-scale decision-making (Renard et al., 1997; Sahli et al., 2019).

III.3.7.1. Spatial Distribution and Interpretation

The average Annual Soil Loss by Sub-Watershedis show in table III.6.

Table III.6. Average Annual Soil Loss by Sub-Watershed

Class Range (t/ha/year)	Color Code	Sub-WatershedBehavior	Interpretation
2.92 – 3.43	Light blue to light pink	Southern& western sub-watersheds	Lower soil loss due to flatter terrain, better vegetative cover or soil types.
3.44 – 4.61	Green to orange	Central sub-watersheds	Moderate erosion likely linked to mixed land use and moderate slopes.
5.06 – 5.87	Darker hues (e.g. purple, brown)	Eastern and northern sub-watersheds	High erosion pressure—likely caused by steep slopes, intense rainfall, and land use degradation.

These spatial patterns correlate strongly with the topography**,** climate, and land use heterogeneity across the watershed. According to Mazour&Roose (2002), erosion-prone zones often coincide with sedimentary hills and deforested lands in northwestern Algeria.

III.3.7.2. Hydrological and Management Implications

- Priority areas for intervention are those with average soil loss >5 t/ha/year, particularly in eastern and central-eastern sub-watersheds, where high LS and R factors overlap.
- These areas may be targeted for conservation measures, such as:
  - Check dams and terraces in steep areas (Parveen & Kumar, 2012),
  - Reforestation and controlled grazing zones (Sahli et al., 2019),
  - Promotion of contour farming and mulching.
- Sub-watershed-level results also support sediment routing and sediment yield analysis in hydrological models like HEC-HMS or SWAT (Hamdan et al., 2021).

The Classification and Analysis of Estimated Soil Loss Across Sub-Watersheds in the Tafna Basin is show in table III.7

Table III.7. Classification and Analysis of Estimated Soil Loss Across Sub-Watersheds in the Tafna Basin

Sub-Watershed	Average Soil Loss (t/ha/year)	Classification	Remarks
SW1	2.92	Very Low	Likely dominated by forest or low-slope terrain.
SW2	2.95	Very Low	Stable areas with minimal human disturbance.
SW3	3.22	Low	Moderatelysloped, possiblycultivated.
SW4	3.43	Low	Likely affected by mixed land uses and moderate LS factor.
SW5	4.04	Moderate	Steeper slopes or more degraded cover.
SW6	4.06	Moderate	Possibly deforested or cultivated hillsides.
SW7	4.25	Moderate	Area requiring monitoring, depending on rainfall intensity.
SW8	4.32	Moderate	Could be affected by upstream sediment transport.
SW9	4.43	Moderate	Could be affected by upstream sediment transport.
SW10	4.61	High	High erosion potential; overlaps with high R-factor zones.
SW11	5.06	High	Priority area for conservation measures.
SW12	5.26	High	Erosion may be driven by slope and rainfall erosivity.
SW13	5.6	Very High	Erosion hotspot — likely degraded soils and steep terrain.
SW14	5.73	Very High	Confirmed erosion-prone area, intervention urgently needed.
SW15	5.87	Very High	Peak erosion rate; coincides with critical sediment yield zones.

- These averages were extracted using zonal statistics tools in GIS, based on the pixel-level soil loss layer derived from RUSLE.
- Use of remote sensing (e.g., CHIRPS rainfall, land cover data) and platforms like ArcGIS and GEE helps automate and validate such basin-wide assessments (Kateb et al., 2020).

III.4. Conclusion

This chapter has shown how combining the RUSLE model with modern mapping tools like ArcGIS and Google Earth Engine can give us a powerful window into the soil erosion risks facing the Tafna watershed. By analyzing factors like rainfall intensity, slope steepness, soil

properties, vegetation cover, and land practices all using reliable spatial data we've been able to paint a detailed picture of where the land is most at risk of washing away.

The results aren't just academic. They reveal specific zones particularly in the southern and central parts of the basin where steep slopes, erodible soils, and limited vegetation combine to drive serious soil loss, sometimes exceeding 45 tons per hectare each year. These findings echo earlier studies (Kateb et al., 2020; Sahli et al., 2019) and underline the urgency of intervention in these areas. Whether through reforestation, terracing, or better agricultural practices, targeted action here could make a real difference.

We've also shown how erosion patterns can be averaged at the sub-watershed scale, which is useful for broader watershed planning and hydrological modeling, particularly with tools like HEC-HMS or SWAT (Hamdan et al., 2021; Ranzi et al., 2012). Perhaps most importantly, this approach is not only robust, but also flexible it can be updated as conditions change, whether through new land use trends or shifts in climate. In short, this kind of modeling helps turn complex data into clear, actionable insight something every land manager and policymaker needs when facing the real-world challenges of erosion and water resource management.

CHAPTER IV:

MODELING SUSPENDED SOLID TRANSPORT USING MACHINE LEARNING

CHAPTER IV:

MODELING SUSPENDED SOLID TRANSPORT USING MACHINE LEARNING

IV.1. Introduction:

Artificial neural networks (ANNs) are computational models inspired by the structure and function of biological neural systems. An ANN is composed of interconnected processing units (neurons), where each node performs a simple computation. The links between these nodes represent synapses, and each carries a weight that adjusts as learning progresses (Haykin, 1994).

The structure of an artificial neuron was derived from the basic functioning of a biological neuron. In biological systems, dendrites receive signals from neighboring neurons. These signals, whether excitatory or inhibitory, are integrated by the soma. When a threshold is reached, the neuron fires an action potential via the axon. This principle inspired early neuron models like the McCulloch–Pitts neuron, which introduced a binary model capable of implementing logic gates (McCulloch & Pitts, 1943; Haykin, 1994).

Despite the biological roots, the field of artificial intelligence rapidly shifted focus toward empirical performance. As a result, most artificial neurons became simplified approximations, optimized for computational efficiency rather than strict biological realism (Gupta et al., 2009). Feedforward structures and multilayer perceptrons (MLPs), trained through backpropagation algorithms, now form the foundation of many AI systems (Althoff & Rodrigues, 2021).

Neurons in the brain communicate on a millisecond scale, far slower than artificial circuits that operate in nanoseconds. However, the brain compensates with enormous parallelism, capable of performing an estimated 10^16 operations per second (Haykin, 1994; Ghorbani et al., 2021). Artificial neural networks, while faster in clock speed, remain limited in parallelism, but achieve strong results through model optimization and efficient learning algorithms.

The modern use of ANNs reflects a pragmatic evolution. Though early researchers aimed to mirror biological processes, the field today prioritizes models that generalize well, optimize efficiently, and scale effectively. The transition underscores a key philosophy in machine learning: practical utility outweighs biological fidelity (He et al., 2022; Ikram et al., 2022).

IV.2. History

The origins of deep learning can be traced to 1943, when Warren McCulloch and Walter Pitts proposed a simplified model of the human brain using logical calculus. They introduced the McCulloch–Pitts neuron, the first mathematical representation of a neural network, laying the theoretical foundation for artificial neural networks (McCulloch & Pitts, 1943).

In 1957, Frank Rosenblatt developed the perceptron, the first algorithm capable of learning through weight adjustments. He simulated it on an IBM 704 and later presented the custom-built Mark I Perceptron in 1960. While it generated significant attention, Rosenblatt's claims were considered overly optimistic, triggering skepticism in the early AI community (Rosenblatt, 1958).

Backpropagation, a core mechanism of modern deep learning, was first introduced conceptually in the 1960s by Henry J. Kelley (1960) and Stuart Dreyfus (1962), though it remained computationally inefficient. In 1986, Rumelhart, Hinton, and Williams made it viable by demonstrating its effectiveness for training multilayer perceptrons (Rumelhart, Hinton &Williams, 1986). This breakthrough allowed deep networks to learn useful distributed representations.

The field experienced cycles of intense enthusiasm and disappointment, commonly referred to as "AI winters." The second AI winter in the late 1980s was partly triggered by Minsky and Papert's (1969) critique of the perceptron's limitations, particularly its inability to solve nonlinear problems like XOR. Despite reduced funding and interest, some researchers continued to refine neural network methods.

A major turning point came with the rise of powerful hardware. From 1999 onwards, the development and use of graphics processing units (GPUs) enabled faster training of deep neural networks. Over a decade, GPUs improved training speeds by over 1000×, making it possible to train complex networks without layer-wise pretraining (He et al., 2022; Ikram et al., 2022).

Several milestones marked the modern deep learning era:

- **ImageNet (2009)**: Created by Fei-Fei Li and her team, this large-scale visual database with over 14 million labeled images enabled breakthroughs in computer vision (Deng et al., 2009).
- **AlexNet (2012)**: Krizhevsky, Sutskever, and Hinton trained this deep convolutional network using GPUs, winning the ImageNet competition and reducing error rates by a large margin (Krizhevsky et al., 2012).
- **GANs (2014)**: Ian Goodfellow introduced generative adversarial networks, where two networks (generator and discriminator) compete, greatly advancing generative modeling (Goodfellow et al., 2014).

These milestones reflect a shift from biologically inspired models toward performance-focused architectures. While artificial neural networks were inspired by biology, practical deep learning has evolved based on computational efficiency, mathematical optimization, and empirical success (Haykin, 1994; Ghorbani et al., 2021).

IV.3. Fundamental Components: Neurons, Layers, Weights, Biases, and Activation Functions

An artificial neural network (ANN) consists of **neurons** organized into **layers**. It typicallyincludes:

- An **input layer** that receives external data
- One or more **hidden layers** for intermediate computations
- An **output layer** that delivers final results

Neurons connect across layers through **links**, each with an associated **weight**. The weight reflects the influence of one neuron on another. These weights are trainable and are tuned

during learning (Haykin, 1994). Except for input neurons, most neurons also include a **bias** a constant input adjustable during training enabling the network to shift its activation threshold and improve flexibility (Gupta et al., 2009).

For eachneuron:

1. Compute the **weighted sum** of its inputs plus bias:
$$z = \sum w_i x_i + b$$
2. Apply a **nonlinear activation function**f(z) to produce the neuron's output

This process allows the network to respond only when sufficient weighted input is received.

IV.4. Activation Functions

Activation functions are critical as they introduce **nonlinearity**. Without them, the network behaves like a simple linear regression model, incapable of capturing complex relationships in data (Gupta et al., 2009). Common activation functionsinclude:

- **Sigmoid**: maps input to (0, 1); traditionally used in early MLPs
- **Tanh**: outputs between -1 and +1, zero-centered
- **ReLU (Rectified Linear Unit)**: activates only positive inputs; effective in mitigating the vanishing gradient problem (He et al., 2022)

These nonlinear functions empower networks to model complex hydrological patterns and non-linear sediment-runoff relationships.

IV.5. Learning vs Hyperparameters

- **Parameters** (weights and biases): learned during training via algorithms like backpropagation (Rumelhart et al., 1986)
- **Hyperparameters**: set before training, include:
    - Learning rate
    - Number of hiddenlayers
    - Batch size
    - Number of epochs

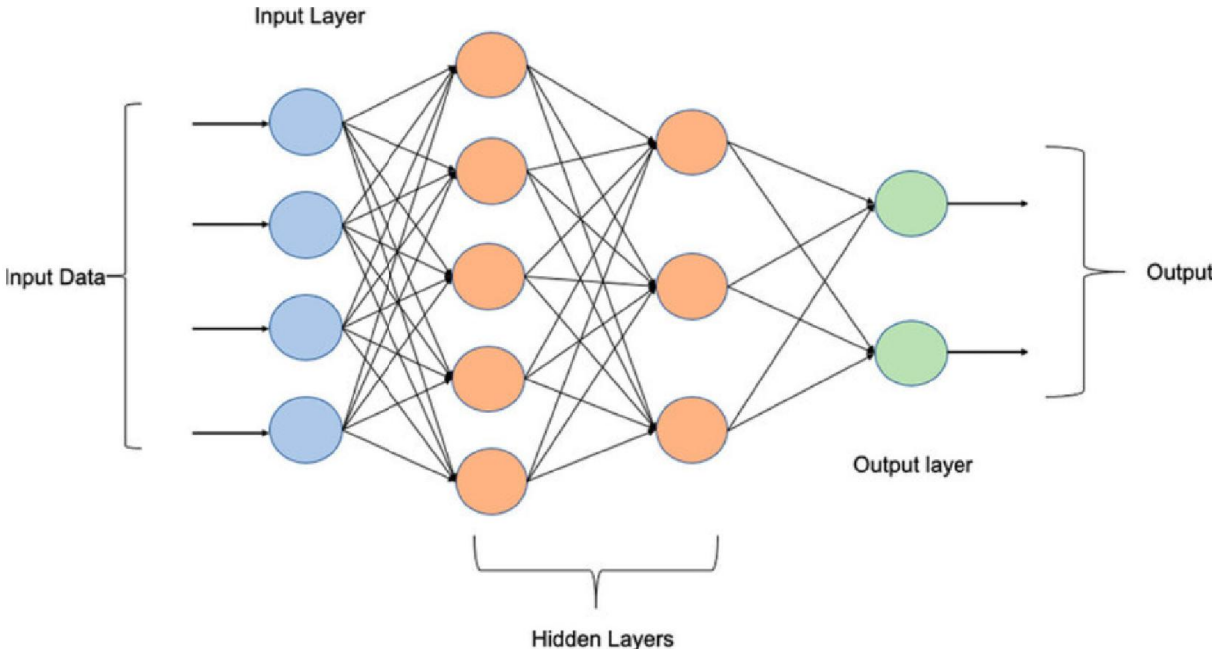
Tuning these hyperparameters is crucial. Limiting them hampers learning; excessive values may cause overfitting or inefficiency (Althoff& Rodrigues, 2021).

IV.6. Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) is a classic feedforward neural network essential to modern deep learning.

CHAPTER IV:

MODELING SUSPENDED SOLID TRANSPORT USING MACHINE LEARNING



FigureIV.2. Multi-layer perceptron (MLP-NN) basic Architecture(adapted from Haykin, 1998)

IV.6.1. Architecture: Input, Hidden, and Output Layers

An MLP consistsof:

- **Input Layer**: Receives raw data features. Each neuron corresponds to an input variable. Example:fertilitypredictionusing 9 input neurons (Haykin, 1994).
- **Hidden Layers**: One or more layers of neurons that process inputs using nonlinear activation, enabling the network to learn complex relationships (McCulloch & Pitts, 1943; Haykin, 1994).
- **Output Layer**: Produces the final result. The number of neurons aligns with the task (e.g., one neuron for regression). Outputs use functions like softmax for classification confidence (Gupta et al., 2009; Althoff& Rodrigues, 2021).

Neurons are fully connected forward across layers. Inputs flow from input to output without recursion—this structure defines feedforward connectivity.

The **Universal Approximation Theorem** states that an MLP with sufficient hidden units can approximate any continuous function to arbitrary accuracy (Haykin, 1994).

IV.6.2.OperationalPrinciples:Forward Propagation

Forward propagation executesdeterministiccalculations:

1. **Inputs Received**: A data vector is supplied to input neurons.
2. **Weighted Transmission**: Each value is multiplied by the weight of its connection to hidden-layer neurons.

3. **Summation with Bias**: Hidden and output neurons compute a weighted sum plus a bias term:
$$z = \sum w_i x_i + b$$
4. **Activation Function**: Applies a nonlinear function (e.g., sigmoid, tanh, ReLU) to produce neuron output (He et al., 2022).
5. **Layer to Layer**: Output of one layer becomes input to the next.
6. **Prediction Output**: Final values are produced. For classification, the output layer may use softmax to generate probabilities (Gupta et al., 2009).

Although deterministic, adding softmax provides meaningful confidence scores, critical for informed decision-making.

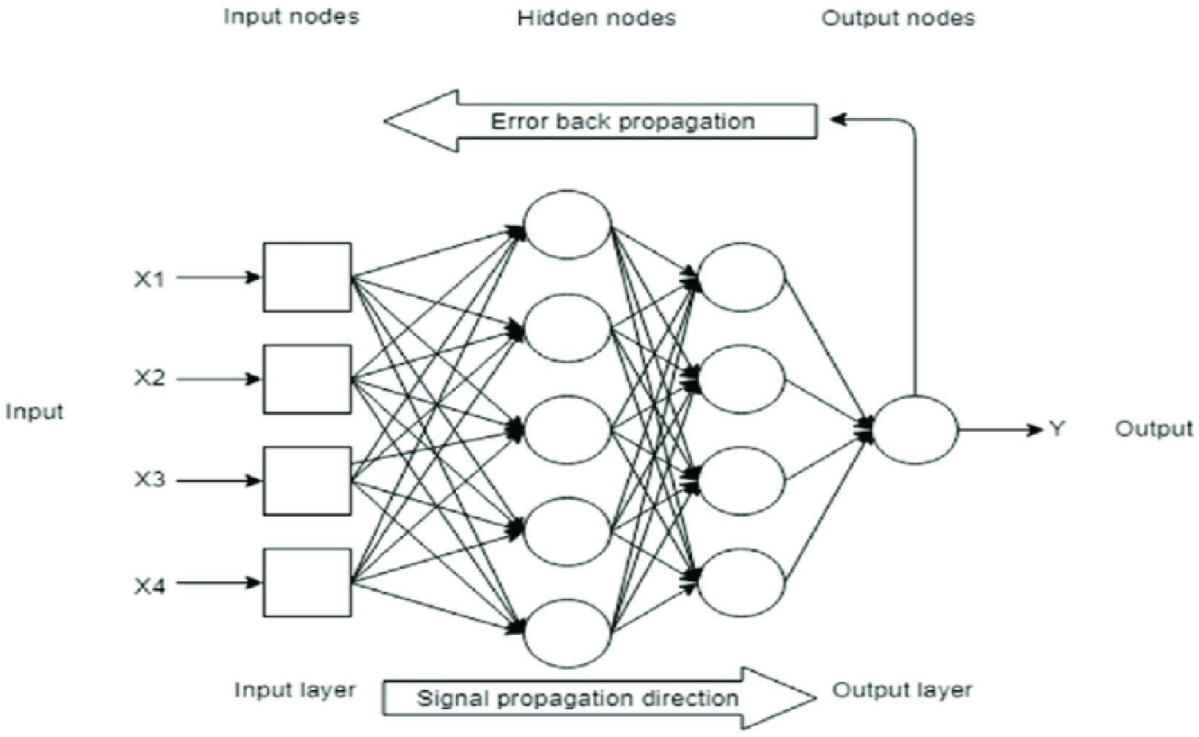
IV.6.3. Supervised Learning Paradigm

MLPs learn through **supervised learning**: each input is paired with a known output ("ground truth"). The goal is to approximatethis mapping.

Training involves:

- **Error Computation**: Comparing network output to the true value using a loss function (e.g., mean squared error, cross-entropy).
- **Backpropagation**: Errors propagate backward to update weights and biases via gradient descent (Rumelhart et al., 1986).
- **Iterative Optimization**: Through repeated epochs, the network refines its parameters to minimize error (Althoff& Rodrigues, 2021; He et al., 2022).

IV.7. Feed-Forward Backpropagation (FFBP)



FigureIV.1. Feed-forward back-propagation mechanism in artificial neural network

IV.7.1. Algorithm: Objective and Mathematical Foundation

Backpropagation (BP) computes gradients of a loss function with respect to network weights and biases. It applies the chain rule in reverse through the network layers, enabling efficient optimization without redundant calculations. Early concepts were introduced by Kelley (1960) and Dreyfus (1962), but the method became practical after Rumelhart, Hinton, and Williams (1986) demonstrated its ability to train deep, nonlinear networks. FFBP is essential for handling multilayer networks.

IV.7.2. Training Process: Four Iterative Steps

1. **ForwardPass**
    - Input vector passes through the network using initial (often random) weights and biases.
    - Each neuron computes a weighted sum plus bias, applies a nonlinear activation, and forwards output to the next layer.
2. **Loss Calculation**
    - Loss functions like Mean Squared Error (MSE) for regression or cross-entropy for classification quantify the error between network output and ground truth labels.
3. **BackwardPass (ErrorBackpropagation)**
    - Gradient of loss with respect to each parameter is computed via reverse-mode differentiation.
    - This step determines each weight's contribution to the total error.
4. **Weight Update (Gradient Descent)**
    - Use optimization methods like stochastic gradient descent or its variants to update weights.
    - Learning rate controls update magnitude; too large may overshoot minima, too small slows convergence.

Training with FFBP is an iterative optimization in a complex loss landscape, where hyperparameters such as learning rate and batch size critically influence convergence and generalization.

IV.7.3. Role of FFBP in Training Multilayer Perceptrons

FFBP is the standard algorithm for training MLPs. It enables them to learn complex nonlinear mappings by adjusting connection strengths to minimize error, surpassing what single-layer perceptrons can accomplish.

The combination of:

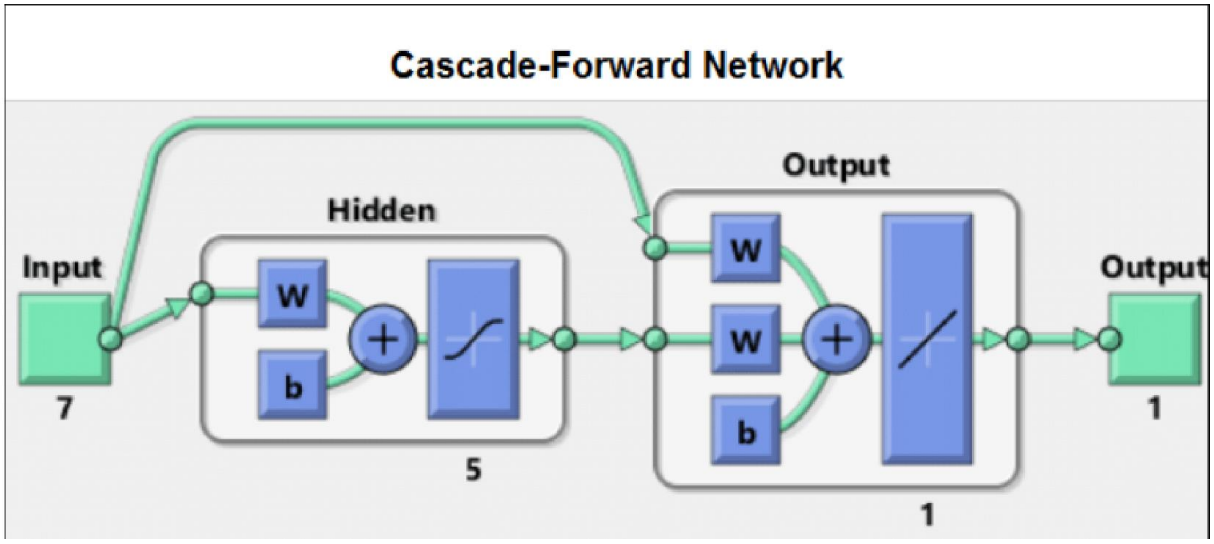
- **Flexible MLP architecture** (multiple layers and nonlinear activations)

- **Efficient gradient-based learning** via FFBP

forms the cornerstone of modern neural networks. FFBP unlocks the representational power of MLPs, enabling them to solve real-world tasks that demand nonlinear modeling capacity.

IV.8. Cascade Forward Backpropagation (CFBP)

Cascade forward backpropagation networks (CFBP) are a variant of multilayer perceptrons (MLPs) designed with enhanced inter-layer connectivity to improve convergence and learning efficiency.



FigureIV.3. Feed-Forward and Cascade-Forward Network Architectures (MATLAB Image)

IV.8.1.Architecture: Enhanced Connectivity and Direct Links

CFBP networks consist of input, hidden, and output layers like standard MLPs. However, they differ in that each layer is connected not only to the subsequent layer but also directly to all following layers, including the output. This includes:

- Direct connections from the input layer to each hidden layer
- Direct connections from the input layer to the output layer
- Standard forward connections between consecutive hidden layers

This architecture introduces more weights and biases than typical MLPs. The increased connectivity allows deeper layers to access raw input features directly. This can preserve important feature information and support more robust learning, especially when certain inputs carry high predictive value across the network (Haykin, 1994; He et al., 2022).

IV.8.2. Key Differences and Advantages Over Standard Feedforward Models

CFBP networks, when coupled with cascade correlation learning (Fahlman &Lebiere, 1990), exhibit several benefits:

- **Faster learning and real-time adaptation**: CFBP allows continuous online learning by providing direct access to inputs throughout the architecture, facilitating faster convergence (Ikram et al., 2022).
- **Improved convergence and gradient flow**: Shortcut connections help gradients reach deeper layers without vanishing, reducing training time and improving accuracy (Gupta et al., 2009).
- **Ability to escape local minima**: Cascade designs help the network avoid poor local solutions in the loss landscape, enhancing optimization robustness (Althoff& Rodrigues, 2021).
- **Support for VLSI implementation**: Cascade forward architectures have been used in real-time hardware designs, making them suitable for embedded systems (Ghorbani et al., 2021).

IV.8.3. Training Algorithms and Performance Characteristics

CFBP networks are commonly trained using variants of the backpropagation algorithm. Key algorithmsinclude:

- **Resilient Backpropagation (RBP)**: Offers faster convergence by adapting weight step sizes (He et al., 2022)
- **Levenberg–Marquardt (LM)**: Provides fast and stable optimization for moderate-sized networks (Haykin, 1994)
- **Bayesian Regularization (BR)**: Improves generalization, especially on noisy datasets

Metaheuristic approaches such as Genetic Algorithms (GA), Grey Wolf Optimizer (GWO), and Hunger Games Search (HGS) have also been applied to fine-tune hyperparameters and improve model robustness (Ikram et al., 2022).

When compared to MLPs and traditional feedforward networks, CFBP models often converge more quickly and can deliver competitive accuracy. However, performance still depends heavily on the nature of the data, the model's complexity, and the chosen optimizer (Gupta et al., 2009; Althoff& Rodrigues, 2021).

IV.9. Random Forest Models

Random Forest (RF) is a powerful ensemble learning method that leverages the collective intelligence of multiple decision trees to produce more robust and accurate predictions.

IV.9.1. Principles of Ensemble Learning and Decision Trees

RF belongs to the class of ensemble methods that aim to enhance predictive performance by combining several “weak learners” (typically decision trees). While individual decision trees are prone to overfitting and high variance, Random Forest mitigates this by aggregating the outputs of many deep trees trained on various bootstrapped samples of the dataset (He et al., 2022; Ikram et al., 2022).

This technique builds on the concept that an ensemble of weak models, when trained with diversity and aggregated properly, can outperform a single strong learner. It's an embodiment of the "wisdom of crowds" principle in machine learning (Haykin, 1994).

IV.9.2. Architecture and Operational Mechanisms: Bagging, Feature Randomness, Aggregation

The operational flow of a Random Forest model includes:

- **Bootstrap Aggregating (Bagging)**: Multiple subsets are created from the training data using sampling with replacement. Each subset is used to train an individual decision tree (He et al., 2022).
- **Random Feature Selection**: At each decision node, a random subset of features is chosen for splitting, ensuring decorrelation between trees (Ikram et al., 2022).
- **Tree Construction**: Trees are grown to maximum depth without pruning. Overfitting is prevented not by restricting tree complexity but by averaging across diverse trees.
- **Prediction Aggregation**: For classification tasks, majority voting is used; for regression, predictions are averaged.
- **Out-of-Bag Error Estimation**: Data not included in a bootstrap sample serves as a validation set for that tree, enabling unbiased error estimation without cross-validation (Ghorbani et al., 2021).

This dual-randomness—both in data sampling and feature selection—is the cornerstone of Random Forest's strength and generalization ability (Gupta et al., 2009).

TableIV.1. Comparative Analysis of Random Forests and Neural Networks

Criteria	**Random Forest**	**Neural Networks**
Data Type	Structured/tabular	Structured, images, text, audio
Training Complexity	Low (few hyperparameters)	High (architecture, learning rate, optimizer)
ComputationalNeeds	Moderate	High (especiallydeepmodels)
Interpretability	High (feature importance available)	Low ("black box")
Overfitting Risk	Low (due to ensemble)	Medium-High (needsregularization)

Neural networks are more flexible and powerful for complex data types like images or sequences, but Random Forests are often more efficient and interpretable for structured datasets (Althoff& Rodrigues, 2021). Neural networks require careful architecture design and hyperparameter tuning, while Random Forests tend to perform well with minimal tuning (Haykin, 1994; He et al., 2022).

In research contexts involving hydrology or sediment modeling, Random Forests have been successfully applied due to their ability to capture nonlinear relationships without complex training regimes, making them ideal for initial modeling trials or when computational resources are limited (Touaibia, 1999).

IV.10. Learning Paradigms in Neural Networks

Neural networks learn from data via several paradigms, each suited to differing data availability and problem types.

IV.10.1. Supervised Learning

Principle
Supervised learning uses labeled datasets each input has a correct output (ground truth). The model learns a mapping from inputs to outputs. Performance depends on the quality and quantity of labels, making data annotation a critical bottleneck (Althoff& Rodrigues, 2021).

Mechanism
Duringtraining:

- The model predicts outputs.
- It computes error between predictions and true labels using a loss function (e.g., MSE, cross-entropy).
- Backpropagation adjusts weights and biases to minimize this error (Rumelhart et al., 1986; Althoff& Rodrigues, 2021).

Common in classification (image recognition, spam detection, medical diagnosis) and regression (real estate pricing, financial forecasting).

IV.10.4. Semi-Supervised Learning

Principle
Semi-supervised learning (SSL) uses both labeled and unlabeled data. This is valuable when labels are scarce or expensive, but unlabeled data is abundant (Ikram et al., 2022).

Mechanism

- Train on labeledsamples.
- Use the model to generate pseudo-labels on unlabeled data.
- Retrain on combined datasets. Techniques include self-training, co-training, and graph-based SSL.

IV.11. Evaluation Metrics for Machine Learning Models

Evaluating machine learning models is essential to assess their performance and generalizability. Different metrics offer unique insights into model accuracy and reliability.

IV.11.2.Model Efficiency Metrics: Nash–Sutcliffe Efficiency (NSE)

IV.11.2.1. Nash–Sutcliffe Efficiency (NSE)

$$NSE = 1 - \frac{\sum_{t=1}^{n}(Q_{o,t} - Q_{m,t})^2}{\sum_{t=1}^{n}(Q_{o,t} - \overline{Q}_o)^2}$$

- $Q_{o,t}$: observed flow at time t

- $Q_{m,t}$: modeled flow at time t
- $\bar{Q}_o$: mean observed flow

- Range:]−∞,1]
- 1:perfect match
- 0: as good as the observed mean
- < 0: worse than using the observed mean as prediction (Althoff & Rodrigues, 2021)
- Sensitive to extreme values.

IV.12. Machine Learning with MATLAB

IV.12.1 Overview of MATLAB

MATLAB (Matrix Laboratory) is a high-level programming language designed for numerical computation, developed by The MathWorks. It enables users to perform matrix manipulations, visualize data through plots and graphs, and develop complex algorithms. MATLAB is one of the most widely used tools in engineering and applied sciences.

The software has a broad range of applications across various fields, including:

- Automotive systems,
- Health monitoring,
- Smart power grids,
- Machine learning,
- Robotics,
- Signal and image processing,
- Communication systems,
- Finance,

  and manyothers.

MATLAB can be used independently or in combination with specialized add-on packages known as toolboxes, which expand its functionality for specific domains such as advanced mathematics, statistics, and optimization.

Among the available toolboxes are:

- the Partial Differential Equation Toolbox,
- the CurveFitting Toolbox,
- and the Neural Network Toolbox, which will be discussed in more detail later in this work.

The figure IV.4 shows the main MATLAB interface

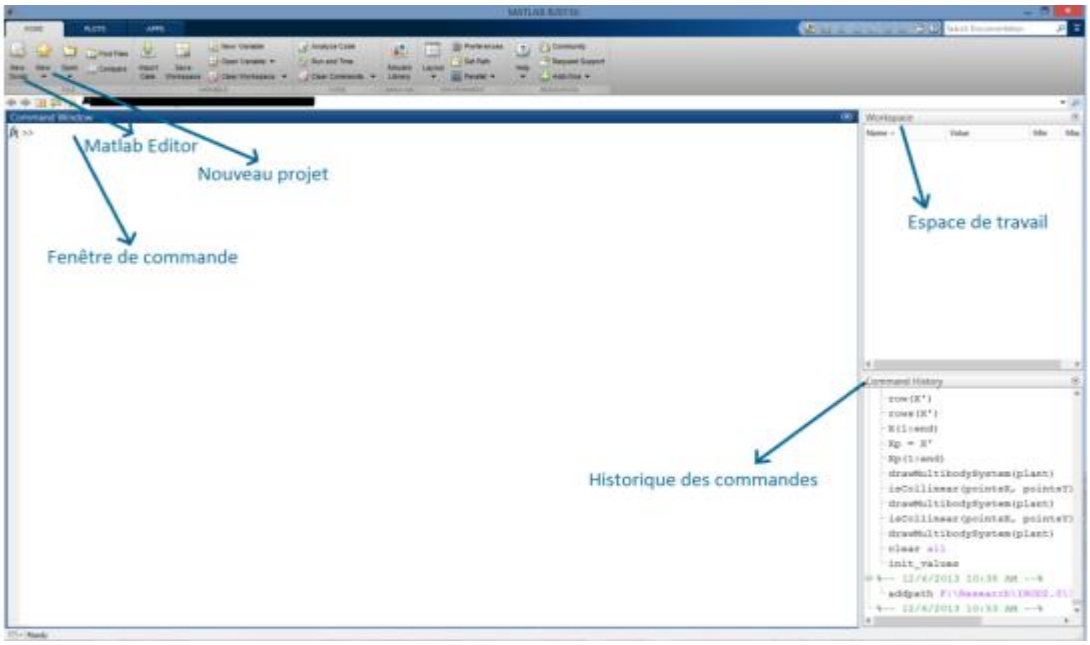


Figure IV.4. Main MATLAB interface (Demuth et Beale, 2002).

New Project (New): As its name suggests, this button is used to open a new *Command Window*, effectively starting a new project.
Command Window: This is where all program instructions are entered. Commands are written and executed one after another in this area, allowing real-time interaction with MATLAB.
MATLAB Editor: Unlike the Command Window, here the instructions are written all at once in the form of a script or function. Once completed, the entire block of code can be executed as a single program.
Workspace: This section displays all the variables and constants currently in use in the Command Window. It provides information such as variable names, formats, sizes, and types.
Command History: This area shows a log of all commands executed during the current and previous MATLAB sessions, providing a useful history for review or reuse.

IV.13. Results and Discussion

IV.13.1. Application of the MLPNN Model

In this study, the **Multi-Layer Perceptron Neural Network (MLPNN)** model was used to predict solid discharge based on input hydrological variables. The MLPNN is a type of feedforward artificial neural network that is well-suited for nonlinear regression problems. It consists of an input layer, one or more hidden layers, and an output layer, with each layer containing interconnected neurons. The model is trained using historical data to learn the underlying patterns and relationships between inputs and outputs, which are then used for prediction.

The dataset used in this modeling was collected from the **SIDI AISSA hydrometric station** (station code: 16 06 14). The data were divided into two phases for training and validation. In the first configuration, **70% of the data were used for training** the model, while the

remaining **30% were used for validation**. In a second configuration, **80% of the data were used for training** and **20% for validation**, allowing for a comparison of model performance under different training ratios.

Several simulation attempts were carried out using the Multi-Layer Perceptron NeuralNetwork (MLPNN) to predict solid transport. To achieve this, the number of hidden layersand thenumber of neurons per layer were varied, while keeping the sigmoid transfer function fixed in the hidden layers.

Among all tested configurations, the most relevant architectures, in terms of predictive performance, are the following:

- The **first configuration MLPNN1** is a **(1–9–1)** structure, meaning the input layer contains **1 neuron** representing the input parameter (liquid flow rate, *Ql*), followed by **1 hidden layer with 9 neurons**, and an output layer with **1 neuron** representing the predicted solid flow rate.
- The **second configuration MLPNN2** consists of a **(1–3–2–1)** architecture, with two hidden layers containing **3 and 2 neurons**, respectively.
- The **third configuration MLPNN3** retained is a deeper network: **(1–5–3–1–1)**, with three hidden layers containing **5**, **3**, and **1 neuron**, respectively, between the input and output layers.

IV.13.1.1.Model Configuration with 70% Training and 30% Validation Data

IV.13.1.1.1. MLPNN1 model

The architecture of the MLPNN1 model is shown in figure IV.5

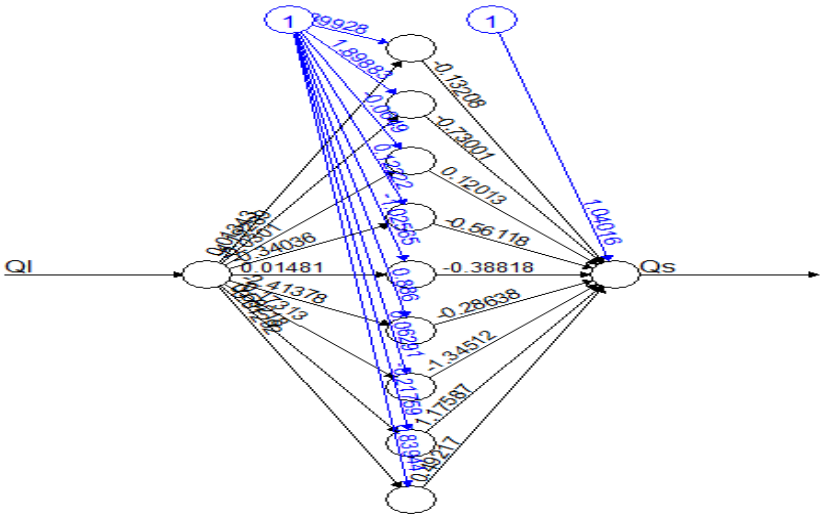


Figure IV.5. Architecture of the MLPNN1 model (70%, 30%)

The regression lines obtained are shown in Figures IV.6.

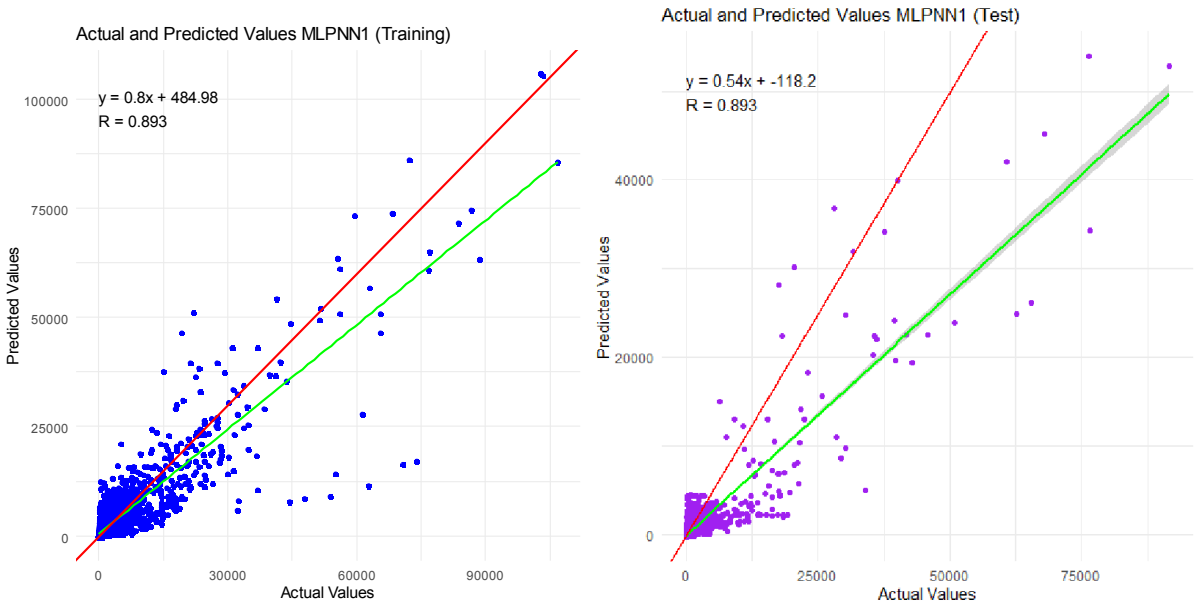


Figure IV.6. Regression Lines of predicted and Measured Solid Flow Rate (kg/s) Using MLPNN1 (70%, 30%)

The Predicted and observedSolid Flow Rate using MLPNN1 is show in figure IV.7.

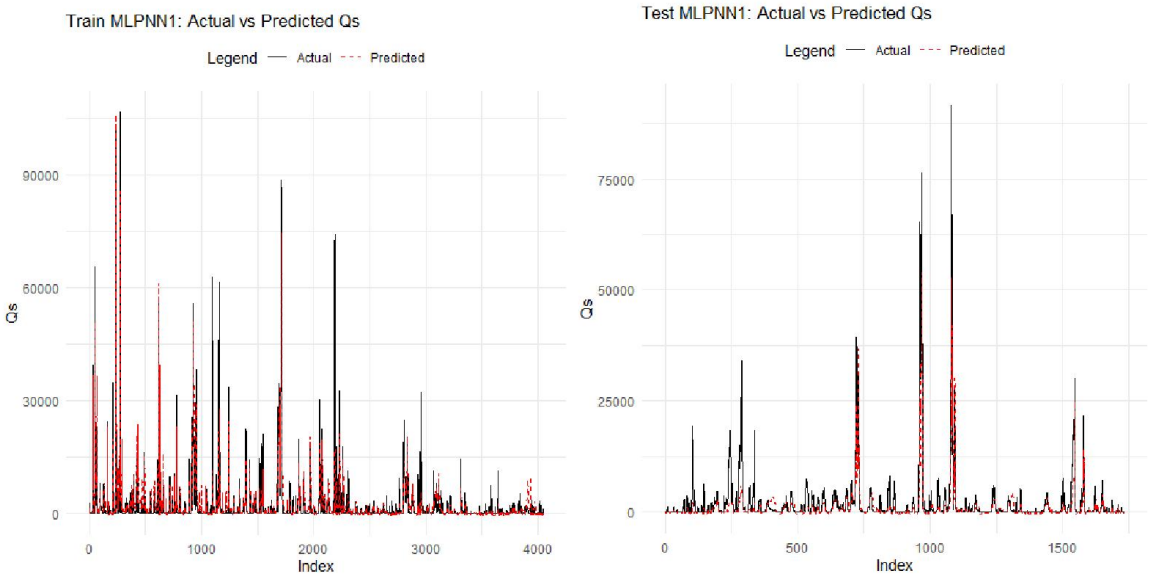


Figure IV.7. Predicted and observedSolid Flow Rate using MLPNN1 (70%, 30%)

The performance parameters are shown in the table IV.2.

Table IV.2. Performance parameters of the MLPNN1 model (70%, 30%)

Model	Hidden Layers	Training	Test

MLPNN	Input	Output		R	NSE	R	NSE
MLPNN1	Ql	Qs	1(9)	0.89	0.79	0.89	0.68

The results shown in table IV.2, figure IV.6 and figure IV.7,presents the performance of the **MLPNN1 model** used to predict **solid flow rate (Qs)** based on a single input: the **liquid flow rate (Ql)**. The neural network architecture consists of **one hidden layer with 9 neurons**, denoted as **1(9)**. The model's performance was evaluated using two statistical indicators: the **correlation coefficient (R)** and the **Nash-Sutcliffe efficiency (NSE)**, for both the training and test (validation) phases.

During the **training phase**, the model achieved an **R value of 0.89**, indicating a strong linear correlation between the **predicted** and **measured solid flow rates**. The **NSE value of 0.79** confirms that the model has a high predictive capability and is able to reproduce the observed data with a good level of accuracy.

In the **test phase**, the model maintained a similar correlation coefficient (**R = 0.89**), which shows that the model generalizes well and retains its prediction performance even on unseen data. However, the **NSE decreased to 0.68**, which is still considered satisfactory but slightly lower than in the training phase. This drop may be attributed to the complexity of the physical process or the limited variability in the test data.

Overall, the results demonstrate that the **MLPNN1 architecture**, despite its relatively simple structure, is capable of capturing the nonlinear relationship between liquid and solid discharge. The consistency of the correlation values across both phases reflects the model's robustness, while the NSE values confirm its efficiency in simulating solid transport behavior based on hydrological input.

IV.13.1.1.2. MLPNN2 model

The architecture of the MLPNN2 model is shown in figure IV.8

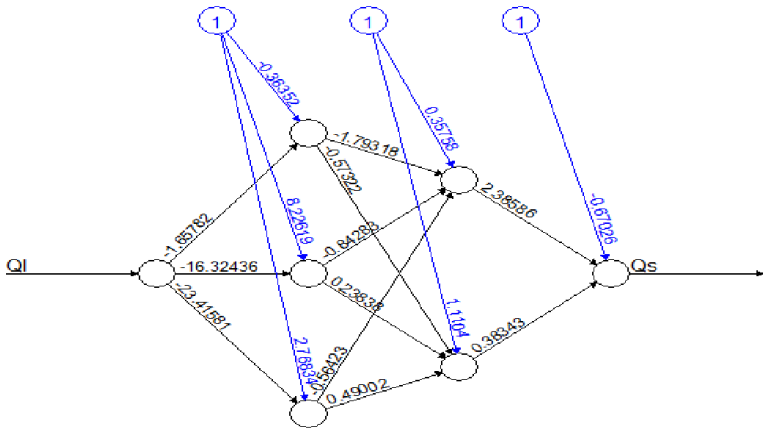


Figure IV.8. Architecture of the MLPNN2 model (70%, 30%)

The regression lines obtained are shown in Figures IV.9.

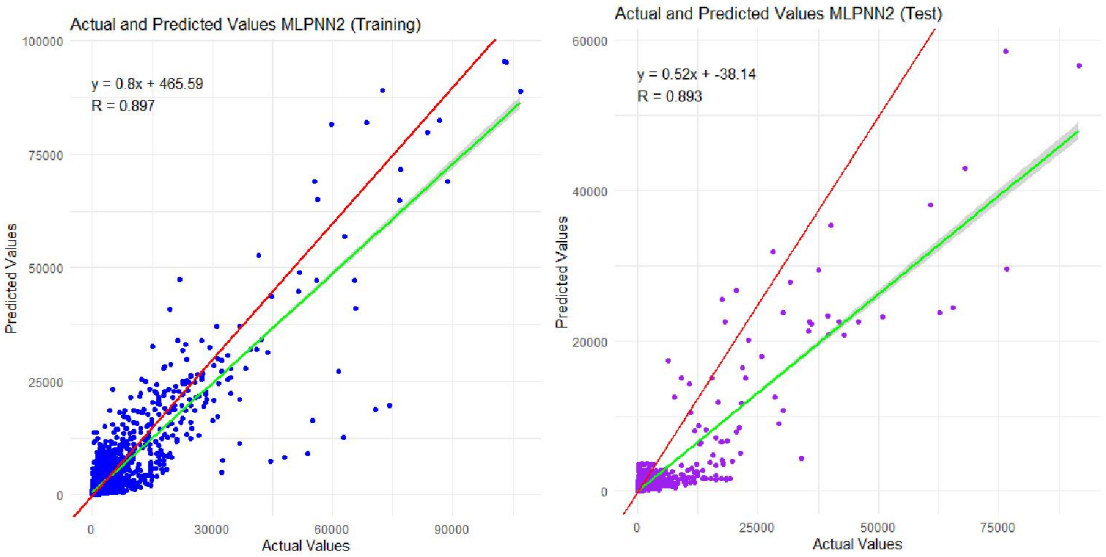


Figure IV.9. Regression Lines of predicted and Measured Solid Flow Rate (kg/s) Using MLPNN2 (70%, 30%)

The Predicted and observedSolid Flow Rate using MLPNN2 is show in figure IV.10.

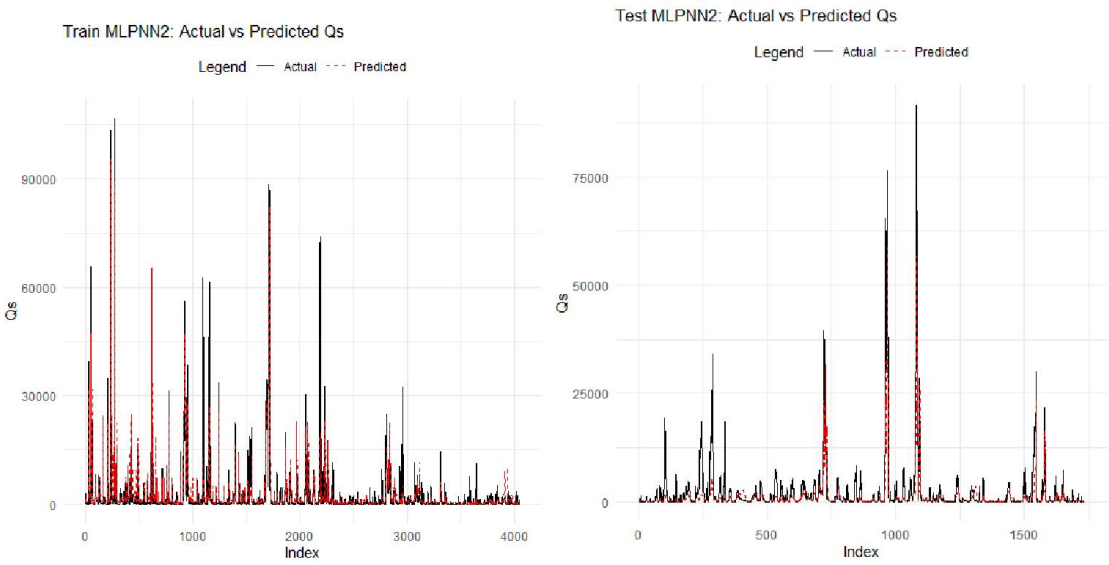


Figure IV.10. Predicted and observedSolid Flow Rate using MLPNN2 (70%, 30%)

The performance parameters are shown in the table IV.3.

Table IV.3. Performance parameters of the MLPNN2 model (70%, 30%)

Model			**Hidden Layers**	**Training**		**Test**	
MLPNN	**Input**	**Output**		**R**	**NSE**	**R**	**NSE**
MLPNN2	**Ql**	**Qs**	**2(3,2)**	**0.89**	**0.80**	**0.89**	**0.67**

The architecture of this model consists of **two hidden layers** with **3 and 2 neurons**, respectively, represented as **2(3,2)**.During the **training phase**, the model achieved a **correlation coefficient (R) of 0.89**, which indicates a strong linear relationship between the predicted and measured values of the solid discharge. The **Nash-Sutcliffe Efficiency (NSE)** value of **0.80** reflects a very good agreement between the model's predictions and the observed data, suggesting that the model captures the underlying dynamics of the system with high accuracy.

In the **test phase**, the model maintained the same correlation coefficient (**R = 0.89**), confirming that it generalizes well and maintains prediction quality on unseen data. However, the **NSE dropped slightly to 0.67**, indicating a small decrease in predictive performance, which is a common occurrence due to variability in testing data or possible overfitting to the training dataset.

The results overall demonstrate that the **MLPNN2 architecture**, with two hidden layers, performs reliably in predicting solid transport from liquid flow rate. The consistency in **R values across both phases** indicates robustness, while the **NSE values above 0.65** suggest good model efficiency and reasonable generalization capability. The slightly improved NSE in training compared to the previous architecture (MLPNN1) may reflect the advantage of a deeper network in capturing more complex nonlinear patterns.

IV.13.1.1.3. MLPNN3 model

The architecture of the MLPNN3 model is shown in figure IV.11

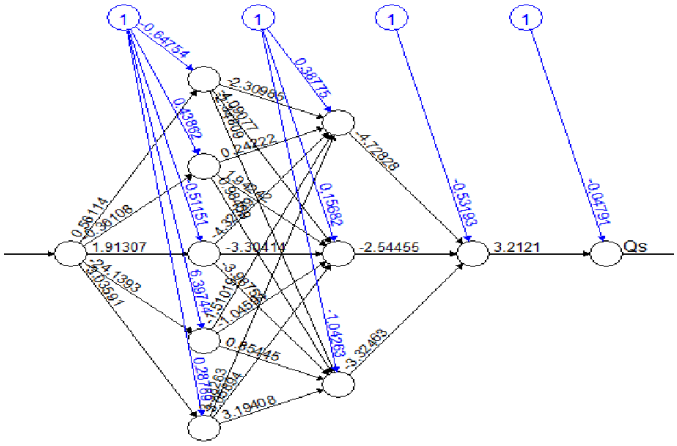


Figure IV.11. Architecture of the MLPNN3 model (70%, 30%)

The regression lines obtained are shown in Figures IV.12.

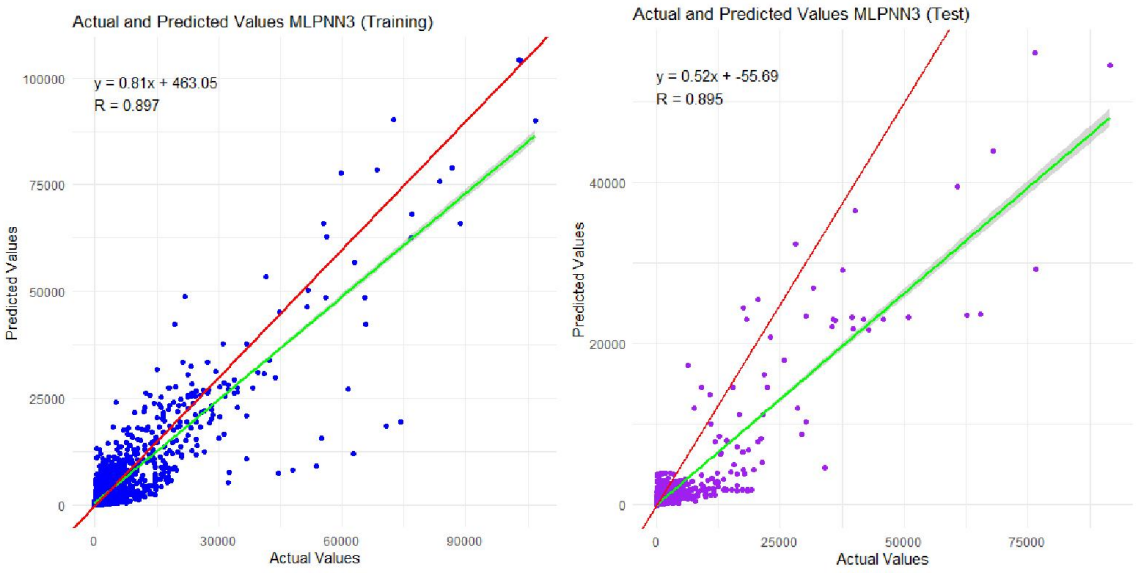


Figure IV.12. Regression Lines of predicted and Measured Solid Flow Rate (kg/s) Using MLPNN3 (70%, 30%)

The Predicted and observedSolid Flow Rate using MLPNN3 is show in figure IV.13.

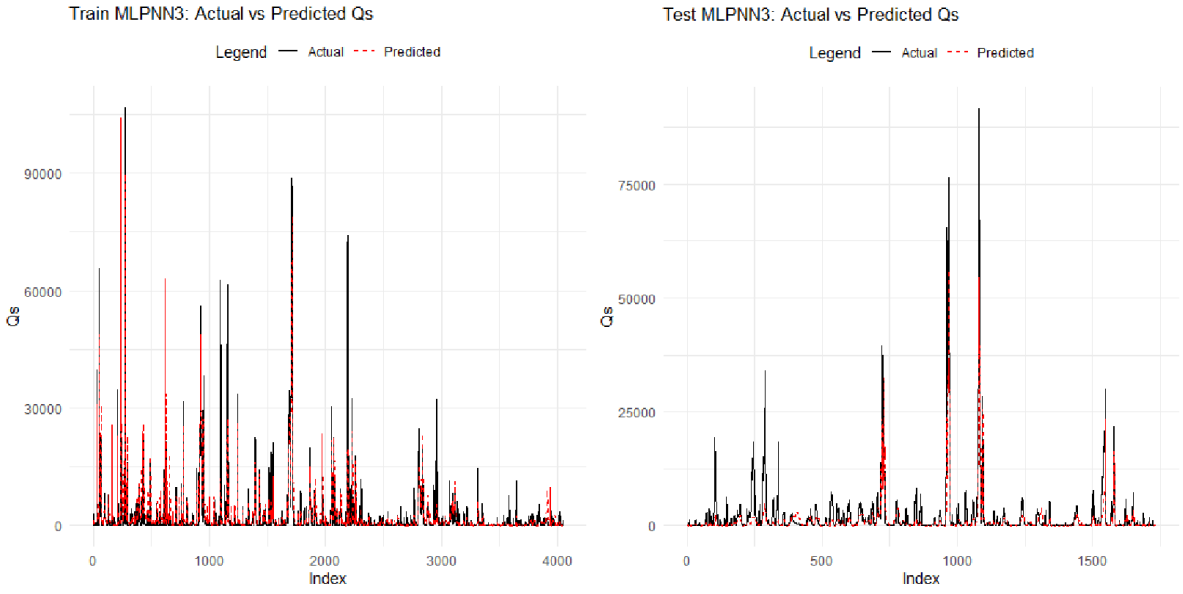


Figure IV.13. Predicted and observedSolid Flow Rate using MLPNN3 (70%, 30%)

The performance parameters are shown in the table IV.4.

Table IV.4. Performance parameters of the MLPNN3 model (70%, 30%)

Model			Hidden Layers	Training		Test	
MLPNN	Input	Output		R	NSE	R	NSE

MLPNN3	**Ql**	**Qs**	**3(5,3,1)**	**0.89**	**0.80**	**0.89**	**0.67**

The network architecture includes **three hidden layers**, with **5, 3, and 1 neurons** respectively, denoted as **3(5,3,1)**. This deeper structure was intended to explore the impact of increased complexity on model performance.

During the **training phase**, the model achieved a **correlation coefficient (R) of 0.89**, which indicates a strong linear correlation between predicted and observed solid flow values. The **Nash-Sutcliffe Efficiency (NSE)** during training was **0.80**, which demonstrates a good ability of the model to replicate the behavior of the target variable.

In the **test phase**, the **R value remained at 0.89**, indicating excellent consistency between training and testing performance. However, the **NSE slightly decreased to 0.67**, similar to the MLPNN2 configuration. This drop suggests a modest loss of accuracy in generalization, which could be due to the higher model complexity leading to slight overfitting.

Overall, the MLPNN3 model shows performance that is **comparable to MLPNN2**, with **identical R values** and **slightly varied NSE values**. Despite having an additional hidden layer, MLPNN3 does not significantly outperform simpler architectures, suggesting that **increasing model depth beyond a certain point may not yield substantial gains** in this particular application. Nonetheless, the model remains robust and effective for predicting solid transport based on liquid flow input.

IV.13.1.2. Model Configuration with 80% Training and 20% Validation Data

IV.13.1.2.1. MLPNN1 model

The architecture of the MLPNN1 model is shown in figure IV.14

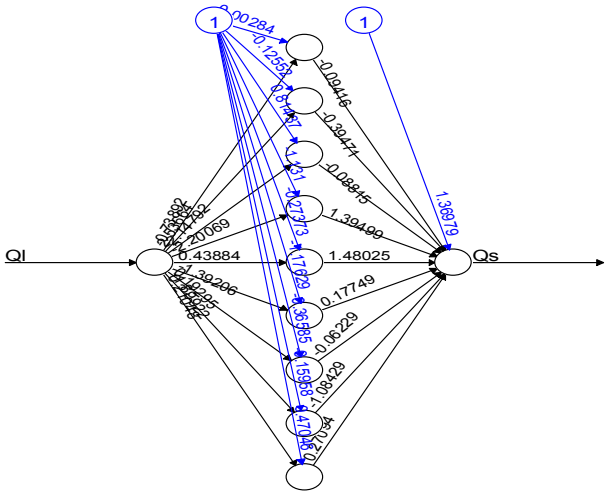


Figure IV.14. Architecture of the MLPNN1 model (80%, 20%)

The regression lines obtained are shown in Figures IV.15.

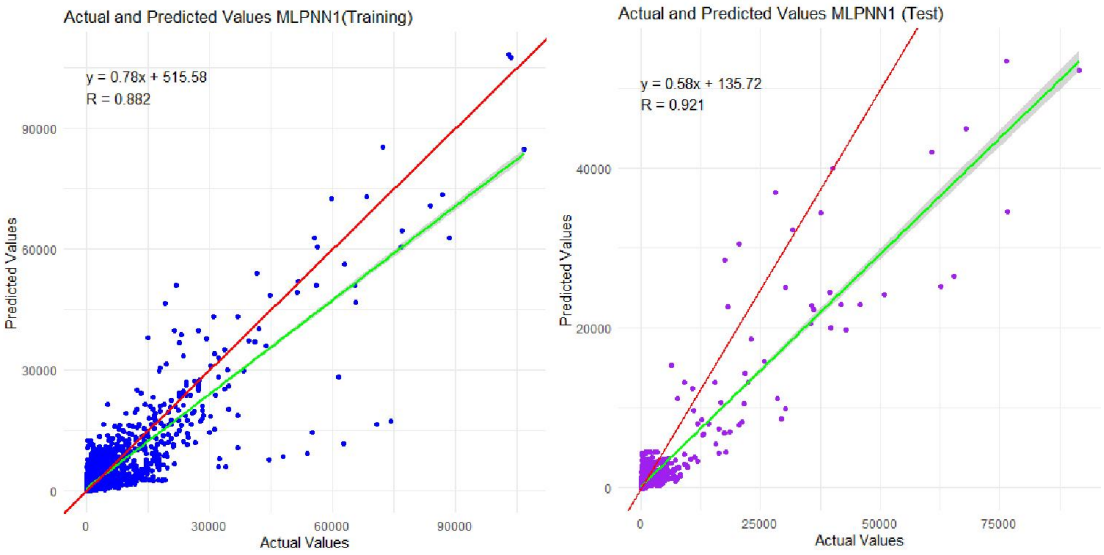


Figure IV.15. Regression Lines of predicted and Measured Solid Flow Rate (kg/s) Using MLPNN (80%, 20%)

The performance parameters are shown in the table IV.5.

Table IV.5. Performance parameters of the MLPNN1 model (80%, 20%)

Model			**Hidden Layers**	**Training**		**Test**	
MLPNN	**Input**	**Output**		**R**	**NSE**	**R**	**NSE**
MLPNN1	**Ql**	**Qs**	**1(9)**	**0.88**	**0.77**	**0.92**	**0.75**

In the **training phase**, the model achieved a **correlation coefficient (R) of 0.88**, indicating a strong correlation between the predicted and the measured solid flow rates. The **Nash-Sutcliffe Efficiency (NSE)** during training was **0.77**, which reflects a good level of agreement between the simulated and observed values, confirming the model's ability to capture the underlying physical process during learning.

In the **test phase**, the model showed **improved performance**, with an **R value of 0.92**, demonstrating an even stronger correlation between the predicted and actual solid flow rates on unseen data. The **NSE also increased to 0.75**, confirming that the model not only generalizes well but may even perform better on the validation dataset. This could be due to the representative nature of the validation data or the network's robustness against overfitting.

Overall, the MLPNN1 model shows excellent performance in both training and testing phases. The results indicate that even with a relatively simple architecture, the model effectively captures the nonlinear relationship between liquid flow and solid transport. The

improvement in both R and NSE during testing highlights the model's **strong generalization capability**, making it a **reliable tool for solid flow rate prediction** based on liquid discharge measurements.

IV.13.1.2.2. MLPNN2 model

The architecture of the MLPNN2 model is shown in figure IV.16

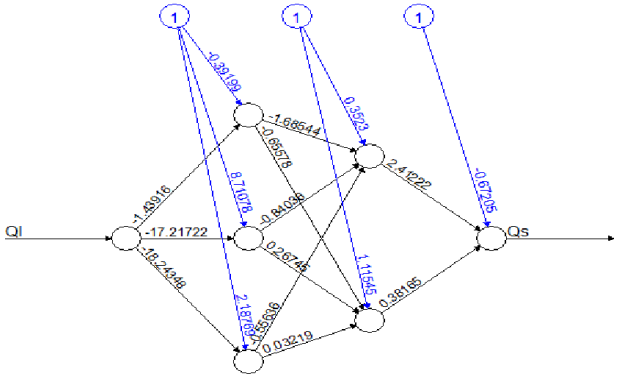


Figure IV.16. Architecture of the MLPNN2 model (80%, 20%)

The regression lines obtained are shown in Figures IV.17.

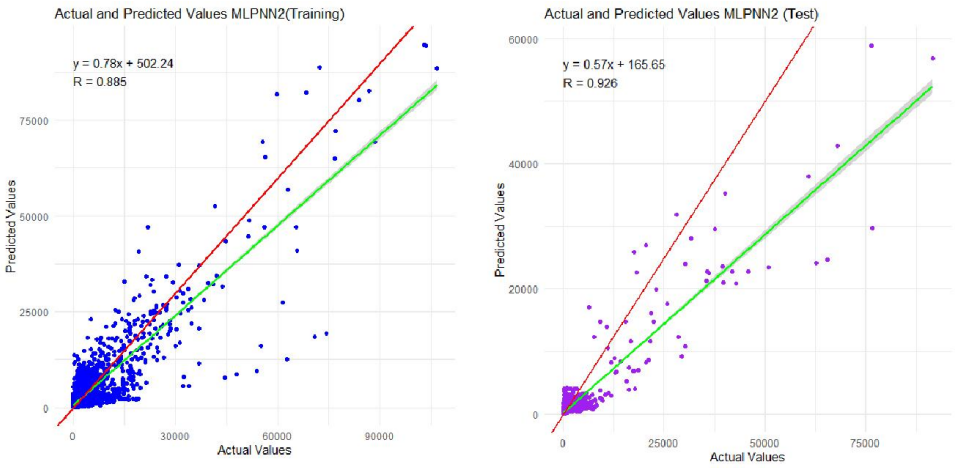


Figure IV.17. Regression Lines of predicted and Measured Solid Flow Rate (kg/s) Using MLPNN2 (80%, 20%)

The performance parameters are shown in the table IV.6.

Table IV.6. Performance parameters of the MLPNN2 model (80%, 20%)

Model	Hidden	Training	Test

			Layers				
MLPNN	**Input**	**Output**		**R**	**NSE**	**R**	**NSE**
MLPNN2	**Ql**	**Qs**	**2(3,2)**	**0.88**	**0.78**	**0.92**	**0.74**

In the **training phase**, the model achieved a **correlation coefficient (R) of 0.88**, suggesting a strong positive correlation between the predicted and observed solid discharge values. The **Nash-Sutcliffe Efficiency (NSE)** was **0.78**, which indicates that the model captures the variance in the observed data well, and that the predictions are close to the measured values.

In the **test (validation) phase**, the model showed a **notable improvement**, with an **R value increasing to 0.92**. This reflects an even stronger correlation in the testing phase, indicating excellent generalization capability. The **NSE reached 0.74**, which is also very good and confirms the reliability of the model when applied to unseen data.

Overall, the MLPNN2 model demonstrates robust and consistent performance across both training and testing. The slight increase in performance during validation suggests that the model is not overfitted and is able to generalize effectively. The use of **two hidden layers** appears to enhance the model's ability to capture the nonlinear behavior of solid transport, making MLPNN2 a **reliable architecture for predicting solid discharge** based on hydrological input.

IV.13.1.2.3. MLPNN3 model

The architecture of the MLPNN3 model is shown in figure IV.18

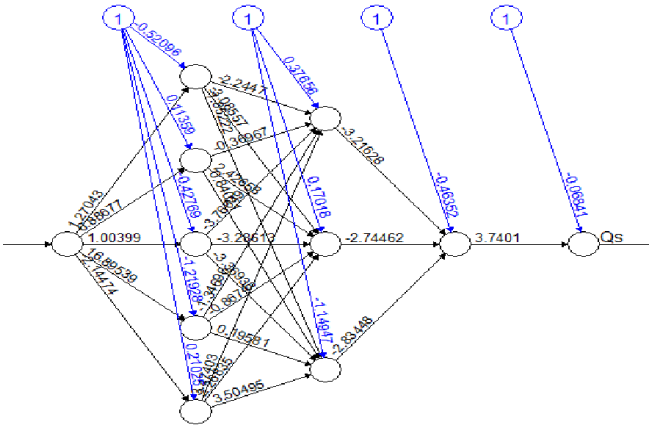


Figure IV.18. Architecture of the MLPNN3 model (80%, 20%)

The regression lines obtained are shown in Figures IV.19.

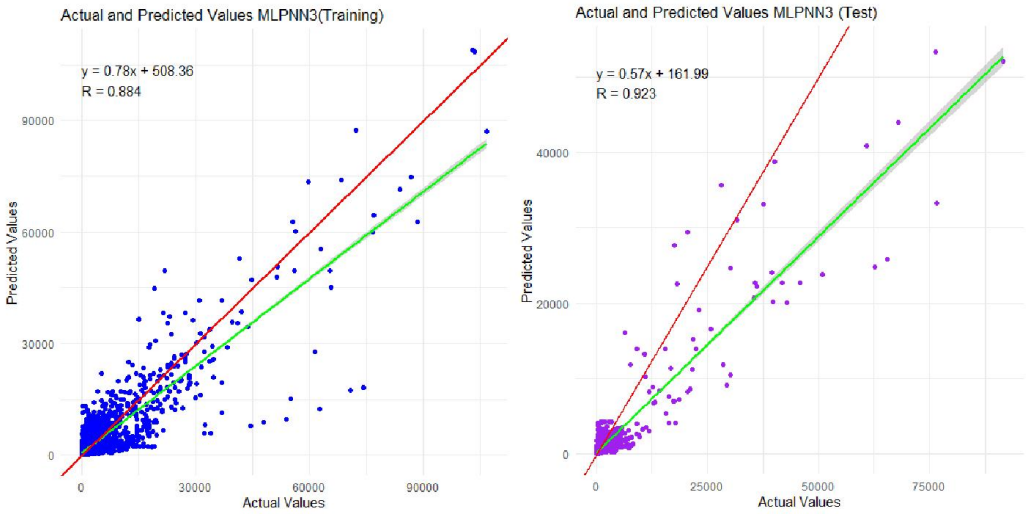


Figure IV.19. Regression Lines of predicted and Measured Solid Flow Rate (kg/s) Using MLPNN3 (80%, 20%)

The performance parameters are shown in the table IV.7.

Table IV.7. Performance parameters of the MLPNN3 model (80%, 20%)

Model			**Hidden Layers**	**Training**		**Test**	
MLPNN	**Input**	**Output**		**R**	**NSE**	**R**	**NSE**
MLPNN3	**Ql**	**Qs**	**3(5,3,1)**	**0.88**	**0.78**	**0.92**	**0.74**

In the **training phase**, the model achieved a **correlation coefficient (R) of 0.88**, indicating a strong relationship between the predicted and measured values of solid discharge. The **Nash-Sutcliffe Efficiency (NSE)** was **0.78**, suggesting that the model can reliably replicate the dynamics of the training data.

During the **test phase**, the model showed a notable improvement in performance, with an **R value of 0.92**, confirming a very strong linear relationship between predicted and observed solid flow rates on unseen data. The **NSE also improved to 0.74**, further supporting the model's ability to generalize effectively without overfitting.

These results demonstrate that the **MLPNN3 architecture**, despite being deeper than MLPNN1 and MLPNN2, performs equally well in terms of correlation and efficiency. The consistent values of **R** and **NSE** in both training and testing phases highlight the robustness of the model. The deeper structure may offer a slight advantage in capturing more complex nonlinearities in the relationship between liquid and solid discharge, though the performance gain compared to simpler architectures is relatively modest.In conclusion, **MLPNN3 provides a strong and stable prediction performance**, making it a valid and reliable model for solid transport modeling based on hydrological input.

IV.13.1.3. Model Performance Comparison Under Two Data Splits: 70%–30% and 80%–20%

The performance of three MLPNN models (MLPNN1, MLPNN2, and MLPNN3) was evaluated under two different training–validation data splits: 70% for training and 30% for testing, and 80% for training and 20% for testing. The objective of this comparison is to assess how the proportion of training data affects the generalization and predictive accuracy of each model.

In the 70%–30% configuration, all three models exhibited consistent results:

- R = 0.89 across training and testing phases for all models, showing strong correlation between predicted and measured solid discharge values.
- NSE values ranged from 0.79 to 0.80 during training and from 0.67 to 0.68 during testing, indicating reliable performance but with a moderate drop in predictive accuracy on unseen data.

Under the 80%–20% configuration, the models showed slightly lower R values during training (R = 0.88), which is expected due to reduced training variability, but higher R values during testing (R = 0.92). This suggests better generalization when more data is available during model learning. The NSE values also improved on the test set, reaching 0.74–0.75, compared to 0.67–0.68 with the 70%–30% split.

These results reveal that increasing the training portion from 70% to 80% leads to slightly improved generalization and prediction accuracy, especially in the testing phase. This is evident for all models, but particularly for MLPNN1, which achieved the highest test NSE of 0.75 under the 80%–20% split, along with a very strong R = 0.92.In conclusion, the 80%–20% configuration is more favorable for all MLPNN architectures tested, as it enhances the models' ability to generalize and predict solid transport behavior. This comparison highlights the importance of optimizing data partitioning to improve model robustness and performance.

IV.13.2. Random forest (RF)

The regression lines obtained for two configuration (70%-30%) and (80%-20%) are shown in Figures IV.20 and figure IV21.

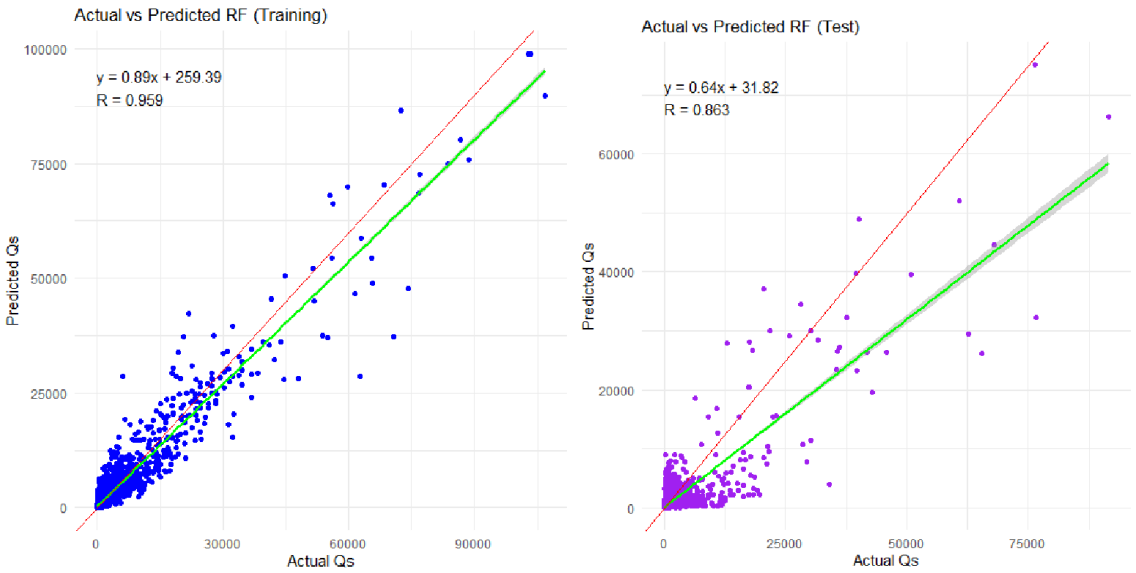


Figure IV.20. Regression Lines of predicted and Measured Solid Flow Rate (kg/s) Using RF (70%, 30%)

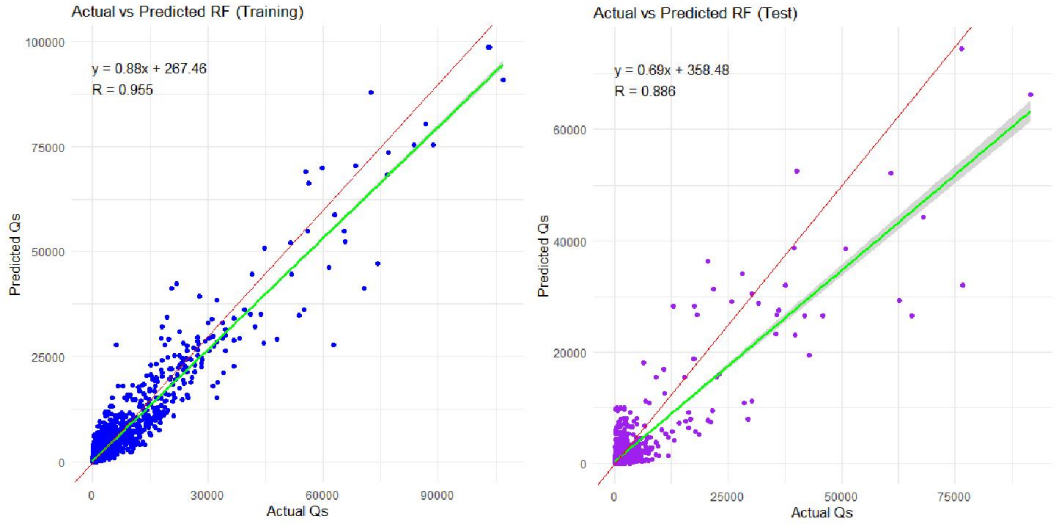


Figure IV.21. Regression Lines of predicted and Measured Solid Flow Rate (kg/s) Using RF (80%, 20%)

The performance parameters are shown in the table IV.8.

Table IV.8. Performance parameters of the RF model

Model			**Training**		**Test**	
			70%		**30%**	
	Input	**Output**	**R**	**NSE**	**R**	**NSE**
RF	**Ql**	**Qs**	**0.95**	**0.91**	**0.88**	**0.77**
			80%		**20%**	
			R	**NSE**	**R**	**NSE**
			0.95	**0.91**	**0.86**	**0.71**

The performance of the Random Forest (RF) model in predicting solid flow rate (Qs) from liquid flow rate (Ql) was evaluated using two different data partitioning strategies: 70% for training and 30% for testing, and 80% for training and 20% for testing. The goal is to assess the influence of the training set size on the model's predictive accuracy and generalization.

For the first configuration, the model was trained on 70% of the dataset and validated on the remaining 30%. The results show a very high correlation coefficient (R) of 0.95 and an NSE of 0.91 during training, indicating that the model fits the training data extremely well. In the testing phase, the RF model achieved R = 0.88 and NSE = 0.77, which reflects a strong predictive capability and effective generalization to unseen data. These results suggest that the model captures the underlying relationship between input and output variables with high accuracy.

In the second configuration, with 80% of the data used for training and 20% for testing, the model again achieved R = 0.95 and NSE = 0.91 during training — identical to the previous configuration, demonstrating the model's stability and reliability. However, the test performance slightly decreased, with R dropping to 0.86 and NSE to 0.71. This indicates a small decline in generalization, which might be due to the smaller test set containing data with higher variability or complexity.

Both configurations show that the RF model is highly robust and accurate in modeling solid transport. The training results remain constant across both splits, reflecting the model's ability to learn consistently from larger datasets. The 70%–30% configuration yielded slightly better test performance (R = 0.88, NSE = 0.77) compared to the 80%–20% configuration (R = 0.86, NSE = 0.71). This suggests that a slightly larger test set may have provided a more representative evaluation of model generalization.

Overall, the RF model demonstrates excellent performance in both configurations, with only a minor trade-off between training size and validation accuracy. Its consistently high correlation and efficiency metrics make it a reliable and effective tool for predicting solid discharge from liquid flow in hydrological systems.

IV.13.3. Comparison of MLPNN and Random Forest (RF) Models

The comparative analysis between the Multi-Layer Perceptron Neural Network (MLPNN) models and the Random Forest (RF) model highlights the strengths and limitations of each in predicting solid flow rate (Qs) from liquid flow rate (Ql). Three MLPNN architectures were tested: MLPNN1 with one hidden layer (1(9)), MLPNN2 with two hidden layers (2(3,2)), and MLPNN3 with three hidden layers (3(5,3,1)). Under the **70%–30% configuration**, all MLPNN models demonstrated identical training correlation coefficients (**R = 0.89**) with NSE values ranging from **0.79 to 0.80**, and testing R values also equal to **0.89**, with NSEs between **0.67 and 0.68**. These results suggest that MLPNNs, regardless of depth, provided stable but moderate generalization performance.

In contrast, the RF model significantly outperformed MLPNNs during training, achieving **R = 0.95** and **NSE = 0.91**, which indicates an excellent fit to the training data. During testing, RF maintained high performance with **R = 0.88** and **NSE = 0.77**, outperforming all MLPNN configurations in terms of predictive accuracy and robustness on unseen data in the 70%–30% setup.

In the **80%–20% configuration**, MLPNN models showed slightly lower R values during training (**R = 0.88**) but improved testing performance, with R values rising to **0.92** and NSEs up to **0.75**. This suggests better generalization when more data is allocated to training. On the other hand, the RF model maintained consistent training performance (**R = 0.95**, **NSE = 0.91**), but test performance decreased slightly compared to the 70%–30% case (**R = 0.86**, **NSE = 0.71**), likely due to the smaller and potentially less diverse validation set.

In summary, the **RF model consistently achieved the highest training accuracy and strong generalization**, especially under the 70%–30% split. However, **MLPNNs, particularly under the 80%–20% configuration**, demonstrated **better generalization** in terms of **test correlation** (R = 0.92). Thus, while RF excels in accuracy and robustness,

MLPNNs may offer better generalization under specific training conditions. The choice between these models should therefore consider both predictive strength and sensitivity to training/validation split strategies.

IV.14. Conclusion

The results obtained for solid transport modeling based on liquid flow demonstrate the effectiveness of both **Random Forest (RF)** and **MLPNN** models. The RF model achieved **superior overall performance**, with high R and NSE values, indicating excellent accuracy during both training and testing. MLPNN models, although slightly less accurate in training, showed **strong generalization ability**, especially with the 80%–20% data split. This highlights the **reliability of both approaches** in predicting solid discharge, with RF being more precise and MLPNN offering **better predictive stability** under certain configurations. These findings confirm the value of machine learning techniques in hydrological modeling.

General Conclusion

This research study focused on evaluating soil erosion and suspended sediment transfers in the Tafna watershed (Algeria, North-West). The study presented an integrated framework by combining the empirical RUSLE model with machine learning models (MLPNN and Random Forest), which represents a new usable paradigm to quantify the risk of erosion, as well as incorporating potential sediment yield through using both spatial and temporal dimensions.

This project also can push forward watershed management by merging GIS and remote sensing, and artificial intelligence addressing, both the spatial aspect of erosion in terms of physical characteristics of erosion along with the temporal nature of sediment transport as an additional addressable ecosystem service.

Watershed Characteristics

• The Tafna Basin has an approximate area of 7,700 km² and is characterized by highly heterogeneous topography

• Steep slope (>25%) occupy approximately 2.45% the basin, while moderately steep terrain (10–25%) occupies 22.1% of the area – confirming the naturally erosion susceptible conditions of this region.

The Revised Universal Soil Loss Equation (RUSLE) implemented by using the RUSLE equation directly in ArcGIS and incorporating Google Earth Engine, developed multi on the erosion risks using five spatially derived descriptive factors including: rainfall erosivity factor (R), soil erodibility factor (K), topographic slope and length factor (LS), cover management factor (C) and conservation practices factor (P). Key Findings from RUSLE modeling are:

• The average annual soil loss at the sub-watershed scale ranged from 2.92 to 5.87 t/ha/year and remained below the tolerable value of 7 t/ha/year.

• However, pixel-level average values identified hotspots with soil erosion values exceeding 146 t/ha/year seems to be limited to the southern and centre eastern sub-basins with steep slopes and high degrading vegetation.

The dominant zones for RUSLE classes representing soil loss throughout the Tafna basin are:

- ✓ Low erosion (<5 t/ha/year) in forested and flat areas
- ✓ Moderate erosion (5-15 t/ha/year) in transitional slope areas
- ✓ Severe erosion zones (>30 t/ha/year) in steep exposed zones with high values of Conservation practices (P ≈ 1)

These key findings highlight a need for better sourcing land management practices to assess and mitigate erosion in specific hotspot zones, while recognizing that the basin-wide average is moderate.

To complement erosion mapping, sediment discharge prediction was performed using Artificial Neural Networks (MLPNNs) and Random Forest (RF**)**, trained on rainfall and streamflow data from hydrometric stations. The models performances are:

- **MLPNNs** achieved good predictive accuracy, with:
  - **R values**around**0.88–0.89**,
  - and **NSE values** up to **0.75** on test datasets.
- **Random Forest** outperformed neural networks, yielding:
  - **R = 0.95** in training and **0.86–0.88** in testing,
  - **NSE = 0.91** (train) and **0.71–0.77** (test), confirming its robustness for sediment prediction in complex hydrological settings.

This study highlights the valuable contribution of integrating physical modeling, specifically the RUSLE model, with data-driven machine learning approaches to improve the accuracy of erosion and sediment transport forecasting. It offers practical outcomes, including high resolution erosion risk maps, predictive sediment transport tools, and the identification of priority areas requiring urgent intervention. These results equip decision-makers with robust and reliable tools to support soil conservation programs, promote sustainable land-use planning, and guide watershed restoration efforts effectively.

Outlook and Future Perspectives

This research opens the way for several future developments and improvements, including the following directions:

1. Application of Alternative Erosion Models
   Extend the erosion modeling approach in the Tafna watershed by applying other empirical or data-driven models to compare results and improve accuracy.
2. Development of an Erosion Control Planning Strategy
   Establish a comprehensive anti-erosion land management plan based on the erosion sensitivity map. Priority zones should be classified and treated first using suitable vegetative or mechanical conservation methods.
3. Updating and Refining RUSLE Results
   In future studies, update RUSLE outputs with more recent land use, rainfall, and soil data to improve monitoring and support the design of revised soil conservation strategies.
4. Field Validation of RUSLE Outputs
   Validate model results through on-site measurements of soil loss using monitoring plots distributed across different zones of the watershed.
5. Transfer of Neural Network Models to Other Watersheds
   Apply the trained MLPNN and Random Forest models to neighboring or comparable watersheds to test their generalizability and support regional erosion prediction.
6. Testing Other Machine Learning Models
   Explore the use of other advanced neural or machine learning models to improve the precision of sediment yield predictions in the Tafna basin.

References

1. Bach, E.M., Ramirez, K.S., Fraser, T.D., Wall, D.H. (2020). Soil Biodiversity Integrates Solutions for a Sustainable Future. Sustainability, 12, 2662.

2. Bhat, S.A., Kumar, R., Vihwakarma, D.K., Bhat, O., n.d. Runoff and Sediment Monitoring Using the WEPP Model. In: Erosion Measurement, Modeling, and Management. Apple Academic Press, pp. 101–121.

3. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

4. Borrelli, P., Alewell, C., Alvarez, P., Anache, J.A.A., Baartman, J., Ballabio, C., Bezak, N., Biddoccu, M., Cerdà, A., Chalise, D. (2021). Soil erosion modelling: A global review and statistical analysis. Science of the Total Environment. Elsevier.

5. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

6. Chesire, A.K. (2022). Soil erosion prediction using modified universal soil loss equation (MUSLE) in Tugen hills, Baringo, Kenya (PhD Thesis). University of Eldoret.

7. Chollet, F. (2018). Deep Learning with Python. Manning Publications.

8. Demuth, H., & Beale, M. (2002). Neural Network Toolbox User's Guide - Version 4. Natick: The MathWorks.

9. Ganasri, B.P., Ramesh, H. (2016). Assessment of soil erosion by RUSLE model using remote sensing and GIS—A case study of Nethravathi Basin. Geoscience Frontiers. Elsevier.

10. Gandhi krishi Vishwavidyalaya, I., n.d. A Review on Soil Erosion: Causes, Prevention, and Control.

11. Ghosal, K., Das Bhattacharya, S. (2020). A Review of RUSLE Model. J. Indian Soc. Remote Sens., 48, 689–707. https://doi.org/10.1007/s12524-019-01097-0

12. Gong, W., Liu, T., Duan, X., Sun, Y. (2022). Estimating the Soil Erosion Response to Land-Use Land-Cover Change Using GIS-Based RUSLE and Remote Sensing: A Case Study of Miyun Reservoir, North China. Water, 14, 742.

13. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

14. Hamdan, A., Rahman, M.T.A., & Abdullah, R. (2021). Hydrological modeling and sediment yield analysis using HEC-HMS and RUSLE integrated in GIS: A case study of tropical catchments. Water, 13(11), 1567. https://doi.org/10.3390/w13111567

15. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

16. Hasanuzzaman, M., Islam, A., Bera, B., & Shit, P.K. (2022). A comparison of performance measures of three machine learning algorithms for flood susceptibility

mapping of river Silabati (tropical river, India). Physics and Chemistry of the Earth, Parts A/B/C, 127, 103198.

17. Haykin, S. (1998). Neural Networks: A Comprehensive Foundation. Prentice Hall.

18. Ismael, O., Joseph, S., & Patrick, K.H. (2017). HEC-HMS model for runoff simulation in Ruiru reservoir watershed. American Journal of Engineering Research (AJER), 6, 1–7.

19. Jahun, B.G., Ibrahim, R., Dlamini, N.S., Musa, S.M. (2015). Review of soil erosion assessment using RUSLE model and GIS. J. Biol. Agric. Healthc., 5, 36–47.

20. Jasinski, M.F. (1990). Sensitivity of the normalized difference vegetation index to subpixel canopy cover, soil albedo, and pixel scale. Remote Sensing of Environment, 32, 169–187.

21. Jimeno-Sáez, P., Martinez-Espana, R., Casali, J., et al. (2022). A comparison of performance of SWAT and machine learning models for predicting sediment load in a forested basin, Northern Spain. Catena, 212, 105953.

22. Kateb, S., Meddi, M., & Mouffok, F. (2020). Soil erosion risk assessment using RUSLE integrated with GIS and remote sensing in the Beni Haroun watershed (Algeria). Environmental Earth Sciences, 79, 1–14. https://doi.org/10.1007/s12665-020-09273-9

23. Kateb, Z., Bouchelkia, H., Benmansour, A., & Belarbi, F. (2020). Sediment transport modeling by the SWAT model using two scenarios in the watershed of Beni Haroun dam in Algeria. Arabian Journal of Geosciences, 13, 1–17.

24. Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology, 424–425, 264–277.

25. Lei, Q., Zhang, T., An, M., Luo, J., Qin, L., Zhu, A.-X., Qiu, W., Du, X., Liu, H. (2024). Sensitivity analysis of SWAT streamflow and water quality to the uncertainty in soil properties generated by the SoLIM model. Journal of Hydrology. Elsevier.

26. MAHRUKH, S.A.B., Kumar, R., Vihwakarma, D.K., Bhat, O. (2025). Runoff and Sediment Monitoring Using the WEPP Model. Eros. Meas. Model. Manag. Chall. Solut., 101.

27. Mazour, M., & Roose, E. (2002). Influence de la couverture végétale sur le ruissellement et l'érosion des sols sur parcelles d'érosion dans les bassins versants du Nord-ouest de l'Algérie. Bulletin Réseau Erosion, 21, 320–330.

28. McCool, D.K., Brown, L.C., Foster, G.R., Mutchler, C.K., & Meyer, L.D. (1987). Revised slope steepness factor for the Universal Soil Loss Equation. Transactions of the ASAE, 30, 1387–1396.

29. McCool, D.K., Foster, G.R., Mutchler, C.K., & Meyer, L.D. (1989). Revised slope length factor for the Universal Soil Loss Equation. Transactions of the ASAE, 32, 1571–1576.

30. Meddi, M., & Toumi, S. (2015). Rainfall erosivity estimation in Northern Algeria using daily rainfall data. Arabian Journal of Geosciences, 8, 6857–6869. https://doi.org/10.1007/s12517-014-1687-z

31. Mohammadi, A.A., Akgoz, R., Deviren Saygin, S., Erpul, G. (2025). Validation and comparison of event-based sediment yield estimations using MUSLE and RUSLE/SDR models: The case of çarşambasuyu sub-basin, southwest central Anatolia, Türkiye. Nat. Hazards. https://doi.org/10.1007/s11069-025-07328-7

32. Moore, I.D., & Burch, G.J. (1986). Physical basis of the length-slope factor in the Universal Soil Loss Equation. Soil Science Society of America Journal, 50, 1294–1298.

33. Muauz, A., Berehanu, B., Bedru, H. (2024). Spatial Analysis of Water Budget Components in the Upper Awash River Sub-basin, Ethiopia Using SWAT Model.

34. Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10(3), 282–290.

35. Pal, P., Dandasena, N.K., Bhattacharya, P. (2025). Addressing Soil Erosion: Approaches for Sustainable Agriculture and Environmental Preservation. Cultiv. Prog., 39.

36. Parveen, R., & Kumar, U. (2012). Integrated approach of Universal Soil Loss Equation (USLE) and Geographical Information System (GIS) for soil loss risk assessment in Upper South Koel Basin, Jharkhand. Journal of Geographic Information System, 4, 588–596.

37. Prakash, C., Ahirwar, A., Lohani, A.K., Singh, H.P. (2024). Comparative analysis of HEC-HMS and SWAT hydrological models for simulating the streamflow in sub-humid tropical region in India. Environmental Science and Pollution Research, 31, 41182–41196. https://doi.org/10.1007/s11356-024-33861-2

38. Ranzi, R., Bochicchio, M., & Bacchi, B. (2012). Evaluation of soil erosion in alpine catchments by RUSLE and GIS: A comparison with sediment yield. Hydrological Sciences Journal, 57(7), 1305–1322.

39. Ranzi, R., Le, T.H., & Rulli, M.C. (2012). A RUSLE approach to model suspended sediment load in the Lo river (Vietnam): Effects of reservoirs and land use changes. Journal of Hydrology, 422, 17–29.

40. Reda, Y., Moges, A., Kendie, H. (2024). Application of the Modified Universal Soil Loss Equation (MUSLE) for the prediction of sediment yield in Agewmariam experimental watershed, Tekeze River basin, Northern Ethiopia. Heliyon. Elsevier.

41. Renard, K.G., Foster, G.R., Weesies, G.A., McCool, D.K., & Yoder, D.C. (1997). Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE). USDA Agriculture Handbook No. 703.

42. Renard, K.G., Foster, G.R., Weesies, G.A., & Porter, J.P. (1991). RUSLE: Revised Universal Soil Loss Equation. Journal of Soil and Water Conservation, 46, 30–33.

43. Roose, E. (1994). Introduction à la gestion conservatoire de l'eau, de la biomasse et de la fertilité des sols (GCES). FAO.

44. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

45. Sader, S.A., & Winne, J.C. (1992). RGB-NDVI colour composites for visualizing forest change dynamics. International Journal of Remote Sensing, 13, 3055–3067.

46. Sahli, A., Kalla, M., Lakhouili, A., & Messouci, C. (2019). Application of the RUSLE model using GIS techniques for soil erosion assessment in the northeast of Algeria. Geoderma Regional, 17, e00216. https://doi.org/10.1016/j.geodrs.2019.e00216

47. Sahli, Y., Mokhtari, E., Merzouk, B., et al. (2019). Mapping surface water erosion potential in the Soummam watershed in Northeast Algeria with RUSLE model. Journal of Mountain Science, 16, 1606–1615.

48. Seutloali, K.E., Dube, T., Sibanda, M. (2018). Developments in the remote sensing of soil erosion in the perspective of sub-Saharan Africa. Remote Sensing Applications: Society and Environment, 9, 100–106.

49. Somura, H., Arnold, J., Hoffman, D., Takeda, I., Mori, Y., et al. (2009). Impact of climate change on the Hii River basin and salinity in Lake Shinji: A case study using the SWAT model and a regression curve. Hydrological Processes, 23(13), 1887–1900.

50. Tamrabet, Z., Marouf, N., & Remini, B. (2019). Quantification Of Suspended Solid Transport In Endja Watercourse [Dehamecha Basin-Algeria]. Geoscience Engineering, 65, 71–91.

51. Telkar, S. G. (2015). Accelerated soil erosion. International Journal of Economic Plants, 2(4), 178–180.

52. Telkar, S. G., Solanki, S. P. S., Dey, J. K., & Kant, K. (2015). Soil erosion: Types and their mechanism. International Journal of Economic Plants, 2(4), 178–180.

53. Toy, T.J., Foster, G.R., & Renard, K.G. (2002). Soil Erosion: Processes, Prediction, Measurement, and Control. Wiley-Blackwell.

54. Ugwu, I.E., Emmanuel, K., Vintus, O. (2024). Evaluating the Impacts of Soil Erosion on Selected Hillslopes in UNN Using WEPP Model. UNIZIK Journal of Engineering and Applied Science, 3, 567–572.

55. Wang, J., Yang, J., Li, Z., Ke, L., Li, Q., Fan, J., & Wang, X. (2024). Research on Soil Erosion Based on Remote Sensing Technology: A Review. Agriculture, 15(1), 18.

56. Wischmeier, W.H., & Smith, D.D. (1978). Predicting Rainfall Erosion Losses: A Guide to Conservation Planning. USDA Agriculture Handbook No. 537.

57. Wu, Y., Shi, H., Yang, X. (2024). Estimating the CSLE Biological Conservation Measures' B-Factor Using Google Earth's Engine. Remote Sens., 16, 847.

58. Zhang, B., Guo, J., Fang, H., Wu, S., Feng, H., Siddique, K.H.M. (2024). Soil erosion projection and response to changed climate and land use and land cover on the Loess Plateau. Agric. Water Manag., 306, 109187.

59. Zhao, J., Zhang, N., Liu, Z., Zhang, Q., Shang, C. (2024). SWAT model applications: From hydrological processes to ecosystem services. Science of The Total Environment. Elsevier.