A Unified Framework for Advancing Soil Erosion and Flood Assessment Through Deep Learning and Process-Based Modeling

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Hadi Shokati
aus Maragheh/Iran

Tübingen
2026

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	21.07.2026
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Thomas Scholten
2. Berichterstatter/-in:	PD Dr. Harald Neidhartdt

Disclaimer: Generative AI tools, including ChatGPT (GPT-5.3) and DeepL, were utilized to support improvements in readability, grammar, spelling, and overall writing style, while all scientific content, analysis, and interpretations remain the sole responsibility of the author.

Acknowledgments

I owe my deepest gratitude to my supervisor, Prof. Dr. Thomas Scholten. His trust, generosity, and unwavering support shaped not only this dissertation but also my development as a researcher and as a person. He consistently believed in my ideas, gave me the freedom to explore them, and stood behind me in challenging moments. Words are ultimately insufficient to express my appreciation; his mentorship went far beyond scientific supervision and has been a lasting example of integrity, humanity, and academic leadership.

I am equally grateful to my second supervisor, Prof. Dr. Peter Fiener, whose expertise and guidance were indispensable throughout this project. He was never just a second supervisor; whenever I needed scientific input, critical feedback, or practical advice, he made time and engaged with my work in depth. His constructive questions and careful reading of my manuscripts greatly improved the scientific quality of this dissertation.

I would also like to extend my sincere thanks to PD Dr. Harald Neidhartdt for meticulously reviewing this dissertation, and to Prof. Dr.-Ing. Hendrik P. A. Lensch and Prof. Dr. Volker Hochschild for their valuable presence and insightful contributions as members of my examination committee.

My sincere thanks go to Kay Seufferheld, with whom I had the pleasure of collaborating closely during this project. Our countless discussions were intellectually stimulating and personally very enjoyable.

I gratefully acknowledge the technical support and infrastructure provided by the Cluster of Excellence – Machine Learning for Science for granting me the essential cloud computing access required to run my models. I am also highly thankful to the Deutscher Wetterdienst (DWD) for providing the crucial German meteorological data that formed the foundation of this study.

I would like to extend my thanks to all members of our research group at the Department of Soil Science and Geomorphology. Working in this group truly felt like being part of a large, supportive family. The friendly atmosphere, the open exchange of ideas, and the willingness to help each other made my time in Tübingen both scientifically stimulating and personally very happy.

I also wish to acknowledge all those who contributed their time and effort to the manual segmentation of the orthophoto imagery required for this study. Their painstaking work provided the essential data without which this research could not have proceeded.

Last, but most importantly, my deepest, most heartfelt gratitude goes to my wonderful wife, Zahra. This achievement would have been absolutely impossible without her. Whether we were separated by miles or standing side by side, her unconditional love, endless patience, and unwavering support have been my ultimate anchor. She has constantly pushed me to grow, sacrificed so much for my progress, and believed in me every single day. Thank you, Zahra, for being my strength and my home.

Thank you!

Hadi Shokati
Tübingen, May 15, 2026

Abstract

Soil erosion by water remains one of the most pressing environmental challenges for sustainable land management, agricultural productivity, and water quality protection. In Central Europe, and particularly in Germany, the problem is becoming more urgent as climate change intensifies rainfall extremes, agricultural landscapes become more homogenized, and traditional empirical standards increasingly fail to represent contemporary erosive conditions. Accurate assessment of rainfall erosivity, reliable simulation of sediment transport, and rapid identification of erosion- and flood-affected areas are therefore essential for evidence-based soil conservation planning. This dissertation addresses these challenges by integrating process-based modeling, deep learning, and remote sensing into a unified framework for erosion assessment and environmental monitoring in Germany.

The first part of the dissertation focuses on rainfall erosivity, the climatic driver of water erosion. Conventional approaches in Germany, in the absence of high-resolution precipitation data, are based on empirical relationships derived from historical precipitation conditions and do not adequately capture recent shifts in rainfall intensity. To address this limitation, a deep learning-based framework was developed to reconstruct historical rainfall erosivity and project future changes across Germany at high spatial resolution. By combining RADKLIM radar precipitation data with a Hypernetwork-Enhanced CNN model for historical reconstruction and a ConvLSTM encoder-decoder for near-term projection, the study produced a continuous century-scale erosivity dataset from 1931 to 2043. The results revealed a clear long-term increase in erosive forcing. This work demonstrates that data-driven models can serve as effective alternatives for estimating and forecasting rainfall erosivity in the absence of high temporal resolution data.

The second part of the dissertation evaluates the performance of the spatially distributed erosion and sediment transport model WaTEM/SEDEM under soil conservation conditions. Six highly instrumented micro-scale watersheds in Southern Germany, monitored over eight years, were used to test the model's ability to reproduce low sediment yields in landscapes characterized by conservation-oriented management and structural sediment control measures. A Generalized Likelihood Uncertainty Estimation framework (GLUE) was applied to assess parameter uncertainty and model plausibility across different spatiotemporal scales. The results show that WaTEM/SEDEM can reproduce the approximate magnitude of long-term sediment yields, but its performance is limited at annual and micro-watershed scales, especially where sediment fluxes are very low and landscape structures strongly influence connectivity. These findings highlight the importance of scale-aware model evaluation and demonstrate that model suitability depends not only on statistical fit, but also on the intended management application.

The third part of the dissertation addresses the challenge of automated environmental monitoring through deep learning-based image segmentation. Two related studies were conducted: the first

developed Erosion-SAM, a fine-tuned adaptation of the Segment Anything Model for delineating water erosion and deposition features in high-resolution aerial imagery; the second compared fine-tuned SAM and U-Net architectures with ResNet backbones for rapid flood mapping from UAV and helicopter imagery. Both studies show that transfer learning and foundation models can substantially reduce the need for large labeled datasets while maintaining strong segmentation performance. For erosion mapping, the adapted SAM successfully identified visually heterogeneous erosion structures across different land cover types. For flood mapping, both SAM-based and U-Net-based models achieved high accuracy, with strong practical potential for rapid post-event assessment and disaster response. Together, these studies demonstrate the value of modern computer vision methods for generating spatially explicit environmental observations that can support both scientific analysis and operational decision-making.

Taken together, the dissertation presents an integrated approach that links climatic forcing, process-based erosion modeling, and automated image-based monitoring. The overarching contribution lies in showing how deep learning can be used not as a replacement for environmental theory, but as a powerful complement to classical soil erosion science. The results provide improved tools for rainfall erosivity estimation, a critical evaluation of model performance under conservation conditions, and new methods for the rapid mapping of erosion and flood impacts. In doing so, the dissertation contributes to a more adaptive, data-rich, and spatially explicit framework for soil conservation planning, climate change impact assessment, and natural hazard management.

Zusammenfassung

Bodenerosion durch Wasser bleibt eines der drängendsten Umweltprobleme für nachhaltige Landnutzung, landwirtschaftliche Produktivität und den Schutz der Wasserqualität. In Mitteleuropa und insbesondere in Deutschland verschärft sich das Problem, da der Klimawandel extreme Niederschläge verstärkt, landwirtschaftliche Flächen zunehmend homogenisiert werden und traditionelle empirische Standards die gegenwärtigen Erosionsbedingungen immer weniger angemessen abbilden. Eine präzise Erfassung der Niederschlagserosivität, eine zuverlässige Simulation des Sedimenttransports sowie die schnelle Identifikation von erosions- und hochwasserbetroffenen Flächen sind daher essenziell für eine evidenzbasierte Boden- und Landschaftsschutzplanung. Diese Dissertation adressiert diese Herausforderungen durch die Integration von prozessbasierter Modellierung, Deep Learning und Fernerkundung in einen einheitlichen Rahmen für die Erosionsbewertung und das Umweltmonitoring in Deutschland.

Der erste Teil der Dissertation konzentriert sich auf die Niederschlagserosivität, dem klimatischen Treiber der Wassererosion. Da in Deutschland hochauflösende Niederschlagsdaten fehlen, basieren konventionelle Ansätze auf empirischen Zusammenhängen, die aus historischen Niederschlagsbedingungen abgeleitet wurden und jüngste Veränderungen der Niederschlagsintensität nicht ausreichend erfassen. Um diese Einschränkung zu überwinden, wurde ein Deep-Learning-basiertes Framework entwickelt, das die historische Niederschlagserosivität rekonstruiert und zukünftige Veränderungen deutschlandweit mit hoher räumlicher Auflösung projiziert. Durch die Kombination von RADKLIM-Radarniederschlagsdaten mit einem Hypernetwork-gestützten CNN-Modell zur historischen Rekonstruktion und einem ConvLSTM-Encoder-Decoder zur kurzfristigen Projektion entstand ein kontinuierlicher Datensatz der Erosivität über ein Jahrhundert hinweg (1931–2043). Die Ergebnisse zeigen einen deutlichen langfristigen Anstieg der erosiven Kraft. Diese Arbeit zeigt, dass datengestützte Modelle als effektive Alternativen zur Abschätzung und Vorhersage der Niederschlagserosivität dienen können, wenn keine Daten mit hoher zeitlicher Auflösung verfügbar sind.

Der zweite Teil der Dissertation bewertet die Leistungsfähigkeit des räumlich verteilten Erosions- und Sedimenttransportmodells WaTEM/SEDEM unter Bedingungen des Bodenschutzes. Sechs intensiv überwachte kleinräumige Einzugsgebiete in Süddeutschland, die über acht Jahre beobachtet wurden, dienten der Untersuchung der Fähigkeit des Modells, geringe Sedimentfrachten in Landschaften mit konservierender Bodenbewirtschaftung und strukturellen Sedimentrückhaltemaßnahmen abzubilden. Ein Generalized Likelihood Uncertainty Estimation-Framework (GLUE) wurde eingesetzt, um die Parameterunsicherheit und Modellplausibilität über verschiedene raum-zeitliche Skalen zu bewerten. Die Ergebnisse zeigen, dass WaTEM/SEDEM die ungefähre Größenordnung langfristiger Sedimentfrachten reproduzieren kann, seine Leistungsfähigkeit jedoch auf jährlicher und kleinräumiger Ebene eingeschränkt ist, insbesondere

dort, wo die Sedimentflüsse sehr gering sind und die Landschaftsstrukturen die Vernetzung stark beeinflussen. Diese Ergebnisse unterstreichen die Bedeutung einer skalenbewussten Modellbewertung und zeigen, dass die Eignung eines Modells nicht nur vom statistischen Fit, sondern auch vom geplanten Managementeinsatz abhängt.

Der dritte Teil der Dissertation behandelt die Herausforderung der automatisierten Umweltüberwachung durch Deep-Learning-basierte Bildsegmentierung. Zwei zusammenhängende Studien wurden durchgeführt: Erstens wurde Erosion-SAM entwickelt, eine feinabgestimmte Adaption des Segment Anything Models zur Abgrenzung von Wassererosions- und Sedimentablagerungsstrukturen in hochauflösenden Luftbildern; zweitens wurden feinabgestimmte SAM- und U-Net-Architekturen mit ResNet-Backbones für die schnelle Hochwasserkartierung aus UAV- und Hubschrauberbildern verglichen. Beide Studien belegen, dass Transfer Learning und Foundation Models den Bedarf an großen annotierten Datensätzen erheblich reduzieren können, ohne die Segmentierungsleistung zu beeinträchtigen. Für die Erosionskartierung identifizierte das angepasste SAM erfolgreich visuell heterogene Erosionsstrukturen über verschiedene Landnutzungstypen hinweg. Für die Hochwasserkartierung erreichten sowohl SAM- als auch U-Net-basierte Modelle hohe Genauigkeiten mit großem praktischem Potenzial für eine schnelle Lagebeurteilung nach einem Ereignis sowie für die Katastrophenhilfe. Zusammen zeigen diese Studien das Potenzial moderner Computer-Vision-Methoden für die Erstellung räumlich expliziter Umweltbeobachtungen, die sowohl wissenschaftliche Analysen als auch operative Entscheidungsprozesse unterstützen können.

Insgesamt stellt die Dissertation einen integrierten Ansatz vor, der klimatische Einflussfaktoren, prozessbasierte Erosionsmodellierung und automatisiertes bildgestütztes Monitoring miteinander verknüpft. Der übergreifende Beitrag besteht darin, zu zeigen, wie Deep Learning nicht als Ersatz für umweltwissenschaftliche Theorien, sondern als leistungsstarke Ergänzung der klassischen Bodenerosionsforschung eingesetzt werden kann. Die Ergebnisse liefern verbesserte Werkzeuge zur Abschätzung der Niederschlagserosivität, eine kritische Bewertung der Modellleistung unter Bodenschutzbedingungen und neue Methoden zur schnellen Kartierung von Erosions- und Hochwasserschäden. Damit leistet die Dissertation einen Beitrag zu einem anpassungsfähigeren, datenreichen und räumlich expliziteren Rahmen für die Bodenschutzplanung, die Bewertung von Klimawandelfolgen und das Management von Naturgefahren.

Table of Contents

List of Figures

List of Tables

List of Publications

Manuscripts in preparation

Manuscript I

Shokati, H., Scholz, F., Seufferheld, K. D., Fiener, P., Scholten, T. (2026): “From 1931 to 2043: deep learning based high resolution rainfall erosivity maps for Germany.

Published manuscripts

Manuscript II

Seufferheld, K. D., Batista, P. V., Shokati, H., Scholten, T., & Fiener, P. (2026). A GLUE-based assessment of WaTEM/SEDEM for simulating soil erosion, transport, and deposition in soil conservation optimised agricultural watersheds. Soil, 12(1), 301-319, https://doi.org/10.5194/soil-12-301-2026.

Manuscript III

Shokati, H., Engelhardt, A., Seufferheld, K., Taghizadeh-Mehrjardi, R., Fiener, P., Lensch, H. P., & Scholten, T. (2025). Erosion-SAM: Semantic segmentation of soil erosion by water. *Catena*, *254*, 108954, https://doi.org/10.1016/j.catena.2025.108954.

Manuscript IV

Shokati, H., Seufferheld, K. D., Fiener, P., & Scholten, T. (2026). Rapid flood mapping from aerial imagery using fine-tuned SAM and ResNet-backboned U-Net. *Hydrology and Earth System Sciences*, *30*(3), 743-756, https://doi.org/10.5194/hess-30-743-2026.

Abbreviations

Abbreviation	Full form
SE	Soil erosion
e.g.	For example
RADOLAN	Rain gauge-adjusted radar rainfall data
DWD	German weather service
SAM	Segment Anything Model
CWSAM	ClassWise-SAM-Adapter
SAR	Synthetic aperture radar
ViT	Vision Transformer
DiceCELoss	Combination of dice and cross-entropy losses
R^2	Coefficient of determination
MAE	Mean absolute error
RMSE	Root Mean Square Error
r	Correlation coefficient
IoU	Intersection-over-union
DFG	Deutsche Forschungsgesellschaft
UAVs	Unmanned aerial vehicles
SVM	Support Vector Machine
RF	Random Forest
MLC	Maximum Likelihood Classifier
CNN	Convolutional neural network
ReLU	Rectified Linear Unit
DRRAiNN	Distributed Rainfall-Runoff Artificial Neural Network
ResNet	Residual neural network
SAM-Bbox	SAM model with bounding box prompts
SAM-Points	SAM model with point prompts
GLUE	Generalized Likelihood Uncertainty Estimation
USLE	Universal Soil Loss Equation
RUSLE	Revised Universal Soil Loss Equation
WaTEM/SEDEM	Water and tillage erosion model and the sediment delivery model
WRB	World Reference Base for Soil Resources
DEM	Digital elevation model
TC	Transport capacity
P factor	Support practices factor
K factor	Soil erodibility
R-factor	Rainfall erosivity factor
ABAG	German adaptation of the USLE

BAdCORD-EUR11	Bias-adjusted CORDEX-EUR11 RCP4.5 projections
DIN 19708	Deutsches Institut für Normung, standard 19708
RNN	Recurrent Neural Networks
LSTM	Long Short-Term Memory
RADKLIM	Radarklimatologie
Z–R	Reflectivity–rain rate relationship
RW	Quality-controlled hourly precipitation sums
YW	5-minute precipitation rates
CDC	Climate Data Center
PWConv	Point-wise convolution
DWConv	Depth-wise convolution
MLP	Multi-Layer Perceptron
SiLU	Sigmoid Linear Unit
ConvLSTM	Convolutional LSTM
R/P ratio	Erosivity-to-precipitation ratio
RCP4.5	Moderate emission pathway RCP4.5
C	Cover management factor
k_{TC}	Transport capacity coefficient
$k_{TC/A}$	Transport capacity coefficient for arable land
$k_{TC/G}$	Transport capacity coefficient for grassland
p_{con}	Parcel connectivity parameter
e_{sur}	Error surface
PBIAS	Percent bias
SY	Sediment yield
Max.	Maximum
Min.	Minimum
X_{main}	Main inputs
X_{ext}	Extended inputs
X_{loc}	Localized inputs
X_{hyper}^{ext}	Extended hyper inputs
X_{hyper}^{loc}	Localized hyper inputs
NJ	Mean annual precipitation
R_{event}	Erosivity of a single event
E	Event's total kinetic energy
I_{max30}	Maximum rainfall intensity over any 30-min interval
W	Weight W for a given layer
φ(·)	Hypernetwork projection
A	Learnable scaling factor
LS	Slope length and steepness factor
SLR	Soil loss ratio

SLR_{month}	Average soil loss ratio of the respective month
$R_{prop,month}$	Proportion of annual rainfall erosivity for the respective month
Co_{tot}	Total soil cover
Co_{crop}	Cover of the growing crop on the respective field
Co_{res}	Measured soil cover of the residues
S_{IR}	Interrill slope gradient factor
S_m	Slope
B_{dep}	border deposition
A_i	Potential soil erosion at raster cell i
$A_{new,i}$	Potential soil erosion with uncertainty at raster cell i
L_i	Likelihood of realization i
MAE_i	Mean absolute error of realization i
Sim_i	Simulated values for behavioral runs of realization i
Obs_i	Observed sediment value for realization i
TP	True positives
FP	False positives
TN	True negatives
FN	False negatives
P_{gt10}	Number of days with precipitation greater than 10 mm
P_{gt20}	Number of days with precipitation greater than 20 mm
P_{gt30}	Number of days with precipitation greater than 30 mm
P_{gt50}	Number of days with precipitation greater than 50 mm
P_{max}	Maximum daily precipitation
P_{mean}	Mean daily precipitation
PI	Precipitation Intensity
PA	Total annual precipitation
R95P	Total rainfall from the heaviest 5% of rainy days
R_{DIN}	R-factor derived from empirical equation of DIN 19708
T_{mean}	Mean air temperature
T_{max}	Maximum air temperature
T_{min}	Minimum air temperature
log	Logarithm
tanh	Hyperbolic tangent function
σ	Sigmoid activation function
X_t	Input tensor at time t
h_{t-1}	Hidden state from previous time step
c_{t-1}	Cell state from previous time step
h_t	Current hidden state
c_t	Current cell state

W_{x*}	Convolutional kernels acting on the input
W_{h*}	Convolutional kernels acting on the hidden state
W_{c*}	Peephole convolutions from the cell state to the gates
b*	Bias terms in the ConvLSTM gates
*	Convolution operator
$\circ$	Hadamard (element-wise) product
f_t	Forget gate at time t
i_t	Input gate at time t
o_t	Output gate at time t
N	Samples in the evaluation set
E_i	Reference annual erosivity at sample i
$\hat{E}_i$	Predicted annual erosivity at sample i
g_i^c	Ground truth binary indicator of the class label c of pixel i
S_i^c	Predicted segmentation probability of class c for pixel i
n	Number of data
y_i	Observed value at sample i
$\hat{y}_i$	Predicted value at sample i
$\bar{y}$	Mean of observed values over all samples ($\bar{y}$)
L_{CE}	Cross-entropy loss
L_{Dice}	Dice loss

Unit	**Meaning**
N	Newton (unit of force)
h^{-1}	Per hour
yr^{-1}	Per year
mm	millimeters
kJ	Kilojoules
t	Tons
t ha^{-1}	Tons per hectare
t ha^{-1} yr^{-1}	Tons per hectare per year

1. Overview

1.1 Introduction

Soil erosion by water is a major threat to agricultural sustainability because it removes fertile topsoil, weakens soil structure, and reduces the long-term productivity of land (Quinton & Fiener, 2024; Scholten & Seitz, 2019; D. Yang et al., 2003). In Europe, the problem is especially relevant because erosion risk is shaped not only by land use and topography but also by rainfall erosivity, which determines how strongly precipitation can detach and transport soil particles (Panagos et al., 2015; Wischmeier & Smith, 1978). At the same time, climate change is intensifying short-duration rainfall extremes, which increases the likelihood of erosive events and makes past climatic averages a weaker guide for present and future risk (Auerswald & Fiener, 2024; Hosseinzadehtalaei et al., 2020).

This matters because soil erosion assessment still depends on data and models that are often difficult to obtain at the scale where decisions are made. Rainfall erosivity, for example, is ideally computed from high-frequency precipitation records, but such records are rare and unevenly distributed in space and time (Fiener et al., 2013; Wischmeier & Smith, 1978). Similarly, erosion and sediment transport models can simulate important landscape processes, yet their performance becomes harder to judge when sediment yields are very low, measurements are uncertain, and conservation structures alter connectivity in ways that are not easy to represent explicitly (Beven & Lane, 2022; Oost et al., 2000). In parallel, field-scale monitoring of erosion and flood impacts still relies heavily on labor-intensive mapping and manual interpretation, although these tasks are exactly the ones where timely information is most needed. Recent advances in machine learning, especially segmentation models such as the Segment Anything Model and U-Net-based architectures, now make it possible to extract environmental features from imagery with much less manual effort than before (He et al., 2016; Kirillov et al., 2023; Ronneberger et al., 2015).

This dissertation brings together these interconnected research strands—rainfall erosivity estimation, erosion and sediment transport modeling, and deep learning-based image segmentation for erosion and flood detection—into a unified body of work that advances both the scientific understanding and the practical tools available for soil erosion assessment and natural hazard management. All four studies share a common geographic focus on Germany and collectively demonstrate how modern computational techniques, ranging from spatiotemporal deep learning architectures to transfer learning with foundation models, can be leveraged to address longstanding data limitations and methodological challenges in the earth and environmental sciences.

The central premise of this dissertation is that erosion science now requires tighter integration across three domains that have too often been treated separately. First, climatic forcing must be represented with sufficient temporal detail to capture changes in rainfall intensity and therefore erosive power. Second, process representation at the landscape scale must be critically evaluated under actual conservation-oriented land management, where connectivity is intentionally reduced and sediment yields become small but highly consequential. Third, the observational bottleneck that limits both validation and operational monitoring must be alleviated through automated image analysis. By addressing these three domains jointly, the dissertation seeks not only to improve isolated components of erosion assessment but also to establish the basis for a more integrated, data-rich, and future-oriented framework for soil conservation planning.

1.2 Research Questions

The overarching objective of this dissertation is to advance the quantification, modeling, and monitoring of soil erosion and related hydrological hazards by integrating physically meaningful climatic forcing, spatially distributed erosion modeling, and modern deep learning approaches for environmental image segmentation. This objective is operationalized through four interlinked manuscripts, each addressing a different component of the broader erosion assessment chain.

The dissertation is guided by the following research questions:

- **RQ1:** How has rainfall erosivity evolved across Germany over the past century, and how accurately can deep learning architectures reconstruct and project these trends?
- **RQ2:** How effective is the WaTEM/SEDEM model in simulating erosion, sediment transport, and deposition in micro-scale watersheds optimized for soil conservation, and how does its performance depend on spatial and temporal aggregation?
- **RQ3:** What are the structural and parameter-related limitations of WaTEM/SEDEM when applied to field-dominated and structure-dominated watersheds featuring grassed waterways, retention ponds, and conservation-oriented crop management?
- **RQ4:** Can foundation models such as SAM be successfully fine-tuned for the segmentation of water erosion and deposition features in high-resolution aerial orthophotos of agricultural landscapes?
- **RQ5:** How does the performance of fine-tuned SAM compare with U-Net architectures for rapid flood mapping from UAV and helicopter imagery, and what role does prompting strategy play in segmentation quality?

Taken together, these questions move from climatic forcing to landscape process representation and finally to automated observation. They are therefore designed not as independent inquiries, but as components of a single research strategy. RQ1 is addressed primarily by Manuscript I in Chapter 3 through the development of historical reconstruction and future forecasting frameworks for rainfall erosivity in Germany. RQ2 and RQ3 are addressed by Manuscript II in Chapter 4 through the GLUE-based assessment of WaTEM/SEDEM in conservation-optimized agricultural watersheds. RQ4 and RQ5 are addressed by Manuscripts III and IV in Chapter 5 through the Erosion-SAM and flood segmentation studies, while the integration of all six questions is synthesized explicitly in Chapter 6.

1.3 Outline of the Dissertation

This dissertation is structured in three parts, preceded by this overview chapter.

Part I: Conceptual and Technical Foundations

Chapter 2 provides the conceptual, methodological, and technical foundations of the dissertation. It reviews the development of rainfall erosivity estimation, the empirical and distributed modeling of soil erosion and sediment transport, the role of soil conservation practices and landscape structures in reducing sediment connectivity, and the emergence of machine learning and foundation models in environmental remote sensing. In doing so, it establishes the common theoretical framework that connects the four studies and identifies the key research gaps they address.

Part II: Research Contributions

Chapter 3 is based on **Manuscript I**, currently being finalized for publication. The study presents the development of a comprehensive, intensity-aware rainfall erosivity dataset for Germany spanning 1931–2043. Using 5-minute RADKLIM radar precipitation data, a Hypernetwork-Enhanced CNN model for historical reconstruction, and a ConvLSTM encoder-decoder for near-term forecasting under RCP4.5 climate projections, this study produces a continuous, century-scale, 1 km × 1 km gridded R-factor product for Germany. The results reveal a pronounced long-term increase in erosivity since the 1960s and show that extreme erosivity events now exhibit markedly shorter return periods, indicating a substantial rise in their frequency.

Disclaimer 1.1 Chapter 3 is based on the under-submission journal publication with the following contributions:

Shokati, H., Scholz, F., Seufferheld, K. D., Fiener, P., & Scholten, T. (2026) From 1931 to 2043: Deep learning based high-resolution rainfall erosivity maps for Germany.

	Ideas	Analysis	Writing
H. Shokati	60 %	50 %	75 %
F. Scholz	10 %	10 %	5 %
K. D. Seufferheld	5 %	5 %	5 %
P. Fiener	5 %	15 %	5 %
T. Scholten	20 %	15 %	10 %

Chapter 4 is based on **Manuscript II**. This study evaluates the capacity of the spatially distributed erosion and sediment transport model WaTEM/SEDEM to simulate sediment yields in six highly instrumented micro-scale watersheds in Southern Germany, monitored over eight years under optimized soil conservation management. Employing a Generalized Likelihood Uncertainty Estimation (GLUE) framework, the study rigorously assesses model performance across multiple spatiotemporal scales and demonstrates that while the model captures the magnitude of very low measured sediment yields, it is fit for long-term conservation planning at larger spatial scales rather than for precise annual predictions at the micro-watershed level.

Disclaimer 1.2 Chapter 4 is based on the peer-reviewed journal publication with the following co-author contributions:

Seufferheld, K. D., Batista, P. V., **Shokati, H.,** Scholten, T., & Fiener, P. (2026). A GLUE-based assessment of WaTEM/SEDEM for simulating soil erosion, transport, and deposition in soil conservation optimised agricultural watersheds. Soil, 12(1), 301-319. https://doi.org/10.5194/soil-12-301-2026

	Ideas	Analysis	Writing
K. D. Seufferheld	60 %	60 %	60 %
P.V.G. Batista	15 %	20 %	15 %
H. Shokati	5 %	5 %	5 %
T. Scholten	5 %	5 %	5 %
P. Fiener	15 %	10 %	15 %

Chapter 5 brings together two deep learning-based image segmentation studies (Manuscripts III and IV). The first presents Erosion-SAM, a fine-tuned version of the Segment Anything Model for automatic segmentation of water erosion and deposition features in high-resolution aerial orthophotos of agricultural fields across different land cover types. The second compares fine-tuned

SAM with U-Net architectures using ResNet backbones for rapid flood area detection and mapping from UAV and helicopter imagery. Together, these studies demonstrate the versatility and effectiveness of modern deep learning architectures for automated environmental feature detection, with practical applications ranging from erosion monitoring and model validation to emergency flood response and damage assessment.

Disclaimer 1.3 First part of the Chapter 5 is based on the peer-reviewed journal publication with the following contributions:

Shokati, H., Engelhardt, A., Seufferheld, K., Taghizadeh-Mehrjardi, R., Fiener, P., Lensch, H. P., & Scholten, T. (2025). Erosion-SAM: Semantic segmentation of soil erosion by water. Catena, 254, 108954. https://doi.org/10.1016/j.catena.2025.108954

	Ideas	Analysis	Writing
H. Shokati	50 %	60 %	70 %
A. Engelhardt	10 %	5 %	5 %
K. D. Seufferheld	0 %	5 %	5 %
R. Taghizadeh-Mehrjardi	0 %	5 %	5 %
P. Fiener	20 %	10 %	5 %
H.P.A. Lensch	0 %	5 %	5 %
T. Scholten	20 %	10 %	5 %

Disclaimer 1.4 Second part of the Chapter 5 is based on the peer-reviewed journal publication with the following contributions:

Shokati, H., Seufferheld, K. D., Fiener, P., & Scholten, T. (2026). Rapid flood mapping from aerial imagery using fine-tuned SAM and ResNet-backboned U-Net. Hydrology and Earth System Sciences, 30(3), 743-756, https://doi.org/10.5194/hess-30-743-2026

	Ideas	Analysis	Writing
H. Shokati	60 %	60 %	75 %
K. D. Seufferheld	10 %	10 %	5 %
P. Fiener	10 %	10 %	5 %
T. Scholten	20 %	20 %	15 %

Part III: Conclusion and Future Directions

Chapter 6 provides an integrated synthesis of the findings from all four studies and links them back to the research questions introduced in this chapter. It discusses the broader implications of the dissertation for rainfall erosivity estimation, erosion modeling under soil conservation conditions, and automated environmental image analysis. The chapter also reflects on methodological limitations and outlines promising future research directions, including probabilistic erosivity projections, improved process representation in sediment transport models, explainable artificial intelligence, and the creation of integrated erosion assessment pipelines that combine climatic forcing, process-based modeling, and image-based monitoring.

Part I.

Conceptual and Technical Foundations

2. Conceptual and Technical Foundations

2.1 The Global Challenge of Soil Erosion by Water

Soil erosion by water is one of the most widespread and consequential forms of land degradation, threatening agricultural productivity, ecosystem services, and long-term food security across the globe (Montanarella et al., 2016; Quinton & Fiener, 2024; Rickson et al., 2015). The process encompasses the detachment of soil particles by raindrop impact and overland flow, their transport across hillslopes, and their eventual deposition in downslope positions, waterways, or reservoirs (Wischmeier & Smith, 1978). Global estimates suggest that average soil erosion rates on agricultural land substantially exceed natural rates of soil formation, leading to an irreversible net loss of soil resources on human timescales (Feng et al., 2010). Beyond on-site productivity losses, eroded sediment generates severe off-site impacts including reservoir siltation, eutrophication of surface waters, and degradation of aquatic habitats (Pimentel et al., 1995).

The drivers of water erosion are well understood in principle. Rainfall provides the erosive energy that initiates detachment and transport, while topography governs the acceleration and accumulation of overland flow (Desmet & Govers, 1996). Soil properties—particularly texture, organic matter content, and aggregate stability—determine how readily particles are detached, a characteristic captured in erosion models through the soil erodibility factor (Auerswald et al., 2014). Land use and vegetation cover regulate the degree to which the soil surface is shielded from raindrop impact and flowing water (Schwertmann et al., 1987). In recent decades, these drivers have been shifting in ways that intensify erosion risk: agricultural intensification, including field enlargement and soil compaction, has reduced landscape roughness and increased hydrological connectivity (Brus & Akker, 2018; Foucher et al., 2014; Keller et al., 2019), while anthropogenic climate change is increasing the frequency and intensity of extreme precipitation events across large parts of Europe and globally (Auerswald & Fiener, 2024; Hosseinzadehtalaei et al., 2020; Myhre et al., 2019). These converging pressures make the accurate quantification of erosive forcing and the rigorous evaluation of conservation strategies more urgent than ever.

At the same time, soil erosion is not merely a biophysical process but also a management challenge embedded in agricultural systems, land tenure, policy, and monitoring capacity. In many regions, erosion remains under-managed not because its consequences are unknown, but because the design of effective interventions depends on reliable information about where erosion occurs, how intense it is, and which measures are most likely to reduce it under changing climatic conditions. This creates a strong need for methods that can bridge scales—from event-scale rainfall intensity to long-term regional risk patterns, and from field-level erosion signatures to landscape-scale sediment delivery.

2.2 Rainfall Erosivity: From Concept to Continental Mapping

Rainfall erosivity, expressed through the R-factor in the Universal Soil Loss Equation (USLE) and its revised forms (Renard, 1997; Wischmeier & Smith, 1978), quantifies the capacity of precipitation to detach and transport soil particles. It is defined as the long-term average annual sum of the product of event kinetic energy and maximum 30-minute intensity (EI30) over all erosive events (Fiener et al., 2013). This formulation captures both the energy available for particle detachment and the potential for surface runoff generation, making the R-factor the most physically meaningful climatic input to erosion modeling (Panagos et al., 2015). A comprehensive review by Yin et al. (2017) summarized the state of the art in R-factor estimation methods, mapping approaches, and observed temporal trends, noting that the original RUSLE formulation underestimates R-factor values by approximately 10% relative to updated calculation procedures.

Direct computation of the R-factor requires precipitation data at high temporal resolution, ideally 5 to 30 minutes, over extended periods (Wischmeier & Smith, 1978). However, such high-frequency records have historically been scarce, restricting reliable erosivity estimates to limited station networks and short time periods (Borrelli et al., 2016; Diodato et al., 2017; Fiener et al., 2013). To address this gap, empirical regressions relating the R-factor to aggregated precipitation statistics—daily, monthly, or annual totals—have been widely employed (Beguería et al., 2018; Yin et al., 2007). In Germany, the national standard DIN 19708 (Deutsches Institut für Normung, 2005) provides a linear equation estimating mean annual erosivity from mean annual precipitation. However, this regression was calibrated in the early 1990s using data reflecting climatic conditions from the 1960s through the 1980s, and the latest revision of the standard (DIN 19708:2022-08) explicitly advises limiting its use to historical datasets (Uber et al., 2024).

At the European scale, a landmark effort by Panagos et al. (2015) assembled the Rainfall Erosivity Database at European Scale (REDES), comprising 1,541 precipitation stations with temporal resolutions of 5 to 60 minutes, and produced the first continent-wide R-factor map at 1 km resolution using Gaussian Process Regression. This was followed by monthly erosivity maps (Ballabio et al., 2017; Panagos et al., 2016) that revealed pronounced seasonality, with summer erosivity in central Europe being nearly four times higher than in winter. Building on REDES, Panagos et al. (2017) projected future rainfall erosivity under climate change scenarios, estimating an 18% increase in mean European erosivity by 2050 relative to the 2010 baseline, with highly heterogeneous spatial patterns. Bezak et al. (2020) extended this work by reconstructing past rainfall erosivity across Europe for 1961–2018 and documented a tendency toward increasing erosivity trends, with 15% of stations showing statistically significant positive trends versus only 7%

with negative trends. More recently, Matthews et al. (2025) evaluated the European ground-radar climatology EURADCLIM for continental erosivity assessment, demonstrating both the promise and the remaining challenges of radar-based erosivity products at pan-European scales.

In Germany specifically, long-term gauge records have documented significant increases in rainfall erosivity since the mid-20th century. Fiener et al. (2013) reported a 4.4% per-decade acceleration in summer erosivity for Western Germany between 1937 and 2007, while Auerswald et al. (2019) confirmed that German rainfall erosivity increased substantially between 1960 and 2017, attributing this trend explicitly to climate change rather than methodological artifacts. These findings motivate the development of new computational frameworks that can reconstruct historical erosivity from available climate archives and project its evolution under future climate scenarios.

A further challenge lies in the fact that rainfall erosivity is not simply a function of annual rainfall volume. Two years with similar annual precipitation totals may differ greatly in erosivity if one year is dominated by frequent low-intensity rainfall and the other by a smaller number of short, high-intensity storms. This distinction is particularly important under climate change, where shifts in rainfall intensity distributions may outpace changes in annual totals. Consequently, any method that attempts to estimate the R-factor from coarse precipitation aggregates alone risks missing the very dimension of rainfall most relevant for erosion. This conceptual limitation provides a major rationale for the use of deep learning in Chapter 3, where the objective is not merely to interpolate rainfall patterns but to infer sub-daily intensity-related erosive signatures from broader climatic information.

2.3 Soil Erosion Modeling: Empirical Basis and Spatially Distributed Approaches

For a better understanding of transport processes and the links between land use and soil erosion, erosion models have been used for decades (Borrelli et al., 2021). In principle, conceptual models such as SWAT (Arnold & Fohrer, 2005) and WaTEM/SEDEM (Rompaey et al., 2001), and physically based erosion models such as WEPP (Laflen et al., 1991), EUROSEM (Morgan, 1996), and EROSION-3D (Schmidt et al., 1999) can be used to analyze both the redistribution of soil across landscapes and the sediment input into water bodies.

The physically-based models are particularly important for detailed process studies. Hydrological processes and erosion processes are mapped together, usually with very short time steps (s-min) and high spatial resolution (grid cells with 1 - 10 m edge length) (Baartman et al., 2020; Wilken et al., 2017). There are mixed forms of the conceptual models (with and without parallel modeling of the hydrology), whereby the modeling usually has a lower spatial (> 10 m) and temporal resolution (1 day - annual mean). Due to the large computational effort and the lack of high-resolution input data, the physically-based models are used almost exclusively in small "research catchment areas" (< 1 km²) for which dynamic measurements are available (Fiener et al., 2019a). In contrast, conceptual models are also used in larger catchment areas or regions (Nadeu et al., 2015; Verstraeten et al., 2006) (Figure 2.1), which is mainly due to the significantly lower data requirements of these models. To estimate the potential erosion of a unit area (e.g. grid cell), the Universal Soil Loss Equation (USLE, Wischmeier & Smith (1978)) and its further developments are usually used and combined with approaches of sediment transport. The empirical USLE expresses mean annual soil loss as the product of six factors—rainfall erosivity (R), soil erodibility (K), slope length (L), slope steepness (S), cover management (C), and support practices (P)—and remains the most widely used erosion model globally (Borrelli et al., 2021; Karydas et al., 2014; Wischmeier & Smith, 1978). A major advantage of USLE-type approaches is that the required data are often available or can be estimated from standard environmental datasets, which has enabled broad application in national and regional erosion assessments (Fiener et al., 2020). The RUSLE (Renard, 1997) and regional adaptations such as the German ABAG (Schwertmann et al., 1987) have been developed to improve local applicability while preserving the model's conceptual simplicity and modest data requirements.

A comprehensive review by Benavidez et al. (2018) analyzed the strengths and limitations of the USLE family of models across different geoclimatic regions, identifying uncertainty in factor estimation, inability to account for gully erosion, and difficulties in predicting sediment delivery as key limitations. USLE-type models estimate only potential soil loss at the point of detachment and do not account for sediment transport and deposition processes that govern the actual delivery of eroded material to waterways (Oost et al., 2000). At the landscape scale, sediment is intercepted and re-deposited along hillslopes, at field boundaries, within vegetated buffers, and in engineered structures, so that only a fraction of the eroded material reaches the watershed outlet.

To represent these processes, the Water and Tillage Erosion Model and Sediment Delivery Model (WaTEM/SEDEM) was developed by combining USLE-based erosion calculation with spatially distributed sediment routing using a transport capacity concept and a multiple flow routing algorithm (Oost et al., 2000; Rompaey et al., 2001; Verstraeten et al., 2002). Unlike more complex process-based models such as WEPP or LISEM that require extensive parameterization and computational resources, WaTEM/SEDEM requires input data comparable to the USLE while additionally modeling sediment redistribution and deposition patterns across the landscape. The model has been tested against reservoir sedimentation data (Hlavčová et al., 2018), mesoscale catchment sediment yields (Batista et al., 2022; Rehm & Fiener, 2024), and radionuclide-derived erosion patterns (Oost et al., 2000; Wilken et al., 2020).

Despite its wide application, important challenges persist. Erosion models typically exhibit systematic biases, overpredicting low sediment yields and underpredicting high yields (Kinnell, 2007; Nearing, 1998; Risse et al., 1993). The evaluation of model performance is itself complicated by measurement uncertainties in observational data, particularly at low erosion rates where natural variability is disproportionately large relative to the signal (Nearing, 2000). Formal uncertainty frameworks such as the Generalized Likelihood Uncertainty Estimation (GLUE) approach (Beven & Binley, 1992) provide a principled way to account for these uncertainties by acknowledging that multiple parameter sets may produce equally acceptable model realizations within the bounds of observational error (Beven, 2006; Beven & Lane, 2022). Critically, the capacity of WaTEM/SEDEM to represent erosion, transport, and deposition within soil conservation settings—where both in-field practices and landscape structures drastically reduce sediment connectivity—has received limited attention in the literature, despite the practical importance of such assessments for conservation planning.

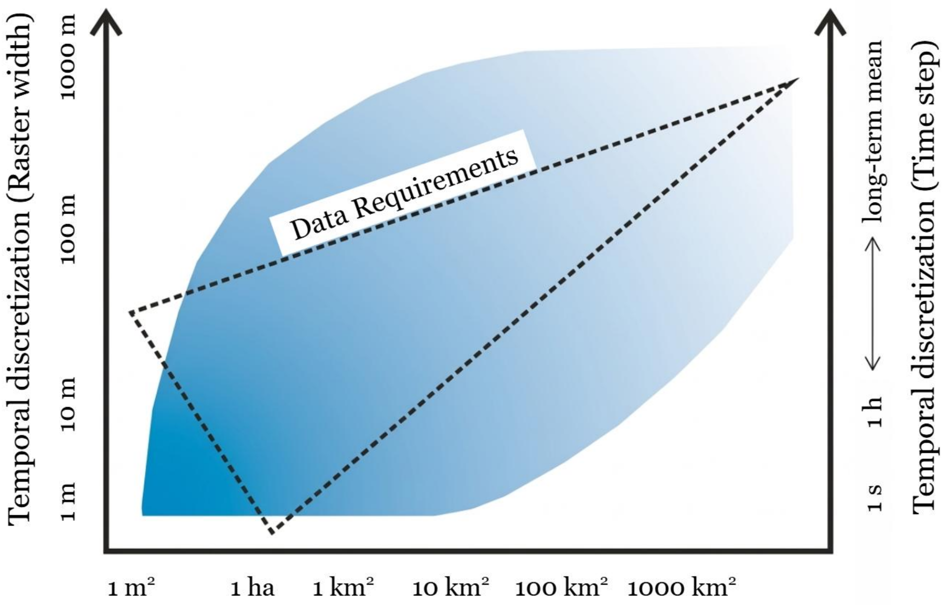


Figure 2.1. Schematic representation of the spatial and temporal discretization of different erosion models and their application in catchment areas of different sizes; Data basis (J., 2012).

2.4 Soil Conservation: In-Field Practices and Landscape Structures

Effective soil conservation combines two complementary strategies: reducing soil detachment through in-field practices and intercepting eroded sediment through landscape-scale transport control structures (Andersson & D'Souza, 2014; Oost et al., 2000). In-field practices aim to maximize soil surface protection by maintaining continuous vegetation or residue cover. Key measures include no-till or reduced tillage, which preserves crop residues on the surface, optimized crop rotations that minimize the duration of bare-soil exposure, and the use of cover crops during fallow periods (Fiener et al., 2019a; Schwertmann et al., 1987). In erosion models, the protective effect of vegetation is captured through the cover management factor (C factor), which integrates the temporal dynamics of soil cover weighted by the seasonal distribution of rainfall erosivity (Wischmeier & Smith, 1978).

At the landscape scale, sediment transport control structures intercept and trap eroded material along runoff pathways. Grassed waterways along thalwegs are among the most effective of these structures. Fiener & Auerswald (2003) demonstrated in a long-term experiment in Southern Germany that grassed waterways reduced sediment delivery by 77–97% and runoff by 10–90%, depending on their morphological design, with flat-bottomed waterways outperforming V-shaped cross sections due to greater infiltration areas. Event-based modeling by Abaci et al. (2010) confirmed these findings, showing that grassed waterways reduced sediment yield by 3.5–59% during individual storm events, with higher efficiency for smaller events and lower-gradient contributing slopes. Retention ponds at watershed outlets provide additional sediment trapping, with measured efficiencies averaging 70 ± 14% in micro-scale watersheds in Bavaria (Fiener et al., 2005).

The combined implementation of in-field conservation and landscape structures can reduce watershed sediment yields dramatically. At the Scheyern experimental farm in Southern Germany, mean yields of only 0.16 t ha^{-1} yr^{-1} were measured under optimized conservation management, compared to typical rates of 3–10 t ha^{-1} yr^{-1} in the same region under conventional practices (Auerswald et al., 2009; Fiener et al., 2019a). Garbrecht & Starks (2009) documented analogous long-term sediment yield reductions of approximately 86% over a 60-year period of progressive conservation implementation in the Fort Cobb Reservoir watershed in Oklahoma. Despite these documented successes, the representation of soil conservation effects within modeling frameworks remains challenging, as conservation practices reduce sediment yields to very low levels where the signal-to-noise ratio in measurements is unfavorable and model biases are amplified.

These issues are especially important for this dissertation because one of its core arguments is that model adequacy depends strongly on the system state to which the model is applied. A model that performs acceptably in conventionally managed watersheds with high sediment export may fail under conservation-optimized conditions precisely because the remaining erosion signal is small, episodic, and strongly shaped by trapping structures. This does not necessarily invalidate the model in general; rather, it raises the critical fit-for-purpose question of whether a model can still support the type of management decision for which it is intended.

2.5 Machine Learning for Spatiotemporal Environmental Prediction

Machine learning methods open up a new way of solving the problem of integrating land use patterns and structures into soil erosion modeling (LeCun et al., 2015). While established erosion models allow a valid estimation of soil erosion in the area, ML models offer the possibility to extract small-scale patterns and new insights from remote sensing and geodata (Reichstein et al., 2019). Machine learning approaches have been established for several years in the field of soil sciences (Behrens & Scholten, 2006; MCBRATNEY, 2003; Shokati et al., 2023, 2024, 2025) and remote sensing

(Lamprecht et al., 2017; Yuan et al., 2020) and are increasingly being used in environmental and geosciences (Ebrahimi-Khusfi et al., 2021; Lary et al., 2016). Initial attempts with Machine learning in the field of soil erosion research have so far been limited to data-based regression approaches to predict individual erosion forms or factors. The occurrence of erosion phenomena is learned based on relationships between known control and target variables. For example, the occurrence of gully erosion in India (Gayen et al., 2019) and in Iran (e.g. W. Chen et al. (2021)) could be predicted with a high degree of accuracy on the basis of topographical properties and distance measures, as could surface and gully erosion in Taiwan (Nguyen et al., 2021), the K-factor for the RUSLE modeling in the Northern Great Plains of the USA (Taghizadeh-Mehrjardi et al., 2019) and wind erosion rates in Iran (Mina et al., 2022).

The rapid maturation of deep learning over the past decade has created transformative opportunities for environmental modeling and monitoring. Convolutional Neural Networks (CNNs) have demonstrated exceptional capabilities for spatial pattern recognition (LeCun et al., 2015), while Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have become standard tools for temporal sequence modeling due to their ability to capture long-range dependencies through gating mechanisms (Hochreiter & Schmidhuber, 1997). For geoscientific applications involving gridded spatiotemporal data—such as the prediction of rainfall erosivity fields across space and time—neither architecture alone is sufficient: CNNs lack temporal memory, while LSTMs ignore spatial structure.

The Convolutional LSTM (ConvLSTM), proposed by Shi et al. (2015), bridges this gap by replacing the fully connected transformations in LSTM gates with convolution operations, enabling simultaneous extraction of spatial features and retention of temporal state across gridded sequences. ConvLSTM has been successfully applied to precipitation nowcasting (W. Liu et al., 2022) and other spatiotemporal forecasting tasks, consistently outperforming standalone CNN and LSTM approaches (Hong et al., 2017; Kim et al., 2017). For spatial reconstruction tasks requiring adaptation to local environmental conditions—such as mapping rainfall erosivity from coarse-resolution predictors across climatically heterogeneous regions—hypernetwork-enhanced architectures represent a further advance. Unlike standard CNNs that apply fixed learned weights across all locations, hypernetwork-enhanced models dynamically generate convolutional weights conditioned on auxiliary input variables, allowing the network to adjust its processing logic to local regimes of rainfall intensity and climate (Scholz et al., 2025).

These architectural innovations are particularly relevant for the challenge addressed in Chapter 3. The task there is not a conventional short-term weather forecast, but the reconstruction and forecasting of a derived climatic hazard variable—rainfall erosivity—across a long historical and near-future time horizon. This requires models that can infer hidden structure from heterogeneous climatic inputs while preserving spatial consistency across Germany's diverse climatic regions. In this sense, deep learning is not deployed as a black-box replacement for erosion science, but as a tool for learning complex mappings where direct observations are incomplete and where conventional empirical formulations are increasingly inadequate.

2.6 Foundation Models and Transfer Learning for Environmental Image Segmentation

A persistent bottleneck in applying deep learning to environmental monitoring is the scarcity of labeled training data. Producing pixel-level annotations for tasks such as soil erosion feature mapping or flood extent delineation is labor-intensive, subjective, and difficult to scale (Safonova et al., 2023). Transfer learning—adapting models pre-trained on large general-purpose datasets to domain-specific tasks—offers a powerful solution by leveraging learned feature representations that generalize across domains (Rawat & Wang, 2017). Pre-trained networks such as ResNet (He et al., 2016), originally trained on the ImageNet dataset (Krizhevsky et al., 2012), have been successfully transferred to remote sensing tasks including plant mapping (Onojeghuo et al., 2023) and water body segmentation (Ghaznavi et al., 2024).

The U-Net architecture (Ronneberger et al., 2015), with its encoder-decoder structure and skip connections, has become a standard framework for pixel-level segmentation in remote sensing. When combined with pre-trained backbones such as ResNet-50 or ResNet-101, U-Net achieves strong segmentation performance even with limited training data, as the encoder provides robust low-level feature representations while the decoder is adapted to the target task. For flood mapping from aerial imagery, U-Net and related architectures have shown promising performance in capturing complex water boundaries from UAV data (Gebrehiwot et al., 2019; Kumar et al., 2025; Munawar et al., 2021; Sinha et al., 2024).

The emergence of foundation models has further expanded the transfer learning paradigm. The Segment Anything Model (SAM), developed by Meta AI and pre-trained on over one billion masks from eleven million images, provides a general-purpose segmentation capability with strong zero-shot generalization (Kirillov et al., 2023). SAM's three-component architecture—a Vision Transformer-based image encoder, a prompt encoder processing spatial cues such as points or bounding boxes, and a lightweight mask decoder—enables flexible, prompt-guided segmentation. Osco et al. (2023) provided an early assessment of SAM for remote sensing applications, demonstrating both its promise and its limitations with lower spatial resolution imagery. Subsequent studies have proposed various fine-tuning strategies to bridge the domain gap between natural images and remote sensing data, including adapter modules, multiscale feature enhancement, and parameter-efficient fine-tuning (Sahay & Savakis, 2024; Zhou et al., 2024).

For environmental applications, fine-tuning only the lightweight mask decoder while keeping the heavyweight image encoder frozen has proven particularly effective, enabling efficient adaptation with minimal computational overhead (Sun et al., 2024). These advances motivate the application of fine-tuned SAM to two challenging segmentation tasks addressed in this dissertation: the detection of soil erosion features in high-resolution aerial orthophotos and the rapid mapping of flood-affected areas from UAV and helicopter imagery. In both cases, the central question is not simply whether a powerful vision model can segment a class of objects, but whether it can do so under the specific visual complexity of environmental scenes, where boundaries may be diffuse, class definitions may be ambiguous, and labeled datasets are small.

2.7 Floods: A Growing Natural Hazard and the Need for Rapid Mapping

Floods rank among the most destructive and costly natural hazards worldwide (Smith et al., 2014; Tingsanchali, 2012), causing the greatest number of fatalities among all natural disaster types (Kuenzer et al., 2013). Global warming is expected to increase both the frequency and intensity of flood events, amplifying their devastating impacts on communities, infrastructure, and agricultural landscapes (Kamilaris & Prenafeta-Boldú, 2018). Recent catastrophic floods in Germany (Lehmkuhl et al., 2022), Spain (Ezzatvar & López-Gil, 2024), Australia (Kelly et al., 2023), and Japan (Lin et al., 2020) underscore that effective flood management is a universal imperative.

Remote sensing plays a critical role in the response and recovery phases of flood management by providing spatial information for damage assessment, resource allocation, and insurance claims (De Leeuw et al., 2014). While satellite imagery from platforms such as Sentinel-2 and Landsat has been widely used for flood detection(Portalés-Julià et al., 2023), revisit time limitations and cloud cover often prevent timely data acquisition. UAVs and helicopters offer a flexible alternative, capturing high-resolution imagery at lower altitudes with less susceptibility to cloud interference (Hashemi-Beni et al., 2018; Klemas, 2015). However, the operational value of such imagery depends on rapid processing, which is difficult to achieve through manual delineation alone.

The application of deep learning to flood extent mapping has advanced rapidly. Gebrehiwot et al. (2019) demonstrated that fully convolutional networks fine-tuned on UAV imagery achieved water classification accuracy of 97.5%, substantially outperforming traditional support vector machine classifiers. Munawar et al. (2021) developed an automated flood detection system using UAV imagery and deep learning, achieving 91% overall accuracy in distinguishing flooded from non-flooded regions. More recently, U-Net and its variants have been widely applied to flood segmentation from aerial imagery (Kumar et al., 2025; Sinha et al., 2024), demonstrating strong performance in capturing complex flood boundaries. However, many of these models still require large labeled datasets for effective training, which are often unavailable in emergency situations.

Transfer learning with foundation models such as SAM and encoder-decoder architectures with pre-trained backbones therefore provides a promising approach to achieving high-performance flood segmentation from the small aerial image datasets that can realistically be collected during or immediately after flood events.

2.8 Motivation and Connecting Framework

The research strands reviewed above—rainfall erosivity estimation, erosion and sediment transport modeling, and deep learning-based image segmentation—have traditionally been pursued as separate disciplines. Yet they are deeply interconnected. Rainfall erosivity is the primary climatic input to erosion models; erosion models require validation data that can increasingly be generated through automated image analysis; and the flood events mapped by deep learning systems are themselves driven by the extreme precipitation processes that also control erosivity trends.

This dissertation brings these strands together into a unified body of work, sharing a common geographic focus on Germany and a common methodological thread in the application of advanced computational techniques to overcome longstanding data limitations. The first study addresses the climatic driver of erosion by developing a deep learning framework that combines high-resolution radar observations (RADKLIM), historical climate archives, and regional climate projections to produce a century-scale, intensity-aware, 1 km gridded rainfall erosivity product for Germany. The second study evaluates the fitness for purpose of the established WaTEM/SEDEM erosion and sediment transport model for representing the combined effects of in-field conservation practices and landscape-scale sediment control structures, using a unique long-term monitoring dataset from six micro-scale watersheds in Southern Germany within a rigorous GLUE uncertainty framework. The third and fourth studies extend the methodological spectrum to image-based environmental monitoring, using transfer learning and foundation model adaptation to automate the segmentation of erosion features and flood extents from high-resolution aerial imagery.

The connecting logic across these studies can be understood as a chain of climatic forcing, process representation, and observational feedback. Updated erosivity maps provide more realistic climatic input to erosion models than outdated empirical standards. Erosion models, in turn, provide a spatially distributed process interpretation that can be tested, calibrated, or challenged through image-derived evidence of erosion patterns. Deep learning-based segmentation therefore serves not merely as a standalone application of artificial intelligence, but as a mechanism for improving the empirical basis of erosion science. In this sense, the dissertation argues for a more integrated environmental modeling framework in which climate data, landscape process models, and remote sensing analytics are mutually informative rather than methodologically isolated.

Part II.

Research Contributions

3. Manuscript I: Deep learning-based rainfall erosivity estimation

Chapter 3 presents work from **Manuscript I**, currently being finalized for publication:

Shokati, H., Scholz, F., Seufferheld, K. D., Fiener, P., Scholten, T. (2026): "From 1931 to 2043: deep learning-based high-resolution rainfall erosivity maps for Germany.

This study addresses the climatic driver of soil erosion by quantifying how rainfall erosivity has evolved across Germany from 1931 to 2043 and by demonstrating how deep learning can overcome longstanding data limitations in erosivity estimation. Building on the conceptual foundations in Chapter 2, the study combines 5-minute RADKLIM radar data, long-term climate archives and regional climate projections to construct an intensity-aware, 1 km rainfall erosivity product that is consistent across space and time. In doing so, it directly responds to Research Question 1 on the reconstruction and projection of rainfall erosivity trends using modern deep learning architectures.

Methodologically, the study develops a Hypernetwork-Enhanced CNN for historical reconstruction of annual erosivity fields from coarse climatic predictors and a ConvLSTM encoder–decoder for near-term forecasting forced with bias-adjusted CORDEX-EUR11 RCP4.5 projections. Together, these models produce a continuous erosivity series that reveals both the magnitude of long-term changes and shifts in the frequency of extreme erosivity years, thereby providing improved climatic input for erosion modeling and conservation planning in Germany.

The remainder of the chapter is organized as follows. Section 3.1 introduces the scientific background, outlines the specific objectives and reviews related work on rainfall erosivity estimation. Section 3.2 describes the data basis, the computation of the radar-derived R-factor and the design and training of the Hypernetwork-Enhanced CNN and ConvLSTM models. Section 3.3 presents and discusses the reconstruction and forecasting results, including spatial and temporal patterns of erosivity and model performance across contrasting regimes, while Section 3.4 synthesizes the main conclusions.

Manuscript I: From 1931 to 2043: deep learning-based high-resolution rainfall erosivity maps for Germany

Abstract

Rainfall erosivity (R-factor) is a key driver of soil erosion yet remains poorly quantified in space and time because direct computation of the R-factor requires long, high-frequency precipitation records that are rarely available beyond recent decades. Leveraging the national 5-minute RADKLIM radar archive for Germany, this study derives intensity-aware annual rainfall erosivity maps for 2001–2025 on a 1 km × 1 km grid. To extend erosivity estimation beyond the radar era, a Hypernetwork-Enhanced Convolutional Neural Network (CNN) model is developed that learns spatially varying, nonlinear mappings between coarse-resolution climatic predictors (daily precipitation indices, annual precipitation and temperature, and DIN-based R-factor) and the RADKLIM-derived R-factor. Building on this reconstruction, a ConvLSTM encoder–decoder is trained on annual maps of erosivity, precipitation, and temperature and forced with bias-adjusted CORDEX-EUR11 RCP4.5 projections (BAdCORD-EUR11) to forecast annual erosivity for 2026–2043. The resulting framework produced a continuous, intensity-aware 1-km rainfall erosivity dataset for Germany spanning 1931–2043. It shows mean annual R-factor values rising from 117 N h^{-1} yr^{-1} (1931–2004) to 147 N h^{-1} yr^{-1} (2005–2043), representing an increase of about 26%. Extreme years exceeding 150 N h^{-1} yr^{-1} shifted from a 24.7 return period (1931–2004; 3 events) to a 2.4-year return period in recent observations and future projections (2005–2043; 16 events). The ConvLSTM forecasts indicated that the elevated erosivity regime observed in recent decades is likely to persist over the coming years rather than reverting to the low-erosivity conditions of the mid-20th century, implying sustained or rising erosion risk for German agricultural landscapes. The proposed methodology offers a transferable template for combining high-resolution observations, long-term climate archives, and regional climate projections into a unified, intensity-aware erosivity product that can support climate-resilient soil conservation planning and the revision of national standards.

3.1. Introduction

Soil erosion by water is widely recognized as one of the principal constraints on sustainable and productive agriculture at the global scale, because it strips away nutrient-rich topsoil and accelerates long-term soil degradation (Feng et al., 2010; Quinton & Fiener, 2024; Scholten & Seitz, 2019; Shokati et al., 2025; D. Yang et al., 2003). Consequently, reliable spatial and temporal estimates of soil erosion are essential for effective land-use planning, targeted soil conservation, and the reduction of erosion-related risks in agricultural systems (Senanayake et al., 2022). In Europe, soil loss potential is most commonly assessed using the Revised Universal Soil Loss Equation (RUSLE) (Panagos et al., 2014). RUSLE represents soil loss as a function of six factors: rainfall erosivity, soil erodibility, slope length, slope steepness, cover-management, and support or conservation practices

(Renard, 1997). Among these, rainfall erosivity (R-factor) plays a key role, because precipitation provides the driving energy for soil particle detachment, aggregate breakdown, and subsequent transport by runoff (Panagos et al., 2015). At the annual scale, rainfall erosivity is typically derived by summing the erosivity of individual rainfall events, thereby serving as a critical proxy for the combined effects of high rainfall intensities and event totals (Fiener et al., 2013; Renard, 1997).

The formal definition of rainfall erosivity presupposes precipitation records at high temporal resolution (Wischmeier & Smith, 1978). However, the scarcity of such data across large scales has historically restricted direct computation of the R-factor to short monitoring periods, localized study sites, or extrapolations from sparse gauge networks (Angulo-Martínez et al., 2009; Borrelli et al., 2016; Diodato et al., 2017; Fiener et al., 2013; Janeček et al., 2013; Panagos et al., 2015, 2016; Petan et al., 2010). To address this gap, numerous investigations resort to empirical regressions that relate the R-factor to aggregated precipitation statistics (daily, monthly, or annual totals), thereby enabling extended retrospective analyses (Beguería et al., 2018; Yin et al., 2007).

Soil erosion across Europe is expected to intensify due to the increasing frequency and intensity of extreme rainfall events (Deelstra et al., 2011; Pruski & Nearing, 2002). Contemporary records based on point/gauge measurements already demonstrate marked strengthening of rainfall erosivity; for example, Belgium's data revealed a 31% rise over the 1991–2002 period relative to historical baselines (Verstraeten et al., 2006), and Western Germany documented a 4.4% per-decade acceleration in summer erosivity between 1937 and 2007 (Fiener et al., 2013). These gauge-based studies are complemented by spatially contiguous radar climatologies that capture the fine-scale heterogeneity of erosive storms and indicate substantially higher long-term erosivity levels than earlier gauge-based maps (e.g. Auerswald et al., 2019). Together, these lines of evidence encompass (i) analyses of point/gauge measurements and (ii) analyses of spatial radar data, providing a consistent picture of intensifying rainfall erosivity under a changing climate. Understanding historical rainfall erosivity variations is thus critical for establishing baseline conditions, estimating recurrence intervals, and tracking temporal evolution, particularly regarding climate change scenarios (Angulo-Martínez & Beguería, 2012; S. Chen & Zha, 2018; Fiener et al., 2013; Hanel et al., 2016; Hoomehr et al., 2016; Lai et al., 2016; Panagos et al., 2017; Qin et al., 2016; Y. Wang et al., 2017). Constructing long-term, high-resolution datasets of rainfall erosivity not only deepens knowledge of its spatial and temporal dynamics but also provides a reliable foundation for robust predictive models aimed at mitigating future soil erosion risks.

Machine learning methods have gained attention as a means to enhance the accuracy of erosivity predictions derived from temporally aggregated precipitation records. Lee et al. (2021), for example, constructed and evaluated seven distinct algorithms— eXtreme and Deep Neural Network, Gradient Boost, Random Forest, K-Nearest Neighbors, Gradient Boosting, Decision Tree, Multilayer Perceptron—to forecast monthly R-factors from daily and hourly precipitation maxima together with monthly precipitation totals. Nevertheless, such models tend to be calibrated for specific regions, underscoring the necessity for broader validation across varied climatic and geographic settings and for the adoption of more advanced modeling frameworks. In a related development, Scholz et al., (2025) introduced a hypernetwork-enhanced CNN model for rainfall–runoff modeling and intensity-aware mapping tasks, which dynamically adapts its behavior to local environmental conditions.

Recent progress in deep learning has spawned diverse architectures for time series analysis, with continual improvements in their effectiveness (Kathuria, 2018). When addressing prediction tasks based on sequential information, Recurrent Neural Networks (RNN), particularly Long Short-Term Memory (LSTM) units (Hochreiter & Schmidhuber, 1997), are commonly employed due to their capacity to retain information from preceding temporal steps (Moskolaï et al., 2020). LSTM networks demonstrate considerable strength in capturing nonlinear patterns within time series and have shown success in rainfall forecasting applications (Khosravi et al., 2023; Senanayake et al., 2022). However, their application to imagery is hindered by a pixel-wise processing approach that ignores the broader spatial context necessary for accurate image analysis (Arslan & Sekertekin, 2019). To address this limitation on the spatial side, Convolutional Neural Networks (CNNs) are frequently utilized to extract rich spatial features from gridded or image-like data, as demonstrated for example in crop-yield estimation from UAV imagery (Q. Yang et al., 2019). For truly spatiotemporal problems, however, a more robust solution was proposed by Shi et al., (2015), who hybridized these architectures to create the Convolutional LSTM (ConvLSTM). By embedding convolutional structures directly within the recurrent transitions, the ConvLSTM facilitates simultaneous spatial feature extraction and temporal state retention (W. Liu et al., 2022), resulting in superior predictive accuracy compared to either component used independently (Hong et al., 2017; Kim et al., 2017).

Despite substantial advancements in soil erosion studies, critical gaps remain in how rainfall erosivity is quantified over space and time. Most existing studies still rely on point-based station records or relatively short analysis periods, and therefore do not provide the long-term, spatially continuous (gridded) information needed to support regional land-use planning and erosion control. To address this limitation, this study exploits a national, high-resolution precipitation archive for Germany to generate a century-scale, 1-km gridded reconstruction of annual rainfall erosivity (1931–2043) that explicitly accounts for sub-hourly rainfall intensity. We compute annual erosivity directly from 5-minute gridded precipitation (RADKLIM) for the period 2001–2025 and use those computations as the reference target. As comparable sub-hourly raster data are unavailable before 2001, we construct a deep-learning transfer framework that links these reference erosivity fields to predictors available over much longer periods, including annual and daily precipitation statistics, derived indices, and temperature. For reconstructing historical erosivity, we apply a Hypernetwork-Enhanced CNN model that learns spatially varying relationships between these predictors and the intensity-dependent erosivity field, capitalizing on its ability to adapt to local environmental conditions. On the basis of the resulting 1931–2025 erosivity ensemble, we then analyze temporal variability and long-term trends and, in a final step, generate a short-term forecast of annual erosivity for 2026–2043 using a ConvLSTM architecture that explicitly represents spatiotemporal dependencies and is driven by precipitation and temperature projections from regional climate change scenarios. This study therefore addresses three main questions: (1) How did rainfall erosivity vary across Germany over the historical period 1931–2025, and what are its long-term trends? (2) To what extent can sub-hourly rainfall intensity information be inferred from daily and annual precipitation statistics for long-term erosivity reconstruction? (3) How is annual erosivity projected to evolve over Germany under mid-century climate-change scenarios up to 2043? By delivering a continuous, century-scale (1931–2025) 1-km erosivity product for Germany together with a 2026–2043 outlook, the study seeks to enhance both retrospective and forward-looking assessments of erosion risk and to offer a reproducible template for other regions facing analogous data limitations.

3.2. Materials and methods

3.2.1 Data basis

3.2.1.1 Radar precipitation data

The German Weather Service (Deutscher Wetterdienst, DWD) operates a nationwide network of 17 C-band weather radars (Appendix, Figure 3.S1) that together provide quantitative precipitation estimates on a 1 km × 1 km grid over Germany (Bartels et al., 2004). The radars perform volumetric scans of the atmosphere at frequent intervals (5 min), from which reflectivity is retrieved and converted to precipitation rates using an empirically derived reflectivity–rain rate (Z–R) relationship. These individual radar images are composited into a country-wide mosaic (approximately 1100 km in the north–south and 900 km in the west–east direction in the historical configuration), providing spatially continuous fields of precipitation intensity. To obtain best-estimate precipitation, the radar-derived rain rates are accumulated to hourly totals and adjusted against measurements from a dense network of more than 1000 rain gauges, yielding the gauge-calibrated RADOLAN products that are routinely used for hydrometeorological applications such as flood forecasting (Bartels et al., 2004). For climatological analyses, DWD has reprocessed these radar and gauge data to produce the RADKLIM dataset, which offers quality-controlled hourly precipitation sums (RW) and 5-minute precipitation rates (YW) on the same 1 km × 1 km grid over the effective radar coverage (~128 km radius per radar) from 2001 onwards. RADKLIM is updated annually and is specifically designed to provide a homogeneous, high-resolution basis for studying extreme rainfall and long-term precipitation statistics. In this study, we use the 5-minute RADKLIM YW product for the period 2001–2025 to compute annual rainfall erosivity, taking advantage of its high spatial and temporal resolution to resolve short-duration, high-intensity precipitation. The YW data were obtained from the DWD open data server:

(https://opendata.dwd.de/weather/radar/radolan/yw/).

3.2.1.2 Multi-scale climatic predictors

3.2.1.2.1 Rainfall erosivity from the DIN empirical model

Where high-temporal-resolution precipitation data are available, R-factors can be derived directly from event-based rainfall time series. In many regions, however, such data are sparse or only available for limited periods, so empirical regressions between rainfall erosivity and aggregated precipitation statistics are commonly used instead. In Germany, the national standard DIN 19708 specifies region-specific equations that approximate the mean annual R-factor from long-term precipitation records. At the national scale, DIN proposes a linear relationship between mean annual rainfall erosivity R (N h^{-1} yr^{-1}) and mean annual precipitation NJ (mm):

$$R = 0.078\,8 \times NJ - 2.82 \tag{3.1}$$

In the present study, the DIN-based R-factor is computed and incorporated as one of several large-scale climatic predictor variables within our deep-learning framework.

3.2.1.2.2 Annual temperature data

Although rainfall erosivity is formally defined from precipitation alone, we include annual temperature as a spatially distributed covariate because it reflects large-scale thermodynamic and phase-related controls on erosive rainfall, consistent with previous studies showing that temperature influences both precipitation extremes and erosivity patterns (Panagos et al., 2015; Sharma & Mujumdar, 2019). In colder upland regions, lower temperatures favor snowfall, snow cover, and frozen soils, which effectively shorten the erosive rainfall season and reduce direct rainfall-driven erosivity compared to warmer lowlands, where the same annual precipitation occurs predominantly as rain (Wu et al., 2018).

Three annual gridded temperature variables including minimum, mean, and maximum near-surface air temperature, were obtained for 1931–2025 from the Climate Data Center (CDC) of the German Weather Service (Deutscher Wetterdienst, DWD) at 1 km × 1 km spatial resolution. The grids are derived from observations at DWD stations and legally and qualitatively equivalent partner stations, interpolated to a regular grid using procedures that account for elevation and other topographic effects, and provided in units of 1/10 °C. The annual minimum and maximum temperature products represent, respectively, the coldest and warmest near-surface conditions recorded each year, thereby characterizing thermal extremes, whereas the annual mean temperature field describes the average climatic background state. All three datasets are based on homogenized station records and have undergone standardized quality control and gridding procedures within the CDC framework, ensuring consistency with other DWD climate products.

3.2.1.2.3 Daily rainfall data

Daily gridded precipitation for Germany was taken from the HYRAS-DE-PR v6-1 product provided by the Climate Data Center (CDC) of the German Weather Service (Deutscher Wetterdienst, DWD). This dataset supplies daily precipitation sums (mm) on a 1 km × 1 km grid for the period 1931 to the present and is based on gauge measurements of daily precipitation height across the national station network.

The gridded fields are generated by interpolating station-based daily totals using an anomaly approach relative to long-term monthly mean background fields. These background fields are obtained from multiple linear regression that relates mean monthly precipitation to geographical coordinates, elevation, and slope/exposure, thereby accounting for orographic and regional effects. Daily anomalies at the stations are then distance-weighted and mapped to the grid, with basic threshold and consistency checks applied to the daily updates and full quality control performed during subsequent monthly reprocessing to ensure high data reliability.

3.2.1.2.4 Spectral indices derived from daily precipitation

Because daily precipitation sums do not resolve sub-daily rainfall intensity, we derived a suite of spectral indices from the daily data to partially capture intensity-related characteristics that are relevant for rainfall erosivity. In total, 10 daily-scale precipitation indices were computed, representing different aspects of the precipitation regime such as wet-day frequency and occurrence of heavy and very heavy precipitation. These indices were calculated annually on the 1 km × 1 km grid and subsequently used as additional predictor variables in the reconstruction model, alongside the other climatic predictors described above. A complete overview of all input variables, including the precipitation indices, temperature data, and DIN-based R-factor is provided in Table 3.1.

Table 3.1. List of the data used for modelling.

data		Description	Temporal coverage
Radar Precipitation	$R_{RADKLIM}$	RADKLIM YW product	2001 to 2025
Temperature	T_{max}	Maximum air temperature	1931 to 2025
	T_{min}	Minimum air temperature	1931 to 2025
	T_{mean}	Mean air temperature	1931 to 2025
Interpolated precipitation	P_A	Total annual precipitation	1931 to 2025
	Day_{wet}	Number of days with precipitation	1931 to 2025
	P_{gt10}	Number of days with precipitation greater than 10 mm	1931 to 2025
	P_{gt20}	Number of days with precipitation greater than 20 mm	1931 to 2025
	P_{gt30}	Number of days with precipitation greater than 30 mm	1931 to 2025
	P_{gt50}	Number of days with precipitation greater than 50 mm	1931 to 2025
	PI	Precipitation Intensity (Annual precipitation/number of rainy days)	1931 to 2025
	P_{max}	Maximum daily precipitation	1931 to 2025
	P_{mean}	Mean daily precipitation	1931 to 2025
	R95P	Total rainfall from the heaviest 5% of rainy days	1931 to 2025
DIN R-factor	R_{DIN}	Derived from empirical equation of DIN 19708	1931 to 2025

3.2.2 Rainfall erosivity calculation

Within the RUSLE framework, the rainfall erosivity factor (R-factor) quantifies the ability of rainfall to detach and transport soil particles by combining information on event depth and intensity (Fiener et al., 2013). When precipitation is available at sufficiently high temporal resolution, the annual R-factor is obtained by summing the erosivity of all storms identified as erosive within a given year. For an individual storm, we follow the classical EI_{30} formulation and define event erosivity as (Wischmeier & Smith, 1978):

$$R_{event} = E \times I_{max30} \tag{3.2}$$

where R_{event} is the erosivity of a single event (N h−1), E is the event's total kinetic energy (kJ m−2), and I_{max30} (mm h−1) is the maximum rainfall intensity over any 30-min interval during the event.

In Germany, erosive events are defined according to DIN 19708 as rainfall events that either produce at least 10 mm of precipitation or reach a maximum 30 min intensity of at least 10 mm h−1. Distinct events are separated by dry intervals of more than 6 hours (Wischmeier & Smith, 1978), so that up to three erosive storms may occur within a single day, while a single event may also span multiple days. The kinetic energy E is computed as a function of rainfall intensity using an established empirical relationship (Equation 3) and accumulated over all 5-min intervals belonging to the erosive event (Wischmeier & Smith, 1978).

$$\begin{aligned} &\mathrm{E} = (11.89 + 8.73 \times \log \mathrm{I}) \times 10^{-3} && \text{for } 0.5\ \mathrm{mm\ h^{-1}} \leq \mathrm{I} \leq 76.2\ \mathrm{mm\ h^{-1}} \\ &\mathrm{E} = 0 && \text{for } \mathrm{I} < 0.05\ \mathrm{mm\ h^{-1}} \\ &\mathrm{E} = 28.33 \times 10^{-3} && \text{for } \mathrm{I} \geq 76.2\ \mathrm{mm\ h^{-1}} \end{aligned} \tag{3.3}$$

In this study, we used the 5-min RADKLIM YW radar product for the period 2001–2025 to identify erosive events, calculate event-based erosivity according to Equation 2, and sum over all events to derive annual R-factors, thereby fully exploiting the high spatiotemporal resolution of the dataset. To account for spatial, methodological and temporal differences between 5-min, 1 km × 1 km RADKLIM radar data and the 1-min, point-scale rain-gauge data on which USLE erosivity was developed, we applied the spatial, method and temporal scaling factors proposed by Fischer et al., (2018). For RADKLIM YW (μ = 1 for radar, σ = 1 km pixel width, τ = 5 min), this yields a combined multiplicative correction factor of 1.51, which was applied to all RADKLIM-based annual R-factors. The RADKLIM-derived erosivity fields exhibit pronounced small-scale noise and occasional unrealistically high annual values, which are attributable to a combination of genuine convective extremes and residual radar artefacts and therefore require systematic quality control. In the national RADKLIM-based erosivity map of Germany, Auerswald et al., (2019) reported a maximum long-term mean annual erosivity of approximately 454 $N \cdot h^{-1} \cdot yr^{-1}$ after temporal smoothing, winsorizing of extreme years, and geostatistical interpolation; accordingly, in the present study we adopt a conservative upper threshold of 700 $N \cdot h^{-1} \cdot yr^{-1}$ and remove annual values exceeding this limit to reduce the undue influence of noisy extremes on subsequent analyses.

3.2.3 Reconstructing historical rainfall erosivity data

To reconstruct historical rainfall erosivity for 1931–2000, a period without sub-hourly gridded precipitation, we adopt and extend a Hypernetwork-Enhanced CNN architecture originally proposed for fully distributed rainfall–runoff modeling (Scholz et al., 2025). In contrast to the original DRRAiNN formulation, which combines a temporally recurrent point-wise LSTM with a graph-based river network, our implementation focuses exclusively on the spatial rainfall–runoff component and omits both the explicit temporal processing and the river routing module. Trained on the 2001–2025 period, the model learns a non-linear relationship between coarse-resolution climate predictors and annual erosivity maps derived from 5-minute RADKLIM data.

3.2.3.1 Hypernetwork-Enhanced CNN model architecture

The core of the reconstruction framework is a custom Hypernetwork-Enhanced CNN architecture.

In this study, the architecture is used in a purely spatial configuration, with the temporal LSTM block and graph-based river network of the original Distributed Rainfall-Runoff Artificial Neural Network (DRRAiNN) setup removed, because the target variable is annual erosivity rather than discharge time series. Unlike standard CNNs that apply static learned weights across all spatial locations, our architecture employs a hypernetwork mechanism to dynamically generate synaptic weights for the main convolutional layers based on local environmental conditions. This allows the model to adapt its processing logic to specific regimes of rainfall intensity and climate. The backbone of the network consists of a sequence of two HyperConvNeXtBlocks with 15 hidden channels. Each block follows an inverted bottleneck design, comprising a depth-wise convolution (DWConv, kernel size 3 × 3), layer normalization, and two point-wise convolutions (PWConv1 for channel expansion and PWConv2 for projection) with a Sigmoid Linear Unit (SiLU) activation. Instead of static parameters, the weights of the depth-wise and point-wise convolutions are generated by hypernetworks conditioned on climatic covariates: a convolutional hypernetwork with the same receptive field produces the spatially varying DWConv kernels, while lightweight Multi-Layer Perceptrons (MLPs) predict the channel-mixing PWConv weights from selected hyper-inputs. Formally, the weight W for a given layer is computed as:

$$W = W_{base} + \alpha \cdot \phi\left(X_{hyper}\right) \tag{3.4}$$

where W_{base} represents a static weight matrix shared across all samples, X_{hyper} denotes the auxiliary hyper-inputs, ϕ (·) is the hypernetwork projection, and α is a learnable scaling factor. This formulation allows the network to modulate its sensitivity to precipitation features based on the magnitude of extreme events.

The model utilizes a stratified input strategy to handle different categories of meteorological variables derived from daily and annual aggregations, organized as follows:

3.2.3.1.1 Main inputs (X_{main})

Variables representing the broad climatic baseline and general wetness were processed through the primary convolutional path to capture spatial gradients. This set includes total annual precipitation, days with precipitation, moderate precipitation frequency indices (>10mm, >20mm, >30mm), and mean temperature (T_{mean}).

3.2.3.1.2 Extended inputs (X_{ext})

Secondary climatic context was provided as auxiliary conditioning to the network, specifically minimum and maximum temperatures (T_{min}, T_{max}) and the empirical DIN-19708 erosivity estimate. Including the DIN estimate allows the network to learn a residual correction map on top of the traditional empirical standard.

3.2.3.1.3 Localized inputs (X_{loc})

To prevent the spatial smoothing inherent in convolution operations from blurring critical peak events, extreme value indicators were injected directly into the feature maps after the convolutional stage. These include maximum daily precipitation and the frequency of heavy storms (>50 mm), ensuring the model retains high sensitivity to localized extremes.

3.2.3.1.4 Extended hyper inputs (X_{hyper}^{ext})

The 95th percentile of daily precipitation ($R95p$) serves as a control signal for the hypernetwork to generate dynamic weights for the spatial convolution kernels, allowing the network to adapt to the regional climatic regime.

3.2.3.1.5 Localized hyper inputs (X_{hyper}^{loc})

Precipitation intensity metrics (Simple Daily Intensity Index and mean daily precipitation) modulate the weights of the point-wise layers, adapting the non-linear transfer function to the local rainfall intensity characteristics.

3.2.3.2 Hypernetwork-Enhanced CNN model training and optimization

The Hypernetwork Enhanced CNN model was implemented using PyTorch Lightning and trained on the 2001–2025 reference period. The dataset was processed as spatial patches of size 64 × 64 pixels with a 25% overlap to ensure spatial continuity. To solve the artifact problem in the overlapping regions of these predictions, we implemented a weighted averaging strategy based on the Hamming window (Hamming, 1956). This bell-shaped tapering function is applied to each 64 × 64 patch to smoothly reduce the weight of pixels toward the edges. By doing so, the transitions between overlapping regions are effectively faded together, which suppresses boundary artifacts and ensures a seamless 1-km erosivity map.

The 2001–2025 annual dataset was partitioned into training (60%), validation (20%), and test (20%) subsets using a stratified sampling approach. To ensure the model generalizes across diverse climatic extremes, the splits were engineered to represent both the full temporal trajectory and the entire precipitation spectrum (dry, average, and wet years). This method preserves the inter-annual variability within each subset, thereby mitigating temporal bias and enhancing the predictive robustness of the modeling framework.

Although extreme rainfall events appear as statistical outliers, they play a disproportionate role in driving rainfall erosivity and associated soil erosion (Dai et al., 2023). We therefore retained them in the training data, accepting an outlier-contaminated (noisy) regression setting rather than smoothing them away. Rainfall erosivity additionally exhibits heavy-tailed distributional characteristics, wherein rare extreme events contribute disproportionately to cumulative erosional flux. Accordingly, model training used the Huber loss function (Huber, 1992), which improves robustness to outliers and heavy-tailed errors by combining quadratic and linear penalization (Tong, 2023). Optimization was performed using the Adam optimizer (Kingma & Ba, 2017) with a learning rate of 10-3 and a batch size of 16 patches. Training proceeded for 50 epochs with validation-set performance monitoring to detect and arrest overfitting. Post-training, the trained model was applied to reconstruct historical gridded erosivity fields over the extended period (1931–2000) by propagating historical predictor variables through the learned model. Figure 3.1 presents the schematic architecture of the modeling framework.

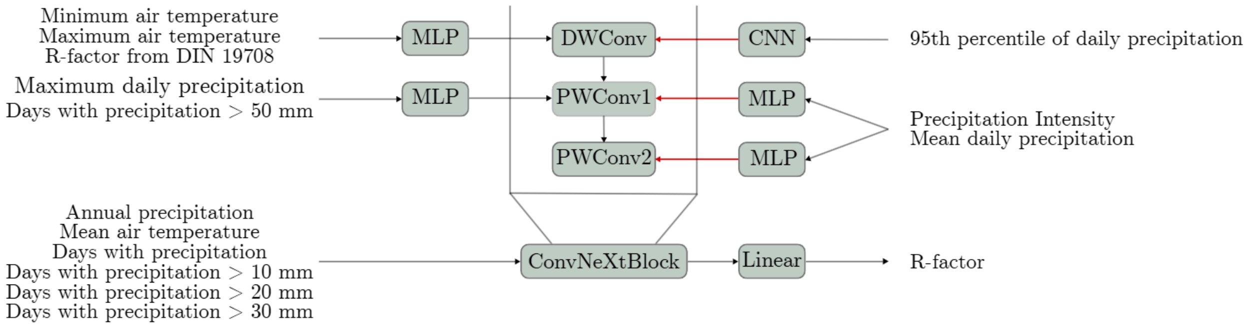


Figure 3.1. Schematic overview of the hypernetwork-enhanced ConvNeXt architecture for annual rainfall erosivity reconstruction. The ConvNeXtBlock consists of two types of convolutions: a depth-wise convolution (DWConv), which is spatially extended but operates on each channel independently, and two point-wise convolutions (PWConv1 and PWConv2), which are not spatially extended but mix the channels. The weights of these convolutions are produced by hypernetworks, where the DWConv weights are produced by a convolutional neural network (CNN) with the same receptive field as the DWConv, and the PWConv weights are produced by Multi-Layer Perceptrons (MLPs).

3.2.4 ConvLSTM-based rainfall erosivity forecasting

3.2.4.1 ConvLSTM architecture

Long Short-Term Memory (LSTM) networks are a specialized type of recurrent neural network (RNN) designed to capture long-range dependencies in sequential data while mitigating vanishing and exploding gradient issues (Moskolaï et al., 2020). Each LSTM unit contains a memory cell regulated by input, forget, and output gates that control how information is stored, discarded, or passed to the next time step. This gating mechanism, combined with sigmoid and hyperbolic tangent activations, allows effective retention and transfer of contextual information, enabling LSTMs to perform robustly on tasks such as time series analysis and natural language processing.

Conventional LSTM architectures are primarily suited for one-dimensional sequential data and thus are limited in capturing spatial correlations. To address this constraint, Shi et al., (2015) proposed the Convolutional LSTM (ConvLSTM), which replaces the fully connected transformations in standard LSTM gates with convolution operations (Figure 3.2). This modification allows the model to maintain the spatial dimensions of input data across layers, enabling simultaneous learning of spatial and temporal dependencies. Consequently, ConvLSTM effectively combines the feature extraction strengths of CNNs with the sequential modeling capabilities of LSTMs, leading to improved predictive performance in spatiotemporal applications such as precipitation nowcasting (W. Liu et al., 2022). The basic computational unit of ConvLSTM is still an LSTM cell composed of an input gate i_t, a forget gate f_t, an output gate o_t, the cell state c_t, and the hidden state h_t. Compared with the classical LSTM, ConvLSTM replaces matrix multiplications with convolution operations and incorporates peephole connections from the cell state to the gates. At each time step t, the gates and states are computed as follows:

$$i_t = \sigma(W_{Xi} * x_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \quad (3.5)$$

$$f_t = \sigma\left(W_{Xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f\right) \quad (3.6)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_{Xc} * x_t + W_{hc} * h_{t-1} + b_c) \quad (3.7)$$

$$o_t = \sigma(W_{Xo} * x_t + W_{ho} * h_{t-1} + W_{co} \circ c_t + b_o) \quad (3.8)$$

$$h_t = o_t \circ \tanh(c_t) \quad (3.9)$$

Here, σ denotes the sigmoid activation function and tanh the hyperbolic tangent function. The tensor x_t is the input at time t; h_{t-1} and c_{t-1} are the hidden and cell states from the previous time step; and h_t and c_t are their current counterparts. The convolutional kernels W_{X*} and W_{h*} operate on the input and hidden state, respectively, while W_{c*} represents peephole convolutions from the cell state to the gates; b_* are bias terms. The operator $*$ denotes convolution, and $\circ$ indicates the Hadamard (element-wise) product. In this formulation, the forget gate f_t controls how much of the previous cell state c_{t-1} is retained, the input gate i_t regulates the incorporation of new information, and the output gate o_t determines which parts of the updated cell state contribute to the hidden state h_t at time t.

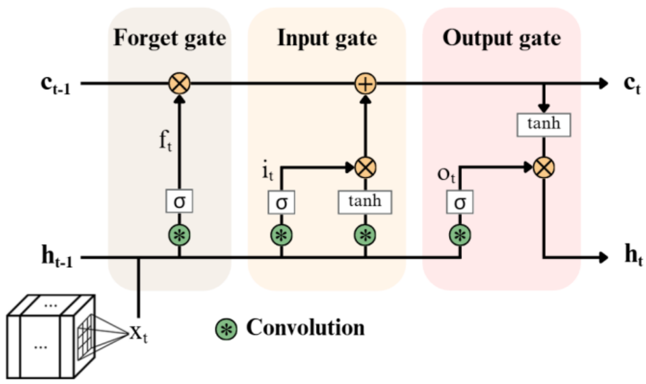


Figure 3.2. The inner structure of ConvLSTM.

3.2.4.2 Sample preparation and forecasting methodology

In this study, three annual gridded datasets defined on an identical spatial grid over the study domain are used as inputs to the ConvLSTM model: rainfall erosivity maps (historical 1931–2025 maps derived from the Hypernetwork Enhanced CNN model), precipitation maps (1931–2043), and mean temperature maps (1931–2043), with post-2025 data from bias-adjusted RCP4.5 projections of the CORDEX-EUR11 regional climate model ensemble (BAdCORD-EUR11; Menz, 2023).

Sample preparation (Figure 3.3A): Temporal forecasting using deep learning requires transforming raw time series into structured input–output pairs. Here, each training sample consists of an input sequence and a corresponding output sequence (frames), both derived from the historical time series. The input and output windows are concatenated along the temporal dimension to form a single sample, enabling the model to learn the mapping from past observations to future erosivity values. A sliding-window approach with a one-year step size is used to traverse the historical record, producing overlapping samples that expose the model to a wide range of temporal patterns and reduce sensitivity to particular calendar years. The input and output window lengths are set to 12 and 6 years, respectively, providing a compromise between a sufficiently long temporal receptive field and computational feasibility, while capturing both multi-decadal climate variability and shorter-term erosivity dynamics (Figure 3.3A).

The period 1931–2019 is partitioned chronologically into training (80%) and validation (20%) subsets to enable robust model selection without temporal leakage. The independent test period 2020–2025 spans exactly one 6-year forecast window, allowing an out-of-sample evaluation of model performance under recent climate conditions. During training, validation, and testing, mean temperature, precipitation and erosivity inputs are derived from observations. During forecasting (beyond 2025), the decoder precipitation and mean temperature inputs are replaced by BAdCORD-EUR11 RCP4.5 projection for the corresponding target years.

The forecasting model adopts a ConvLSTM encoder–decoder architecture tailored to the spatiotemporal nature of rainfall erosivity. The encoder processes the 12-year input sequence of annual precipitation, mean temperature and erosivity fields using two stacked ConvLSTM2D layers, each with 32 convolutional filters, 3×3 kernels, and same padding to preserve spatial dimensions. Dropout ($p = 0.1$) and recurrent dropout ($p = 0.1$) are applied to both layers to mitigate overfitting. The first ConvLSTM layer returns full sequences, while the second outputs only its final hidden state, which serves as a compressed context vector summarizing the 12-year input history.

Model architecture (Figure 3.3B): The decoder generates future erosivity predictions by combining this encoded historical context with bias-adjusted precipitation and mean temperature RCP4.5 projections from BAdCORD-EUR11. A bridge mechanism repeats the encoded context across the 6-year forecast horizon and concatenates it with the corresponding sequence of projected precipitation and mean temperature fields, thereby providing the decoder with both the learned historical state and the future atmospheric forcing. A ConvLSTM2D layer (32 filters, 3×3 kernels, same padding) processes these concatenated spatiotemporal inputs and produces hidden states for all six forecast years. Finally, a Conv3D layer with a single filter and 1×3×3 kernel projects these hidden states to a single output channel, yielding annual rainfall erosivity predictions for each year in the 6-year window (Figure 3.3B).

Multi-stage iterative forecasting strategy (Figure 3.3C): This architectural design reflects the physical linkage between precipitation, mean temperature and erosivity. The encoder learns how historical temperature-precipitation-erosivity relationships manifest in space and time, while the decoder extrapolates these learned patterns into the future conditioned on externally provided precipitation and mean temperature projections. Long-term forecasting up to 2043 requires extending predictions beyond a single 6-year output window while respecting the model's 12-year temporal receptive field. To achieve this, we employ a three-stage iterative forecasting strategy, each stage producing a 6-year erosivity forecast (Figure 3.3C):

In the first stage, covering the 2026–2031 period, the model uses observed annual precipitation, annual mean temperature, and annual erosivity from 2014 to 2025 as the 12-year input window. This historical context is combined with projected precipitation and temperature for 2026–2031 to generate erosivity forecasts for that same window.

In the second stage, spanning 2032–2037, the input window shifts to the 2020–2031 period. This window consists of observed precipitation, temperature, and erosivity from 2020 to 2025 concatenated with projected precipitation, projected temperature, and the forecasted erosivity derived from the first stage for 2026–2031. These inputs, along with projected precipitation and temperature for 2032–2037, allow the model to produce erosivity forecasts for the second 6-year period.

In the third stage, covering 2038–2043, the input window spans 2026–2037 and is composed entirely of projected precipitation, projected temperature, and the erosivity values forecasted during the previous two stages. By processing this 12-year sequence alongside projected precipitation and temperature for 2038–2043, the model generates the final erosivity forecasts for the 2038–2043 horizon.

This recursive strategy allows the model to exploit its full 12-year temporal receptive field at each stage while progressively propagating information forward in time. It maintains temporal consistency across successive forecast windows and enables long-range erosivity projections conditioned on physically plausible mean temperature and precipitation trajectories from regional climate models.

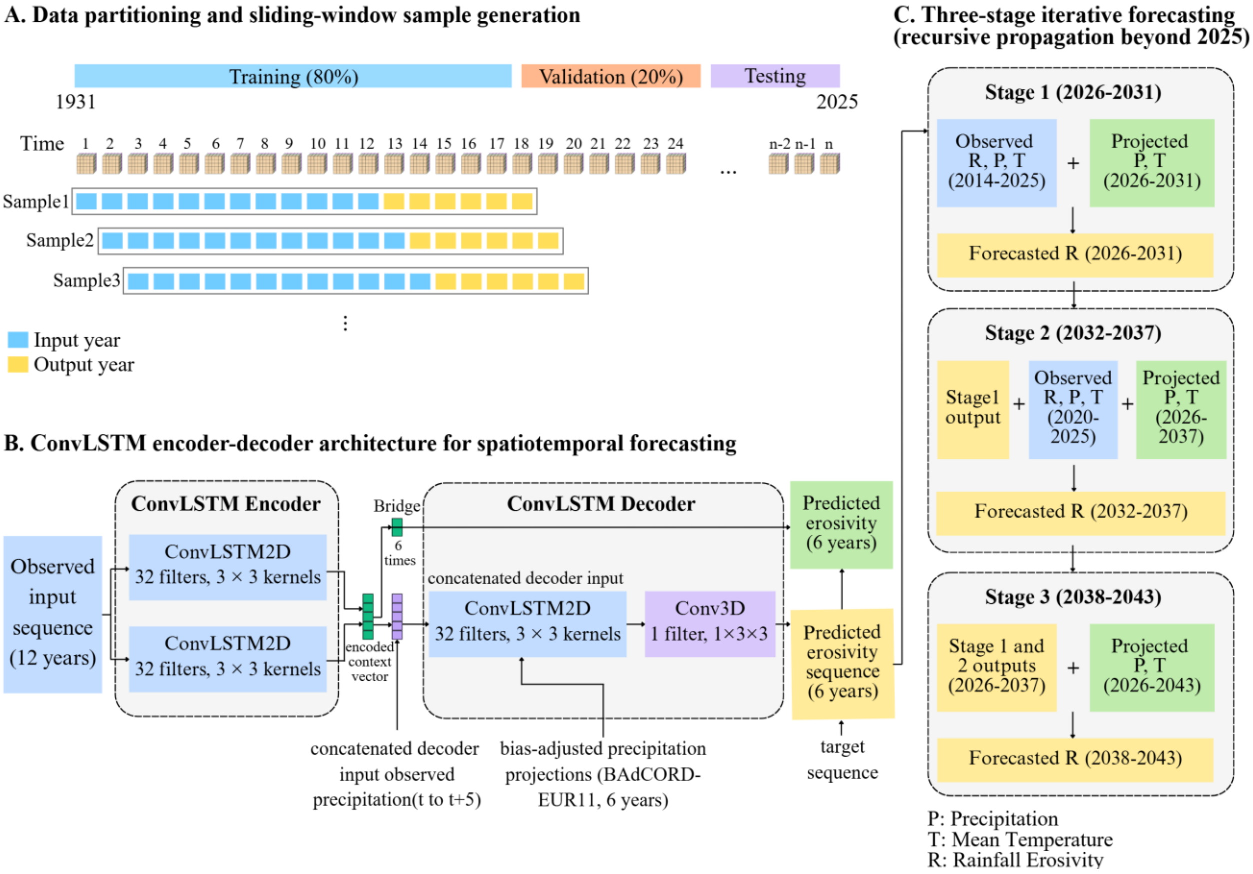


Figure 3.3. Overview of the ConvLSTM rainfall erosivity forecasting framework, detailed by A) sample preparation, B) model architecture, and C) multi-stage iterative forecasting strategy.

3.2.4.3 ConvLSTM model training and optimization

Model training minimized a masked Huber loss function that emphasizes prediction accuracy in well-observed regions while gracefully handling missing or invalid data through spatial masking. The Huber loss (Huber, 1964) combines robustness with differentiability, effectively reducing the influence of outliers while maintaining smooth gradients for optimization. Optimization was carried out with the Adam algorithm (Kingma & Ba, 2017) using an initial learning rate of 10-3, and a learning-rate scheduler reduced the learning rate by a factor of 0.1 whenever the validation loss failed to improve over 10 consecutive epochs. During training, the spatial domain was partitioned into 64×64-pixel patches with 25% overlap, and a batch size of 8 patches was used per update.

The model was trained for up to 50 epochs with early stopping implicitly enforced via model selection based on validation performance to limit overfitting on the relatively short climate record. At the end of training, the network weights corresponding to the lowest validation loss were retained and used for independent testing and subsequent forecasting. All experiments were implemented in Python using the TensorFlow deep learning framework and executed on a Windows 11 workstation equipped with an NVIDIA GeForce RTX 4070 Ti GPU.

3.2.4.4 Model evaluation

Model evaluation comprises (i) a quantitative split-sample assessment against RADKLIM-based annual erosivity during the radar era and (ii) additional consistency checks against independent long-term station records and climate-model-based projections of rainfall erosivity.

For both the historical reconstruction and the near-term forecasting experiments, predictive skill is first quantified on independent evaluation data by comparing modeled annual rainfall erosivity against the corresponding RADKLIM-derived reference fields. We summarize goodness of fit using the root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2), which jointly characterize the typical error magnitude (RMSE, MAE) and the fraction of observed variance captured by the models (R^2). For an evaluation data set with n samples and reference values y_i, model predictions $\hat{y}_i$, and mean reference $\bar{y}$, these statistics are defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{3.10}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |, \tag{3.11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.12}$$

For the Hypernetwork-Enhanced ConvNeXt reconstruction, these metrics are computed by comparing reconstructed annual erosivity maps with RADKLIM-based erosivity for years withheld from model fitting. This split-sample evaluation quantifies how well the transfer framework can recover the intensity-dependent RADKLIM signal from coarser-scale predictors and thus provides a strictly data-driven benchmark for the reconstruction skill. Beyond this internal validation, we will additionally benchmark the reconstructed time series against independent gauge-based estimates of historical rainfall erosivity from Western Germany to assess whether reconstructed trends are consistent with long-term station records. Specifically, we will compare our Germany-wide reconstructions, aggregated over the North Rhine–Westphalia region, with the nine high-resolution station series analyzed by Fiener et al., (2013), who derived summer erosivity for 1937–2007 and used summer erosivity as a proxy for annual erosivity in Western Germany. Because Fiener et al., (2013), work with point-scale summer erosivity at a small set of stations whereas our product provides 1 km × 1 km annual erosivity including all months over the whole of Germany, this comparison will focus on the direction and relative magnitude of temporal trends in R-factor rather than on absolute R-factor values or one-to-one raster agreement.

The ConvLSTM-based forecasting model is analogously evaluated using RMSE, MAE, and R^2 by comparing forecasts to RADKLIM-based annual erosivity for held-out years in the recent period, thereby providing a quantitative, split-sample test of near-term predictive skill under observed climate conditions. To place the projected German-wide erosivity trends under RCP4.5 into a broader context, we will further carry out a qualitative comparison with the multi-model projections of Auerswald and Fiener., (2024). In their study, future erosivity for Germany was derived from seven EURO-CORDEX regional climate models under the high-end RCP8.5 scenario and showed substantial future increases in rainfall erosivity and its interannual variability. Given the differences in emission scenario (RCP8.5 vs. RCP4.5), ensemble size (seven RCMs vs. a single bias-adjusted RCM), and methodological details, our evaluation will again be restricted to the comparison of trends rather than to a direct comparison of absolute R-factor values.

3.3. Results and discussion

3.3.1. RADKLIM-based rainfall erosivity and its climatic controls (2001–2025)

Annual rainfall erosivity was calculated for each year from 2001 to 2025 using RADKLIM 5-minute precipitation data and the DIN empirical formula based on annual precipitation totals.

The spatial correlation matrix presents Pearson correlations between the resulting annual erosivity maps and 14 additional climate and precipitation raster datasets (for abbreviations see heading of Table 3.2), each aggregated over the full 25-year period to characterize their mean spatial patterns (Figure 3.4). Among all variables examined, R95P (total rainfall from the heaviest 5% of rainy days) exhibits the strongest spatial correlation with RADKLIM-based erosivity (r = 0.69), followed closely by indices of PI and P_{gt20} and RADKLIM-based erosivity (with r = 0.67, 0.67 and 0.61, respectively). However, all three annual temperature metrics exhibit very low, partly negative spatial correlations with erosivity ($|r| \leq 0.21$), indicating that the climatologically warmest regions are not the most erosive. However, when correlations are evaluated in the temporal domain using Germany-wide annual means of each variable and erosivity for 2001–2025, a different picture emerges (Table 3.2). Heavy-precipitation indices such as days with precipitation greater than 30 mm, days with precipitation greater than 50 mm, maximum daily precipitation, and days with precipitation greater than 20 mm exhibit very strong year-to-year co-variability with erosivity (r = 0.87, 0.85, 0.84, and 0.82, respectively), followed by R95P and precipitation intensity (r = 0.74 and 0.70). Even variables with weak spatial associations, notably minimum and mean annual temperature, show moderate positive temporal correlations with erosivity (r = 0.45 and 0.31), suggesting that warmer years tend to be more erosive even though warmer regions are not necessarily more erosive in space.

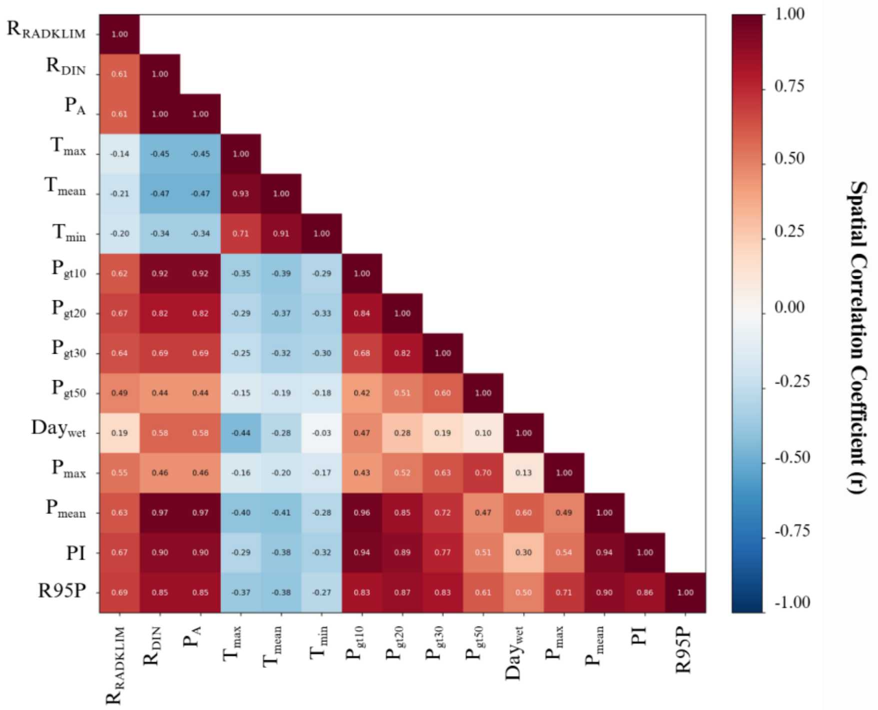


Figure 3.4. Spatial correlation between rainfall erosivity and input covariates across Germany, 2001–2025.

Table 3.2. Temporal correlations between annual rainfall erosivity and input covariates across Germany, 2001–2025.

Variable	abbreviation	Correlation
Number of days with precipitation greater than 30 mm	P_{gt30}	0.87
Number of days with precipitation greater than 50 mm	P_{gt50}	0.85
Maximum daily precipitation	P_{max}	0.84
Number of days with precipitation greater than 20 mm	P_{gt20}	0.82
Total rainfall from the heaviest 5% of rainy days	R95P	0.74
Daily precipitation intensity	PI	0.70
Number of days with precipitation greater than 10 mm	P_{gt10}	0.65
Mean daily precipitation	P_{mean}	0.56
Total annual precipitation	P_A	0.54
R-factor derived from empirical equation of DIN 19708	R_{DIN}	0.54
Mean air temperature	T_{mean}	0.45
Minimum air temperature	T_{min}	0.31
Number of days with precipitation	Day_{wet}	0.24
Maximum air temperature	T_{max}	0.18

3.3.2 Performance of the Hypernetwork-Enhanced CNN model across contrasting erosivity regimes

The model's ability to reconstruct historical R-factors was evaluated over five independent test years

(2002, 2012, 2017, 2022, and 2024), selected to cover a broad range of hydro-climatic regimes. These years represent contrasting erosivity conditions, from low-erosivity years (2012 and 2022; mean R-factor approximately 124 and 131 N h^{-1} yr^{-1}, respectively) to moderate conditions (2017; mean R-factor = 177 N h^{-1} yr^{-1}) and high-erosivity years marked by extreme events (2002 and 2024; mean R-factor approximately 208 and 195 N h^{-1} yr^{-1}, respectively).

A visual inspection of the spatial patterns (Figure 3.5, top and middle rows) demonstrates a strong congruence between the RADKLIM-derived reference maps and the model predictions across all test years. The model successfully reproduces the primary erosivity gradients within Germany, correctly identifying the pre-alpine region and the central low mountain ranges as high-erosivity zones, while assigning lower values to the North German Plain. This spatial distribution is consistent with the long-term erosivity climatologies for Germany described by Auerswald et al., (2019) and broader European assessments by Panagos et al., (2015). This spatial fidelity is quantitatively supported by the correlation coefficients (r), which range from 0.76 to 0.85, and Coefficient of Determination (R^2) values between 0.57 and 0.73.

Particular attention must be paid to the year 2002, which exhibited the highest mean erosivity (208 N h^{-1} yr^{-1}) and the greatest hydrological stress among the tested years. This year is historically significant due to the catastrophic Elbe Flood in August 2002, triggered by a Vb-cyclone system that delivered extreme precipitation intensities to Central Europe. The model's performance during this extreme year highlights its robustness in capturing localized high-magnitude events. As shown in the 2002 prediction map (Figure 3.5, red rectangle), the model successfully identified the intense erosivity hotspot in the Eastern Ore Mountains (Erzgebirge). This region includes the Zinnwald-Georgenfeld station, which recorded a 24-hour rainfall total of 312 mm during the event, a historical record for Germany (Rudolf & Rapp, 2003; Ulbrich et al., 2003). The fact that the model could spatially resolve this anomaly, despite the erratic nature of such extreme convective-stratiform complexes, suggests that the input predictors effectively proxy the kinetic energy dynamics of heavy precipitation even in the absence of direct high-resolution radar data. However, the stochastic nature of these extremes also imposes limitations. The year 2002 yielded the lowest R^2 (0.57) and highest RMSE (71.66 N h^{-1} yr^{-1}) among all test years. This reduction in predictive skill relative to calmer years is attributable to the high spatial variability of extreme cells, where slight spatial misalignments between the predictor variables and the actual storm tracks can lead to larger residuals (Fischer, Winterrath, et al., 2018).

Analyzing the predicted values of different years reveals a clear dependency of model performance on the prevailing erosivity level. For the low erosivity years (2012, 2022), the model achieved its highest accuracy (R^2 = 0.73 and 0.67, respectively; MAE = 29.23 and 36.02 N h^{-1} yr^{-1}, respectively). In the absence of widespread heavy convective storms, the erosivity distribution is driven more by large-scale frontal systems and topographic forcing, which are easier to capture using standard meteorological covariates. In contrast, highly erosive years (2002, 2024) exhibited slightly lower R^2

values (0.57 and 0.68, respectively) and higher error magnitudes (MAE = 46.47 and 45.34 N h^{-1} yr^{-1}). The year 2024, despite having a similar mean erosivity to 2002, showed a better fit (R^2=0.68), potentially indicating that the high erosivity in 2024 was more spatially distributed or less driven by singular outlier events compared to the singular catastrophe of August 2002. This systematic dependence of model skill on the prevailing erosivity level implies that prediction uncertainty increases in years when the annual R-factor is dominated by a small number of highly localized extreme storms. In such regimes, small spatial or temporal mismatches between the predictors and the actual storm tracks can lead to disproportionately large residuals, foreshadowing increased uncertainty in future projections if rainfall erosivity becomes even more strongly controlled by individual extreme events.

The scatter plots (Figure 3.5, bottom row) elucidate the structural behavior of the model's predictions. Across all years, the main cluster of data points aligns well with the 1:1 identity line for low-to-moderate erosivity values (< 200 N h^{-1} yr^{-1}), confirming that the model provides unbiased estimates for the majority of the domain. However, a distinct smoothing effect is observed in the upper tail of the distribution. As indicated by the regression slopes (blue lines) deviating from the 1:1 line, the model tends to underestimate the most extreme erosivity values (e.g., pixels where RADKLIM > 300 N h^{-1} yr^{-1}). This regression to the mean is a known characteristic of statistical and machine learning models applied to stochastic climate variables; the input predictors, often spatially or temporally aggregated, cannot fully resolve the peak kinetic energy of short-duration cloudbursts (Panagos et al., 2017). Despite this underestimation of maxima, the high correlation coefficients (0.76 to 0.85) confirm that the relative spatial ranking of erosivity risk is preserved, validating the model's utility for regional-scale reconstruction.

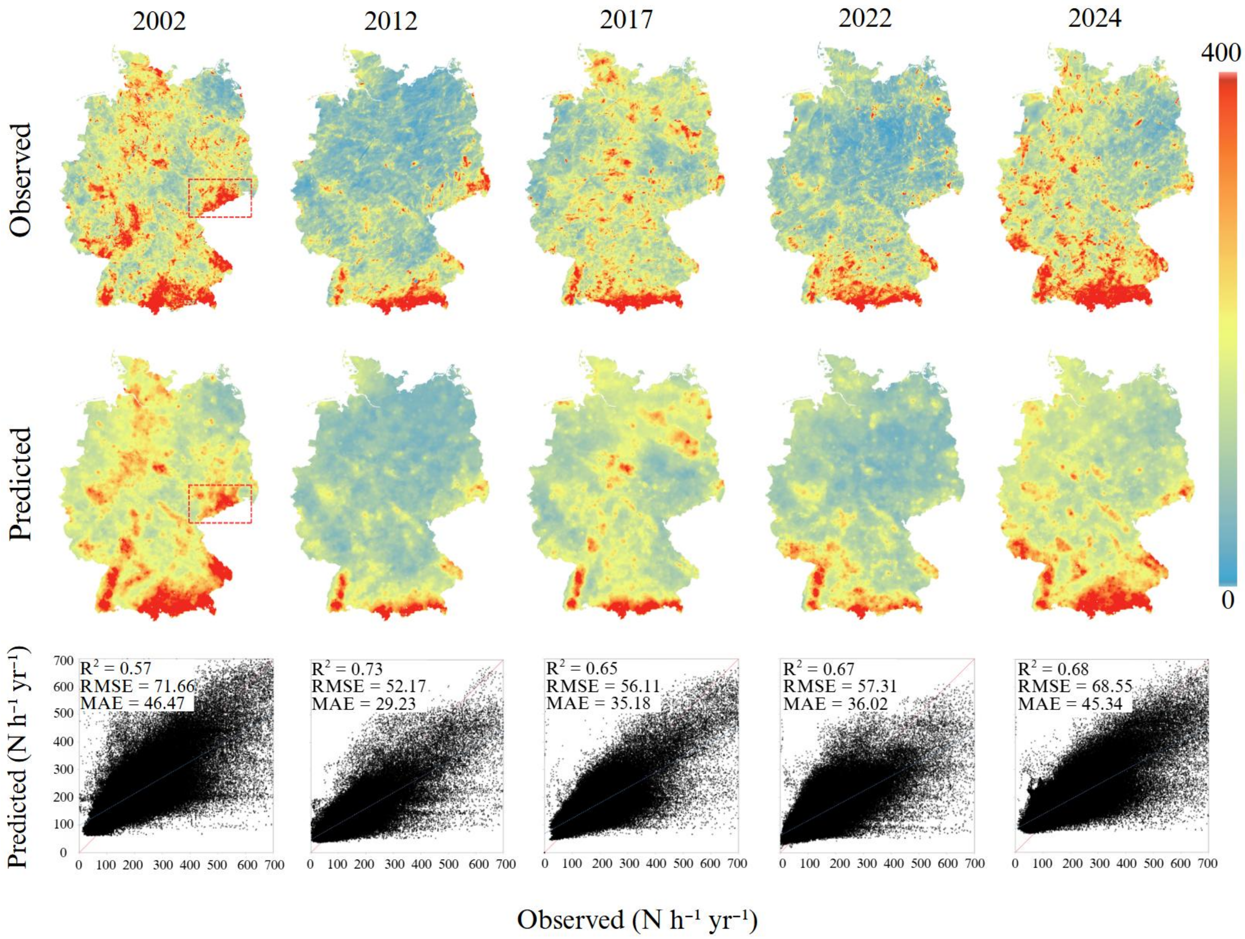


Figure 3.5. Model performance across five test years with different erosivity regimes. Top row: Annual rainfall erosivity maps derived from 5-minute RADKLIM radar data (N h^{-1} yr^{-1}). Middle row: Corresponding spatially predicted erosivity values generated by the model (N h^{-1} yr^{-1}). Bottom row: Scatter plots of predicted versus RADKLIM R-factor values (N h^{-1} yr^{-1}), showing the 1:1 identity line (red) and linear regression fit (blue).

3.3.3 Historical reconstruction and long-term erosivity dynamics (1931–2000)

The reconstruction of annual rainfall erosivity for Germany spanning 1931 to 2000 represents a critical extension of the empirical RADKLIM record into a period where high-resolution radar precipitation data are unavailable. This 70-year hindcast allows for a comprehensive assessment of long-term erosivity trends across the pre-digital observational era, providing essential context for understanding contemporary changes and future projections. The reconstructed time series reveals pronounced multi-decadal variability and a striking secular increase in erosivity, particularly from the 1960s onward (Figure 3.6), consistent with broader patterns of climate change intensification documented across Central Europe (Auerswald, Fischer, et al., 2019; Uber et al., 2024).

The reconstructed blue line (Figure 3.6) exhibits considerable interannual variability throughout the 1931–2000 period, with mean annual erosivity values notably lower than the RADKLIM-derived observations from 2001–2025 (orange line). This discrepancy is physically consistent with independent empirical evidence. Auerswald et al., (2019) demonstrated that German rainfall erosivity significantly increased between 1960 and 2017. This increase was confirmed by long-term rain gauge records and explicitly attributed to climate change rather than methodological artifacts. Our reconstruction successfully captures this ascending trajectory, with reconstructed values in the early period (1931–1960) exhibiting lower mean erosivity compared to the late 20th century (1970–2000), thereby providing a plausible baseline against which the magnitude of contemporary change can be quantified.

A critical feature of the reconstruction is the apparent inflection in the long-term trend occurring around the 1960s–1970s. Visual inspection of the blue line suggests that erosivity remained relatively stable or exhibited modest fluctuations during the 1930s–1950s, followed by a gradual but persistent increase from approximately the mid-1960s onward. This pattern aligns closely with paleoclimatic and observational evidence. Hänsel et al., (2022) documented that the 1930s–1940s in Central Europe were characterized by pronounced drought conditions and anomalously dry summer half-years, which would have suppressed convective precipitation and thus erosivity. In contrast, the post-1960 period witnessed increasing precipitation and, more importantly for erosivity, an increase in extreme precipitation frequency and intensity. Madakumbura et al., (2021) applied artificial neural networks to detect anthropogenic signals in extreme precipitation and found that the anthropogenic fingerprint became statistically detectable from the 1970s onward. This temporal correspondence between our reconstructed erosivity inflection and the emergence of anthropogenic forcing in precipitation extremes strongly supports the physical plausibility of the model-based reconstruction.

Critically, the inclusion of temperature and precipitation-derived indices as model predictors allows the reconstruction to capture these regime shifts dynamically. Temperature serves as a physically-based proxy for atmospheric moisture-holding capacity, while precipitation-derived indices provide direct information about the evolving structure of precipitation delivery. This multi-variate approach ensures that the model does not impose a fixed, contemporary erosivity-precipitation relationship onto the past, but rather adapts to the changing thermodynamic and dynamic state of the atmosphere across the reconstruction period. The model can therefore distinguish between eras when erosivity increases were driven primarily by precipitation volume changes (e.g., increased total summer rainfall) versus periods when intensification of individual storm characteristics dominated (e.g., higher I_{30} values for similar precipitation totals).

The green line in Figure 3.6, representing the erosivity-to-precipitation ratio (R/P ratio), provides a proxy for changes in precipitation intensity structure independent of total rainfall amount. This metric is particularly diagnostic because erosivity is a nonlinear function of both rainfall kinetic energy and intensity, whereas total precipitation is simply an integrated volume. Consequently, an increasing R/P ratio indicates a shift toward more intense, erosive rainfall regimes, even if total annual precipitation remains constant or increases modestly (Panagos et al., 2015).

The temporal pattern of the R/P ratio across the 1931–2025 period reveals important climatological transitions. An upward trend in this ratio would indicate a fundamental shift toward more intense, erosive rainfall regimes consistent with thermodynamic expectations under atmospheric warming. The Clausius-Clapeyron equation indicates that atmospheric water vapor content increases at a rate of approximately 7% K^{-1} of warming, providing additional moisture for more intense precipitation events (Asadieh & Krakauer, 2015). This intensification signal is particularly diagnostic because it cannot be explained by changes in total precipitation volume alone; it specifically reflects changes in the kinetic energy delivery and sub-daily intensity distribution of rainfall events. The R/P ratio serves as an internal validation metric for the reconstruction's physical plausibility. If the reconstructed ratio remained constant or declined over time, it would raise concerns about model specification or predictor adequacy. Conversely, a gradual increase in the ratio, particularly accelerating from the 1970s onward, coinciding with the emergence of anthropogenic climate signals in precipitation extremes, would provide strong evidence that the model successfully captures the evolving nature of erosive rainfall in Germany. This capacity to resolve intensity regime shifts, enabled by the mechanistically-informed choice of temperature and precipitation indices as predictors, represents a key strength of the modeling framework and enhances confidence in the reconstruction's ability to represent past erosivity dynamics that cannot be directly observed.

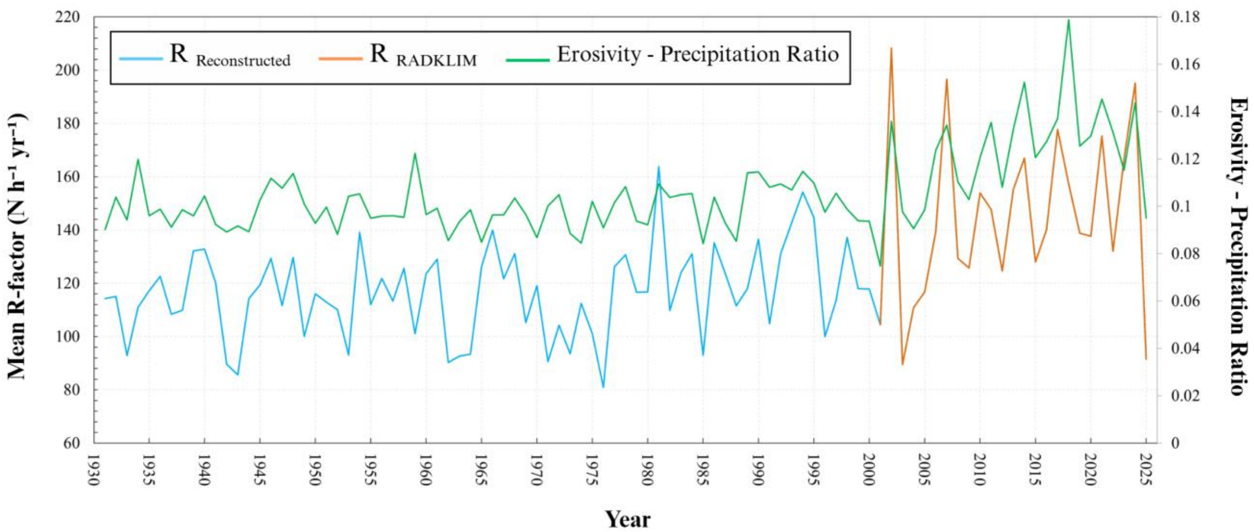


Figure 3.6. Reconstructed (1931–2000, blue) and RADDKLIM (2001–2025, orange) annual mean rainfall erosivity for Germany, with erosivity-to-precipitation ratio (green).

To assess the external plausibility of the reconstructed erosivity series, we compared our model output with the station-based long-term analysis of Fiener et al., (2013) for nine stations in western Germany. Their dataset provides an independent historical benchmark derived from high-resolution precipitation records for 1937–2007; however, because their annual erosivity definition is based on the summer half-year (April–November), whereas our reconstruction considers the full annual cycle, this comparison is not intended as a direct one-to-one validation of absolute magnitudes. Instead, the comparison is used to evaluate whether the reconstructed series reproduces the same interannual variability and long-term tendency observed in the reference record.

The results indicate a consistent positive association between the reconstructed values and the station-based observations, with the scatter distribution showing a clear linear relationship and a correlation coefficient of $r = 0.72$ (Figure 3.7). This level of agreement suggests that the model successfully captures the dominant large-scale signal in erosivity variability, even though the two datasets differ in spatial support, temporal coverage, and seasonal definition.

The temporal comparison further supports the physical realism of the reconstruction. Both time series exhibit broadly coherent decadal fluctuations and a shared upward tendency toward the end of the observation period, indicating that the model reproduces the direction and timing of long-term change rather than merely matching mean conditions. This agreement is particularly important because the comparison is based on an independent historical dataset derived from point-scale precipitation records, whereas the present reconstruction is built from gridded predictors and a different seasonal integration framework. Taken together, the correlation structure and trend coherence provide evidence that the reconstructed erosivity field is not only internally consistent but also externally plausible when evaluated against an established regional benchmark.

Importantly, the comparison should be interpreted conservatively. Since the dataset of Fiener et al., (2013) is based on summer erosivity only, some differences in amplitude are expected and do not undermine the reconstruction. The key result is that both approaches identify the same broad historical evolution of erosive rainfall in western Germany, thereby supporting the use of the reconstructed series for long-term trend analysis and subsequent forecasting. In this sense, the comparison confirms that the model preserves the climatological signal of erosivity change, even under a stricter validation setting where exact agreement in values is neither expected nor required.

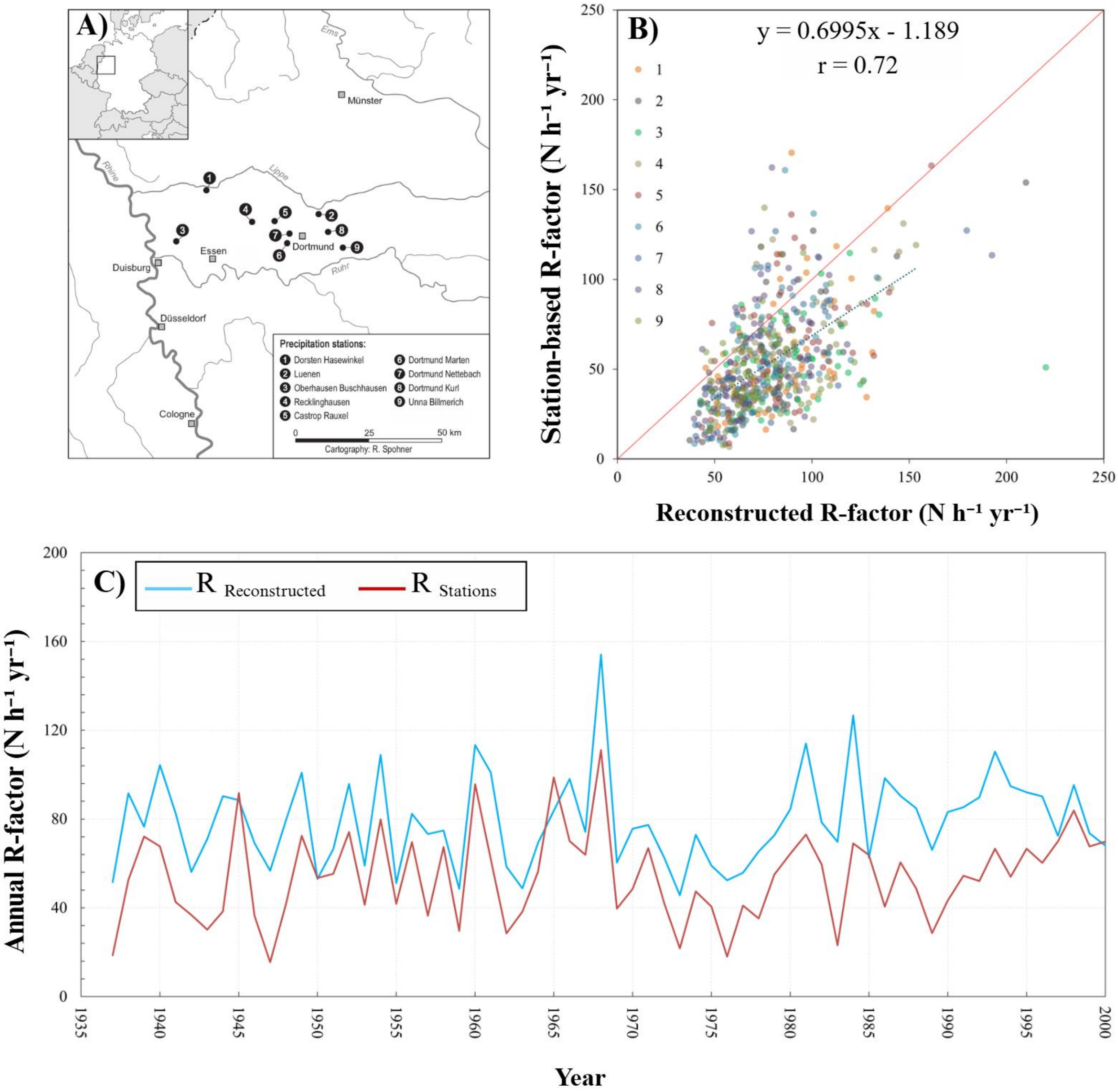


Figure 3.7. Validation of the reconstructed rainfall erosivity series against the station-based benchmark of Fiener et al., (2013). (A) Precipitation stations adapted from Fiener et al., (2013), (B) Scatter comparison between station-based annual erosivity values and the corresponding model predictions for the nine stations, (C) Temporal evolution of erosivity at the station sites, comparing the reference series from Fiener et al., (2013) with the reconstructed values, illustrating broadly consistent interannual variability and long-term trends.

3.3.4 Future rainfall erosivity forecasting using ConvLSTM model (2026–2043)

The ConvLSTM encoder–decoder was first evaluated over six independent test years (2020–2025) (Figure 3.8). The selected years span a wide hydro-climatic range, with annual precipitation totals between about 962 and 1447 mm and mean annual erosivity from 60 N h^{-1} yr^{-1} (2025) to 195 N h^{-1}

yr^{-1} (2024). Across this period, the model achieved R^2 between 0.58 and 0.70 and RMSE/MAE values mostly in the ranges 45–70 and 31–47 N h^{-1} yr^{-1}, respectively, indicating a stable and generally high level of skill throughout the test window.

A visual inspection of the spatial patterns in Figure 3.8 reveals a strong correspondence between observed and predicted erosivity maps for all six years. The ConvLSTM reliably estimated the main erosivity gradients within Germany, identifying the pre-Alpine region and central low mountain ranges as hotspots while assigning lower values to the North German Plain, consistent with long-term radar-based climatologies and European-scale assessments (Auerswald, Fischer, et al., 2019; Panagos et al., 2015). The scatter plots confirm that the bulk of grid cells cluster around the 1:1 line, showing that the relative spatial ranking of erosivity risk is preserved even in years with comparatively low mean erosivity such as 2025. Minor year-to-year fluctuations in R^2 (with slightly lower values towards the end of the 6-year output window) are within an acceptable range and are likely related to both differences in annual erosivity levels and the fact that later years are predicted without access to intervening realized erosivity fields.

Systematic errors are mainly confined to the upper tail of the distribution, where the model tends to underestimate the most extreme R-factor values in southern Germany. This behavior is expected given the predictor design: while the target erosivity fields are computed from 5-min RADKLIM intensities, the ConvLSTM is driven only by annual precipitation sums and mean temperatures, which cannot explicitly represent the sub-hourly kinetic energy peaks of convective storms (Auerswald, Fischer, et al., 2019; Fischer, Winterrath, et al., 2018). Similar underestimation of high erosivity has been reported for empirical and machine-learning models that relate R-factor to coarse-resolution rainfall statistics in regions lacking high-frequency data (Oliveira et al., 2013; Yin et al., 2015). In this sense, the slight smoothing of extremes is a physically consistent consequence of the available predictors rather than a deficiency of the ConvLSTM architecture.

Taken together, the results indicate that the ConvLSTM captures much of the spatial and interannual variability of annual rainfall erosivity from aggregated climate covariates, although it tends to underestimate the most extreme R-factor values. We therefore use the model to generate 2026–2043 forecasts, while interpreting the highest projected extremes with caution.

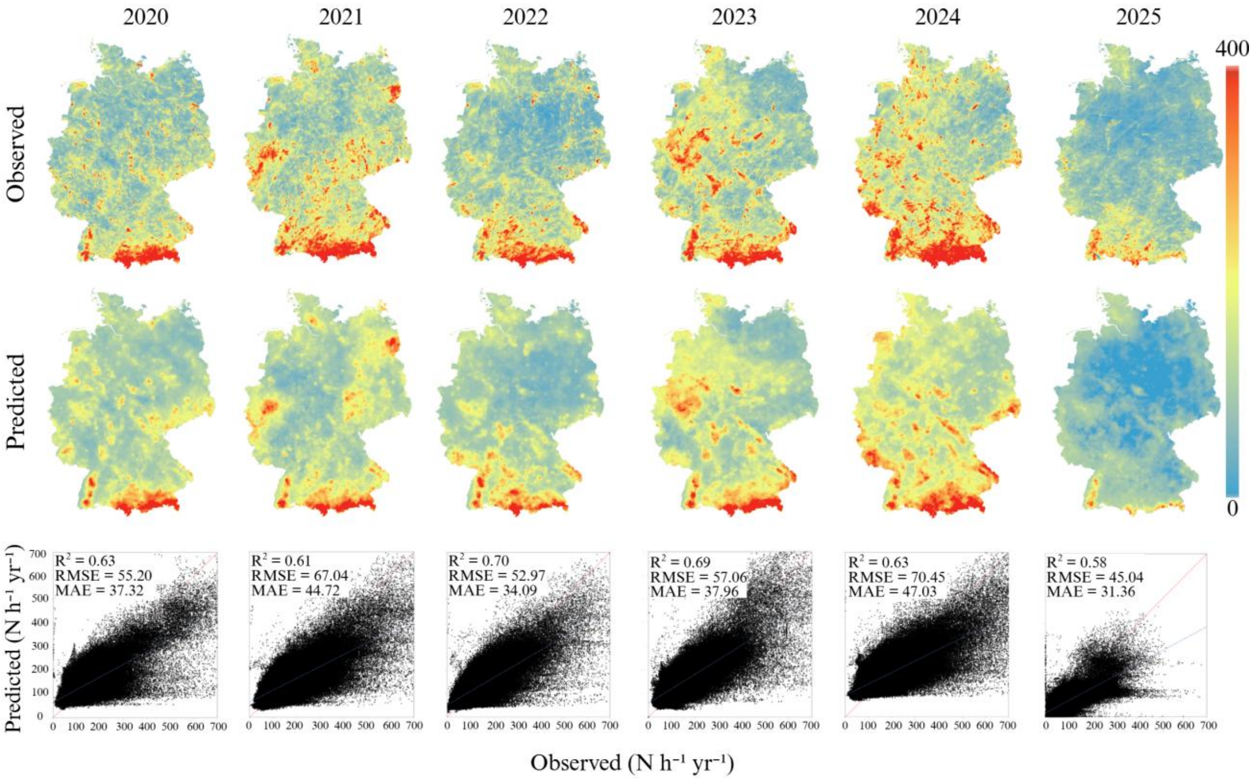


Figure 3.8. Spatiotemporal patterns of observed and ConvLSTM-predicted annual rainfall erosivity across five test years. Top row: Annual rainfall erosivity maps derived from 5-minute RADKLIM radar data (N h^{-1} yr^{-1}). Middle row: Corresponding spatially predicted erosivity values generated by the ConvLSTM model (N h^{-1} yr^{-1}). Bottom row: Scatter plots of predicted versus observed R-factor values (N h^{-1} yr^{-1}), showing the 1:1 identity line (red) and linear regression fit (blue).

The full time series of mean annual rainfall erosivity for Germany (Figure 3.9) combines the CNN-based reconstruction (1931–2000), the RADKLIM-derived observations (2001–2025) and the ConvLSTM forecasts (2026–2043). The corresponding spatial forecast maps for each 6-year window are provided in the Appendix (Figure 3.S2). A non-parametric change-point analysis based on Pettitt's test (Pettitt, 1979) identified 2005 as the most likely breakpoint, indicating a statistically distinct shift from a phase of comparatively lower erosivity to a later phase characterized by persistently higher values. Consistent with the analysis in Section 3.3.3, the blue reconstruction exhibits relatively low and variable erosivity in the early and mid-20th century, followed by a marked rise from the 1960s onward. This upward tendency is amplified in the RADKLIM era (orange line), where the cluster of high-erosivity years after 2000 confirms that recent decades have been characterized by substantially more erosive conditions than the pre-1960 baseline, in agreement with independent gauge- and radar-based studies for Germany and Europe (Auerswald, Fischer, et al., 2019; Panagos et al., 2015). From a hazard-relevant perspective, years with an annual R-factor exceeding 150 N h^{-1} yr^{-1} were rare in the first phase, occurring only three times between 1931 and 2004 (1981, 1994, and 2002), corresponding to a return period of approximately 24.7 years. In contrast, 16 such exceedances were recorded between 2005 and 2043, equivalent to a return period of about 2.4 years. This pronounced shortening of the return period for very high-erosivity years indicates that extreme erosive conditions have become substantially more frequent since the detected change point. The mean R-factor also rises from 117 N h^{-1} yr^{-1} in 1931–2004 to 147 N h^{-1} yr^{-1} in 2005–2043, representing an increase of about 26%. This relative change is broadly consistent with pan-European assessments under RCP4.5, which project order-of-magnitude increases of about 15–20% in mean rainfall erosivity over large parts of Northern and Central Europe by mid-century (Panagos et al., 2017). Together, these changes indicate that not only the mean erosivity but also the frequency of very high-erosivity years has increased markedly, consistent with recent evidence for more frequent and intense heavy-rainfall events across Central Europe (Ehmele et al., 2020; Zeder & Fischer, 2020).

The ConvLSTM projections (green line) indicate that this elevated erosivity regime is likely to persist under the RCP4.5 forcing scenario, with mean R-factor values remaining well above the reconstructed pre-2005 level and broadly comparable to, or slightly higher than, the recent observed period. Quantitatively, the projected values do not suggest any return to the low-erosivity conditions that characterized much of the early 20th century. Interannual variability in the forecasts is of similar magnitude to that observed since the 1990s, suggesting continued alternation of relatively quiet years and strongly erosive years rather than a monotonic rise. It is noteworthy that the projected increases emerge despite a tendency of the ConvLSTM configuration to underestimate the highest R-factor values relative to RADKLIM and despite the use of a moderate emission pathway (RCP4.5); higher-end scenarios and convection-permitting simulations for Central Europe under RCP8.5 indicate substantially larger future increases, particularly in river basins with strong convective activity (Uber et al., 2024). This behavior is consistent with regional and continental assessments that project further increases or sustained high levels of rainfall erosivity in Central Europe during the 21st century, driven primarily by intensification of heavy precipitation under a warming climate (Fiener et al., 2013; Panagos et al., 2017).

Viewed together, the three segments form a coherent narrative of gradual but persistent strengthening of erosive rainfall from the mid-20th century onward, with no indication in the ConvLSTM forecasts of a return to the low-erosivity conditions that characterized the 1930s–1950s. In magnitude, the projected German-wide mean for 2026–2043 lies at the lower but still substantial end of the range reported for Central Europe, where increases of up to ~80% in rainfall erosivity have been simulated for specific river basins under high-end emission scenarios (Uber et al., 2024), suggesting that our results represent a conservative estimate under moderate warming. The long, model-supported perspective from 1931 to 2043 therefore reinforces the interpretation that anthropogenic warming has already shifted Germany into a more erosive rainfall regime and that, under a moderate emissions pathway, this regime is likely to be maintained in the coming decades. At the same time, the forecast segment should be interpreted as a conditional projection rather than a deterministic continuation: it inherits uncertainties from the regional climate model ensemble, from the CNN-based reconstruction that defines the pre-radar baseline, and from the ConvLSTM's iterative use of its own predictions in later forecast windows. In this sense, the agreement between reconstructed, observed, and projected trends is encouraging, but it should be viewed as supportive rather than definitive evidence that the model captures the full magnitude of future change in rainfall erosivity over Germany.

Direct quantitative comparison with Auerswald and Fiener, (2024) is precluded by fundamental differences in emission pathway (RCP4.5 versus RCP8.5), ensemble design (single bias-adjusted RCM versus seven EURO-CORDEX models), and methodological pipeline; pixel-wise or year-to-year correlation between the two datasets would therefore not constitute a meaningful validation metric. At the qualitative level, however, both studies project sustained or increasing rainfall erosivity for Germany through the mid-21st century, accompanied by pronounced interannual variability relative to the mid-20th-century baseline. The projected increase of approximately 26% between the 1931–2004 and 2005–2043 periods is expectedly more conservative than the changes reported by Auerswald and Fiener, (2024) under the more strongly forced RCP8.5 scenario, consistent with the lower radiative forcing imposed by RCP4.5. In this respect, the present projections align with the broader literature in the direction of change, even where a direct numerical comparison of magnitudes and spatial patterns is not warranted.

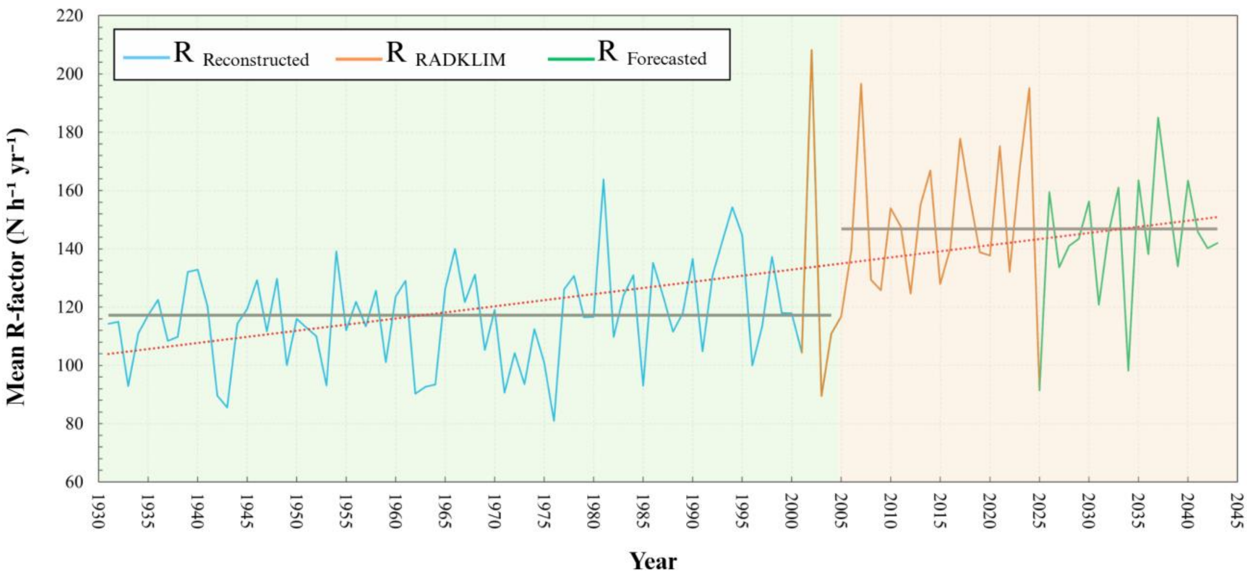


Figure 3.9. Long-term evolution of mean annual rainfall erosivity in Germany from 1931 to 2043, combining the CNN-based reconstruction for 1931–2000 (blue line), the RADKLIM-derived observations for 2001–2025 (orange line), and the ConvLSTM projections for 2026–2043 (green line). The dotted red line indicates the overall linear trend, and the grey horizontal segments denote the mean erosivity levels of statistically distinct phases identified by Pettitt's change-point test. The change point separates an earlier lower-erosivity regime (1931-2004) from a later higher-erosivity regime (2005-2043), highlighting the shift toward more erosive conditions in recent decades.

3.3.5. Limitations and uncertainties

A structural limitation of the present framework concerns the treatment of snowfall in both the observational and modelling components. By construction, the R-factor is defined for liquid precipitation events and does not apply to snowfall, because solid precipitation has very low raindrop kinetic energy and therefore contributes little direct erosive forcing. In practice, this implies that wintertime precipitation only enters the erosivity budget when it falls as rain with sufficient intensity to satisfy the DIN event criteria, while snowfall is effectively ignored. At the same time, quantitative precipitation estimates from weather radar are typically more uncertain during snowfall than during rain, due to differences in particle characteristics and in the underlying reflectivity–precipitation relationships, so that RADKLIM-based winter precipitation may carry larger systematic and random errors than rainfall-dominated seasons. Under climate change, a warming-induced shift from snowfall to rainfall at near-freezing temperatures is expected to increase the fraction of cold-season precipitation that contributes to erosivity, effectively lengthening the erosive season. The inclusion of annual temperature and precipitation-derived indices as predictors partly accounts for such phase-related changes, because they capture both thermodynamic controls and evolving intensity statistics; however, uncertainties in the historical and projected snow–rain partitioning remain and are not resolved explicitly by the present framework. Consequently, the long-term trends and future increases reported here should be interpreted as conditional on the assumed phase partitioning in the underlying climate and radar products, with the possibility that the true shift in erosive forcing is somewhat larger in regions where snowfall has already substantially transitioned to rainfall.

3.4. Conclusion

By combining high-resolution radar information, long-term climate data, and near-future projections into a consistent framework, this study provides an updated, intensity-aware view of rainfall erosivity that spans both the historical and projected climate regimes. In this sense, the work offers a century-scale perspective on how erosive forcing has evolved and is likely to evolve further under ongoing warming. While powerful, deep learning approaches remain fundamentally empirical and data dependent. The reconstruction and forecasts inherit uncertainties from the training data period, the choice of predictors, and the extrapolation beyond historical experience. Future model iterations would benefit from incorporating explainable artificial intelligence techniques to better elucidate the learned relationships between climatic covariates and erosivity patterns. Future work should therefore aim to (i) develop explainable and interpretable deep learning architectures that provide clearer insights into the physical drivers captured by the models; (ii) extend the framework with convection-permitting climate simulations; and (iii) couple the erosivity products with erosion, runoff, and land-management models to assess downstream impacts. Further validation against more independent long-term station records and experimental plots would also help to refine the models, quantify uncertainties, and support the development of revised national guidelines that better reflect the emerging erosivity regime in Germany and comparable regions.

Funding

This work was supported by the Deutsche Forschungsgesellschaft (DFG) [No. SCHO 739/25-1].

4. Manuscript II: GLUE assessment of WaTEM/SEDEM

Chapter 4 is based on co-authored **Manuscript II,** the peer-reviewed journal publication:

Seufferheld, K. D., Batista, P. V., Shokati, H., Scholten, T., & Fiener, P. (2026). Soil, 12(1), 301-319, https://doi.org/10.5194/soil-12-301-2026.

This Chapter shifts the focus from climatic forcing to landscape-scale process representation by evaluating the spatially distributed erosion and sediment transport model WaTEM/SEDEM under soil-conservation-optimized conditions in micro-scale agricultural watersheds. Using a unique eight-year monitoring dataset from six adjacent experimental watersheds in Scheyern, Southern Germany, the study assesses how well WaTEM/SEDEM can reproduce very low sediment yields in a system where both in-field conservation practices and landscape structures intentionally reduce sediment connectivity. In this way, it directly addresses Research Questions 2 and 3 concerning the effectiveness and structural limitations of WaTEM/SEDEM in conservation-oriented settings and the dependence of model performance on spatial and temporal aggregation.

The study applies a Generalized Likelihood Uncertainty Estimation (GLUE) framework to account explicitly for observational uncertainties and parameter equifinality, thereby evaluating model behavior across a large ensemble of parameter sets rather than a single calibration. This allows a nuanced assessment of model plausibility from annual plot scales up to multi-year watershed scales and highlights the conditions under which WaTEM/SEDEM is fit for long-term conservation planning, as opposed to detailed event- or year-scale prediction.

The chapter is structured as follows. Section 4.1 introduces the study context, research objectives and the rationale for using WaTEM/SEDEM and GLUE in conservation-optimized watersheds. Section 4.2 describes the test site, monitoring data, model setup, parameter ranges and the implementation of the GLUE framework, including model conditioning and evaluation metrics. Section 4.3 reports the results on model performance across scales, the behavioral parameter space and spatial patterns of erosion and deposition, followed by a detailed discussion in Section 4.4 on uncertainties, model limitations, spatial dynamics and parameter distributions. Section 4.5 concludes with implications for the fit-for-purpose use of WaTEM/SEDEM in conservation-oriented agricultural landscapes.

Manuscript II: A GLUE-based assessment of WaTEM/SEDEM for simulating soil erosion, transport, and deposition in soil conservation optimised agricultural watersheds

Abstract

Soil erosion models are important tools for soil conservation planning. Although these models are generally well-tested against plot and field data for in-field soil management, challenges arise when scaling up to the landscape level, where sediment trapping along landscape features becomes increasingly critical. At this scale, a separate analysis of model performance for representing erosion, sediment transport, and deposition processes is both challenging and often lacking. Here, we assessed the capacity of the spatially distributed erosion and sediment transport model WaTEM/SEDEM to simulate sediment yields in six highly instrumented micro-scale watersheds ranging from 0.8 to 7.8 ha, monitored over eight years from 1994 to 2001, in Southern Germany. The watersheds were composed of two groups: four field-dominated watersheds characterised by arable land with minimal landscape structures, and two structure-dominated watersheds featuring a combination of arable land and linear landscape structures (mainly grassed waterways along thalwegs) that minimise sediment connectivity. Arable fields in both watershed groups were managed for soil conservation, including no-till and optimised crop rotations. A Generalised Likelihood Uncertainty Estimation (GLUE) framework was employed to account for measurement and model uncertainties across multiple spatiotemporal scales. Our results show that while WaTEM/SEDEM captured the magnitude of the very low measured sediment yields in the monitored watersheds, the model did not meet our pre-defined limits of acceptability when operating on annual time steps. Model performance improved substantially when outputs were averaged over the eight-year monitoring period, with mean absolute errors of 0.14 t ha^{-1} yr^{-1} for field-dominated and 0.29 t ha^{-1} yr^{-1} for structure-dominated watersheds. Our findings demonstrate that WaTEM/SEDEM can represent the influence of soil conservation practices on reducing soil erosion and sediment yield in our study area. However, the model is fit for long-term conservation planning at larger spatial scales and not for precise annual predictions for individual micro-scale watersheds with specific conservation practices even if high-resolution, high-quality input data are available for parameterisation.

4.1. Introduction

Soil erosion by water is a major threat to global soil health and associated ecosystem functions and services, endangering agricultural sustainability and food security (Montanarella et al., 2016; Quinton & Fiener, 2024; Rickson et al., 2015). Although the problem of accelerated soil erosion has been known for a long time and a wide variety of soil conservation practices have been tested and implemented locally for many decades, adoption remains limited due to economic constraints, lack of technical knowledge, and insufficient policy support (Aghabeygi et al., 2024; Quinton & Fiener, 2024). This is particularly problematic in regions where agricultural intensification (e.g. field consolidation, soil compaction Brus and Van Den Akker, 2018; Keller et al., 2019; Foucher et al., 2014; Wang et al., 2022) and the increase in frequency and intensity of extreme precipitation events due to climate change (Auerswald and Fiener, 2024; Hosseinzadehtalaei et al., 2020; Myhre et al., 2019) are exacerbating the erosion hazards.

Overall, effective soil conservation relies on two complementary strategies: (i) in-field soil conservation and (ii) sediment transport control structures along the flow pathways. In-field practices focus on increasing soil surface cover by vegetation to prevent soil detachment by raindrop impact and overland flow. Such practices include optimised crop rotations, using cover crops, and soil residue management (Andersson & D'souza, 2014). Sediment transport control practices consist of structures installed along the runoff pathway to increase infiltration, sediment trapping, and hence minimise sediment connectivity. Typical structures are vegetative filter strips (Gumiere et al., 2011), grassed waterways (Fiener & Auerswald, 2003), retention ponds (Fiener et al., 2005), or generally optimised layout of fields along slopes (Oost et al., 2000).

Soil erosion models are potentially valuable tools for identifying erosion-prone areas and developing what-if scenarios, allowing stakeholders to assess different configurations of on- and off-site soil conservation practices. This enables the identification of optimal intervention strategies before implementation. Diverse models have been developed and applied for this purpose, ranging from empirical or conceptual to process-oriented (e.g. Eekhout et al., 2018; Smith et al., 2018; Nearing, 2013; Dymond et al., 2010; Hessel & Tenge, 2008). As indicated by erosion modelling reviews (Batista et al., 2019; Borrelli et al., 2021), the most widely used erosion model is still the empirical Universal Soil Loss Equation (USLE; Wischmeier & Smith, 1978) and its revisions and regional adaptations, such as the revised USLE (RUSLE; Renard, 1997) and the German ABAG (Allgemeine Bodenabtragsgleichung, German for Universal Soil Loss Equation; Din-Normenausschuss, 2022; Schwertmann et al., 1987).

While these USLE-type models have been adapted to calculate spatially distributed erosion rates, they are limited to calculating potential soil loss without considering sediment transport processes and downslope deposition. To overcome this limitation, the Water and Tillage Erosion Model and the Sediment Delivery Model (WaTEM/SEDEM) (Oost et al., 2000; Rompaey et al., 2001; Verstraeten et al., 2002) was developed. WaTEM/SEDEM combines the USLE-technology with spatially distributed sediment transport and deposition modelling. The performance of the model has been tested using sediment trapping in reservoirs (e.g. Hlavčová et al., 2018), sediment yield in small rivers of mesoscale catchments (e.g. Batista et al., 2022; Rehm & Fiener, 2024), and long-term erosion and deposition patterns derived from radionuclides (e.g. Oost et al., 2000; Wilken et al., 2020). However, to the best of our knowledge, the suitability of WaTEM/SEDEM for representing soil erosion, transport, and deposition processes within soil conservation settings (e.g. no-till, cover crop, optimised crop rotations) combined with sediment transport control structures to reduce sediment connectivity (e.g. grassed waterways, retention ponds), has not been thoroughly tested against measured data.

One difficulty is that most plot-scale data do not account for landscape features, as they are typically not included in plots. Conversely, watershed outlet data may integrate the effects of both in-field soil conservation and sediment transport control structures into one lumped measurement, which makes it difficult to disentangle their individual contributions to (dis)connecting the sediment cascade. Long-term monitoring data from micro-scale watersheds (1-10 ha) offer the opportunity to evaluate in-field soil conservation practices separately from sediment transport control structures implemented at field to the landscape scale (Choudhury et al., 2022; Fiener & Auerswald, 2018). However, such datasets are rare (Fiener et al., 2019a) , and erosion and sediment delivery models have hardly been tested under these conditions.

Regardless of the spatial scale in which erosion is monitored, it is important to note that perfect observational data do not exist. All measurements include errors stemming from instrumental precision, temporary malfunctioning, and data handling and processing. These uncertainties have important implications for evaluating erosion models, which cannot be expected to be better than the observational data used for model conditioning and testing (Beven & Lane, 2022; Beven, 2019). One approach for evaluating (uncertain) environmental models is the Generalized Likelihood Uncertainty Estimation (GLUE) framework (Beven & Binley, 1992). GLUE acknowledges that it is not possible to identify a single calibrated parameter set as "correct". Rather, all parameter combinations that produce results within given limits-of-acceptability cannot be rejected (Beven & Lane, 2022). Contrarily, if not a single model realisation encompasses the uncertainty bounds of the observational data, non-behavioural models or model structures can be rejected, which might lead to improvements in terms of understanding and modelling.

Here we employ a rejectionist limits-of-acceptability approach within the GLUE framework to test the widely used WaTEM/SEDEM model for representing soil erosion, transport, and deposition in soil conservation optimised agricultural watersheds, i.e. featuring a combination of in-field practices and sediment transport control structures. Specifically, we aimed to: (i) identify limits of acceptability of model error derived from measurement uncertainty in order to reject non-behavioural model realisations; (ii) develop a two-stage model conditioning process to test the fitness for purpose of WaTEM/SEDEM for representing the effects of both in-field conservation practices and sediment control structures on erosion, transport, and deposition; and (iii) test WaTEM/SEDEM under different levels of temporal (annual vs. eight-year means) and spatial aggregations (individual vs. grouped watersheds). We accomplish these objectives using a comprehensive, long-term monitoring dataset from Southern Germany, which provides high resolution model inputs (e.g. precipitation, crop-specific daily soil cover, etc.), as well as continuous surface runoff and sediment flux data for six micro-scale watersheds under optimised soil conservation and reduced sediment transport (Auerswald et al., 2001; Auerswald & Fiener, 2019; Fiener et al., 2019a).

4.2. Material and methods

4.2.1 Test site

The test site is part of an experimental farm located in Scheyern, Southern Germany (48°29'45.1"N, 11°26'23.6"E; about 470 m above sea level). It is part of Bavaria's tertiary hill region, an important and productive agricultural landscape in Central Europe. The rolling topography is characterised by predominantly east-facing slopes ranging from 0.4° to 11.5° (Wilken et al., 2019b). Climate conditions include a mean annual temperature of 8.4 °C and mean annual precipitation of 834 mm (1994-2001), with the highest precipitation occurring between May and July (Fiener et al., 2019a). Management practices at the farm follow a comprehensive soil conservation philosophy based on two main principles: (i) keeping arable soils covered as long as possible and (ii) reducing hydrological and sedimentological connectivity as far as possible (Fiener et al., 2019a). Within the watersheds, soils consist predominantly of loamy or silty loamy Cambisols (World Reference Base for Soil Resources (WRB), Schad et al., 2022).

The research area comprises six micro-scale watersheds (W01-W06) with a total area of 24 ha and four agricultural fields (F15-F18, Fig. 4.1). The six watersheds exhibit different landscape connectivity characteristics: W01–W04 (0.8 to 4.2 ha) are classified as field-dominated systems. Despite the presence of retention ponds at the outlets of W01 and W02, these watersheds have no internal linear landscape structures, resulting in sediment flux pathways governed primarily by topography and the management of the arable fields. In contrast, W05 and W06 are classified as structure-dominated systems due to the presence of a grassed waterway along the thalweg. Watershed W06 (5.7 ha) constitutes the upper part of the larger watershed W05 (7.8 ha) (Fiener et al., 2019a).

Three key conservation practices were implemented to minimise hydrological and sedimentological connectivity: (i) an optimised field layout with reduced field sizes adapted to the steep slopes, (ii) retention ponds at watershed outlets, and (iii) a grassed waterway along the main thalweg of W05 and W06 (Fiener et al., 2019a). The retention ponds were located at the outlets of watersheds W01, W02, W05, and W06 (Fig. 4.1). Sediment trapping efficiency measurements were conducted for these ponds, revealing an average of 70 ± 14 % (Fiener et al., 2005). Additionally, continuous monitoring systems were installed at the outlet of each micro-scale watershed to measure runoff and sediment yield. The distinction between field-dominated and structure-dominated watersheds will be used consistently throughout this study.

All fields within the watersheds (F15 to F18 in Fig. 4.1) were managed using no-till practices with a crop rotation of winter wheat (*Triticum aestivum* L.), maize (*Zea mays* L.), winter wheat, and potatoes (*Solanum tuberosum* L.). This rotation was staggered across the fields, meaning that while the sequence was identical, the specific crop grown each year varied between fields. After winter wheat, mustard was sown as a cover crop. In the case of potatoes, the mustard was sown into the potato dams built in autumn, while direct seeding into the frost-killed mustard was performed in the following year.

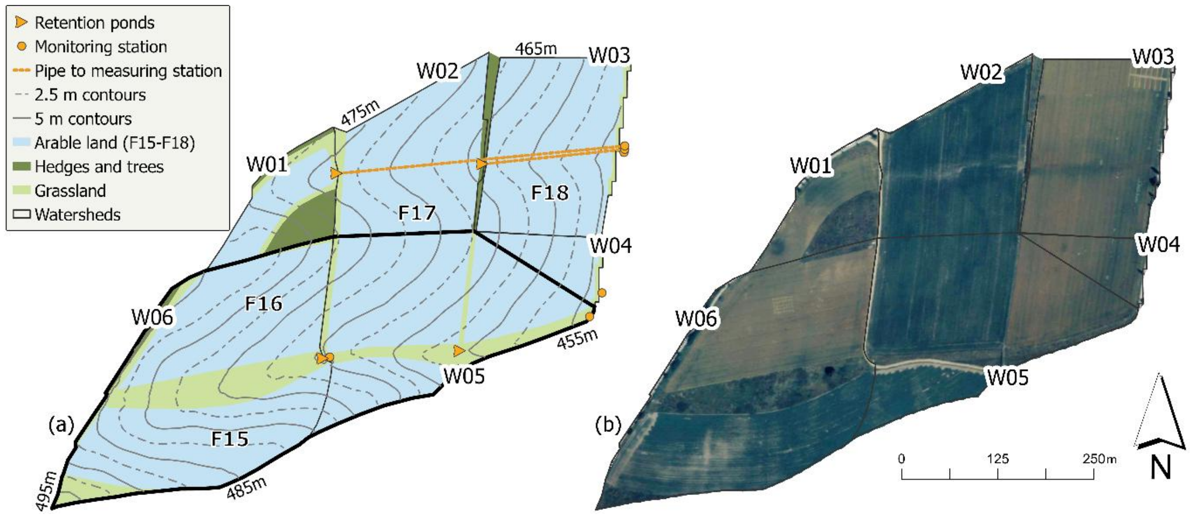


Figure 4.1. (a) Schematic land use and topography of the experimental farm in Scheyern, Bavaria, with flow direction from west to east. W01-W06 are abbreviations for six watersheds; F15-F18 are the fields located in these watersheds. Note: Watershed W05 (thick line) includes the upslope watershed W06, as W06 is cascading into W05. (b) Aerial photograph of the study area shows the land use patterns and field boundaries on the Scheyern farm in 2002.

4.2.2 Erosion monitoring data

The study utilised a unique erosion monitoring dataset acquired between 1994 and 2001. This comprehensive dataset, as well as metadata, are provided by (Fiener et al., 2019a). All spatial data were resampled to a consistent 5 m by 5 m grid resolution, matching the digital elevation model (DEM) provided in the dataset (Wilken et al., 2019b). The temporally dynamic input data included daily soil cover measurements and high-resolution precipitation data recorded at 1-minute intervals from up to 11 monitoring sites (Wilken et al., 2019a). Additional details regarding these input parameters are provided in section 4.2.4 below.

For model testing, we used continuous sediment yield data from the six micro-scale watersheds (W01-W06) between 1994 and 2001 (Fig. 4.1). Runoff and suspended-sediment loads were monitored with a measuring system based on a Coshocton-type wheel sampler (precision ± 10%; Carter & Parsons, 1967; Fiener & Auerswald, 2003). The device continuously diverted an aliquot of approximately 0.5 % from the total flow that left the watersheds through underground-tile outlets with a diameter of 15.6 cm and 29 cm (Fig. 4.1) into storage tanks (1.0 – 3.5 m^3). Runoff volumes were measured after each event. Sediment yield was calculated from runoff volumes and sediment concentrations derived from homogenised tank samples dried at 105°C. At lower rates (< 0.5 L s^{-1}) the system slightly over-estimated runoff, but these small events contributed negligibly to the cumulative water and sediment budgets. Under sampling during very high flows was avoided by (i) employing large wheels (Ø 61 cm) and (ii) the flow-dampening effect of the retention ponds situated immediately upstream of each outlet (Fiener & Auerswald, 2003).

4.2.3 Soil erosion modelling

The WaTEM/ SEDEM version used in this study consists of two main components: (i) WaTEM, which implements a spatially distributed German adaption of the USLE (Schwertmann et al., 1987; Din-Normenausschuss, 2022), and (ii) SEDEM, which incorporates a transport capacity (TC) equation (Eq. 4.4) and a routing algorithm for sediment re-distribution based on the DEM (Oost et al., 2000; Rompaey et al., 2001; Verstraeten et al., 2002). To implement WaTEM/ SEDEM within the GLUE-framework, the original Delphi code-based model was translated to Python 3.12 and was run in PyCharm 2024.1 (Community Edition), which substantially improved computational speed through parallel processing and allowed for easier data handling. To ensure reproducibility and accuracy, we compared the Python implementation against the original Delphi codebase at each individual step of the translation process, verifying that it produced identical outputs for these test cases. Although the Python implementation includes tillage erosion and stream initiation calculations, these components were not utilised in the present study.

The model was applied for the period from April to October of each year from 1994 to 2001, excluding periods potentially affected by snowmelt erosion and prolonged surface runoff from return flow (Fiener et al., 2019a). While the colder months contributed 10.7 % of the total measured sediment yield (Fiener et al., 2019b), our analysis focused on the dominant water erosion period during heavy rainfall months. Each micro-scale watershed was separately modelled.

4.2.4 Potential erosion

In contrast to the original WaTEM/ SEDEM (Oost et al., 2000; Rompaey et al., 2001; Verstraeten et al., 2002), in which the USLE factors are derived according to the RUSLE approach (Renard, 1997), we calculated the USLE factors as calculated according to their German adaptation (Eq. 4.1). This approach is specifically adapted to the soils and climatic conditions of Central Europe and represents the standard methodology for the study region (Schwertmann et al., 1987; Din-Normenausschuss, 2022).

$$A = R * K * LS * C * P \tag{4.1}$$

Where A is the potential erosion (t ha^{-1} yr^{-1}), R the rainfall erosivity factor (N h^{-1} yr^{-1}), K the soil erodibility factor (t ha^{-1} h N^{-1}), LS the slope length and steepness factor (dimensionless), C the cover management factor (dimensionless), and P the agricultural practices factor (dimensionless).

The high-resolution rainfall data from eleven (1994–1997) and two (1998–2001) precipitation monitoring stations located in the research area were used to calculate the rain erosivity factor (R-factor) (Wilken et al., 2019a). Following the German adaptation of the USLE, rainfall events were considered erosive if they met at least one of two criteria: (i) total rainfall amount ≥ 10 mm (in contrast to the 12.7 mm threshold of the standard USLE, Wischmeier & Smith, 1978) or (ii) maximum 30-minute intensity ≥ 10 mm h^{-1}. Individual events were separated by at least 6 hours without rainfall (Schwertmann et al., 1987; Din-Normenausschuss, 2022). The calculated rainfall erosivities per monitoring station were interpolated to 5 m by 5 m resolution maps using inverse distance weighting, and the spatially distributed values ranged between 65.90 and 155.10 N h^{-1} yr^{-1} across the eight-year study period.

Soil erodibility (K factor) values were computed following Auerswald et al. (2014) and already provided in the monitoring data set (Auerswald et al. 2019a). The values, originating from a 50 by 50 m sampling grid, were spatially interpolated using ordinary kriging to generate a continuous surface with a 5 m by 5 m resolution grid. The resulting K factor values across the study area ranged from 1.8 to 4.6 t ha^{-1} h N^{-1}.

The slope length and slope steepness factor (LS factor) was calculated based on the DEM using the approach by Desmet & Govers (1996). When calculating the LS factor for W01, the shrubbed area (Fig. 4.1) was excluded due to its negligible runoff contribution. Additionally, we calculated the LS factors for W02 and W03 separately from their upslope catchments (i.e. W01 and W02), since their runoff was directed via underground pipes to the monitoring stations (see Fig. 4.1).

To account for the temporal dynamics of soil protection, we calculated the C factor based on soil cover and rainfall erosivity. This calculation involved deriving continuous daily soil cover data from field measurements, determining the corresponding soil loss ratios (SLR), and weighting these by the seasonal rainfall erosivity.

From 1994 to April 1997, bi-weekly crop and residue cover measurements were conducted during growing seasons, with monthly measurements during autumn and spring and additional observations before and after soil management operations. To obtain continuous time series, the point data was linearly interpolated to generate daily cover values (Auerswald et al. 2019b). For the subsequent period (April 1997–2001), we applied standardised daily crop development and residue cover curves, which were derived from the daily crop and residue cover values observed during the detailed monitoring phase in combination with management information (mainly sowing and harvest dates) observed between 1997 and 2001(Auerswald et al. 2019b; Fiener et al., 2019a). Total soil cover was calculated with residues protecting portions of the otherwise exposed soil according to:

$$Co_{tot} = Co_{crop} + \left(100 - Co_{crop}\right) * \frac{Co_{res}}{100} \qquad (4.2)$$

Where Co_{tot} is the total soil cover (%), Co_{crop} the cover of the growing crop on the respective field (%), and Co_{res} the measured soil cover of the residues (%).

Figure 4.2 illustrates the total soil cover on the respective fields with monthly rainfall erosivity.

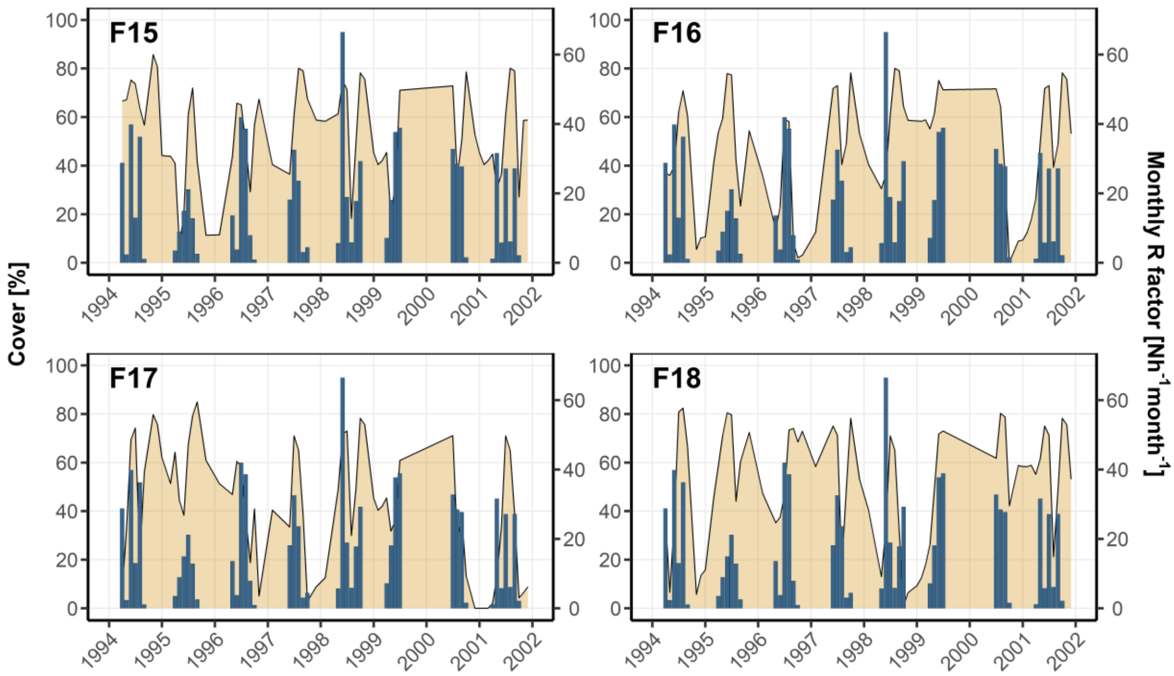


Figure 4.2. Each field's total soil cover (Residues and crops). Blue bar plots (monthly sum) show monthly R-factors.

The soil loss ratio (*SLR*) quantifies the protective effect of soil cover by comparing potential soil loss under a given vegetation condition to that under standardised fallow conditions (Schwertmann et al., 1987; Wischmeier & Smith, 1978). While the *SLR* traditionally considers five crop growth stages, from bare soil (0% cover) to full canopy coverage (75-100% cover), we also considered crop residue cover. Determining field-specific *SLR* values involved categorising soil cover into the five growth stages and assigning corresponding *SLR* values. As no-till was applied at the research farm, lower *SLR* values were assigned than in conventional systems due to increased soil surface protection. These *SLR* values were obtained from Schwertmann et al. (1987) and adapted based on our expert knowledge regarding the soil conservation practices in the Scheyern experimental farm (Fiener & Auerswald, 2007; Fiener et al., 2019a).

The annual C factor was calculated by multiplying the monthly proportion of the R factor with the average *SLR* value of the respective month:

$$C = \sum_{month=1}^{12} SLR_{month} * R_{prop,month} \tag{4.3}$$

Where C is the cover management factor (dimensionless), SLR_{month} the average soil loss ratio value of the respective month and $R_{prop,month}$the proportion of the annual rainfall erosivity factor for the respective month (dimensionless).

Figure 4.3 shows the annual C and R factor throughout the entire study period.

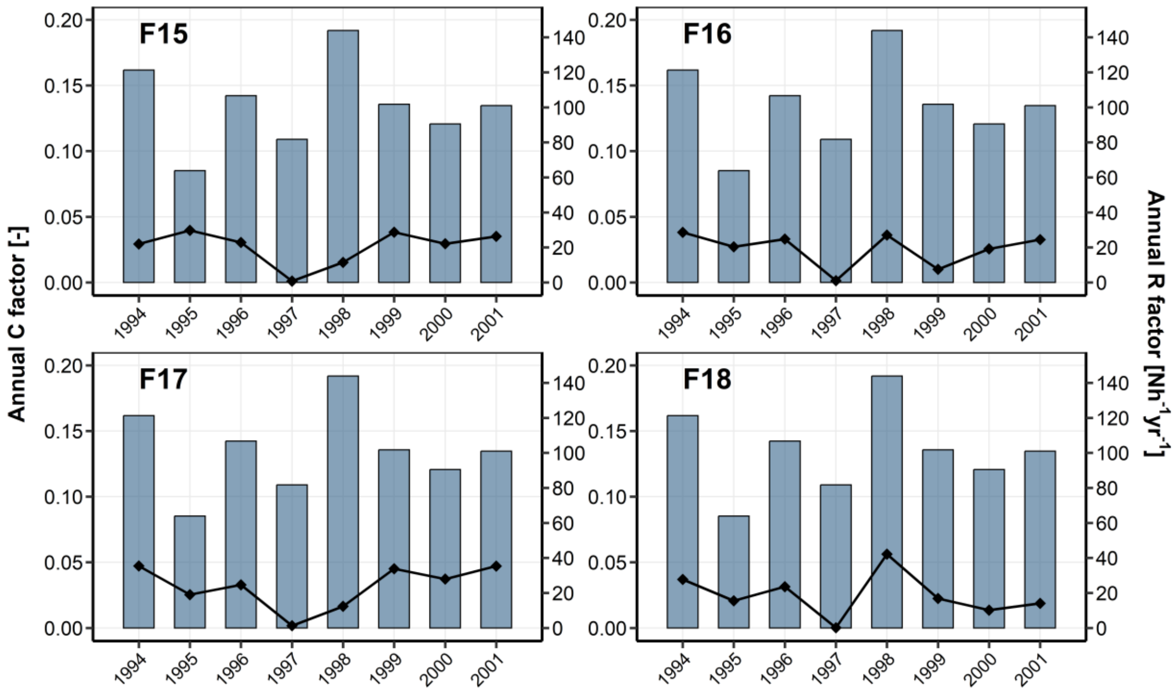


Figure 4.3. Annual C factor (black dots and lines) and R factor (blue bars) for the individual fields F15 to F18 over eight years.

The support practices factor (P factor) was not specifically parametrised for contour-seeding because of field heterogeneity, i.e. not all parts of a single field were contour-seeded, and/or the absence of specific P factor values for structures such as the potato dams. Furthermore, the field geometries often result in high L factors that often exceed the critical slope length limit for effective contouring defined in the German USLE (Schwertmann et al., 1987; Din-Normenausschuss, 2022). Hence, the effective P-factor converges towards 1.0. We accounted for the uncertainty stemming from this lack of parameter representation as part of the model conditioning process (see section 4.2.4 below).

4.2.5 Sediment transport and deposition

Transport capacity (TC) quantifies the maximum amount of sediment transported through a grid cell without deposition. When the incoming sediment load into a raster cell exceeds TC, the excess material is deposited within the cell, whilst the remaining portion continues its downstream movement. TC was calculated with the approach proposed by Rompaey et al. (2001):

$$TC = k_{TC} * R * K * (LS - S_{IR}) \tag{4.4}$$

With:

$$S_{IR} = 4.12 * S_m^{0.8} \quad (4.5)$$

where TC is the transport capacity (t ha^{-1}), k_{TC} the transport capacity coefficient (m) described below, R the rainfall erosivity factor in (N h^{-1} yr^{-1}), K is the soil erodibility factor (t ha^{-1} h N^{-1}), LS the slope length and steepness factor (dimensionless), S_{IR} the interrill slope gradient factor (dimensionless) and S_m the slope (m m^{-1}).

The transport capacity coefficient (k_{TC}) represents the theoretical upslope distance required for sediment generation to reach maximum TC at a given raster cell under the assumption of uniform slope and erosion conditions (Rompaey et al., 2001). The transport capacity coefficient depends on surface roughness and therefore differs according to land use and management. In our model parameterisation, we distinguish between higher values for arable land ($k_{TC/A}$) and lower values for grassland ($k_{TC/G}$; along field borders and in grassed waterways), which were subjected to the GLUE-based analysis (see Section 4.2.7).

WaTEM/SEDEM's hillslope sediment transport module employs a multiple flow routing algorithm, which distributes sediment from individual cells to their downslope neighbours based on Quinn et al. (1991). The algorithm calculates local slopes to eight neighbouring cells and applies specific weighting factors: 0.50 for orthogonal neighbours and 0.35 for diagonal neighbours. The sediment flux is distributed proportionally to the weighted slope values of all cells at equal or lower elevations.

In this study, we implemented the parcel connectivity (p_{con}) parameter specifically at field boundaries. p_{con} reduces the contributing upstream area by a value [%] at these transitions (Notebaert et al., 2006). This reduction has a dual effect: (i) it directly lowers the slope length part of the LS factor, thereby decreasing the potential erosion for subsequent downstream cells, and (ii) it affects TC, which is calculated using the LS factor (Eq. 4.4). Unlike the original WaTEM/SEDEM version (Notebaert et al., 2006), we implemented p_{con} within the multiple flow routing algorithm loop calculating the contributing upstream area, ensuring its effects propagate downstream through the flow network. Consequently, the reduction in sediment transport influences the downstream cells and extends to subsequent agricultural fields and vegetated areas. Moreover, we introduced a border deposition (b_{dep}) parameter, which represents a forced deposition mechanism activated when agricultural field cells contribute sediment to adjacent vegetated areas. Under these conditions, a defined percentage of the transported sediment is deposited directly at the field border within the field.

Retention ponds were implemented within the 5 m by 5 m land use raster map. The locations of the four retention ponds at the outlets of the micro-scale watersheds were mapped, with assigned trapping efficiencies of 54%, 82%, 59%, and 85% for watersheds W01, W02, W05, and W06, respectively, as measured in (Fiener et al., 2005). The standard deviation across all watersheds (± 14%) was applied to account for measurement error in the trapping efficiency values within the

GLUE framework (see Section 4.2.6).

4.2.6 Generalised Likelihood Uncertainty Estimation (GLUE)

We employed the GLUE methodology (Beven & Binley, 1992) to represent model and measurement uncertainties and to identify and analyse behavioural parameter spaces. The GLUE approach recognises that multiple parameter sets may provide equally acceptable simulations of a system within the limitations of a given model structure and observational errors (Beven, 2006).

We established limits of acceptability for the simulated sediment yields by considering multiple sources of uncertainty in the event-based measurements of runoff and sediment concentrations used for calculating annual and mean annual sediment yields. These included Coshocton wheel measurement errors (± 10%, Fiener & Auerswald, 2003), runoff collector storage tank sampling errors (estimated ± 10%), and retention pond uncertainties (± 14 %). For events with data collection issues (flagged in the data set), we assigned an additional ± 50 % error margin. However, for events flagged as "storage tank overflown", we introduced only an upper error boundary since the measurement taken from the storage tank represents a minimum possible sediment yield during a rainfall event. Finally, we propagated the measurement errors using a Monte Carlo simulation with 1,000 realisations and sampling from normal distributions that represented the range of potential errors. The 2.5th and 97.5th percentiles of the resulting aggregated (annual and mean annual) sediment yields were used as the limits of acceptability for simulated values. These uncertainty bounds served as criterion for identifying behavioural model realisations. Hence, only simulations producing outputs within these error margins were classified as behavioural and retained for subsequent analysis.

4.2.7 Model conditioning and evaluation

The model results were evaluated using R-Studio (R 4.4.2; R-Studio 2024.12.1 Build 563) in two phases to account for the different sediment transport processes in field-dominated and structure-dominated watersheds.

Phase 1 - Field-dominated watersheds:

We performed a Monte Carlo simulation with 25,000 realisations for the field-dominated watersheds, sampling parameters from uniform distributions across *a priori* selected ranges (Tab. 4.1). To consider the inherent potential errors in USLE calculations, including uncertainties associated with the parameterisation of the P factor, we modified the potential erosion in individual raster cells through an error surface (e_{sur}) before routing the sediment. This error surface was sampled from a uniform distribution for each realisation, modifying the USLE-calculated potential erosion (Eq. 4.1) within a range of 0 to ± 0.5:

$$A_{new,i} = A_i + A_i \, * e_{sur} \tag{4.6}$$

Where A_i is the potential soil erosion (t ha^{-1} yr^{-1}) calculated by the USLE (Eq. 4.1) at raster cell i, $A_{new,i}$ is the potential soil erosion (t ha^{-1} yr^{-1}) with incorporated uncertainty at raster cell i, and e_{sur} the error surface (dimensionless).

The decision to aggregate the uncertainty of the ABAG factors into a single error surface (e_{sur}), rather than sampling individual factors within the Monte Carlo simulation, stemmed from the nature of our input data. We assumed that parameterisation errors (apart from the P factor) were negligible due to the exceptionally high-quality monitoring data used as input for calculating the ABAG factors (section 4.2.4). However, as the ABAG is based on regressions that carry residual error, we pragmatically used an error surface to evaluate inherent model biases.

To ensure that $k_{TC/G}$ is consistently lower than $k_{TC/A}$, both were sampled with a constrained relationship, where $k_{TC/A}$ values were required to be at least 1.5 but no more than 5 times higher than $k_{TC/G}$ values. Model realisations were classified as behavioural if the simulated sediment yield values fell within the established limits of acceptability (error margins) for the observed data. Likelihoods were calculated only for the spatiotemporal aggregated data for which behavioural model realisations were identified. For these behavioural simulations, we calculated likelihoods by rescaling the mean absolute error (MAE) (Brazier et al., 2000):

$$L_i = \frac{1}{MAE_i} / \sum \frac{1}{MAE_i} \tag{4.7}$$

With:

$$MAE_i = |Sim_i - Obs_i| \tag{4.8}$$

where L_i is the likelihood of one realisation i (dimensionless), MAE_i is the mean absolute error of realisation i (t ha^{-1} yr^{-1}), Sim_i is the simulated values for behavioural runs of realisation i (t ha^{-1} yr^{-1}), and Obs_i is the observed sediment value for realisation i (t ha^{-1} yr^{-1}).

Table 4.1. Parameter ranges used for MC simulation in the WaTEM/SEDEM model. These ranges were selected based on the literature on previous model applications. $k_{TC/A}$ and $k_{TC/G}$ are the transport capacity coefficients for arable land and grassland, p_{con} is the parcel connectivity, e_{sur} is the error surface and b_{dep} the border deposition.

Range	$k_{TC/A}$ [m]	$k_{TC/G}$ [m]	p_{con} [%]	e_{sur}	b_{dep} [%]
low	1	1	50	-0.5	0
high	300	100	90	0.5	20

Phase 2 - Structure-dominated watersheds:

For the structure-dominated watersheds, we used the likelihoods associated with behavioural parameter values conditioned in Phase 1 to represent in-field processes ($k_{TC/A}$ and e_{sur}) in order to generate another 25,000 realisations. In this second phase, model conditioning was focused on the parameters controlling sediment redistribution through landscape structures ($k_{TC/G}$, b_{dep} and p_{con}). The same limits of acceptability approach as in phase one was applied to identify behavioural simulations. We calculated new likelihood values for these simulations to analyse their performance in representing structural erosion control practices.

4.2.8 Spatiotemporal model evaluation

Model outputs were evaluated at multiple spatiotemporal scales through sequential aggregation steps to analyse short-term dynamics against its intended long-term design: First, we temporally aggregated the sediment yields by calculating eight-year means for each individual watershed. Second, we spatially aggregated the simulated sediment yields by calculating their means for each watershed group (field- and structure-dominated), but keeping an annual resolution. Third, eight-year means for each watershed group were calculated (spatial and temporal aggregation).

To further analyse relative errors, the percent bias ($PBIAS$) was calculated by:

$$PBIAS = \left(\frac{Sim_i - Obs_i}{Obs_i}\right) * 100 \tag{4.9}$$

Where Sim_i is the simulated values for behavioural realisation i (t ha^{-1} yr^{-1}), and Obs_i is the observed sediment value for realisation i (t ha^{-1} yr^{-1}).

4.2.9 Spatial analysis

Spatial analysis was performed using R-Studio (R 4.4.2; R-Studio 2024.12.1 Build 563). To quantify the spatial distribution of sediment yield and the associated uncertainty, the cell-wise mean and standard deviation were calculated across all behavioural model realisations.

4.3. Results

4.3.1 Model performance across scales

The modelled annual sediment yields for field-dominated watersheds (W01-W04) were within the same order of magnitude of the measurements. However, the model was not considered behavioural for predicting annual sediment yields according to our pre-established acceptability criterion. The simulated annual sediment yields were predominantly overestimated (22 out of 32 cases; Fig. 4.4a-d), occasionally underestimated (4 out of 32 cases; in the year 1997 and 2000 in W01; 1994 in W02; 1994 in W04; Fig. 4.4a, b and d), with only a small portion of simulations overlapping the observational data (6 out of 32 cases; Fig. 4.4a-d). The tendency to overestimate sediment yield is more pronounced in watersheds W05 and W06. Only in 1994 the model had the tendency to underestimate measured sediment yields in watershed W05 (Fig. 4.5b). In W06, measured sediment yields were the lowest among all watersheds (maximum of 0.02 t ha⁻¹ yr⁻¹ in 2000), with zero sediment yield measurements in 1995, 1997, and 2001, yet the model consistently overestimated sediment yield across all years in this watershed.

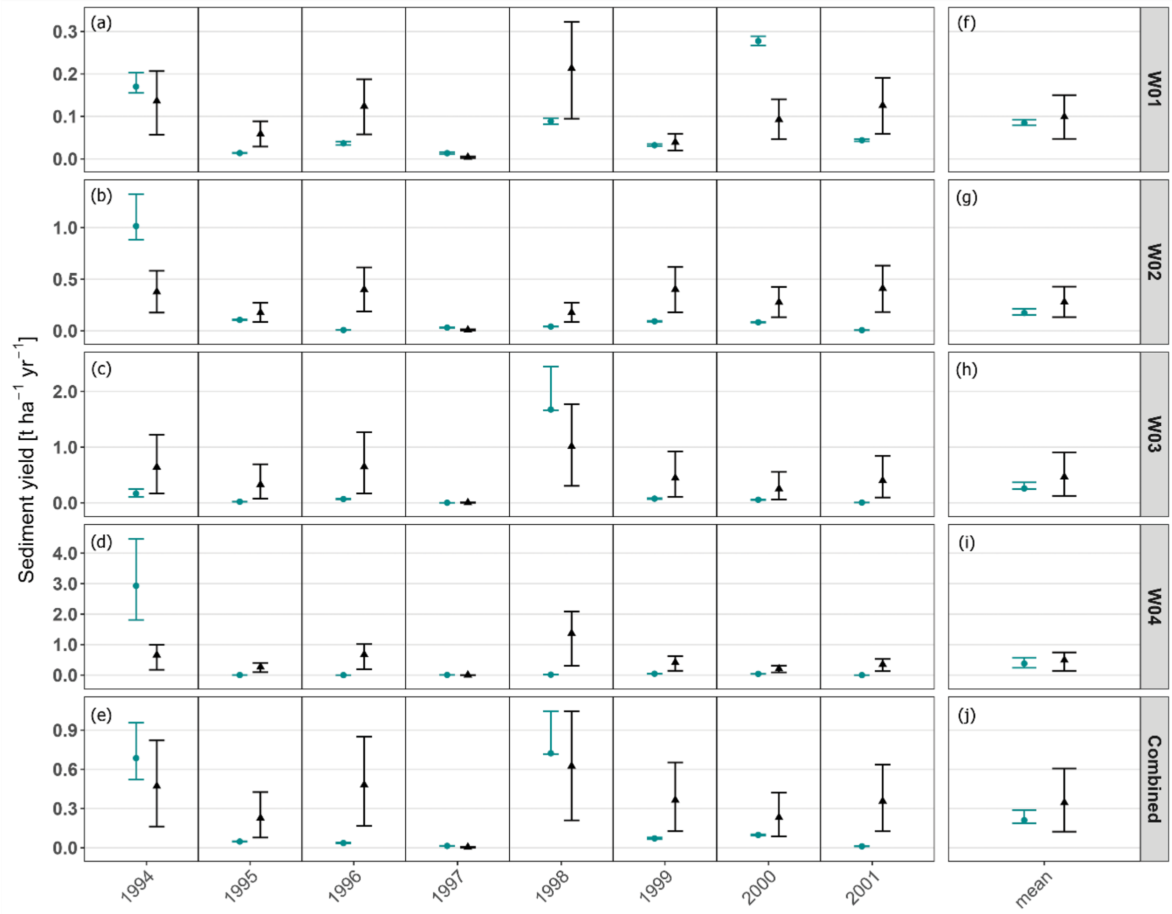


Figure 4.4. Annual and eight-year mean sediment yields in field-dominated watersheds. (a-d) Annual sediment yields: Black triangles indicate the mean and the full range from 25,000 model realisations (black whiskers), while cyan dots represent mean measured sediment yields with computed error ranges (cyan whiskers). (f-i) The watershed-specific eight-year mean measured sediment yields (cyan) and eight-year mean simulated yields (black). (e) Spatially combined annual watershed sediment yields. (j) Spatially aggregated eight-year mean yields. Note: In some years (e.g., 1998), cyan whiskers show larger uncertainties above the mean values; this is because storage tank overflow contributes only to higher uncertainties (see Section 4.2.6).

When evaluated using eight-year mean values per watershed, simulations showed better agreement with observations. The temporal aggregation revealed varying proportions of behavioural model realisations across individual watersheds. W04 had the highest amount with 69 % of all realisations, while other watersheds exhibited lower proportions (W01: 13 %, W02: 22 %, W03: 22 %). W05 exhibited minimal behavioural realisations of 1 %. In W06 (Fig. 4.5c), none of the actual model realisations matched the observational data including measurement errors. Furthermore, no common behavioural realisations were found across all watersheds, indicating that each watershed had a different behavioural parameter space. The analysis of the spatially aggregated watersheds (field-dominated vs. structure-dominated), while maintaining annual temporal resolution, revealed behavioural model realisations in some years but not consistently throughout the entire eight-year period for each watershed group (Fig. 4.4e, 4.4b).

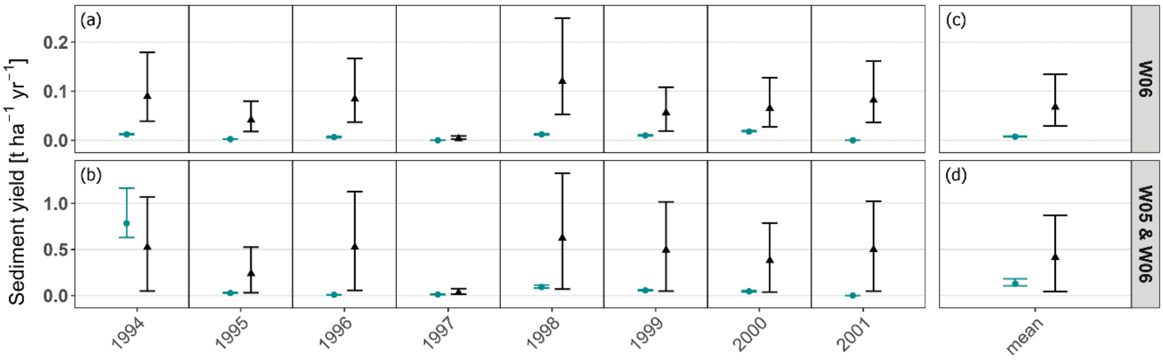


Figure 4.5. Annual and eight-year mean sediment yields in structure-dominated watersheds. (a-b) Annual sediment yields: Black triangles indicate the mean and the full range from 25,000 model realisations (black whiskers), while cyan dots represent mean measured sediment yields with computed error ranges (cyan whiskers). (c-d) The watershed-specific eight-year mean measured sediment yields (cyan) and eight-year mean simulated yields (black).

When combining both spatial and temporal aggregation, behavioural realisations were generated for each watershed group (Fig. 4.4j, 4.5d). Across the entire set of 25,000 realisations, the mean *MAE* values were 0.14 t ha^{-1} yr^{-1} for field-dominated watersheds and 0.29 t ha^{-1} yr^{-1} for structure-dominated watersheds, with maximum *MAE* values of 0.40 t ha^{-1} yr^{-1} and 0.74 t ha^{-1} yr^{-1}, respectively. Table 4.2 presents the model performance metrics specifically for the subset of behavioural model realisations within the watershed groups. The eight-year mean modelled sediment yield across field-dominated watersheds (W01-W04) was 0.35 t ha^{-1} yr^{-1}, compared to the measured eight-year mean of 0.21 t ha^{-1} yr^{-1}. For structure-dominated watersheds W05 and W06, we simulated an eight-year mean of 0.41 t ha^{-1} yr^{-1} (Fig. 4.5c, d), against a measured mean of 0.13 t ha^{-1} yr^{-1}. The 1994 sediment yield peak in W05 strongly influenced the system's overall performance, ultimately leading to an increased number of behavioural model realisations when evaluated across the entire period (Fig. 4.5d).

Table 4.2. Comparison of model performance metrics between micro-scale watershed groups based on eight-year mean of behavioural realisations, including mean sediment yield (SY) as well as error statistics (MAE, PBIAS) with maximum (Max.), minimum (Min.) and mean values.

Unit of measure		**Field-dominated**	**Structure- dominated**
Behavioural realisations [%]		28.70	1.35
Measured SY [t ha^{-1} yr^{-1}]	Mean	0.21	0.13
Simulated SY [t ha^{-1} yr^{-1}]	Mean	0.24	0.15
MAE [t ha^{-1} yr^{-1}]	Min.	4.21×10^{-6}	5.76×10^{-5}
	Mean	0.04	0.03
	Max.	0.08	0.05
PBIAS [%]	Min.	-11.51	-17.70
	Mean	15.19	15.96
	Max.	36.69	42.29

4.3.2 Behavioural parameter space

We analysed the behavioural parameter space for the spatially and temporally aggregated model outputs, as only this aggregation level yielded behavioural realisations for both field-dominated and structure-dominated watershed groups.

For field-dominated watersheds, the analysis focused on the error surface and in-field parameter e_{sur} and $k_{TC/A}$. While behavioural realisations were identified across the entire ranges of all parameters, higher likelihood values concentrated in specific regions. Specifically, e_{sur} values closer to -0.5 exhibited higher likelihood values than lower e_{sur} values (Fig. 4.6b). In contrast, $k_{TC/A}$ showed no discernible pattern across the response surface (Fig. 4.6c). The relationship between these parameters revealed a clear compensation mechanism, where lower $k_{TC/A}$ values required higher e_{sur} values to produce behavioural realisations (Fig. 4.6a).

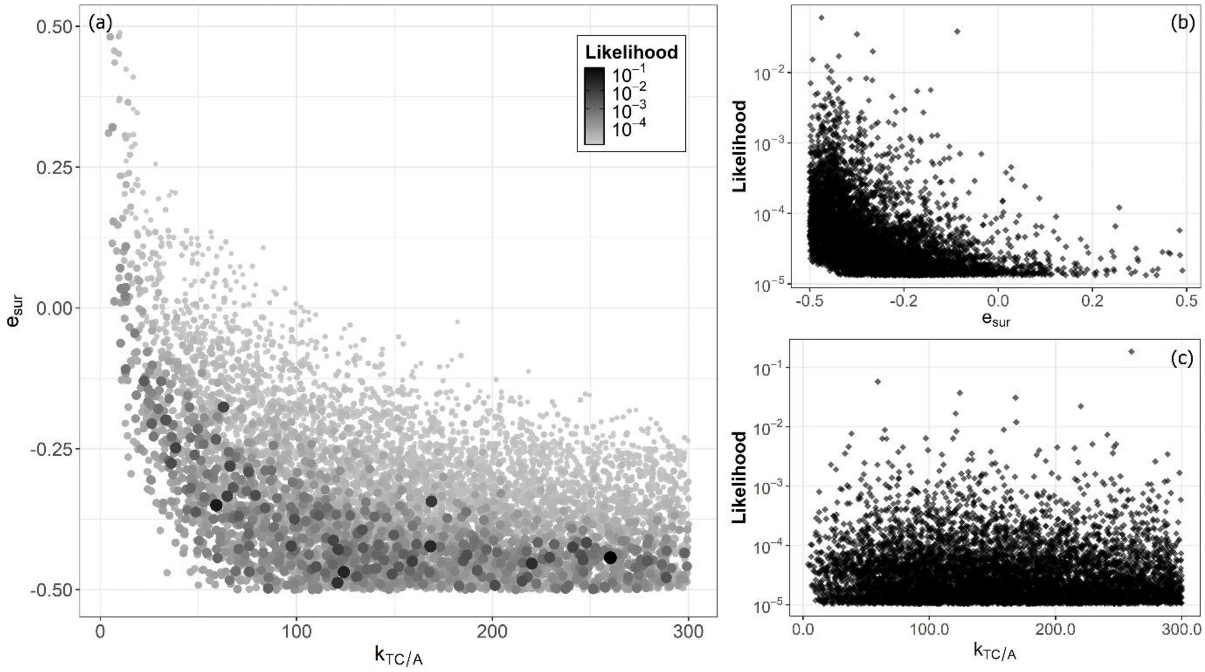


Figure 4.6. Parameter likelihoods across field-dominated micro-scale watersheds, showing only behavioural model realisations. (a) The relationship between e_{sur} and $k_{TC/A}$ parameters. Circle size and shade intensity indicate the likelihood of each parameter combination, with larger and darker circles representing higher likelihood values. (b) The relationship between likelihood and e_{sur}. (c) The relationship between likelihood and $k_{TC/A}$.

In structure-dominated watersheds, the analysis focused on parameters controlling sediment transport and deposition in grasslands and landscape structures ($k_{TC/G}$, b_{dep}, and p_{con}). The $k_{TC/G}$ parameter exhibited a distinct likelihood peak between approximately 9 and 11 m, with behavioural values ranging from approximately 7.5 m to 15 m (Fig. 4.7a), which is notably narrower than the sampled range of up to 150 (Tab. 4.1). In contrast, b_{dep} and p_{con} displayed relatively uniform likelihood distributions across their entire ranges (Fig. 4.7b, c). When plotting b_{dep} against p_{con}, homogeneous likelihood distributions emerged with no apparent dependencies (Fig. 4.7f). Examinations of b_{dep} and p_{con} against $k_{TC/G}$ revealed a horizontal band of high likelihood values at specific $k_{TC/G}$ values, without any directional trends (Fig. 4.7d, e).

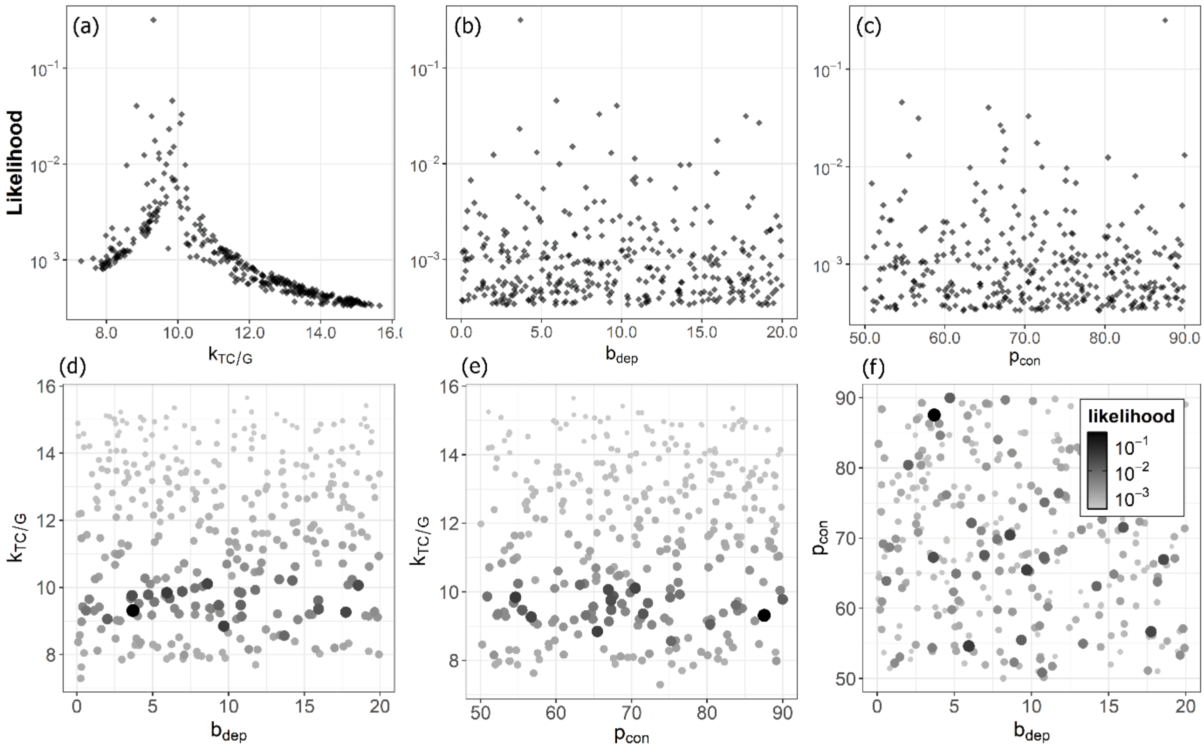


Figure 4.7. Parameter likelihoods across structure-dominated micro-scale watersheds, showing only behavioural model realisations. (a) The relationship between likelihood and $k_{TC/G}$. (b) The relationship between likelihood and b_{dep}. (c) The relationship between likelihood and p_{con}. (d) The relationship between b_{dep} and $k_{TC/G}$. Circle size and colour intensity indicate the likelihood of each parameter combination, with larger and darker circles representing higher likelihood values. (e) The relationship between p_{con} and $k_{TC/G}$. (f) The relationship between b_{dep} and $k_{TC/G}$.

4.3.3 Spatial analysis

In field-dominated watersheds, substantial deposition was primarily confined to retention ponds, while other areas outside arable lands showed minimal deposition, except for W03 (Fig. 4.8a). In W04, negligible to no deposition was observed. Conversely, structure-dominated watersheds exhibited considerably more intense erosion-deposition dynamics. The grassed waterway showed a clear deposition pattern, with W06 exhibiting the most pronounced deposition patterns leading toward the retention pond at the outlet.

The map of the standard deviation (Fig. 4.8b) also displays a spatial shift in model uncertainty between the two watershed groups. In field-dominated watersheds, standard deviation exhibits concentrations along in-field flow pathways, the retention ponds and the small structures next to their outlets. In structure-dominated watersheds a substantial increase in standard deviation is visible along the flow pathways towards and along the grassed waterway.

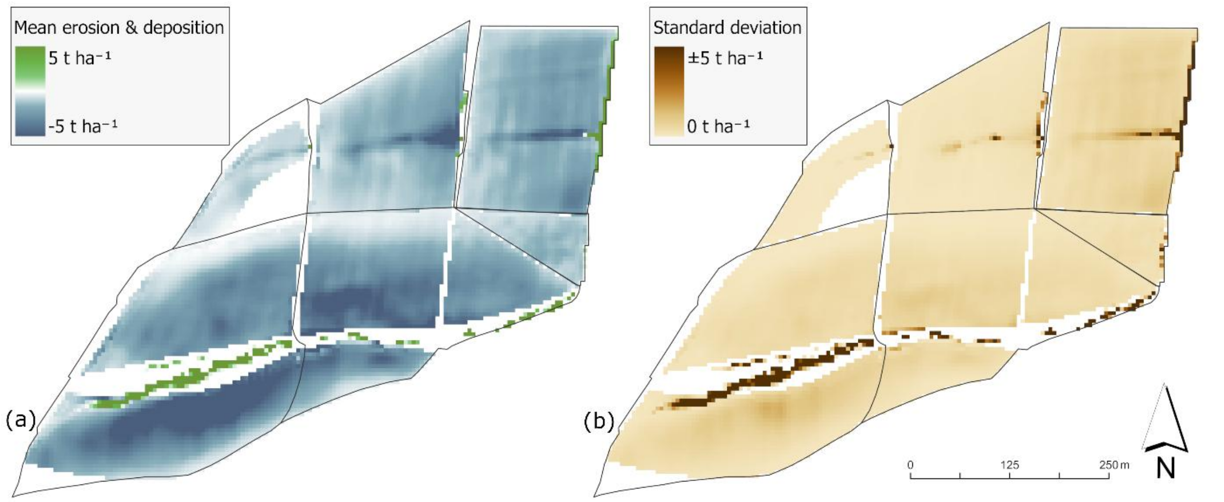


Figure 4.8 (a) The mean of simulated potential erosion and deposition of behavioural model realisations over the eight-year period. Negative values indicate erosion and positive values deposition. (b) The cell-wise standard deviation of behavioural model realisations over the eight-year period.

4.4. Discussion

4.4.1 GLUE framework and uncertainties

We tested WaTEM/SEDEM using a limits of acceptability approach within the GLUE framework. For this, we implemented a two-phase procedure, first conditioning and evaluating field-dominated watersheds, and then using the behavioural parameter space of these watersheds to condition and evaluate structure-dominated systems. Alatorre et al. (2010) demonstrated that soil erosion models often exhibit parameter compensation effects, where different parameter combinations produce similar outputs at the catchment scale - a manifestation of the equifinality concept (Beven, 2006). Our sequential approach helped to minimise these effects by first constraining the simulated erosion ($k_{TC/A}$ and e_{sur}) in field-dominated watersheds and then conditioning the transport parameters ($k_{TC/G}$, p_{con} and b_{dep}) in more complex systems.

The limits of acceptability approach incorporated multiple sources of measurement uncertainty. Nearing (2000) demonstrated through replicated plot studies that natural variability in erosion measurements is particularly pronounced for low-magnitude erosion events, such as the ones observed in this study. While Nearing (2000) proposed a quantitative method for estimating the expected variability of erosion measurements, his approach is specifically developed for plot-scale studies and cannot be extrapolated to watersheds or more complex landscape systems. Given the (to the best of our knowledge) current absence of methodologies for determining error boundaries for low sediment yield measurements at larger scales, we necessarily relied on relative error estimates. An implementation of proper measurement variability-derived error ranges would likely result in a substantially higher number of behavioural model realisations, particularly for low sediment yield measurements where the variability is the highest (Nearing, 2000). This reveals the need for developing robust approaches for defining limits-of-acceptability criteria for sediment yield estimates that account for the full range of uncertainties, e.g. instrument precision, sampling errors, data processing, or site-specific variations.

Erosion models typically exhibit systematic biases, overpredicting low sediment yields while underpredicting high sediment yields (Nearing, 1998; Risse et al., 1993; Kinnell, 2007). This is particularly relevant for our study area, where the implemented watershed-wide soil conservation and sediment trapping resulted in a measured mean sediment yield of only 0.16 t ha^{-1} yr^{-1}, which is substantially lower than erosion rates typically range between 3-10 t ha^{-1} yr^{-1} in the Bavarian Tertiary hill region (Auerswald et al., 2009). To investigate the modelling under/over prediction issue, we used an error surface (e_{sur}) multiplied with the erosion calculated by the USLE (Eq. 4.6). The analysis of behavioural model realisations revealed a concentration of likelihood values near small e_{sur} values, reducing sediment by up to 50 % (Fig. 4.6b). This indicates that in our study WaTEM/SEDEM overestimates soil erosion in landscapes with implemented conservation practices. This is also evident looking at figure 4.3a-d and 4.4a-b, which illustrate a general tendency for overestimation of modelled sediment yields in all watersheds.

4.4.2 Model performance and limitations

WaTEM/SEDEM correctly simulated the magnitude of the very low sediment yields in micro-scale watersheds optimised for soil conservation and reduced sediment transport, with annual values closely aligning with measured data (Fig. 4.4a-d, 4.5a-b). Nonetheless, the model did not consistently meet our limits of acceptability for annual realisations and therefore was rejected for making precise annual simulations. The model's performance improved notably when applied to longer-term means and larger spatial units, where more behavioural model realisations were identified.

Field-dominated watersheds

The model simulated the very low sediment yields resulting from well-established in-field soil conservation practices in field-dominated watersheds, comparable to the measured data. In general, observed sediment yields were overestimated, which can be attributed primarily to difficulties in accurately representing this conservation system, particularly unique practices such as mustard sown onto autumn-built dams where potatoes were later directly planted (Fiener & Auerswald, 2003). Such unconventional approaches are not adequately captured in the *SLR* values for no-till systems as evaluated in the German adaptation of the USLE (ABAG; Schwertmann et al., 1987; Din-Normenausschuss, 2022), even with the use of very low soil loss ratios in the parameterisation of the C factor, which represent the continuous soil cover through the crop rotation in the experimental farm (Fig. 4.2).

Conversely, the model underestimated sediment yields in some years because even optimally managed conservation systems experience short time windows with weak protection. During these short windows with reduced soil protection (Fig. 4.2), substantial erosion events may occur, like in systems not under soil conservation. In general, erosion processes are typically dominated by extreme events (Gonzalez-Hidalgo et al., 2012; Steegen et al., 2000), as exemplified in our study by an April 1994 rainfall event of 114 mm within 66 hours coinciding with low soil coverage in W02 (Fig. 4.2), accounting for approximately 58 % of that year's total sediment yield (see year 1994 in Fig. 4.4b). The model's annual time step fails to capture these critical temporal coincidences, a structural limitation that becomes more pronounced when such events are infrequent. This temporal limitation aligns with findings by Risse et al. (1993), who demonstrated that USLE's model efficiency diminishes at the annual scale. When averaging over the eight-year study period, these extreme events are smoothed out, which explains the model's improved performance at longer timescales (Tab. 4.2). This deficit could not be compensated using high-resolution input data (daily soil cover, high resolution and on-site rainfall measurements). However, it is important to note that the episodic nature of erosion is always difficult to capture even with event-based models, and hence aggregation over time tends to improve any kind of erosion model. This is especially true if events are rare and the overall erosion values are small (Nearing, 1998). Future studies could employ synthetic data generation (Srikanthan & McMahon, 2001), allowing for an assessment of the model's sensitivity beyond our high-resolution input data.

For the temporally aggregated eight-year means, there was no single parameter set that produced behavioural model realisations across all field-dominated watersheds simultaneously when applying our limits of acceptability criterion. This indicates a limitation in parameter transferability within our study context. While Rompaey et al. (2001) recognized technical limitations of WaTEM/ SEDEM in model transferability related to grid size and routing methods, our findings suggest additional challenges in accurately representing processes within micro-scale watersheds with specific in-field soil conservation practices (e.g. no-till farming). The need for watershed-specific calibration, even within relatively homogeneous landscapes with similar crop and soil properties, indicates that parameter calibration compensates for inherent model or data limitations. At such fine scales, WaTEM/SEDEM may struggle to accurately represent the complex interactions between soil conservation practices and erosion processes. However, it is important to note that these difficulties may stem from the low-magnitude nature of the erosion events observed in our study. As demonstrated in paired-plot experiments, natural variability is higher for small events (Wendt et al., 1986). Therefore, areas with higher erosion rates typically yield data that is less noisy and inherently easier for models to reproduce (Nearing, 1998).

Similar calibration challenges seem to exist more broadly in WaTEM/SEDEM applications across different landscape types and research questions, as evidenced by a wide range of calibrated k_{TC} values reported across different studies (Tab. 4.3), with $k_{TC/A}$ values varying from 10 to 174.4 m. As

Beven (2006) argues, such calibration approaches may achieve mathematical fitting while concealing fundamental model inadequacies.

Structure-dominated watersheds

Unlike field-dominated watersheds, structure-dominated systems demonstrated different response patterns to extreme erosion events. In these watersheds, sediment generated during individual large erosion events (as observed in the field-dominated watersheds), is predominantly captured by grassed waterways and retention ponds (Fiener & Auerswald, 2003, 2005), thus reducing the variability of event sediment yields. This buffering effect explains why the model consistently overestimates sediment yield across all years for structure-dominated watersheds (Fig. 4.5a-b), in contrast to the occasional underestimation observed in field-dominated systems (Fig. 4.4a-d).

Only one exception to this pattern was observed: the model underestimated sediment yield in W05 during 1994, when the lower part of the grassed waterway required reseeding after losing its initial grass cover along the thalweg during a spring erosion event (Fiener & Auerswald, 2003). This exceptional case quantitatively demonstrates the role of functional grassed waterways, as the measured sediment yield in 1994 for W06 was substantially higher (0.78 t ha^{-1} yr^{-1}) than in subsequent years when the grassed waterway was fully established (averaging only 0.03 t ha^{-1} yr^{-1} from 1995-2001), representing an approximately 96 % reduction in sediment yield (Fig. 4.5b).

The model's systematic overestimation of sediment yields in structure-dominated watersheds reveals limitations in representing the sediment trapping mechanisms of grassed waterways. The primary limitation is the model's inability to capture re-infiltration processes within the grassed waterway, which is not accounted for in WaTEM/SEDEM's transport capacity formulation. Fiener & Auerswald (2005) demonstrated that grassed waterway effectiveness depends strongly on morphological characteristics, particularly the cross-sectional shape, with flat-bottomed waterways showing substantially higher runoff reduction. The infiltration increases with length and flatter cross-sections of grassed waterways, which provide larger runoff widths and consequently greater infiltration areas. A previous study showed that in the upper part of the grassed waterway (W06), where WaTEM/SEDEM more substantially underestimates the sediment trapping, the long-term runoff and sediment yield reductions were respectively about 90% and 97%, while it was about 10% and 77% in the lower part of the grassed waterway (W05) with a ditch-like cross-section (Fiener & Auerswald, 2003).

4.4.3 Spatial dynamics of erosion and deposition

The main reason for the reduction of sediment yield at the outlet within structure-dominated watersheds is visible in the eight-year mean map of behavioural model realisations (Fig. 4.8a). Depositional patterns are prominent along the flux pathway of the grassed waterway, particularly in W06. Because WaTEM/SEDEM is sensitive to the parameter controlling deposition inside these soil conservation structures ($k_{TC/G}$), the standard deviation in these areas is exceptionally high (Fig. 4.8b).

However, high uncertainty is not limited to the grassed waterways. It is also evident in depositional areas of field-dominated watersheds (most pronounced in W03). This suggests that deposition is generally prone to high uncertainties, regardless of the dominant structures. Furthermore, field-dominated watersheds exhibit high standard deviations within the fields themselves. This likely stems from parameter uncertainty regarding sediment supply, specifically the error surface parameter (e_{sur}, see Fig. 4.6b), which directly influences the sediment supply generated within the arable land.

4.4.4 Spatial aggregation

For the structure-dominated watersheds, spatial aggregation was critical for overcoming watershed-specific model failures. While the model failed to produce any behavioural realisations for W06 (even temporally lumped), the spatially aggregated group achieved 1.35% behavioural realisations. This aligns with the scale-dependency concepts reviewed by de Vente & Poesen (2005), who note that process dominance shifts with spatial scale, often allowing models to perform adequately at larger scales even if they miss finer processes.

For field-dominated watersheds, the spatial aggregation acted primarily as an averaging of varying crop states, smoothing out the heterogeneity of cover conditions, and any given watershed peculiarities. In contrast, for structure-dominated watersheds, spatial aggregation facilitates the identification of behavioural parameter spaces by masking local structural inadequacies, such as the deposition dynamics in grassed waterways. Consequently, the differences between spatially aggregated field-dominated and structure-dominated watersheds can be attributed to their structural components.

4.4.5 Distribution of behavioural model parameter values

The *TC* within agricultural fields is primarily controlled by the transport coefficient $k_{TC/A}$ (Rompaey et al., 2001). Lower $k_{TC/A}$ values reduce TC, promoting in-field deposition and consequently decreasing sediment yield at the watershed outlet. Our analysis revealed behavioural model realisations across the full *a priori* selected range of $k_{TC/A}$ values, with no clear pattern for field-dominated watersheds, demonstrating no sensitivity even at very low $k_{TC/A}$ values near 1 or very high e_{sur} values of 0.5 (Fig. 4.6a). This lack of sensitivity may be attributed to the retention ponds in W01 and W02 and the very low simulated erosion values. Since *TC* remained sufficiently high to transport the low sediment fluxes even with very low $k_{TC/A}$ values, sediment transport is thus supply-limited rather than constrained by transport capacity within the fields.

The low transport capacity coefficient for rougher surfaces, in case of this study for grassland $k_{TC/G}$, usually triggers deposition in these areas (Rompaey et al., 2001). Our analysis identified behavioural values for $k_{TC/G}$ between approximately 7.5 m to 15 m with a notable likelihood spike between approximately 9 m and 11 m, relatively low values compared to other studies (Tab. 4.3). While Onnen et al. (2019) reported similarly low values for Danish landscapes, they explicitly attributed this to sandy soils in Denmark. However, our study area features predominantly silt loam and loamy soils, which are much more comparable to the Belgian soils (Tab. 4.3) where low k_{TC} values for rough surfaces were implemented (Peeters et al., 2008; Rompaey et al., 2001; Verstraeten et al., 2002).

Table 4.3. Comparison of k_{TC} parameter values of behavioural model realisations used in different studies with WaTEM/ SEDEM.

High k_{TC} values mostly used for arable land [m]	Low k_{TC} values mostly used for non-arable land [m]	Country	Source
150	not used	Germany	Wilken et al. (2020)
10 to 24	1 to 12	Denmark	Onnen et al. (2019)
100 & 150	25	Belgium	Peeters et al. (2008)
75	42	Belgium	Verstraeten et al. (2002)
75	42	Belgium	Rompaey et al. (2001)
174.4	not used	Belgium	Oost et al. (2000)

These low $k_{TC/G}$ values can have some possible interpretations: (i) most likely, the model is compensating for its inability to represent re-infiltration processes in grassed waterways, and/or (ii) the model may partly compensate for an overestimation of erosion rates in the draining fields. However, although the model outputs are fully spatially distributed (Fig. 4.8), it is not possible to compare the simulated patterns with spatially distributed observational data (e.g. aerial images, field surveys), because, except for some rare larger events, the effective soil conservation established prevents visible erosion features like rills.

In our study, WaTEM/SEDEM showed no sensitivity to parameters b_{dep} and p_{con} that represent the influence of linear landscape features. These parameters displayed homogenous likelihood distributions across the sampled parameter space (Fig. 4.7b-c). This lack of sensitivity could stem from several factors: (i) sampling an overly narrow parameter space, (ii) limited influence of field borders in the studied watersheds due to the layout of the fields and watersheds with a small number of border situations, and/or (iii) a dominance of $k_{TC/G}$ implemented over a long grass structure, which may nullify the influence of b_{dep} and p_{con} in the model outputs, especially in watershed W05 and W06.

An additional limitation of the current parameterisation approach, particularly for grassed waterways, is its static nature. The effectiveness of grassed waterways and retention structures varies throughout the year due to seasonal vegetation changes (Fiener & Auerswald, 2003). Additionally, there is an important interaction between sediment influx and trapping efficiency—as influx increases, the relative trapping efficiency typically decreases (Dermisis et al., 2010; Fiener & Auerswald, 2018). Dermisis et al. (2010) demonstrated this inverse relationship, showing that grassed waterway trapping efficiency decreases as peak runoff discharge increases, with notable breakpoints in efficiency between different flow rates. The current static connectivity and transport capacity parameters (p_{con}, b_{dep} and $k_{TC/G}$) cannot adequately capture these temporal variations and flux-dependent relationships, suggesting the need for a more dynamic parameterisation approach that accounts for both seasonal changes and influx response if the model is applied to an annual time-step.

4.5. Conclusion

We evaluated WaTEM/SEDEM's capability to simulate sediment yields in micro-scale watersheds optimised for soil conservation and sediment transport reduction using a limits-of-acceptability approach within the GLUE framework. Our investigation examined model performance across different levels of spatiotemporal data aggregation and analysed the sensitivity of the model's response surface to the variability in the behavioural parameter space. Moreover, we used a two-step conditioning process, in which model parameters linked to in-field erosion processes were conditioned in field-dominated watersheds and later applied in structure-dominated watersheds, for which a separate set of connectivity parameters was also conditioned.

The model was unable to produce behavioural realisations for watersheds optimised for soil conservation and sediment transport reduction at annual time steps based on our strict limits of acceptability criterion despite the small absolute prediction errors over all model realisations (eight-year MAE = 0.14 t ha^{-1} yr^{-1} for field-dominated and eight-year MAE = 0.29 t ha^{-1} yr^{-1} for structure-dominated watersheds). For the field-dominated watersheds, the model particularly struggled with simulating annual sediment yields. This is because soil conservation practices reduced the number of erosion events, yet some events (e.g., after potato harvest) retained a similar magnitude as in conventional cultivation.

Aggregating model outputs in time and space worked best for field-dominated systems, which compensated for the underestimation of soil conservation in controlling soil erosion and the model's inability to capture extreme events within an annual time step. While WaTEM/SEDEM is generally better suited for long-term erosion modelling, our findings confirm that this is especially the case for watersheds with optimised soil conservation and reduced sediment transport.

The GLUE framework revealed specific patterns in the sampled parameter space, particularly the compensation mechanism between $k_{TC/A}$ and e_{sur} values for field-dominated watersheds, and the narrow behavioural parameter range of $k_{TC/G}$ values (7.5-15 m) for structure-dominated watersheds. Spatially, this sensitivity is mirrored by high standard deviations concentrated along the grassed waterways. In contrast, the standard deviation within arable fields shows the influence of sediment supply parameterisation (e_{sur}) in field-dominated systems (Fig. 4.8b). The likelihood distributions of $k_{TC/A}$ and especially e_{sur} enabled the pre-conditioning of structure-dominated watersheds, reducing parameter compensation effects that typically mask model structural deficiencies.

Ultimately, our study demonstrates that WaTEM/SEDEM can simulate the magnitude of very low sediment yields observed from soil conservation agricultural systems, provided that high spatiotemporal resolution input data and locally adapted USLE factors (e.g., the ABAG) are available. However, capturing the combined effects of low in-field erosion and linear landscape features like grassed waterways, where concentrated runoff occurs, remains challenging for WaTEM/SEDEM, primarily due to the model's inability to represent re-infiltrating processes that are critical for sediment trapping in such structures. Additionally, our model evaluation approach revealed that model performance strongly depends on the spatiotemporal scale of analysis. While the model produced behavioural realisations for the aggregated eight-year monitoring period, it did not reliably simulate annual sediment yields. For long-term, large-scale soil conservation planning in which the effects of single erosive events on individual fields are less relevant for representing the system behaviour, WaTEM/SEDEM seems to be fit for purpose within our testing conditions.

Code availability

The R code used to compute individual factors and statistics is available at https://zenodo.org/records/18714865. The Python WaTEM/SEDEM code is available upon reasonable request.

Data availability

The input data are openly available and can be downloaded here: https://adgeo.copernicus.org/articles/48/31/2019/adgeo-48-31-2019.html (Last visited 11.07.2025). The specific data used to compute individual factors and statistics can be found at https://zenodo.org/records/18714865.

Author contribution

KDS: Conceptualisation, data curation, formal analysis, investigation, methodology, software, visualisation, writing (original draft preparation); PVGB: Conceptualisation, methodology, validation, writing (review and editing); HS: Writing (review and editing); TS: Supervision, writing (review and editing); PF: Conceptualisation, data curation, project administration, supervision, validation, writing (review and editing).

Competing interests

Pedro V. G. Batista and Peter Fiener serve as Topic Editor and Executive Editor of SOIL, respectively.

Acknowledgements

We acknowledge the valuable dataset from the Forschungsverbund Agrarökosysteme München (FAM). The scientific activities of the FAM research network were financially supported by the German Federal Ministry of Education and Research (BMBF 0339370). We thank all researchers and technical staff involved in collecting and maintaining these long-term datasets, which made this study possible.

During the preparation of this work, the authors used Anthropic Claude 3.7 Sonnet to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Financial support

This research was funded by the German Research Foundation (DFG) through the DYLAMUST project (Project number: 509809226).

5. Manuscripts III and IV: Deep learning-based image segmentation

Chapter 5 is based on following two peer-reviewed journal publications:

Manuscript III

Shokati, H., Engelhardt, A., Seufferheld, K., Taghizadeh-Mehrjardi, R., Fiener, P., Lensch, H. P., & Scholten, T. (2025). Erosion-SAM: Semantic segmentation of soil erosion by water. *Catena*, *254*, 108954, https://doi.org/10.1016/j.catena.2025.108954.

Manuscript IV

Shokati, H., Seufferheld, K. D., Fiener, P., & Scholten, T. (2026). Rapid flood mapping from aerial imagery using fine-tuned SAM and ResNet-backboned U-Net. *Hydrology and Earth System Sciences*, *30*(3), 743-756, https://doi.org/10.5194/hess-30-743-2026.

This chapter advances the dissertation's overarching goal of data-rich, spatially explicit environmental monitoring by focusing on deep learning-based image segmentation for water-related natural hazards. Building on the previous studies, which addressed rainfall erosivity reconstruction, process-based erosion and sediment transport modeling, and their associated uncertainties, the work presented here turns to high-resolution remote sensing as an independent observational pillar for detecting and quantifying erosion and flood impacts. By leveraging recent developments in foundation models and transfer learning, the chapter demonstrates how modern computer vision can close critical data gaps in soil erosion and flood assessment and thereby strengthen both scientific understanding and operational decision-making.

Manuscript III in this chapter introduces Erosion-SAM, a fine-tuned adaptation of the Segment Anything Model (SAM) for automatic delineation of water erosion and deposition features in very high-resolution aerial orthophotos of agricultural landscapes in Bavaria. Erosion-SAM is trained on manually segmented fields that cover bare cropland, vegetated cropland, and grassland, and it systematically compares different preprocessing strategies as well as the role of point prompts during inference.

Manuscript IV extends the use of deep learning segmentation to rapid flood mapping based on UAV and helicopter imagery, again exploiting transfer learning and high-capacity architectures to work with relatively limited labeled data. In this study, fine-tuned SAM variants are compared with U-Net architectures using ResNet backbones to delineate inundated and non-inundated areas under heterogeneous acquisition conditions.

Taken together, the two papers illustrate a coherent deep learning framework for automated, high-resolution monitoring of erosion and flood processes, complementing the climatic forcing reconstruction and process-based modeling presented earlier in the dissertation. Erosion-SAM opens up new possibilities for generating spatially distributed validation and training data for soil erosion models, while the flood segmentation work demonstrates how similar concepts can be translated to disaster-response contexts with stringent timeliness requirements. By integrating foundation models, transfer learning, and domain-specific remote sensing data, this chapter shows how modern computer vision can transform erosion and flood assessment from largely manual, retrospective mapping into scalable, semi-automated workflows that support both scientific inquiry and practical land and risk management.

5.1. Manuscript III: Erosion-SAM: semantic segmentation of soil erosion by water

The first part of **Chapter 5** is based on **Manuscript III** (Shokati et al., 2025), which develops and evaluates Erosion-SAM, a fine-tuned adaptation of the Segment Anything Model (SAM) for automated mapping of water erosion and deposition features in high-resolution orthophotos of agricultural fields in Germany. The study addresses the challenge that conventional soil erosion assessment relies heavily on manual interpretation of aerial imagery and field surveys, which is time-consuming, subjective and difficult to scale to larger areas. By leveraging a foundation model and transfer learning, Erosion-SAM is designed to segment visually heterogeneous erosion structures across different land covers while requiring comparatively small labeled training datasets. Within the overall dissertation framework, this study primarily answers Research Question 4 on how fine-tuned foundation models can be used for automated, high-resolution soil erosion mapping.

The study uses an orthophoto dataset covering agricultural fields with documented erosive rainfall events, accompanied by a detailed erosion classification scheme that distinguishes between erosion and deposition features. It systematically analyzes how model performance responds to key design choices, including different image preprocessing strategies (cropping versus resizing), prompt types, data augmentation and implementation details of the fine-tuning procedure. Model performance is quantified using standard segmentation metrics and further evaluated in a field-based assessment that aggregates pixel-level predictions to field-scale erosion classes.

The remainder of the first part of **Chapter 5** is organized as follows. Subsection 5.1.1 introduces the conceptual background, research objectives and the rationale for using SAM as a basis for erosion mapping. Subsection 5.1.2 describes the study area, orthophoto data, erosion classification, model architecture, fine-tuning strategy, prompt design, preprocessing approaches, data augmentation scheme and evaluation metrics. Subsection 5.1.3 presents and discusses the training behavior, validation performance and segmentation results across land cover types and field-based erosion classes, while Subsection 5.1.4 summarizes the main findings and implications of the Erosion-SAM approach.

Manuscript III: Erosion-SAM: Semantic segmentation of soil erosion by water

Abstract

Soil erosion (SE) by water threatens global agriculture by depleting fertile topsoil and causing economic costs. Conventional SE models struggle to capture the complex, non-linear interactions between SE drivers. Recently, machine learning has gained attention for SE modeling. However, machine learning requires large data sets for effective training and validation. In this study, we present Erosion-SAM, which fine-tunes the Segment Anything Model (SAM) for automatic segmentation of water erosion features in high-resolution remote sensing imagery. The data set comprised 405 manually segmented agricultural fields from erosion-prone areas obtained from the rain gauge-adjusted radar rainfall data (RADOLAN) for bare cropland, vegetated cropland, and grassland. Three approaches were evaluated: two pre-processing techniques— resizing and cropping — and an improved version of the resizing approach with user-defined prompts during the testing phase. All fine-tuned models outperformed the original SAM, with the prompt-based resizing method showing the highest accuracy, especially for grassland (recall: 0.90, precision: 0.82, dice coefficient: 0.86, IoU: 0.75). SAM performed better than the cropping approach only on bare cropland. This discrepancy is attributed to the tendency of SAM to overestimate SE by classifying a large proportion of fields as eroded, which increases recall by covering most of the eroded pixels. All three fine-tuned approaches showed strong correlations with the actual SE severity ratios, with the prompt-enhanced resizing approach achieving the highest R^2 of 0.93. In summary, Erosion-SAM shows promising potential for automatically detecting SE features from remote sensing images. The generated data sets can be applied to machine learning-based SE modeling, providing accurate and consistent training data across different land cover types, and offering a reliable alternative to traditional SE models. In addition, erosion-SAM can make a valuable contribution to the precise monitoring of SE with high temporal resolution over large areas, and its results could benefit reinsurance and insurance-related risk solutions.

5.1.1. Introduction

Soil erosion (SE) by water poses a significant threat to agricultural sustainability worldwide, as it removes fertile topsoil and leads to extensive soil degradation (Eltner et al., 2015; Quinton & Fiener, 2024; Scholten & Seitz, 2019). Beyond its impact on agriculture, SE imposes significant economic burdens and challenges the profitability of farming practices (Pimentel et al., 1995). To address these challenges, numerous traditional SE models have been developed (Jetten & Favis-Mortlock, 2006; Mitasova & Mitas, 2001), ranging from empirical to physically based models, varying in spatial scale from sub-plot to catchment level (Borrelli et al., 2021). The regression-based empirical Universal Soil Loss Equation (USLE, (Wischmeier & Smith, 1978)) and the revised USLE (RUSLE, (Renard, 1997)) are the most commonly used SE models (Karydas et al., 2014). However, collecting data to calibrate and validate such SE models is still a major challenge. Typically, runoff and sediment delivery measurements for individual events taken over extended periods at plot outlets with natural or artificial rainfall (Schindewolf & Schmidt, 2012) or at watershed outlets (Kohrell et al., 2023) are used to compare predicted and observed soil loss rates. However, when using outlet data for model validation, especially in large watersheds, the internal process dynamics of the watershed are often overlooked. To address this, long-term erosion assessments using radionuclides as tracers (Walling et al., 2003; Wilken et al., 2020) or event-based high-resolution remote sensing data (Cândido et al., 2020) can be used to test and validate these internal dynamics. However, high-resolution remote sensing data has primarily been used in SE models to incorporate variables such as vegetation coverage, slope length and slope gradient (Ganasri & Ramesh, 2016) and to identify visible occurrence of erosion processes at the regional scale (Polovina et al., 2024) rather than for model validation and training. Fischer et al., (2018) advanced this field by implementing a targeted monitoring approach that tracks potential SE events over the Bavarian Tertiary Hills. Their study visually classified aerial orthophotos into four soil loss classes representing the spatial extent of SE damage within fields. These classes were used to validate event-based soil loss estimates using soil loss ratios from USLE-technology. The study showed a strong and statistically significant correlation between the visually classified and the predicted soil loss at the field scale. While the use of visually classified orthophotos proved to be effective for validating SE estimates, the manual classification process is still labor-intensive and time-consuming. Machine learning models offer a promising way to automate this process and extract more detailed information from remote sensing data. Machine learning models have been used in various studies to assess and map soil erosion (Agnihotri et al., 2021; Mosavi et al., 2020; F. Wang et al., 2021). However, traditional machine learning models occasionally provide suboptimal predictions (Bui et al., 2020; Gautam et al., 2021; Saha et al., 2021). Recent studies have therefore suggested a shift towards deep learning models, emphasizing that they are likely to be less constrained and more flexible than traditional machine learning models (Prasad et al., 2022). Previous studies have shown the significant potential of deep learning approaches in SE studies (e.g. Khosravi et al., 2023; Liu et al., 2023). However, deep learning models also require a considerable amount of ground-truth data not only for validation but also for effective training, while remotely sensed data labeled with SE features are scarce.

Recent advances in deep learning have introduced several innovative techniques to reduce reliance on big data. These include few-shot learning, self-supervised learning, transfer learning, and data augmentation, which data augmentation and transfer learning have become very popular in the field of deep learning applications in remote sensing (Safonova et al., 2023). In data augmentation, a dataset is artificially enlarged by creating new samples through transformations such as rotation, scaling and flipping. This technique improves the generalizability and robustness of the model and reduces overfitting (Shorten & Khoshgoftaar, 2019; Wagner et al., 2023). Transfer learning, on the other hand, uses previously learned features to improve the generalization ability of the data-based machine learning model. The importance of transfer learning in processing small remote sensing datasets has been demonstrated in various studies (Character et al., 2021; Hou et al., 2022; Lv et al., 2022; Saadeldin et al., 2022; S. Wang et al., 2018; H. Zhang et al., 2019). Building on this foundation, the Segment Anything Model (SAM) has emerged as a pioneering approach in image segmentation, offering substantial potential for transfer learning applications. Trained on an extensive dataset of over 1 billion masks derived from 11 million images, SAM demonstrates strong generalization capabilities across diverse datasets and generates precise segmentation masks (Kirillov et al., 2023). For instance, Song et al., (2024) refined SAM by fine-tuning it for crop segmentation, demonstrating its superior performance compared to five other advanced segmentation methods. Similarly, the ClassWise-SAM-Adapter (CWSAM) was developed by Pu et al., (2024) through fine-tuning SAM for land cover classification using spaceborne Synthetic Aperture Radar (SAR) imagery. Their results showed that CWSAM outperformed traditional state-of-the-art semantic segmentation algorithms.

Despite these advances in deep learning and transfer learning techniques, automated segmentation of SE at the pixel level has not been tested so far. For the first time, this study fine-tunes the pre-trained SAM as a transfer learning technique using aerial orthophotos to automatically identify the erosion and deposition forms at the pixel level. The proposed Erosion-SAM model tests different data preprocessing methods and incorporates pixel-based predictions, thereby enhancing both the processing and prediction accuracy. Overall, this study aims to 1. develop a method to automate pixel-based SE feature segmentation, thereby improving the scalability and accuracy of SE detection, and 2. evaluate the accuracy of the fine-tuned model across different types of land cover, namely grassland, vegetated cropland, and bare cropland. By customizing SAM with aerial images of agricultural fields affected by erosive rainfall, we aim to qualitatively identify and delineate visible traces of SE processes at a high spatial resolution of 20 cm with minimal training data.

5.1.2. Materials and methods

5.1.2.1. Study area

The study area is located in the south-eastern part of Bavaria in south-eastern Germany (Figure 5.1.1). The region was selected based on two main criteria: (i) availability of high-resolution orthophotos (0.2 m x 0.2 m) taken after high-intensity rainfall events, and (ii) generally intensive agricultural practices in hilly terrain comprising cropland and grassland. The test sites within the study area shown in the orthophotos include locations with a mean annual precipitation of 800 to 1200 mm/year (Auerswald et al., 2009) and a mean temperature between 6°C and 11°C (German Weather Service, DWD[1]). Dominant soils are Cambisols and Luvisols (Gocke et al., 2021; Iuss & Fao, 1999), whereas slopes range between 0 % and 40 %. Typical crops are winter wheat, maize, winter barley and potatoes, with the erosion-prone maize fields dominating the data set.

[1] https://www.dwd.de

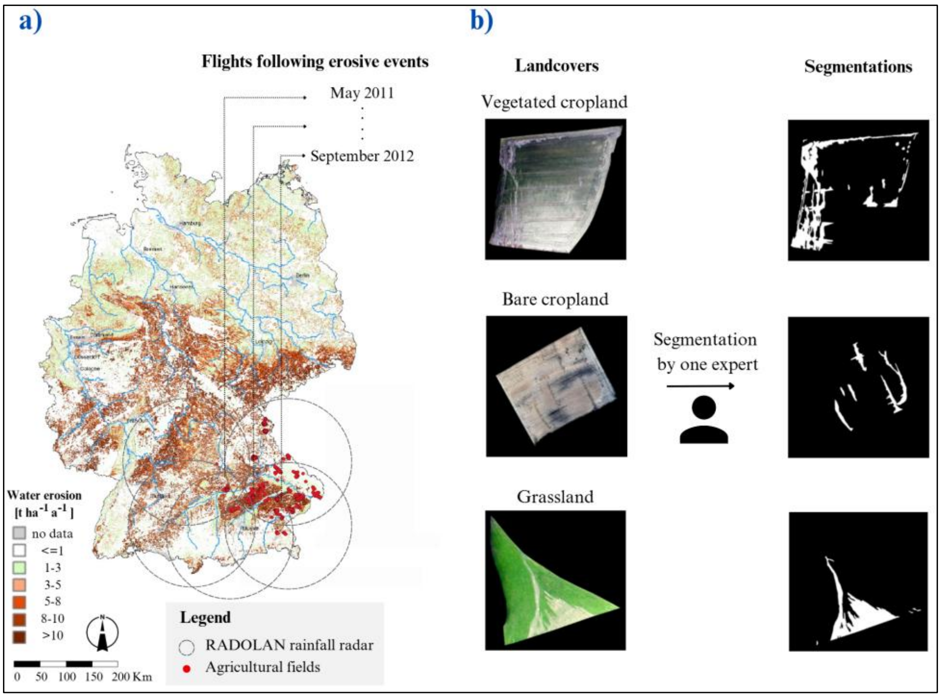


Figure 5.1.1. a) Soil erosion map of Germany in a 75 × 75 m raster (Auerswald et al., 2009) and b) examples of manual segmentation of agricultural fields with erosive rainfall, showing erosion and deposition features across different land covers.

5.1.2.2. Orthophotos

The Orthophotos used in this study were acquired in a research project of the Bavarian Agency for Agriculture, as documented in Fischer et al., (2018). The general idea was to acquire orthophotos after heavy rainfall events in erosion-prone areas for the period between May and September in 2011 and 2012. To decide for the timing of the data acquisition, the RADOLAN rainfall data of the DWD with a spatial resolution of 1 × 1 km² were used. In case of rainfall events (at least 10 mm total rain depth or maximum 30 min rain intensity higher than 10 mm/h) that indicates a high erosion potential, flights with a small airplane were performed not longer than 30 days after the determined erosive rainfall event. The aerial images were taken at a flight height of about 900 m with an onboard Nikon D200 camera. This resulted in an image pixel resolution of 0.2 m x 0.2. The images were georeferenced using ArcGIS Pro 3.1.3 to ensure accurate spatial alignment with InVeKoS data (Integrated Administration and Control System). We manually identified and masked 405 individual parcels (Figure 5.1.1) from the aerial imagery that exhibited erosion and deposition features such as ephemeral gullies, rills and sediment fans. The fields were categorized into grassland (n = 128) and cropland (n = 277), whereas the latter was then subdivided into vegetated cropland (n = 131) and bare cropland (n = 146).

5.1.2.3. Erosion classification

For each field in the categories, an expert performed a visual interpretation to manually segment SE features within the respective field (Figure 5.1.1) using ArcGIS Pro 3.1.3, delineating areas where erosion and deposition had occurred irrespective of the land cover type. Pixels showing erosion or deposition on cropland or grassland were classified and labeled as "erosion-deposition", while all other pixels within the field were labeled as "undisturbed areas". From each land cover type, 12 fields were randomly selected and segmented by three independent human classifiers to quantify the accuracy of the manual segmentation. The area with erosion and deposition features varied 8 % for grassland, 13 % for vegetated cropland and 17 % for bare cropland between the four replications.

5.1.2.4. Model construction

5.1.2.4.1. The Segment Anything Model (SAM)

The Segment Anything Model (Meta AI Inc., USA) marks a significant advancement in image segmentation technology. SAM consists of three main components (Figure 5.1.2). The image encoder, built on a robust Vision Transformer (ViT) backbone, extracts image features. The prompt encoder, a lightweight embedding module, generates both sparse and dense embeddings from various prompt inputs (points, boxes or masks). Finally, the mask decoder combines the outputs of the image encoder and the prompt encoder to generate the final segmentation masks.

SAM is available in three variants: base (ViT-B), large (ViT-L) and huge (ViT-H) with 91, 308 and 636 million parameters respectively (Kirillov et al., 2023). These models differ in terms of computational requirements and architectural complexity. In this study, we used the ViT-B variant of the SAM to reduce computational demands.

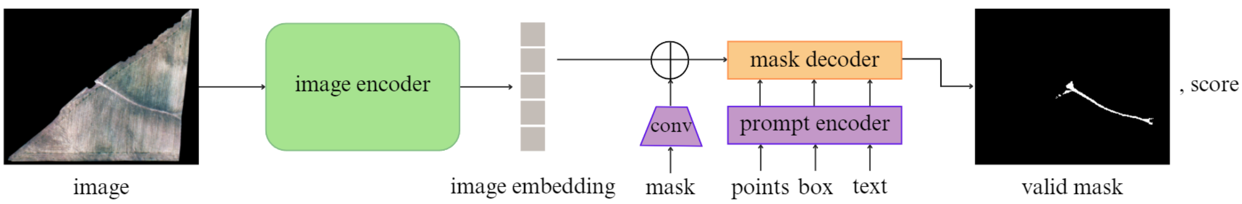


Figure 5.1.2. Overview of the Segment Anything Model (SAM) (adapted from Kirillov et al., (2023)). SAM uses a heavyweight image encoder to create an image embedding, which is combined with input prompts and passed through a prompt encoder. The features are then decoded into object masks, with corresponding confidence scores.

5.1.2.4.2. Fine-Tuning SAM

Despite SAM's high accuracy and little to no training data requirements, its performance can be limited under specific domain conditions. To address this, we employed the fine-tuning technique, a form of transfer learning. Fine-tuning involves adapting a pre-trained model (including its architecture and weights) to a specific task instead of training a new model from scratch. Fine-tuning enhances performance, reduces the need for large training data sets, lowers computational costs, and saves time (Rawat & Wang, 2017). Pre-training on a broad data set is particularly effective when the task-specific data set is small or lacks extensive labeled data.

When fine-tuning SAM, we focused on fine-tuning the mask decoder (Figure 5.1.2) due to the complexity and the high number of parameters of the image encoder. The mask decoder has a lightweight design (Sun et al., 2024), i.e., it contains fewer parameters and a simpler architecture, which makes fine-tuning easier, faster, and more memory efficient.

5.1.2.4.3. Input prompts

The architecture of SAM enables the integration of human prompts, which increases the effectiveness of human-in-the-loop annotation. These prompts can be multimodal: points on the area to be segmented, a bounding box around the object or textual descriptions of what is to be segmented. For segmentation, we used point prompts in the training phase of our fine-tuned SAM. Specifically, 80 points were used, each separated by a minimum radius of 10 pixels. However, in the testing phase, we used two different approaches to comprehensively evaluate the performance of the model. First, we tested the model without prompts, as manual SE segmentation is challenging and in real-world scenarios often no pre-existing masks are available. Second, we considered a scenario where only minimal user input is possible in real-world applications. In this approach, point prompts were used in testing to mirror the process used in training.

5.1.2.4.4. Image pre-processing approaches

To address the variability in the sizes of agricultural fields, we considered two methods of pre-processing to standardize the image sizes for the model.

Uniform Resizing: This method involves resizing all field images to 256 × 256 pixels before inputting them into the model. After the model has made its predictions, the images are resized back to their original dimensions. While this approach contributes to the standardization of the input dimensions, it can lead to distortions, especially in hilly terrain, which may result in unrealistic erosion patterns or the loss of important details during downscaling. These effects can affect the reliability of model predictions. To minimize such problems, it is important to choose a resizing dimension that minimizes distortion and ensures that the adjusted sizes remain close to the actual dimensions of the fields.

Image Cropping: In this approach, the training images are divided into multiple 256 × 256 image patches with a step size of 256 pixels. To ensure compatibility with the specified patch size during image processing, we used zero padding, in which the image size is expanded by appending zeros to facilitate the division into uniform patches. To balance the input between erosion and non-erosion pixels, patches consisting entirely of non-erosion pixels are excluded from the modeling process. Each image patch is processed individually by the model. After prediction, the image patches are reassembled to reconstruct the original image. To avoid jagged predictions at the corners of the test image patches, the step size for cropping the test data was reduced to 128 pixels, creating an overlap between the merged patches. A blending technique using a simple second-order spline window function is then applied. This method weights the pixels in the overlapping areas when merging the patches and provides smoother transitions between neighboring patches.

Both resizing and cropping approaches aim to ensure consistent image input sizes for the model while accommodating the varying dimensions of the agricultural fields. These approaches were compared to determine which yields the best performance and accuracy.

5.1.2.4.5. Data Augmentation

The choice of data augmentation technique depends on the quantity, quality and type of data (Safonova et al., 2023). In this study, due to the pixel values being zero in the corners of the input images, certain data augmentation methods, such as random rotation or random cropping, might result in augmented images consisting entirely of zero values, potentially causing errors. Therefore, we used horizontal and vertical flips methods of data augmentation, which do not encounter this issue.

5.1.2.4.6. Field-based soil erosion classification for model testing

Erosion-SAM labels each pixel of a high-resolution orthophoto as erosion-deposition or undisturbed areas based on the presence or absence of erosion and deposition features. Even though this approach provides a detailed spatial resolution, field-based assessment is a valuable complementary to the analysis. A field-based SE classification provides critical insights by identifying which land cover types are most susceptible to erosion and deposition, evaluating how different management practices affect soil loss and accumulation on fields, and assessing how different environmental conditions influence erosion processes. Field-based classifications have also been explored by researchers such as Fischer et al., (2018). In this study, we build on this concept by implementing a field-based approach to create a data set for testing our segmentation model using 15 % of our images (n = 60). The percentage of eroded areas within each agricultural field are calculated using our labeled ground-truth data pixels, and fields are subsequently categorized into three SE severity classes: Class 1 (0–10% erosion), Class 2 (10–30% erosion), and Class 3 (more than 30% erosion). However, it is important to note that our pixel-based classification cannot directly compare to the original classification based on estimating the area per field affected. The same classification method is then applied to the SE maps produced by the model, allowing a direct comparison between the predicted and manually segmented results.

5.1.2.4.7. Implementation Details

The data from all 405 agricultural fields were randomly divided into training, validation and test sets. Specifically, 70 % of the fields from each land cover category were included in the training set (n = 285), 15 % in the validation set (n = 60) and 15 % in the test set (n = 60). This stratified splitting approach ensured that the distribution of land cover types in the training, validation and test subsets was balanced, resulting in a representative data set for model evaluation. In the cropping approach, to avoid spatial correlations, we first divided the agricultural fields into training, validation and test sets and then performed cropping to ensure that patches from the same field did not appear in different sets. During the training phase, the model adjusts its internal parameters (e.g. weights and biases) to learn patterns in the data using backpropagation. After each iteration, the validation set, which is kept separate from the training data, is used to evaluate the model's performance. This step helps in optimizing hyperparameters (e.g., learning rate and batch size) and preventing overfitting (Vabalas et al., 2019; Ying, 2019). The model with the lowest validation loss is selected as the best performing model, as it is likely to generalize well to unseen data. Finally, the selected model is evaluated using the test set which has remained isolated from all previous stages, providing an unbiased measure of its ability to generalize to new data.

To further increase the robustness of the model, horizontal and vertical flips were applied to the training set as data augmentation methods. To further increase the robustness of the model, horizontal and vertical flips were applied to the training set as data augmentation methods. By applying data augmentation, the number of training data was doubled. Data augmentation was only applied to the training set so that the validation and test data remained unchanged in both resizing and cropping approaches. In the resizing approach, the number of training data increased from 285 to 570 samples after applying the data augmentation methods, while the number of validation and test samples remained unchanged at 60 each. Similarly, in the cropping approach, the number of training samples increased from 1,774 to 3,548, while the number of validation and test samples remained unchanged at 314 and 1,060 respectively. The model was trained for 50 epochs, with each epoch representing a complete pass of the dataset through the segmentation algorithm. To optimize the parameters in the decoder module, the combination of dice and cross-entropy losses (DiceCELoss) was used as the loss function, which uses their average for the total loss and is a standard choice for training (Hadlich et al., 2023). The cross-entropy loss is given by:

$$L_{CE} = -\frac{1}{N}\sum_{c=1}^{C}\sum_{i=1}^{N} g_i^c \, \log S_i^c \tag{5.1.1}$$

where g_i^c is the ground truth binary indicator of the class label c of pixel i, and S_i^c is the corresponding predicted segmentation probability. The dice loss is defined as:

$$L_{Dice} = 1 - \frac{2\sum_{c=1}^{C}\sum_{i=1}^{N} g_i^c\, s_i^c}{\sum_{c=1}^{C}\sum_{i=1}^{N} g_i^c + \sum_{c=1}^{C}\sum_{i=1}^{N} s_i^c} \quad (5.1.2)$$

We used the Adam optimizer, which has been very popular in the field of machine learning in recent years due to its easy fine-tuning and strong adaptability (C. Liu et al., 2023). This optimization algorithm adaptively adjusts the learning rate, enabling efficient and effective model training. The model was trained with a batch size of 4 and an initial learning rate of 0.001. To further refine the training process, we used the learning rate scheduler, which dynamically adjusted the learning rate based on the validation loss. Whenever the validation loss did not improve in five consecutive epochs, the learning rate was reduced by a factor of 0.1. This adaptive approach allowed us to maintain an optimal balance between convergence speed and model performance throughout the training process.

All experiments were conducted using Python and the PyTorch framework on a computer configured with an NVIDIA GeForce RTX 4070 Ti GPU and running the Windows 11 operating system. The parameter settings used during the fine-tuning process are summarized in Table 5.1.1. These parameters were carefully chosen to optimize the performance of the model while overcoming the particular challenges posed by the erosion dataset and computational resources.

Table 5.1.1. The parameter settings used during the fine-tuning process.

Parameter	**Settings**
Data split	Training: 70% (n=285), Validation: 15% (n=60), Testing: 15% (n=60)
Loss function	DiceCELoss (Dice + Cross-Entropy loss)
Optimizer	Adam
Learning rate	Initial: 0.001, Decay: 0.1
Batch size	4
Epochs	50
Data augmentation	Horizontal and vertical flips
Weight initialization	Pretrained weights (SAM)
Framework	PyTorch (Python)
Hardware	NVIDIA GeForce RTX 4070 Ti, Windows 11

5.1.2.5. Evaluation metrics

To evaluate the accuracy of the model in segmenting SE features, four evaluation metrics including precision, dice coefficient, recall and intersection-over-union (IoU) were utilized as defined by the following equations (Vinayahalingam et al., 2023):

$$\text{precision} = \frac{TP}{TP + FP} \tag{5.1.3}$$

$$\text{dice coefficient} = \frac{2\,TP}{2\,TP + FP + FN} \tag{5.1.4}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{5.1.5}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{5.1.6}$$

Where TP, FP and FN represent the number of true positives, false positives and false negatives pixel labels, respectively.

Precision, dice coefficient, recall and IoU are key performance metrics in image segmentation. Precision describes how well the model identifies erosion pixels as positive predicted values, indicating the proportion of correctly predicted erosion pixels. In addition, the dice coefficient measures the similarity between the ground truth and the predicted data set and is commonly used in image segmentation. It ranges from 0 to 1, where 1 represents a perfect overlap and 0 represents no overlap. Similarly, IoU, also known as the Jaccard index, includes the data of TP, FP and FN. Our segmentation process for erosion features involves two steps, localization of the area in an orthophoto and classification of an area as erosion or non-erosion. In addition, IoU is scale-invariant and evaluates how well the predicted erosion area matches the ground truth. Finally, recall, which reflects the sensitivity of the model, describes the proportion of correct erosion pixels in the total number of erosion pixels in the ground truth data. Figure 5.1.3 shows the flowchart of the work.

To assess the field-based predictions, two metrics including the coefficient of determination (R^2) and mean absolute error (MAE) were also calculated as follows:

$$R^2 = 1 - \frac{Residual\ Sum\ of\ Squares\ (RSS)}{Total\ Sum\ of\ Squares\ (TSS)} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5.1.7}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5.1.8}$$

Where n is the number of data, y_i is the observed value, $\hat{y}_i$ is the corresponding predicted value and $\bar{y}$ is the mean of the observed values.

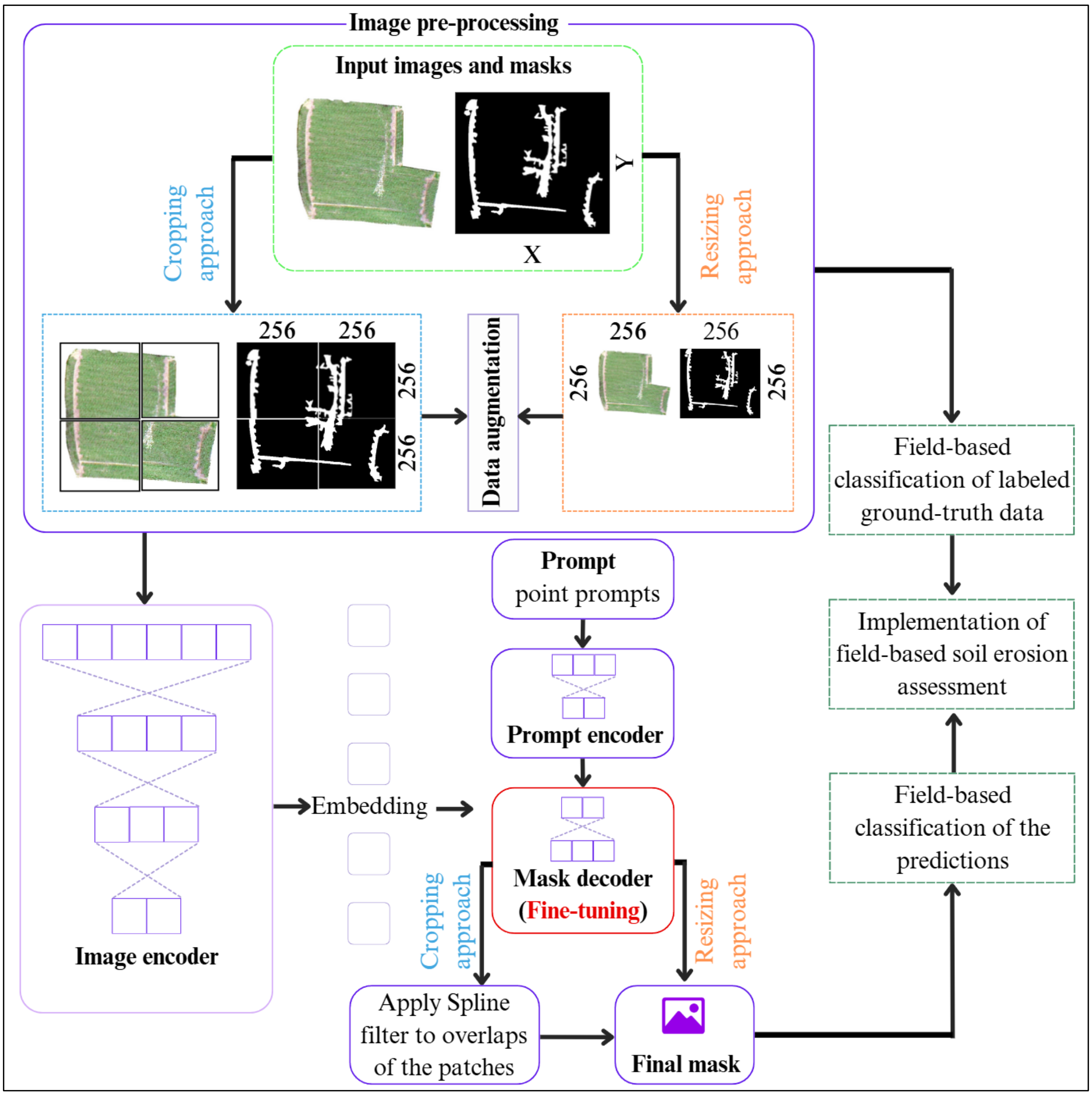


Figure 5.1.3. Overview of the network for fine-tuning the mask decoder of the Segment Anything Model. The input images and masks are preprocessed using two approaches: Cropping to 256 × 256 patches (blue box) or resizing to 256 × 256 pixels (orange box). Both approaches are followed by data augmentation. The images are then passed through the image encoder, where they are transformed into embeddings. These embeddings, in combination with point prompts, are fed into the fine-tuned mask decoder (red box) to generate the final segmentation mask. In the cropping approach, after generating a mask for each patch, they are merged, and a spline filter is applied to the overlaps of the patches to provide smoother transitions between neighboring patches.

5.1.3. Results and discussion

5.1.3.1. Training and validation losses

The model was trained and validated over 50 epochs (Figure 5.1.4) reflecting the number of times the data set passes through the segmentation algorithm. The results showed that for the resizing approach, the training loss dropped significantly during the first epochs. After this sharp decrease, the rate of decrease slowed down and finally reached a value below 0.23 in the last epoch. The validation loss for the resizing approach also showed a general downward trend until epoch 20, indicating that the model generalizes well to unseen data, while it stabilized after this point, indicating an overfitting to the training data after this inflection point (Shorten & Khoshgoftaar, 2019).

In the cropping approach, training losses followed a similar pattern to the resizing approach, with a general decrease across epochs, reaching approximately 0.26 at epoch 50. Similarly, validation losses displayed a consistent downward trend with a minimum of 0.25 at epoch 30. Following epoch 30, the validation loss stabilized, indicating that additional training would likely not improve performance and suggesting overfitting to the training data beyond this point. The validation loss curve showed signs of overfitting later for the cropping method (after 30 epochs) than for the resizing method (after 20 epochs). This discrepancy stems from the different pre-processing approaches. In the resizing method, the images were resized to 256 × 256 pixels, whereas in the cropping method, each image was divided into multiple 256 × 256 patches depending on the original dimensions. As a result, the cropping approach generated a larger training dataset (3548 patches for the cropping approach compared to 570 patches for the resizing approach, both after applying data augmentation techniques), delaying the occurrence of overfitting (Elhage et al., 2022; Safonova et al., 2023; Shorten & Khoshgoftaar, 2019).

The lowest validation loss for the resizing approach was achieved in epoch 20 with a value of 0.24, while it was achieved for the cropping approach in epoch 30 with a value of 0.28. Consequently, the models from these respective epochs were selected as optimal models for the resizing and cropping approaches. Overall, the training and validation losses for the resizing approach were slightly lower than for the cropping approach, indicating a better generalization performance for the resizing method.

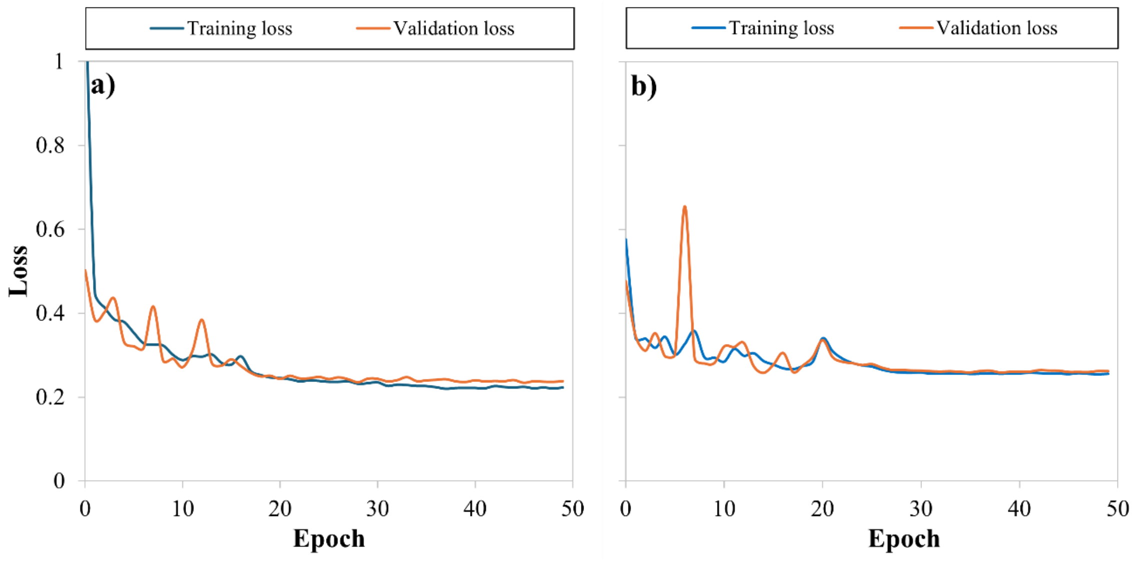


Figure 5.1.4. Training and validation losses over the training epochs for a) resizing and b) cropping approaches.

5.1.3.2. Performance of the model on the validation set

The values of the performance metrics across the training epochs for both the resizing and cropping approaches are shown in Figure 5.1.5. Both approaches showed a consistent ascending course in their metrics, indicating a progressive improvement in model accuracy throughout the training process. Remarkably, the improvement reached a plateau in middle training epochs and behaved similarly to the loss function. This stabilization indicates that the models have achieved optimal performance and further training beyond this point would likely not yield significant improvements.

In the first few epochs, considerable fluctuations were observed in all metrics, which can be attributed to the weight adjustments of the model and the extensive learning phases. As training progressed, the fluctuations decreased, indicating that the learning process became more refined and the model gradually recognized patterns in the data. For the resizing approach, precision was consistently higher than the other metrics throughout training, reaching a value of 0.78 at the best epoch (epoch 20). At the same point, the dice coefficient and recall were 0.77 and 0.76, respectively. IoU remained the lowest with a value of 0.63, indicating a relatively high overlap between predicted and actual SE pixels. For the cropping approach, precision also remained the highest of all evaluation metrics, reaching 0.81 in the best epoch (epoch 30), followed by dice coefficient, recall and IoU at 0.79, 0.76 and 0.66, respectively.

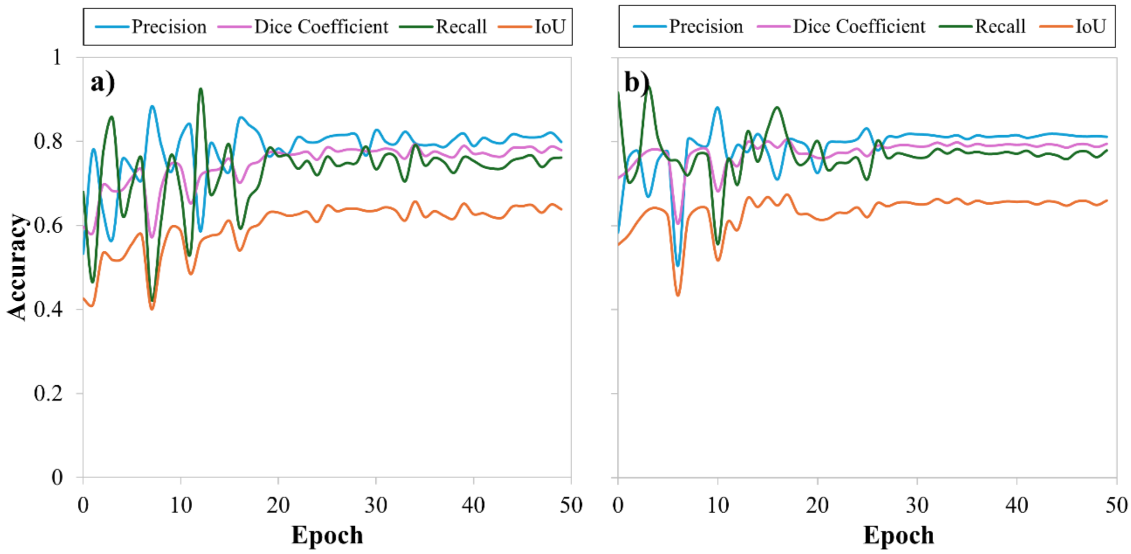


Figure 5.1.5. Performance of the fine-tuned model on the validation set over the training epochs for a) resizing and b) cropping approaches.

5.1.3.3. Performance analysis of the SAM and fine-tuned models across different land covers

The performance of the models on test data for grassland, vegetated cropland and bare cropland and the overall performance are shown in Table 5.1.2. In vegetated cropland, SAM showed its weakest performance compared to bare cropland and grassland, with a recall of 0.52, a precision of 0.19, a dice coefficient of 0.28 and an IoU of 0.16. The significant discrepancy between recall and precision indicates that although SAM detected a large proportion of eroded areas on vegetated cropland, many of the identified regions were false positives, as the model often misclassified non-eroded areas as erosion areas, such as crops, bare patches in between or tractor tracks. This leads to an overestimation of eroded areas (Ishikawa et al., 2021; Sun et al., 2024) (Figure 5.1.6 a).

The fine-tuned model, which utilized the cropping approach for pre-processing the data, achieved its best performance in grassland with recall, precision, dice and IoU values of 0.77, 0.68, 0.72 and 0.56, respectively. This can be attributed to the more uniform ground surface in grassland compared to bare cropland and vegetated cropland, which facilitates the detection of SE. In addition, the deposition areas in the grassland form a clearer contrast to the vegetated, non-eroded or sediment-covered areas. However, for vegetated cropland, segmentation accuracy decreased in all evaluation metrics, with recall, precision, dice, and IoU values of 0.59 each and an IoU of 0.42. This decrease in performance can be explained by the presence of crops, tractor tracks or bare patches, which the model occasionally misclassified as SE. The lowest performance of the cropping approach was observed for bare cropland where the recall, precision, dice, and IoU values were 0.41, 0.31, 0.35 and 0.21, respectively. This lower accuracy is due to the similarity between eroded and natural soil, which makes it difficult for the model to accurately segment SE in bare cropland.

In the resizing approach without using input prompts in the testing phase, the performance for grassland stood out with the highest score, including a recall of 0.84, a precision of 0.77, a dice coefficient of 0.80 and an IoU of 0.67. These metrics indicate that the model performs well in detecting SE in grassland areas, achieving high sensitivity and decent precision. On the other hand, the resizing approach yielded moderate results in vegetated cropland with a recall of 0.57, precision of 0.65, a dice coefficient of 0.61, and IoU of 0.43. While the higher precision suggests that the model is quite accurate in predicting SE, the lower recall suggests that it may miss certain eroded areas due to complex landscape features such as crops, bare patches or tractor tracks, leading to under-detection rather than misclassification. The lowest performance of the resizing approach was observed in bare cropland with a recall of 0.50, a precision of 0.46, a dice of 0.48 and an IoU of 0.31. This is consistent with previous observations in the cropping approach regarding the difficulty of detecting SE in bare cropland.

The resizing approach, having demonstrated superior performance across all land cover types, was selected as the optimal model for SE segmentation. To further improve the practical applicability of the model, user-specific prompts were integrated during the testing phase, allowing the user's knowledge to be incorporated for better identification of SE sites. This improvement takes advantage of the practical fact that field operators often have valuable contextual information about potential SE areas. The implementation of these prompts showed varying effectiveness across different land cover types. In grassland, the approach reached its optimal performance with impressive values: a recall of 0.90, a precision of 0.82, a dice coefficient of 0.86 and an IoU of 0.75. For vegetated cropland, the performance metrics dropped to 0.80, 0.74, 0.77 and 0.62 for recall, precision, Dice coefficient and IoU, respectively. Bare cropland proved to be the most challenging environment where the approach achieved its lowest performance: a recall of 0.73, a precision of 0.72, a dice coefficient of 0.72 and an IoU of 0.57. Nevertheless, the integration of point prompts significantly improved the performance of the model in detecting SE areas, especially in bare cropland environments where identification was previously a major challenge. This marked difference in performance between land cover types suggests that the environmental context significantly influences the segmentation capabilities of the model, with the resizing approach using point prompts effectively mitigating many land cover-specific challenges.

5.1.3.4. Comparison of the effectiveness of SAM and fine-tuned models in detecting soil erosion

A comprehensive evaluation showed that the resizing approach, when enhanced with point prompts during the testing phase, consistently outperformed all other tested models across different land covers (Figure 5.1.6 b and c). Without prompts, the resizing approach remained in second place across all land covers, followed by the cropping approach, with a notable exception in bare cropland environments, where the original SAM performed better than the cropping approach. Despite the higher accuracy in bare cropland, the main problem of SAM was that it tended to overestimate SE by misclassifying a large proportion of fields as eroded, leading to numerous false positives and over-segmentation of SE (Ishikawa et al., 2021; Sun et al., 2024) (Figure 5.1.6 d). This limitation was emphasized by the significant discrepancy between recall and precision observed in SAM across all land covers. In contrast, the resizing and cropping approaches showed a more balanced relationship between recall and precision, indicating higher accuracy in SE classification.

On average across all land covers, the prompt-enhanced resizing approach delivered the most balanced and effective results, achieving the highest scores for recall (0.81), precision (0.76), dice coefficient (0.78) and IoU (0.64). These results revealed a clear hierarchy of performance: the prompt-enhanced resizing approach ranked first, followed by without prompt resizing, cropping and finally the original SAM. These results are consistent with those of Sun et al., (2024), who had previously demonstrated the superiority of fine-tuned SAM implementations over the original SAM. The higher performance of the resizing approach compared to the cropping approach can be attributed to the ability of the resizing approach to better preserve spatial relationships and important features, as it processes entire fields at once, as opposed to the cropping approach, which divides the field into several smaller patches. Erosion and deposition often cause large, visible features that extend across entire agricultural fields extending from one side of the image to the other. In the cropping approach, these features are usually spread across multiple image patches and not limited to a single patch. Since the predictions in the cropping approach are made independently for each patch and later combined, this method does not preserve spatial relationships and important features despite the expected advantages of increasing the sample size and preserving details. Consequently, the resizing approach proves to be more effective in identifying signs of erosion and sedimentation.

Table 5.1.2. The performance of the models on test data across different land cover types.

Model	Land cover	recall	precision	dice	IoU
Original SAM	grassland	0.63	0.48	0.54	0.37
	vegetated cropland	0.52	0.19	0.28	0.16
	bare cropland	0.53	0.37	0.44	0.28
	total	0.56	0.35	0.43	0.27
Cropping approach	grassland	0.77 (+ 0.14)	0.68 (+ 0.20)	0.72 (+ 0.18)	0.56 (+ 0.19)
	vegetated cropland	0.59 (+ 0.07)	0.59 (+ 0.40)	0.59 (+ 0.31)	0.42 (+ 0.26)
	bare cropland	0.41 (- 0.12)	0.31 (- 0.06)	0.35 (- 0.09)	0.21 (- 0.07)
	total	0.58 (+ 0.02)	0.52 (+ 0.17)	0.55 (+ 0.12)	0.38 (+ 0.11)
Resizing approach (without prompt)	grassland	0.84 (+ 0.21)	0.77 (+ 0.29)	0.80 (+ 0.26)	0.67 (+ 0.30)
	vegetated cropland	0.57(+ 0.05)	0.65 (+ 0.46)	0.61 (+ 0.33)	0.43 (+ 0.27)
	bare cropland	0.50 (- 0.03)	0.46 (+ 0.09)	0.48 (+ 0.04)	0.31 (+ 0.03)
	total	0.63 (+ 0.07)	0.62 (+ 0.27)	0.63 (+ 0.20)	0.46 (+ 0.19)
Resizing approach (with prompt)	grassland	0.90 (+ 0.27)	0.82 (+ 0.34)	0.86 (+ 0.32)	0.75 (+ 0.38)
	vegetated cropland	0.80 (+ 0.28)	0.74 (+ 0.55)	0.77 (+ 0.49)	0.62 (+ 0.46)
	bare cropland	0.73 (+ 0.20)	0.72 (+ 0.35)	0.72 (+ 0.28)	0.57 (+ 0.29)
	total	0.81 (+ 0.25)	0.76 (+ 0.41)	0.78 (+ 0.35)	0.64 (+ 0.37)

* Footnote: The numbers in blue and orange represent the increase and decrease in performance of the fine-tuned models compared to the original SAM values.

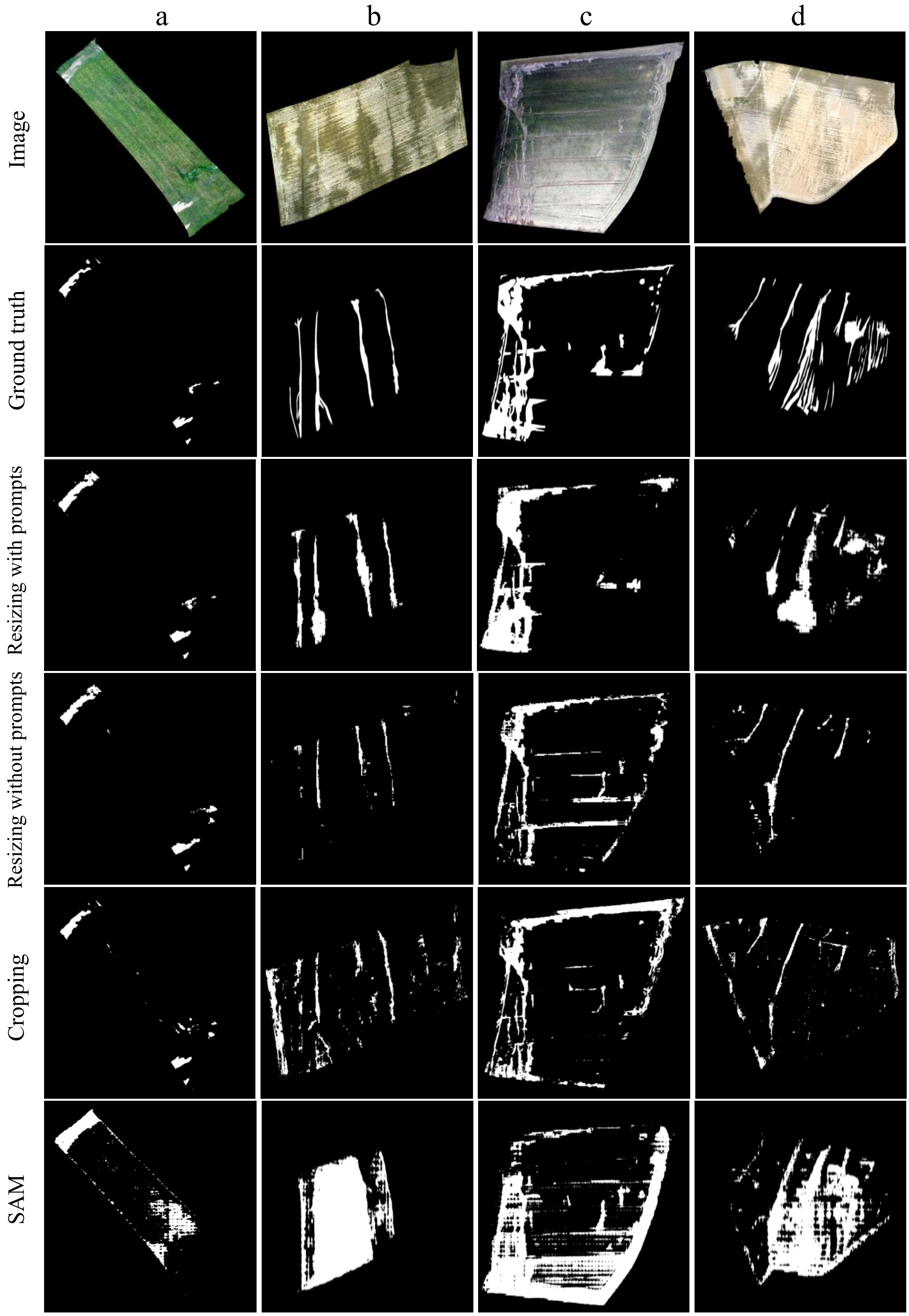


Figure 5.1.6. Overview of soil erosion segmentation with the original SAM and the fine-tuned models.

5.1.3.5. Implementation of field-based soil erosion assessment

The field-based correlation analysis (Figure 5.1.7) revealed that the SE ratios predicted by the resizing approach with point prompts showed the highest agreement with the ground truth data (R^2 = 0.93, MAE = 1.5), followed by the resizing approach without prompts (R^2 = 0.70, MAE = 3.19) and the cropping approach (R^2 = 0.58, MAE = 3.74).

Confusion matrices show the classification performance for SE severity in three different classes (Figure 5.1.8). The prompt-based resizing approach showed superior accuracy, especially for classifying the lower severity classes. For Class 1 (minimal erosion), an accuracy of 97.5% was achieved, with only 2.5% misclassified as Class 2. Class 2 (moderate erosion) also performed solidly, with only 5.9% of fields being misclassified as Class 1. For Class 3 (severe erosion), the model achieved a moderate accuracy of 66.7%, with the remaining 33.3% being misclassified as Class 2.

The standard resizing approach (without point prompts) maintained high accuracy for Class 1 (95.0%), with 5.0% misclassification as Class 2. However, performance for Class 2 decreased, with only 64.7% of fields correctly classified and 35.3% incorrectly classified as Class 1, suggesting that it is difficult to distinguish between Class 2 and Class 1. The performance for Class 3 was similar to the prompt-enhanced approach with 66.7% accuracy and 33.3% misclassification as Class 2.

The cropping approach showed a comparatively lower performance for Class 1 with an accuracy of 90.0%. For Class 2, the accuracy of the model decreased, with 76.5% of fields correctly classified as Class 2 and 23.5% incorrectly classified as Class 1, again indicating some confusion between Class 2 and Class 1. The accuracy for Class 3 remained the same as the other approaches at 66.7%, with 33.3% incorrectly classified as Class 2.

Overall, it can be concluded that both standard resizing and the cropping approaches tend to underestimate the severity of SE compared to the ground truth values. In contrast, the prompt-enhanced resizing approach achieved consistently higher classification accuracy at all severity classes, excelling particularly in distinguishing between Class 1 and Class 2 categories.

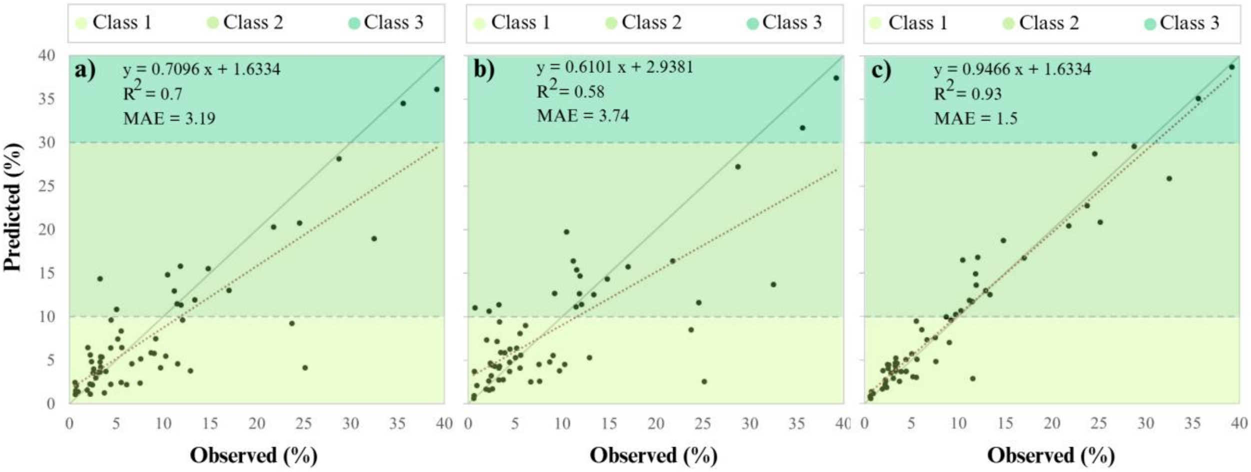


Figure 5.1.7: Prediction of field-based soil erosion ratios using a) without prompt resizing, b) cropping and c) with prompt resizing approaches.

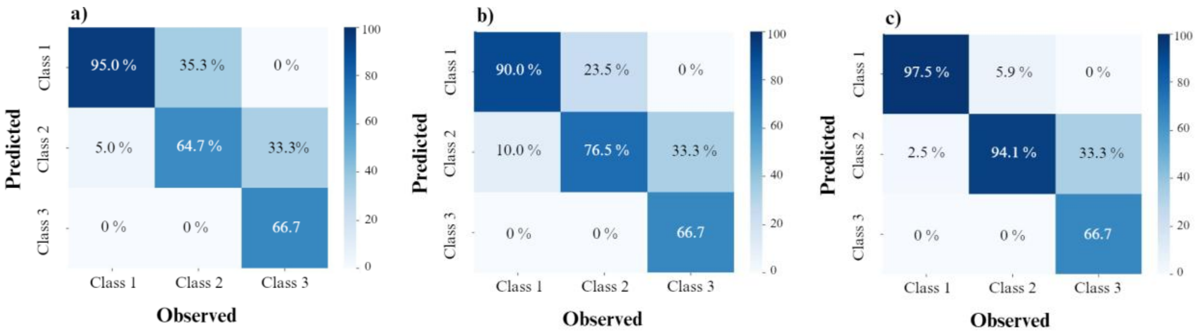


Figure 5.1.8. Confusion matrix showing the percentage of correctly predicted and incorrectly predicted field-based soil erosion classes using a) without prompt resizing, b) cropping and c) with prompt resizing approaches.

5.1.4. Conclusion

This study introduces the innovative Erosion-SAM model, which automatically detects erosion and deposition features for each pixel of high-resolution orthophotos. By fine-tuning the Segment Anything Model for bare cropland, vegetated cropland, and grassland, Erosion-SAM overcomes significant challenges in SE detection. Three methodological approaches were evaluated: resizing, cropping, and an improved resizing method with user-defined prompts during testing. Our comparative analysis showed that fine-tuning significantly improves the accuracy of the SAM model in detecting areas of erosion and deposition and enables accurate segmentation of SE features in remote sensing data. These results highlight the potential of the Erosion-SAM model in overcoming the challenges of SE detection, especially when limited labeled data is available. It provides the unique opportunity to obtain validation data for traditional physically-based and empirical SE models as well as large training data sets for data-based erosion models using machine learning approaches. This innovation enables not only detailed environmental monitoring but also information for policy decisions, agricultural practices and risk assessments for erosion-prone areas, including applications in insurance and land management. Although the results of SE segmentation are promising, this study is the first of its kind and further research is needed. Future studies should investigate the training of the Erosion-SAM model with larger and more accurate data sets. Since our current labeled data were segmented by only one expert, incorporating multiple expert annotations would standardize the data and potentially improve model accuracy. In addition, the relatively small data set used for training underscores the need for expanded data sets to improve the robustness and generalizability of the model. Future research could investigate the integration of high-resolution satellite remote sensing data such as WorldView-3, GeoEye-1 and KOMPSAT-3 which could extend the applicability of Erosion-SAM for large-scale studies. Such efforts would enable monitoring with higher temporal and spatial resolution and provide important insights for long-term environmental management. Further comparative studies are also recommended to evaluate the performance of Erosion-SAM against traditional and machine learning-based SE models in both localization and classification of erosion and deposition areas. These studies would deepen the understanding of the strengths and limitations of the model and support its integration into broader environmental monitoring systems. In summary, Erosion-SAM represents a significant advance in the automatic detection and segmentation of SE. Its potential to revolutionize SE monitoring at fine spatial and temporal scales makes it a valuable tool for addressing global challenges related to land degradation, food security and climate change adaptation.

CRediT authorship contribution statement

Hadi Shokati: Conceptualization, Data curation, Formal analysis, investigation, Methodology, Software, Visualization, Writing – original draft. **Andreas Engelhardt:** Methodology, Software, Validation, Writing – review and editing. **Kay Seufferheld:** Data curation, Writing – review and editing. **Ruhollah Taghizadeh-Mehrjardi:** Conceptualization. **Peter Fiener:** Conceptualization, Data curation, Funding acquisition, project administration, Supervision, Writing – review and editing. **Hendrik P.A. Lensch:** Conceptualization, Supervision. **Thomas Scholten:** Conceptualization, Data curation, Funding acquisition, project administration, Supervision, Writing – review and editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding sources

This work was supported by the Deutsche Forschungsgesellschaft (DFG) [No. SCHO 739/25-1].

Data availability

The aerial images are subject to license restrictions and cannot be shared. However, other data will be made available upon request.

5.2. Manuscript IV: Rapid flood mapping from aerial imagery

The second part of **Chapter 5** is based on **Manuscript IV** (Shokati et al., 2026), which extends the image-based monitoring framework from erosion to flood impacts by comparing fine-tuned SAM and U-Net architectures with ResNet backbones for rapid flood area segmentation from UAV and helicopter imagery. The study targets the growing need for fast, reliable flood extent mapping to support disaster response and damage assessment, where traditional manual delineation is often too slow for operational use. By systematically contrasting a foundation-model-based approach (SAM) with a more classical encoder–decoder segmentation architecture (U-Net), the study investigates the trade-offs between model complexity, data requirements, segmentation accuracy and practical applicability for near-real-time flood mapping. In the context of the dissertation's guiding questions, this study primarily addresses Research Question 5 on the suitability of deep learning segmentation architectures for rapid, operational flood mapping.

The analysis is based on a high-resolution flood image dataset with corresponding ground-truth masks, from which training, validation and test subsets are constructed for model development and evaluation. Both SAM and U-Net variants are fine-tuned under comparable experimental setups, and their performance is assessed using established segmentation metrics as well as qualitative inspection of the resulting flood masks. Particular emphasis is placed on how the models handle heterogeneous flood scenes, such as mixtures of open water, infrastructure and vegetation, and on their robustness under limited training data conditions.

The structure of the second part of **Chapter 5** is as follows. Subsection 5.2.1 introduces the flood mapping problem, the motivation for using deep learning and the specific research questions addressed. Subsection 5.2.2 details the dataset, network architectures for SAM and U-Net, the experimental training setup and the performance evaluation protocol. Subsection 5.2.3 presents and discusses the training and validation behavior, quantitative segmentation results and qualitative examples, including an assessment of model effectiveness in real-world applications and the influence of dataset size and diversity. Subsection 5.2.4 concludes with key insights on the relative merits of SAM and U-Net for operational flood mapping and outlines implications for future model development and data collection.

Manuscript IV: Rapid flood mapping from aerial imagery using fine-tuned SAM and ResNet-backboned U-Net

Abstract

Flooding is a major natural hazard that requires a rapid response to minimize the loss of life and property and to facilitate damage assessment. Aerial imagery, especially images from unmanned aerial vehicles (UAVs) and helicopters, plays a crucial role in identifying areas affected by flooding. Therefore, developing an efficient model for rapid flood mapping is essential. In this study, we present two segmentation approaches for the mapping of flood-affected areas: (1) a fine-tuned Segment Anything Model (SAM), comparing the performance of point prompts versus bounding box prompts, and (2) a U-Net model with ResNet-50 and ResNet-101 as pre-trained backbones. Our results showed that the fine-tuned SAM performed best in segmenting floods with point prompts (Accuracy: 0.96, IoU: 0.90), while bounding box prompts led to a significant drop (Accuracy: 0.82, IoU: 0.67). This is because flood images often cover the image from edge to edge, making bounding box prompts less effective at capturing boundary details. For the U-Net model, the ResNet-50 backbone yielded an accuracy of 0.87 and an IoU of 0.72. Performance improved slightly with the ResNet-101 backbone, achieving an accuracy of 0.88 and an IoU of 0.74. This improvement can be attributed to the deeper architecture of ResNet-101, which allows more complex and detailed features to be extracted, improving U-Net's ability to segment flood-affected areas accurately. The results of this study will help emergency response teams identify flood-affected areas more quickly and accurately. In addition, these models could serve as valuable tools for insurance companies when assessing damage. Moreover, the segmented flood images generated by these models can serve as training data for other machine learning models, creating a pipeline for more advanced flood analysis and prediction systems.

5.2.1. Introduction

Floods are among the world's most common, pervasive, and expensive natural disasters (Smith et al., 2014; Tingsanchali, 2012). Floods cause the greatest number of fatalities, according to the United Nations (Kuenzer et al., 2013). Global warming is expected to increase the frequency and intensity of floods, further amplifying their devastating impacts (Kamilaris & Prenafeta-Boldú, 2018). Recent catastrophic floods in Australia (Kelly et al., 2023), Japan (Lin et al., 2020), Spain (Ezzatvar & López-Gil, 2024) and Germany (Lehmkuhl et al., 2022) underline that flooding is not limited to the countries of the Global South. However, in countries where the infrastructure for flood protection is inadequate, the loss of human life is significantly higher. In the Global North, economic losses are generally higher than in the Global South (Taguchi et al., 2022), as long as no major infrastructure projects such as dams are affected. Effective flood management is therefore essential to mitigate these losses.

Flood management involves four stages: prevention, preparedness, response, and recovery (Plate, 2002). Remote sensing data plays a critical role in the response and recovery stages. This data is particularly valuable for assessing damage and providing actionable information to speed up recovery efforts, such as guiding insurance claims or resource allocation (De Leeuw et al., 2014). In recent years, satellite data from platforms such as Sentinel-2, Landsat 8 and Landsat 9 have been widely used for flood-affected area detection (Portalés-Julià et al., 2023). Satellite imagery provides valuable information to identify flood zones, assess damage and support flood forecasting and risk assessment models. However, due to revisit limitations, such data is not always available immediately after flood events. As a dense cloud cover is often associated with flood events, it can be challenging to obtain cloud-free images after floods. To overcome such limitations, images taken from UAVs and helicopters can provide an effective alternative to satellite data for flood detection (Hashemi-Beni et al., 2018). UAVs and helicopters can capture high-resolution images at lower altitudes and are less susceptible to cloud interference. They are also cost-effective and offer greater flexibility in the scheduling of surveys (Klemas, 2015; Shokati et al., 2023, 2024; Sugiura et al., 2005; Yao et al., 2019).

Rapid damage assessment is crucial for flood response, as delays can exacerbate the humanitarian and economic toll. However, recording and processing large volumes of aerial images in real time requires efficient and accurate automated methods. To counter this, conventional machine learning models such as Support Vector Machine (SVM), Random Forest (RF) and Maximum Likelihood Classifier (MLC) are often used for flood detection (Tanim et al., 2022). However, these models face challenges such as reliance on manual feature engineering, inability to capture complex features such as textural patterns and neglect of spatial correlations in data. To overcome these challenges, deep learning models have proven to be a powerful tool, demonstrating remarkable effectiveness in various computer vision applications such as image classification (Jackson et al., 2023; Qiao et al., 2024), object detection (Ye et al., 2023), and image segmentation (Shokati et al., 2025; Y. Zhang et al., 2023). In particular, numerous studies have leveraged segmentation techniques to identify areas affected by flooding, demonstrating their potential to address real-world challenges. For example, (Pally & Samadi, 2022) developed a flood image classifier using different convolutional neural network (CNN) architectures for segmentation and object detection to calculate water levels and flood areas. In another study, Safavi and Rahnemoonfar, (2023) compared the performance of different encoder-decoder and two-pathway architectures to segment flood-affected areas. Similarly, Wieland et al., (2023) investigated the use of CNNs in detecting water bodies from high-resolution remote sensing imagery. In addition, Wagner et al., (2023) compared 32 CNNs for water segmentation using a dataset of 1128 images depicting river water surfaces.

Deep learning-based approaches in hydrology and especially in flood management have made significant progress in recent years (e.g. Pan et al., 2019; Wang et al., 2020). However, many models still rely on large datasets for training, which can affect their performance if the data are limited, unbalanced or specialized (e.g. Mallah et al., 2022). The model may memorize its features in such cases, leading to overfitting (Safonova et al., 2023). Data augmentation and transfer learning are the most commonly used methods to work with small or highly specialized datasets (Safonova et al., 2023). In data augmentation, new samples are generated by applying different transformations, with the choice of method depending on the type, quality and quantity of data, and transfer learning uses models pre-trained on large datasets, adapting their learned features to new tasks (Safonova et al., 2023). An example of the application of these techniques to soil erosion is presented by (Shokati et al., 2025) who fine-tuned the SAM, a model pre-trained on a large dataset with over 1 billion masks from 11 million images, to segment erosion and deposition features in agricultural fields. Their approach demonstrated that despite the complexity of erosion and deposition processes and their detection, the fine-tuned SAM model achieved high performance. Another transfer-learning approach uses a residual U-Net architecture (Ronneberger et al., 2015), which improves segmentation performance by utilizing pre-trained features. U-Net is widely known for its effectiveness in segmentation tasks such as hydrological streamline detection (Xu et al., 2021) and sea-land segmentation (Shamsolmoali et al., 2019). Incorporating residual connections, as implemented by Onojeghuo et al., (2023), can further improve feature propagation and model convergence. The availability of pre-trained deep neural networks such as ResNet (He et al., 2016) trained on large datasets such as ImageNet (Krizhevsky et al., 2012) facilitates cross-domain knowledge transfer, for example, from natural image classification to remote sensing image segmentation. In particular, combining CNN architectures such as U-Net with ResNet and transfer learning has improved performance on complex tasks such as water body segmentation (Ghaznavi et al., 2024) and plant mapping (Onojeghuo et al., 2023).

Despite significant progress in applying transfer learning in various areas of computer vision, its application to small aerial image datasets for flood management is still largely unexplored. The primary goal in this context is to map the extent and location of flooded areas. Rapid data collection with helicopters and UAVs, for which only a small dataset is required, makes it possible to use this approach to evaluate preventive and preparatory measures before and during flood events. Furthermore, integrating satellite data could eventually lead to a comprehensive flood forecasting and monitoring system.

The central aim of this study is to test the potential of two advanced transfer learning techniques - the fine-tuned SAM and the U-Net architecture with ResNet-50 and ResNet-101 backbones - for automated and fast flood area detection and mapping. By automating the mapping process, our approach aims to speed up damage assessment and enable authorities to respond more efficiently and mitigate the financial and human toll of flooding. The priority is to record the extent and location of the flooded areas to assess the extent of the damage and predict possible further damage. Specifically, we address the following research questions:

1. Which combination of prompt (bounding box or point prompts) for SAM and backbone (ResNet-50 or ResNet-101) for U-Net provides the best performance in flood area detection?

2. How do the fine-tuned SAM and U-Net architecture differ in terms of segmentation accuracy for flood area detection using UAV and helicopter imagery?
3. How do the elements of the sky, such as clouds or open sky, affect segmentation accuracy?

5.2.2 Materials and methods

5.2.2.1 Dataset

This study uses a flood area dataset comprising 290 images with their corresponding masks acquired from Karim et al., (2022). The images were captured using UAVs and helicopters with optical sensors in different regions, including flood events in southern Germany (2013), Karnataka, Kerala and Maharashtra in India (2019), Sabah in Malaysia (2021) and Bangladesh (2022). The dataset includes a variety of scenes, including rural areas, urban areas, peri-urban areas, greenery, buildings, mountains, rivers, sky, and roads, with masks created using Label Studio software (Figure 5.2.1).

Due to the use of different platforms and sensors, the imaging conditions were inherently inconsistent. The camera angle during imaging was inconsistent. In some cases, no clouds or sky elements were visible due to the low camera angle, while in other cases, sky elements such as clouds or open sky were visible due to greater camera rotation. In addition, the images were taken at different heights, resulting in different spatial resolutions. The image dimensions also varied considerably: the width ranged from 219 to 3648 pixels and the height from 330 to 5472 pixels. To ensure the uniformity of the dataset for modeling purposes, all images were resized to 256 × 256 pixels prior to analysis.

Figure 5.2.1. Example images from the Flood Area dataset (top) with their corresponding ground truth masks (bottom). Images and their corresponding masks are from Karim et al., (2022).

5.2.2.2 Network Architecture

5.2.2.2.1 Fine-tuning Segment Anything Model

The Segment Anything Model (SAM) developed by Meta AI Inc., USA is an image segmentation model trained on 1 billion masks extracted from 11 million images. As the name suggests, it can segment any image without the need for additional training data (Kirillov et al., 2023). The architecture of SAM is composed of three primary components: an image encoder, which is built on a robust Vision Transformer (ViT) backbone and extracts features from the input image; a prompt encoder, which uses the input prompts to create embeddings; and a mask decoder, which generates the final mask by combining the outputs of the previous components (Figure 5.2.2).

The SAM architecture enables the integration of human prompts, which increases the efficiency of annotation by the human in the loop. The prompts guide the model to focus on specific regions of the image, improving segmentation accuracy. These prompts can take different forms, e.g., bounding box, point and text. To implement prompt-based interaction with the SAM architecture, we used both bounding box and point prompts in a fully automated manner. For the bounding box prompts, we first identified all foreground pixels in the annotated images (pixels representing flood). The bounding box was then computed as the smallest rectangle enclosing all foreground pixels in each binary mask by determining their minimum and maximum x and y coordinates. For the point prompts, we randomly selected 30 foreground pixels per image. To ensure spatial diversity and avoid clustering, a minimum Euclidean distance of 10 pixels was set between two selected points. This restriction contributed to a more representative coverage of the flooded area, which in turn improves segmentation accuracy.

SAM offers several variants, each tailored to different computational requirements and based on distinct configurations of the Vision Transformer (ViT) backbone: ViT-Base, ViT-Large, and ViT-Huge, containing approximately 91M, 308M, and 636M parameters, respectively (Kirillov et al., 2023). Upon evaluating these variants, we observed that their effectiveness in detecting flood-affected areas was remarkably comparable. To optimize our computational resources, we opted for the ViT-Base variant as it offers a favorable balance between performance and efficiency.

Although SAM can process any image without additional training, its performance may be limited in complex tasks. In such cases, fine-tuning can enhance its segmentation accuracy (Shokati et al., 2025). Fine-tuning is a transfer learning technique that applies a pre-trained model, such as SAM, which has already learned general patterns from a large dataset. A task-specific dataset, such as a flood dataset, is then prepared and formatted to match the model's input requirements. Certain layers of the model are either frozen or modified to regulate the extent to which training updates alter the original knowledge. The model is then trained with the new dataset at a lower learning rate to refine its understanding without overwriting prior knowledge. After training, the model's performance is evaluated using a validation set, and hyperparameters are adjusted for optimization.

Once the model achieves satisfactory accuracy, it is deployed and continuously monitored to ensure robust performance on real-world data. In this study, the mask decoder of SAM (Figure 5.2.2b) is fine-tuned (modified) as it has a simple and efficient design that requires fewer computational resources (Sun et al., 2024). This ensures that the fine-tuning process is fast, efficient, and requires less memory. In this process, the other two components, the image encoder (Figure 5.2.2a) and the prompt encoder (Figure 5.2.2c) are kept fixed (frozen), meaning their parameters are not updated during fine-tuning. This ensures that only the mask decoder is modified.

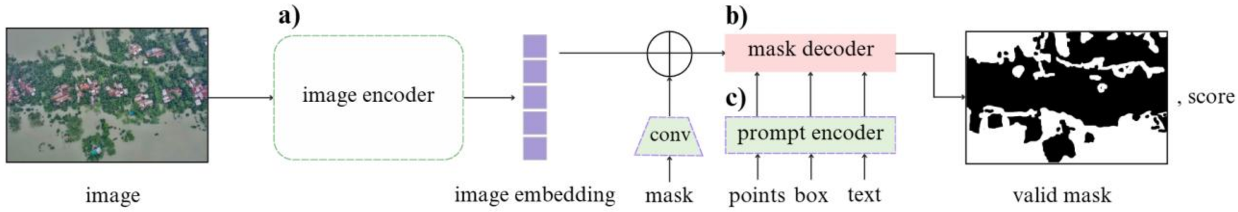


Figure 5.2.2. Schematic figure of the Segment Anything Model architecture, where the image encoder processes the image to extract features, which are combined with the prompt encoder's embeddings and passed to the mask decoder to generate the final segmentation mask. Image and mask are from Karim et al., (2022).

5.2.2.2.2 U-Net architecture with ResNet as the backbone

U-Net is a fully convolutional neural network introduced in 2015 for segmenting biomedical images (Ronneberger et al., 2015). Over time, it has been adapted for other fields, such as remote sensing. With its U-shaped architecture, U-Net comprises a down-sampling path (encoder) and an up-sampling path (decoder). This structure enables the convolutional network to learn and combine features at various levels of detail, which is critical for accurately segmenting small regions and fine details in images.

The down-sampling path follows a typical convolutional network architecture. It involves repeated applications of 3×3 convolutions (without padding), each followed by a ReLU (Rectified Linear Unit) activation and a 2×2 max-pooling operation with a stride of 2 to reduce the spatial dimensions. At each down-sampling stage, the number of feature channels doubles, enabling the network to extract increasingly complex features while progressively losing spatial information. This part of the network is crucial for capturing high-level features from the input image.

The up-sampling path is the reverse of the down-sampling path. This part of the network utilizes deconvolution operations to double the spatial dimensions of the feature maps. Following each deconvolution, a concatenation operation is performed with the corresponding feature maps from the down-sampling path to restore spatial information. This process helps recover the details lost during the down-sampling phase and enhances the network's spatial accuracy for final segmentation. Skip connections are used to concatenate feature maps from the corresponding layers in the encoding path to the decoding layers, ensuring the recovery of information lost during down-sampling. At the end of the architecture, a 1×1 convolutional layer is applied to map the feature maps to the desired number of segmentation classes. This final layer assigns a class label to each pixel, producing the segmented output.

To achieve better results with limited data, we apply knowledge from transfer learning in this study. Specifically, we use a residual neural network (ResNet) (He et al., 2016) pre-trained on ImageNet as the encoder backbone of U-Net. The weights of the pre-trained encoder are kept frozen to utilize the existing low-level feature representations (such as edges, corners, and textures), while only the U-Net decoder is trained on our dataset. This allows the decoder to adapt to our segmentation task without updating the encoder's weights (Figure 5.2.3).

ResNet is a convolutional neural network (CNN) that uses identity skip connections to solve the degradation issue that occurs when accuracy reaches saturation and rapidly deteriorates as network depth grows. Convolutions of 1 × 1, 3 × 3, and 1 × 1 series are used to stack several bottleneck residual blocks to create it. There are several variations of the ResNet model depending on the network's depth, including ResNet18, ResNet34, ResNet50, ResNet101, and others. ResNet50 and ResNet101, which have 50 and 101 layers, respectively, were used in this study.

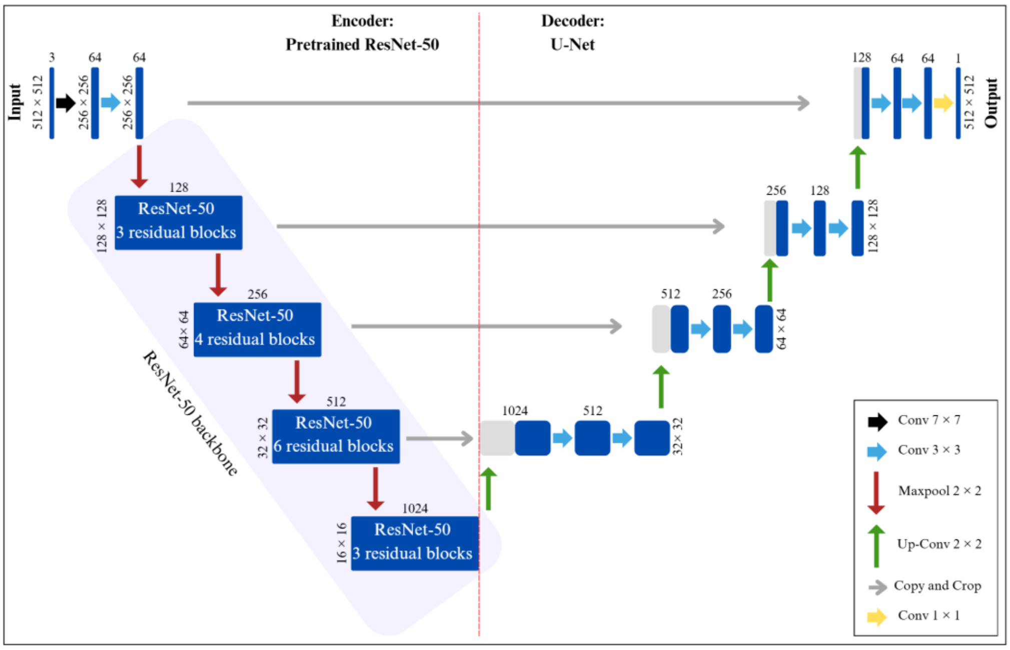


Figure 5.2.3. Architecture of U-Net with ResNet-50 backbone (adapted from Manos et al., (2022)).

5.2.2.3 Experimental Training Setup

For the segmentation model, 70% of the flood area dataset (204 samples) was used for training. Data augmentation was applied exclusively to the training set to increase the diversity of the training data. These included geometric transformations such as random horizontal and vertical flips and rotations of up to 30° as well as color-based transformations such as random grayscale transformations and Gaussian blurs with a kernel size of 3, all applied with a probability of 0.5. The model was trained for 50 epochs, with each epoch representing a complete pass the training data through the model. A batch size of 4 and a learning rate of $1e^{-3}$ were used. A learning rate scheduler was employed to adjust the learning rate if the validation loss did not decrease over 10 consecutive epochs, reducing the previous learning rate by multiplying it by a factor of 0.1. Additionally, 15% of the dataset (43 samples) was used for model validation at the end of each epoch. Among the 50 models generated, the one with the lowest validation loss was selected as the best model and tested on the remaining dataset (15%, 43 samples). The Adam optimizer algorithm was utilized due to its strong adaptability (C. Liu et al., 2023). To minimize the discrepancy between observed and predicted flood extents, we used the Dice-Cross-Entropy Loss, which averages the Dice loss and cross-entropy loss. This composite loss function is widely used in model training, as it balances the strengths of both components (Hadlich et al., 2023; Shokati et al., 2025). It facilitates rapid convergence and often improves final performance, particularly enhancing the Dice coefficient (Hadlich et al., 2023), which is critical for accurately capturing the spatial overlap between predicted and actual flood areas. The dice and cross-entropy losses are calculated as:

$$L_{Dice} = 1 - \frac{2\sum_{c=1}^{C}\sum_{i=1}^{N} g_i^c\, S_i^c}{\sum_{c=1}^{C}\sum_{i=1}^{N} g_i^c + \sum_{c=1}^{C}\sum_{i=1}^{N} S_i^c} \tag{5.2.1}$$

$$L_{CE} = -\frac{1}{N}\sum_{c=1}^{C}\sum_{i=1}^{N} \mathrm{g}_i^c\, \log S_i^c \tag{5.2.2}$$

Where L_{Dice} and L_{CE} are dice and cross-entropy losses, respectively, N is the number of pixels, g_i^c is the ground truth binary indicator of the class label c of pixel i, and S_i^c is the corresponding predicted segmentation probability.

Experiments were carried out using Python and the PyTorch framework on a Windows 11 computer equipped with an NVIDIA GeForce RTX 4070 Ti GPU.

5.2.2.4 Model performance evaluation

The performance of segmentation models is usually evaluated using several metrics, e.g. intersection-over-union (IoU), dice coefficient, recall, precision and accuracy. IoU quantifies the overlap between the predicted and true flood-affected regions, with higher values indicating better model performance in terms of accurately identifying flood-affected areas. The dice coefficient is similar to IoU, but emphasizes correct positive predictions, making it particularly useful in scenarios with class imbalance. Recall measures the model's ability to correctly identify all flood-affected regions, with higher values reflecting fewer overlooked regions. Precision evaluates the accuracy of the model's positive predictions and indicates the proportion of correctly identified flood-affected regions out of all predicted positive outcomes. Finally, accuracy provides an overall measure of the correctness of the model, considering both correct predictions of flood-affected and non-flood-affected regions. However, it may not be as informative for highly imbalanced datasets. Each of these metrics provides different insights into the performance of the model and can be used together for a comprehensive assessment. They are given in the following equations (Vinayahalingam et al., 2023):

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{5.2.3}$$

$$\text{Dice coefficient} = \frac{2\,TP}{2\,TP + FP + FN} \tag{5.2.4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5.2.5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.2.6}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.2.7}$$

Where TP, TN, FP, and FN stand for the pixel labels for true positives, false positives, true negatives, and false negatives, respectively.

5.2.3 Results and discussion

5.2.3.1 Training and validation losses

Both U-Net (with ResNet-50 and ResNet-101 as backbones) and fine-tuned SAM (with bounding box and point prompts) models were trained for 50 epochs (Figure 5.2.4). For the SAM models, whether trained with point or bounding box prompts, the training loss exhibited a sharp decline at the beginning of the training process. This was followed by a slower, more gradual decrease until the final epochs. Similarly, the validation loss initially experienced a steep decline, which then slowed progressively until the point where the downward trend almost stopped, especially for bounding box prompts where the validation loss even increased very slightly, which could be a sign of overfitting (Shorten & Khoshgoftaar, 2019). A comparison of training and validation losses between the point and the bounding box prompts revealed that the model trained with the point prompts generally performed better in both learning the training set and generalizing to the validation set.

For the U-Net model, both training loss and validation loss gradually decreased throughout the training process, whether ResNet-50 or ResNet-101 was used as the backbone. This indicates that the model effectively learned the relationships within the training set while also improving its ability to generalize to unseen data (validation set). The training and validation loss values were nearly identical when using ResNet-50 and ResNet-101 as backbones, suggesting that both backbones performed similarly in terms of learning and generalization capabilities.

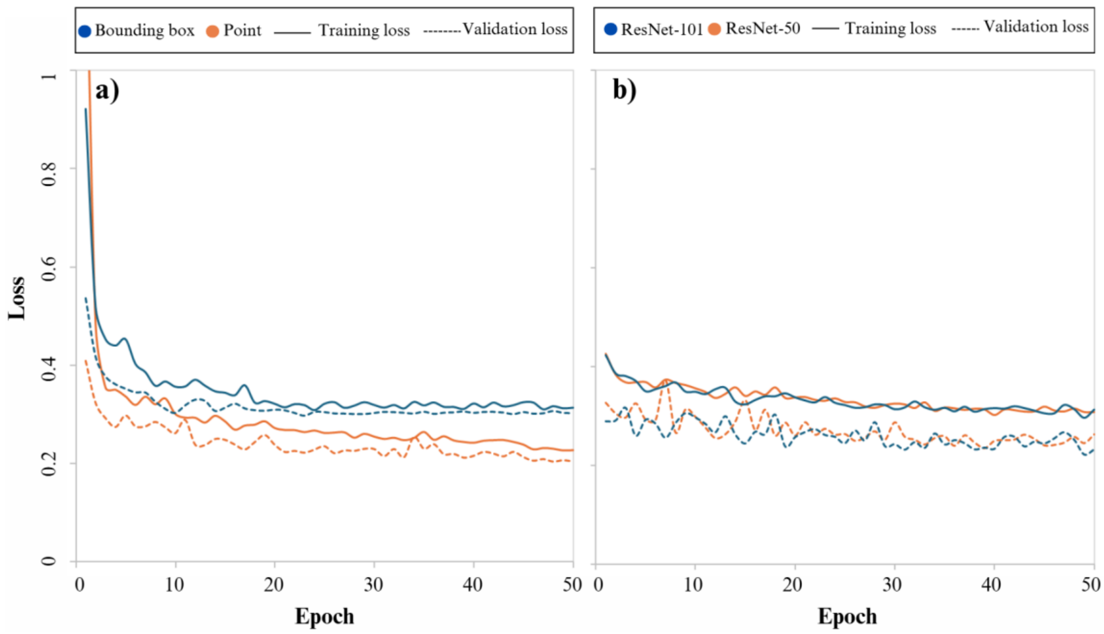


Figure 5.2.4. Training and validation losses for a) Segment Anything Model with point and bounding box prompts and b) U-Net model with ResNet-50 and ResNet-101 backbones.

5.2.3.2 Performance of the model on the validation set

For the SAM model with point prompts (SAM-Points), all evaluation metrics consistently increased during the initial epochs, followed by stabilization in the later epochs (Figure 5.2.5a). The SAM model with bounding box prompts (SAM-Bbox) demonstrated a similar trend, with metrics steadily increasing before reaching stability (Figure 5.2.5b). Comparing these two prompting methods revealed that the SAM-Points (Figure 5.2.5a) outperformed the SAM-Bbox (Figure 5.2.5b) regarding overall validation performance.

For the U-Net model, the metrics for the ResNet-50 backbone showed a steady improvement, eventually reaching stabilization (Figure 5.2.5c). The U-Net model with the ResNet-101 backbone followed a similar trajectory, with only minor differences in terms of stability (Figure 5.2.5d). A direct comparison of the two backbone configurations revealed that both ResNet-50 and ResNet-101 backbones performed similarly in learning and generalization for the U-Net model.

When comparing the performance of the SAM model to the U-Net model, regardless of the prompt type or backbone configuration, the SAM model consistently outperformed the U-Net model on the validation set.

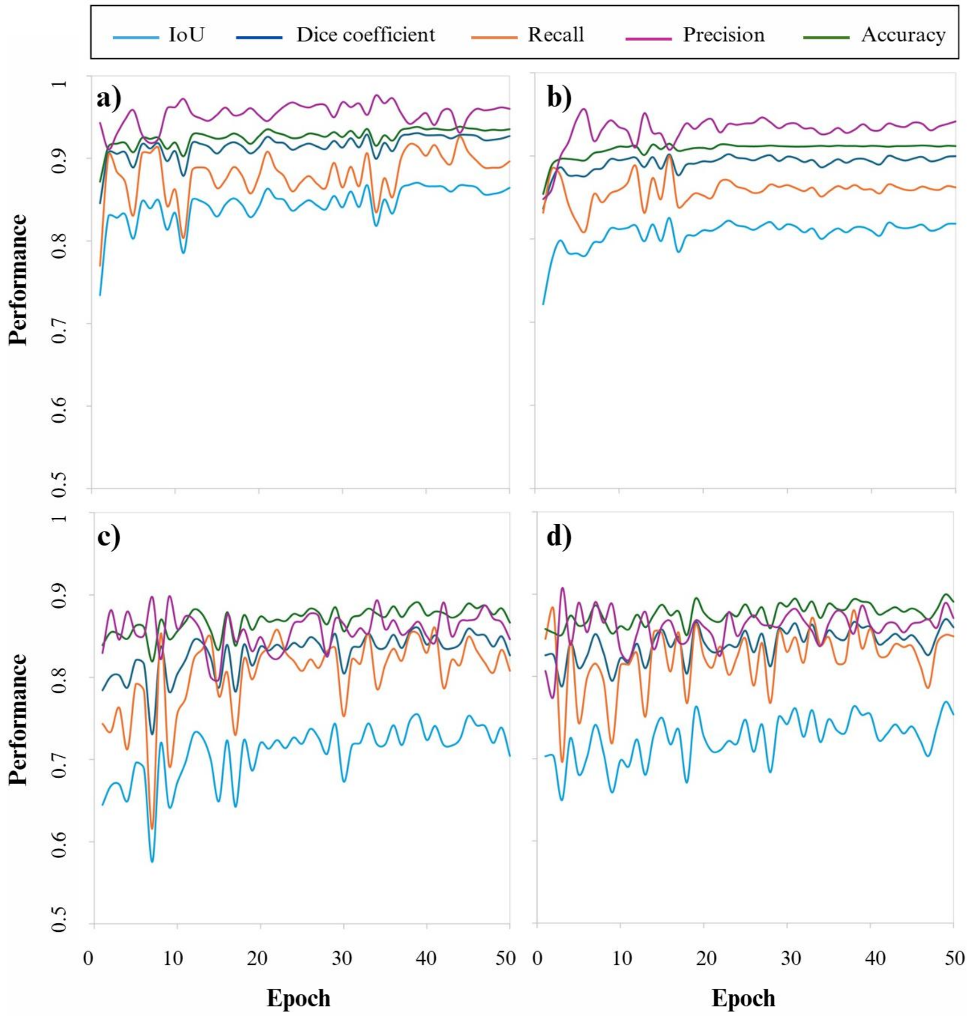


Figure 5.2.5. Performance of the Segment Anything Model on validation set using a) point prompts and b) bounding box prompts and the performance of U-Net model with c) ResNet-50 and d) ResNet-101 backbones.

5.2.3.3 Evaluating segmentation results

In this study, the performance of the SAM and U-Net models was evaluated for segmenting flooded areas. The SAM model was assessed with two types of prompts, point prompts and bounding box prompts, while the U-Net model was tested with two backbone types, ResNet-50 and ResNet-101. The results of the different metrics are shown in Table 5.2.1, and examples of the segmentation results of all models are shown in Figure 5.2.6.

The results indicated that the SAM-Points outperformed all other models (Table 5.2.1). The strong performance of this model can be primarily attributed to the ability of the point prompts to provide precise spatial information. This information allows the model to recognize the exact boundaries of the flood-affected areas and to make accurate predictions. This ability is particularly important in regions where light intensity varies significantly (e.g. Figures 5.2.6e and 5.2.6h), or the boundaries of flooding are unclear. In addition, the architecture of the SAM model uses input prompts, which increases the accuracy of its predictions and contributes to its superior performance.

The SAM-Bbox showed the weakest performance among all other models (Table 5.2.1). Although SAM-Bbox achieved the highest recall (0.96), it performed worse on other metrics compared to SAM-Points. The higher recall indicates the model's strong ability to identify flood-affected areas. However, the lower precision (0.69) shows that it is difficult to delineate the boundaries accurately, resulting in extraneous or non-contiguous areas being part of the flood-affected areas. This limitation arises from the limited information provided by the bounding box prompts, which only provides the general framework of the target area without offering detailed boundary data. In other words, the bounding box prompts provide the model with limited information about the exact boundaries, which increases the probability of false-positive predictions.

While the existing literature suggests that bounding box prompts often perform better than point prompts in different contexts (Cheng et al., 2023; Gaus et al., 2024; Mazurowski et al., 2023; Xie et al., 2024), the choice of prompt is highly dependent on the specific nature of the dataset. Flood segmentation is a rare case where point prompts outperform bounding box prompts. In the data analyzed, the flooding areas often extend across the entire image from one edge to another. Consequently, a bounding box usually covers most of the image and provides the model with less detailed information than point prompts. Theoretically, this lower granularity of information from bounding box prompts leads to poorer performance in such cases. In addition, the inherently diffuse and irregular nature of flood boundaries makes point prompts stronger cues for guiding the model, while bounding box prompts typically include both flooded and non-flooded regions, providing the model with less discriminatory guidance. However, this observation must be interpreted in light of the specific assumptions and methodological choices made in our study. In particular, both types of prompts were generated fully automatically and without manual refinement, and the point prompts were spatially constrained to ensure dispersion across the flooded regions. These design decisions, in combination with the inherent characteristics of flood imagery, such as large, amorphous regions that often cover significant portions of the image, have a direct impact on the relative effectiveness of each prompt type. Therefore, the superior performance of point prompts observed in our experiments should not be generalized to other domains without carefully considering the characteristics of the dataset properties and the prompt generation strategies.

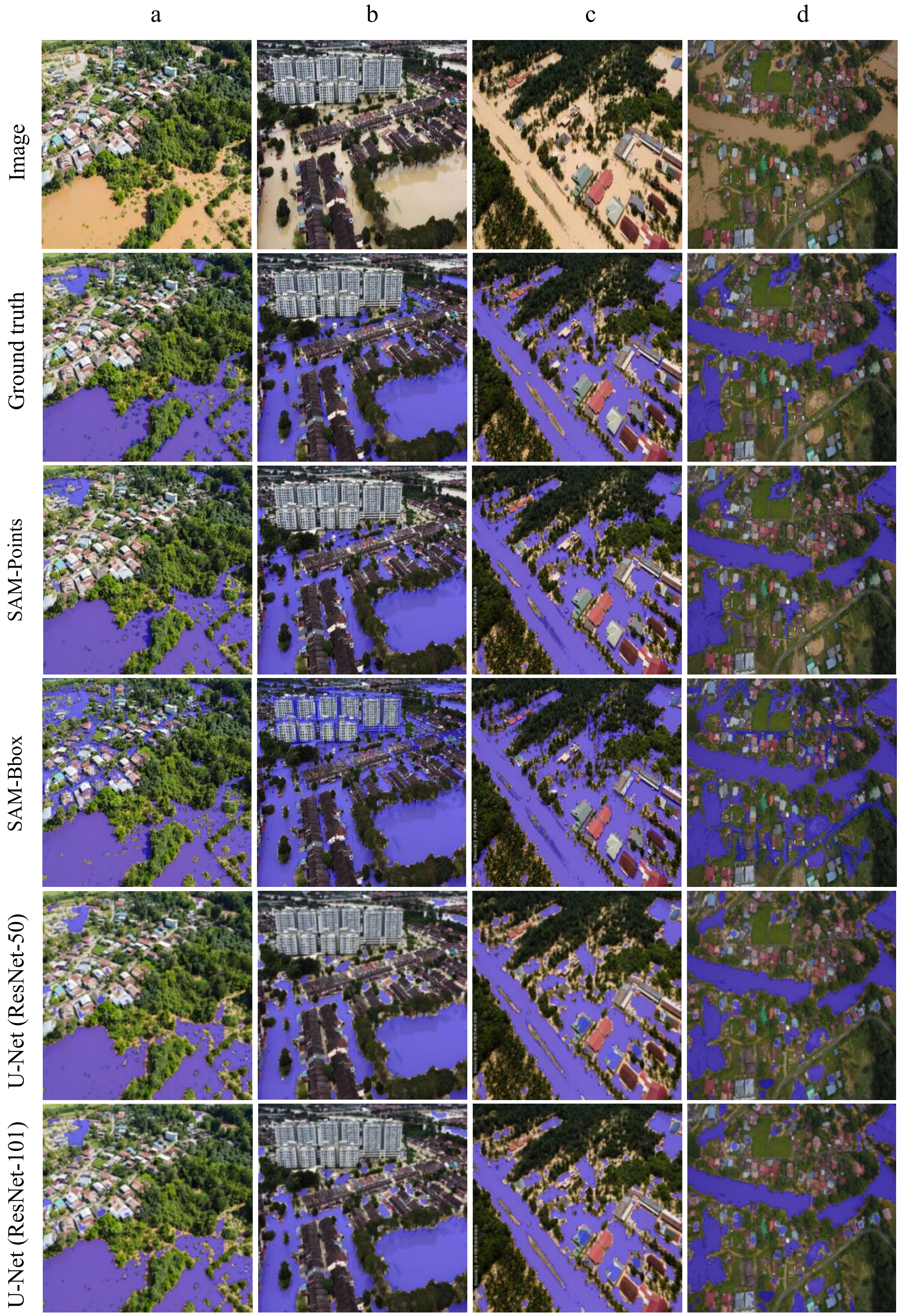
The U-Net models used in this study showed moderate performance in flood segmentation (Table 5.2.1). Notably, the U-Net model with a ResNet-101 backbone outperformed the model with a ResNet-50 backbone. This improved performance of the ResNet-101 backbone was also reported in a study by Ait El Asri et al., (2023) which focused on aerial image segmentation. The superior performance of the ResNet-101 can be attributed to its ability to extract more complex and detailed features from the images due to its deeper layers, allowing U-Net to identify flood-affected areas with higher accuracy. However, the difference between the two models was minimal, and U-Net with ResNet-101 only outperformed ResNet-50 on a few images (e.g. Figure 5.2.6g). The small difference in results suggests that increasing the network depth does not significantly improve performance in the context of flood segmentation. This could be because flooding patterns are not complex enough to require deeper architecture for effective analysis. Nevertheless, the performance of U-Net with both backbones lagged behind that of SAM-Points, highlighting the superiority of the SAM model in utilizing prompt-based information to improve prediction accuracy. The superior performance of the fine-tuned SAM model compared to U-Net was also noted in similar studies by Lehmiani et al., (2023) and Moghimi et al., (2024). However, although SAM demonstrates higher accuracy, it is important to consider the computational trade-offs. Moghimi et al. (2024) observed that SAM requires significantly more computational resources, with approximately three times the parameter count of models like U-Net and longer training times. Therefore, while SAM with point prompts provides the highest precision for flood mapping, U-Net remains a more computationally efficient alternative.

In summary, regarding the first research question, which investigates the optimal combination of prompts for SAM and backbones for U-Net, the results indicated that the SAM-Points performed better than the SAM-Bbox. In addition, the U-Net model with the ResNet-101 backbone performed better than the model with ResNet-50. In response to the second research question on the differences in segmentation accuracy between the fine-tuned SAM model and the U-Net model with ResNet-50 and ResNet-101 backbones, our results showed that the SAM-Points model significantly outperformed both U-Net configurations, while the SAM-Bbox model had the weakest performance among all models tested. This generally aligns with findings by Moghimi et al. (2024), who reported that fine-tuned SAM achieved superior segmentation accuracy compared to established deep learning models such as U-Net, DeepLabV3+, and PSPNet across multiple river scene datasets. However, a key methodological difference exists: Moghimi et al. (2024) fine-tuned SAM without using any input prompts to enable fully automated segmentation. In contrast, our study evaluated the impact of specific prompt types, revealing that while point prompts significantly outperformed U-Net, bounding box prompts failed to do so. This suggests that although SAM's architecture is inherently powerful, its performance in flood mapping is highly sensitive to the prompting strategy used.

To answer the research question regarding the effects of sky elements on segmentation accuracy, our analysis revealed that segmentation accuracy of all models decreased in images where sky elements such as clouds or open sky were present due to significant camera rotation angles, compared to conditions where the camera provided a near top-down view (e.g. Figures 5.2.6e and 5.2.6h). This finding is consistent with the results of Simantiris and Panagiotakis (2024). Moghimi et al. (2024) similarly highlighted that the dynamic nature of water, exhibiting varying colors and reflecting the sky and surrounding structures, poses significant difficulties for segmentation models, although deep learning approaches generally handle these artifacts better than traditional methods. This is because segmentation methods are often based on color information. If the color of objects such as the sky or clouds is similar to the color of the areas affected by the flood, the model may have difficulty distinguishing them from the flood and incorrectly segment them as part of the flood zone. This problem was less pronounced in the SAM-Points than in the other models, while it was more prominent in the SAM-Bbox, as each bounding box usually covers almost the entire image. In other words, sky elements can be falsely considered as potential regions for segmentation.

Table 5.2.1. Segmentation results of the test dataset using the Segment Anything Model with point and bounding box prompts (SAM-Points and SAM-Bbox models, respectively) and the U-Net model with ResNet-50 and ResNet-101 backbones (the highest accuracies are in bold).

Model	Accuracy	IoU	Dice	Precision	Recall
SAM-Points	**0.96**	**0.90**	**0.95**	**0.94**	0.95
SAM-Bbox	0.82	0.67	0.81	0.69	**0.96**
U-Net (ResNet-50)	0.87	0.72	0.84	0.83	0.85
U-Net (ResNet-101)	0.88	0.74	0.85	0.83	0.87


a
b
c
d
Image
Ground truth
SAM-Points
SAM-Bbox
U-Net (ResNet-50)
U-Net (ResNet-101)

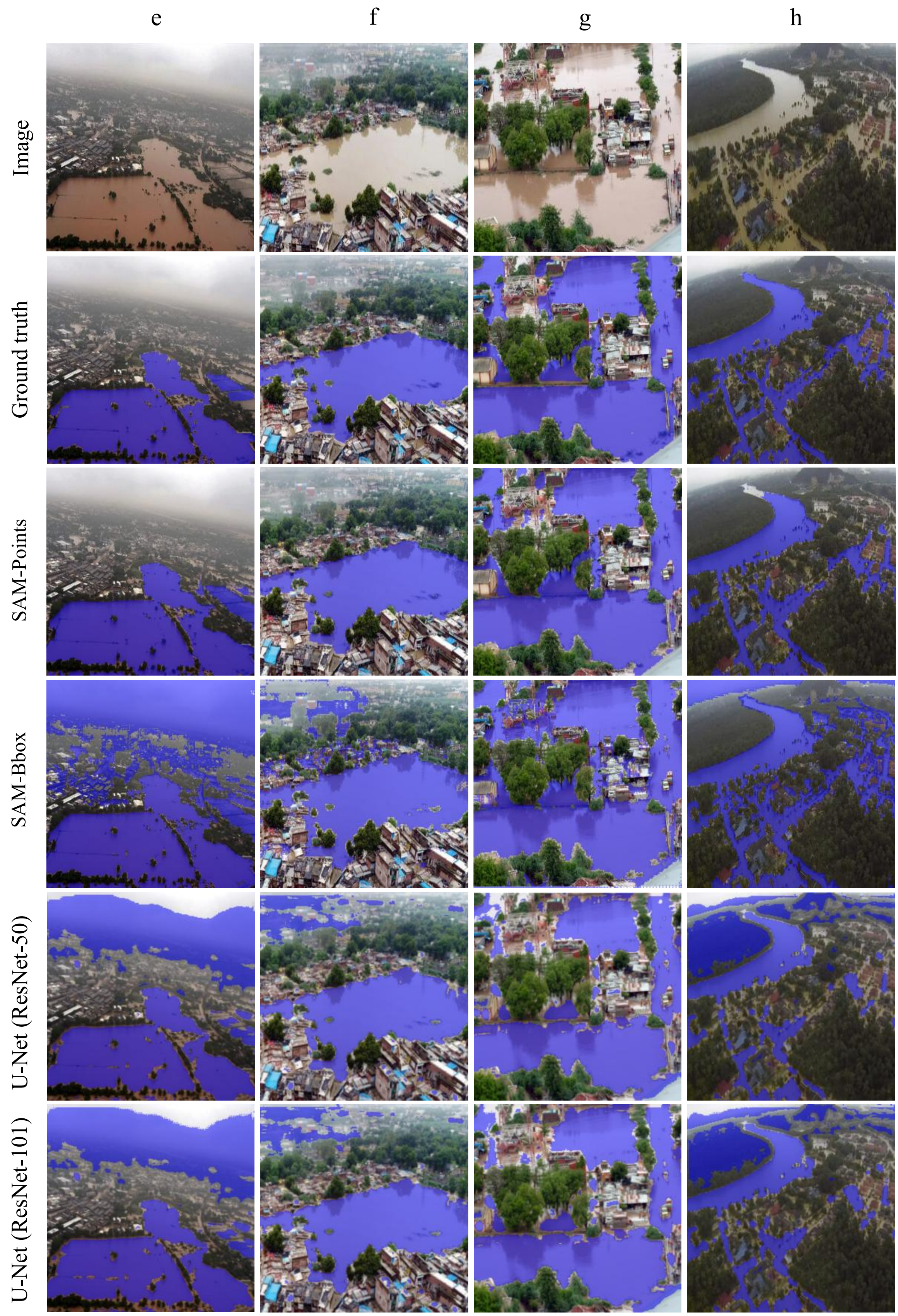


Figure 5.2.6. Example segmented images using the Segment Anything Model with point and bounding box prompts (SAM-Points and SAM-Bbox models, respectively) and the U-Net model with ResNet-50 and ResNet-101 backbones. Subplots (a–h) correspond to different samples from the dataset of Karim et al., (2022).

5.2.3.4 Effectiveness of the models in real-world applications

The results of this study highlight the effectiveness of U-Net (with ResNet-50 and ResNet-101 backbones) and fine-tuned SAM (with bounding box and point prompts) models in segmenting flood-affected areas. Despite variations in flight altitude, image quality, resolution and viewing angles, these models performed remarkably well. In real-world scenarios, where such variations are common, these models can be expected to process aerial images of affected regions quickly and without additional training. This rapid prediction capability can help emergency teams immediately identify flooded areas and take the necessary measures. In addition, these models are of great benefit to insurance companies in assessing damage, speeding up insurance payments and improving post-crisis services. In summary, these approaches offer significant benefits for flood crisis management and help minimize human and economic flood impacts through timely and precise interventions.

5.2.3.5 Dataset size and diversity considerations

Determining the optimal dataset size in transfer learning does not depend on a universal static threshold; rather, it is a function of task complexity, model architecture, and the domain similarity between the source and target tasks. In transfer learning, large-scale pre-trained models such as SAM (Kirillov et al., 2023) and ResNet (He et al., 2016) already capture rich, generalized feature representations from millions of natural images. This is further supported by recent research on water segmentation using SAM, which suggests that the model's robust spatial priors can achieve high performance even with minimal labeled data (Zamboni et al., 2025). As a result, a relatively small number of labeled samples is often sufficient for fine-tuning to achieve high performance in specialized applications.

Our dataset consists of 290 images covering flood events in countries such as Germany, India, Malaysia, and Bangladesh. This geographic diversity ensures variability in environmental conditions, land cover types, flood characteristics, and illumination. The inclusion of both UAV and helicopter imagery from different camera angles and altitudes further increases this variability, providing a robust basis for model generalization. Additionally, data augmentation techniques (such as rotations, flips, grayscale transformations, and Gaussian blur) increased the effective training diversity and reduced the risk of overfitting. Empirically, our results (Table 5.2.1) demonstrate that the fine-tuned SAM model achieved an IoU of 0.90 and an accuracy of 0.96 on unseen data, confirming that the dataset sufficiently captured the variability required for reliable flood segmentation. Comparable studies on environmental and remote sensing tasks (e.g., Ghaznavi et al., 2024; Shokati et al., 2025) have reported strong performance using datasets of similar size, reinforcing the suitability of our sample in the context of transfer learning–based flood segmentation.

5.2.4 Conclusion

In this study, we used two advanced transfer learning techniques: the fine-tuned SAM model and the U-Net architecture with ResNet-50 and ResNet-101 backbones to detect flood-affected areas. Using the Flood Area dataset, which includes UAV and helicopter aerial images of flood-affected areas, we aimed to evaluate and compare the performance of these models in accurately segmenting flood-affected areas. Our results showed that the fine-tuned SAM model outperformed the U-Net model when point prompts were used, while it performed worst among all models when bounding box prompts were used. This emphasizes the crucial role of prompting strategies in influencing the performance of the fine-tuned SAM model in flood segmentation. The results of the study offer practical benefits for emergency response teams as they allow for a faster and more accurate assessment of the areas affected by the floods. In addition, the models are also of great value to insurance companies in the damage assessment phase. Despite these promising results, there is still room for further research. A limitation of the present study is that the Flood Area dataset we used (290 aerial images and associated masks) did not include GPS/georeferenced metadata, so it was not possible to produce georeferenced results. Addressing this dataset limitation in future works would enable more accurate and actionable relief. Future work could also include the development of a user-friendly interface that allows emergency responders and insurance professionals to seamlessly utilize these models and effectively interpret the results. In addition, extending the models to predict high-risk areas before flooding occurs - using inputs such as topographical data, rainfall trends and river flow information - could further enhance their utility.

Author contribution

HS: Conceptualization, Data curation, investigation, Methodology, Formal analysis, Software, Visualization, Writing – original draft. **KDS:** Writing – review and editing. **PF:** Funding acquisition, project administration, Data curation, Conceptualization, Supervision, Writing – review and editing. **TS:** Funding acquisition, Supervision, Conceptualization, project administration, Data curation, Writing – review and editing.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Financial support

This work was supported by the Deutsche Forschungsgesellschaft (DFG) [No. SCHO 739/25-1].

Code and data availability

The codes and data supporting the findings of this study are openly available in the GitHub repository at: https://github.com/hadi1994shokati/Flood-segmentation (last access: 4 February 2026; https://doi.org/10.5281/zenodo.18485640, Shokati, 2026).

Part III.

Conclusion and Future Directions

6. Conclusions, implications, and future perspectives

6.1 Summary and Conclusions

This dissertation addressed key challenges in soil erosion assessment and natural hazard monitoring through four interconnected studies, spanning the full chain from climatic forcing estimation through process-based erosion modeling to automated image segmentation for erosion detection and flood mapping. The research was unified by a common geographic focus on Germany, a shared commitment to rigorous uncertainty quantification, and the systematic application of modern computational techniques—particularly deep learning—to overcome longstanding data limitations in the earth and environmental sciences.

Rainfall Erosivity Estimation (Manuscript I)

The first study developed a comprehensive deep learning framework for reconstructing, observing, and projecting rainfall erosivity across Germany at 1 km spatial resolution. Using 5-minute RADKLIM radar precipitation data for 2001–2025, a Hypernetwork-Enhanced CNN architecture for historical reconstruction back to 1931, and a ConvLSTM encoder-decoder for near-term forecasting to 2043 under the RCP4.5 climate scenario, the study produced a continuous, intensity-aware, century-scale gridded R-factor product for Germany.

Regarding RQ1, Pettitt's test identified a significant change point at 2004/2005, dividing the 1931–2043 series into two distinct regimes: 1943–2004 and 2005–2043. Mean annual R-factor values were 117 N h^{-1} yr^{-1} during 1943–2004, rising to 147 N h^{-1} yr^{-1} during 2005–2043, an increase of 26%. This shift reflects not only higher mean erosivity but also increased frequency of extreme years exceeding 150 N h^{-1} yr^{-1}, from a 24.7-year return period pre-change point to 2.4 years post-change point.

The reliability of the historical reconstruction is supported by both internal model evaluation and external validation. Hypernetwork-Enhanced CNN demonstrated strong skill in reproducing spatial erosivity patterns across diverse hydro-climatic regimes, with correlation coefficients ranging from 0.76 to 0.85 and R^2 values between 0.57 and 0.73. The model accurately captures Germany's primary erosivity gradients, including elevated values in the pre-alpine region and central uplands and lower values in the north German Plain. However, model performance exhibits a systematic dependence on erosivity regime. While accuracy is highest in low-to-moderate erosivity years, extreme years dominated by localized high-intensity events show increased uncertainty, reflecting the inherent stochasticity of convective storms and limitations in resolving sub-grid variability. This is further evidenced by a smoothing effect in the upper tail of the distribution, where extreme erosivity values tend to be underestimated, although their spatial occurrence is still correctly identified.

A key diagnostic of the observed changes is the erosivity-to-precipitation (R/P) ratio, which indicates a shift toward more intense rainfall regimes over time. The increasing trend in this ratio, particularly since the 1970s, suggests that rising erosivity is not solely driven by changes in total precipitation but also by intensification of rainfall events. This behavior is consistent with thermodynamic expectations under atmospheric warming, as described by the Clausius–Clapeyron relationship.

The ConvLSTM forecasting framework indicates that the elevated erosivity regime is likely to persist in the near future under RCP4.5. Importantly, the regime-dependent performance of the reconstruction model suggests that uncertainty may increase in future projections if erosivity becomes increasingly dominated by localized extreme events.

The significance of this study lies not only in the creation of a new dataset but also in its methodological contribution. It demonstrates that deep learning can bridge the gap between limited high-frequency observations and long-term historical reconstruction while explicitly accounting for regime-dependent uncertainty and physically interpretable changes in rainfall intensity structure. This provides a robust foundation for assessing erosion risk under evolving climatic conditions in Germany.

Erosion and Sediment Transport Modeling (Manuscript II)

The second study evaluated the capacity of the spatially distributed erosion and sediment transport model WaTEM/SEDEM to simulate sediment yields in six micro-scale watersheds (0.8–7.8 ha) in Southern Germany, monitored over eight years under optimized soil conservation management. Using a GLUE framework with limits-of-acceptability derived from measurement uncertainty, the study assessed model performance at annual and temporally aggregated scales and distinguished between field-dominated and structure-dominated watershed types.

Regarding RQ2, WaTEM/SEDEM was able to simulate the order of magnitude of the very low sediment yields measured under intensive soil conservation, provided that high-resolution input data and locally adapted USLE factors from the German ABAG were available. However, the model was unable to produce behavioral realizations at the annual time step based on strict limits-of-acceptability criteria, because soil conservation practices reduced the number of erosion events while individual events—such as those occurring after potato harvest—retained magnitudes comparable to conventional cultivation. This temporal intermittency could not be adequately represented by the model's steady-state structure. Model performance improved substantially when outputs were aggregated across the full eight-year monitoring period and when watersheds were grouped by type, confirming that WaTEM/SEDEM is fit for long-term, landscape-scale conservation planning rather than for precise annual predictions at the micro-watershed level.

Regarding RQ3, the evaluation of structure-dominated watersheds containing grassed waterways revealed fundamental challenges in representing the sediment trapping function of linear landscape structures. The model showed a systematic tendency to overestimate sediment yields in these watersheds, and one watershed produced no behavioral realizations at all. This structural limitation was traced primarily to the model's inability to represent re-infiltration processes that are critical for sediment trapping in grassed waterways. The GLUE analysis further revealed pronounced parameter equifinality and compensation mechanisms. In field-dominated watersheds, the erosion surface scaling parameter and the transport capacity coefficient for arable land exhibited strong trade-offs, reflecting the supply-limited nature of sediment transport under conservation conditions. In structure-dominated watersheds, behavioral values of the grassland transport capacity parameter were confined to a narrow range, most likely because the model compensates for its inability to represent re-infiltration by assigning artificially low transport capacities to grassed surfaces.

The broader conclusion is that model performance must be interpreted in relation to decision context and scale. A model may be unsuitable for one purpose—such as annual prediction in very small, conservation-optimized watersheds—yet still highly useful for another, such as long-term planning across multiple fields or landscape units. This fit-for-purpose perspective is one of the key conceptual contributions of the dissertation.

Deep Learning-Based Environmental Image Segmentation (Manuscripts III and IV)

The third and fourth studies applied transfer learning and foundation model adaptation to two environmental segmentation tasks: soil erosion feature detection and flood area mapping from aerial imagery.

Regarding RQ4, the Erosion-SAM model, a fine-tuned version of the Segment Anything Model, demonstrated that SAM can be effectively adapted for automatic pixel-level segmentation of water erosion and deposition features in high-resolution aerial orthophotos. Fine-tuning significantly improved performance over the original SAM across all three land cover types—bare cropland, vegetated cropland, and grassland. The prompt-enhanced resizing approach, which incorporated user-defined point prompts during testing, achieved the best overall results, with mean recall of 0.81, precision of 0.76, Dice coefficient of 0.78, and IoU of 0.64 across all land covers. For grassland specifically, the model achieved recall of 0.90 and a Dice coefficient of 0.86. At the field level, the prompt-enhanced approach yielded an erosion severity ratio agreement of $R^2 = 0.93$ with a mean absolute error of only 1.5%, substantially outperforming alternative preprocessing strategies.

Regarding RQ5, the flood segmentation study compared fine-tuned SAM with point prompts and bounding box prompts, and U-Net with ResNet-50 and ResNet-101 backbones, using 290 UAV and helicopter images from flood events in Germany, India, Malaysia, and Bangladesh. SAM with point prompts achieved the highest performance (Accuracy 0.96, IoU 0.90), while bounding box prompts led to a dramatic decline (Accuracy 0.82, IoU 0.67). This finding is explained by the nature of flood imagery: floodwater often extends from edge to edge, causing bounding boxes to encompass nearly the entire image and thereby providing poor boundary guidance to the decoder. U-Net with ResNet-101 marginally outperformed ResNet-50, suggesting that the additional depth of ResNet-101 provides only limited benefit for flood patterns that are not sufficiently complex to require deeper feature extraction.

Together, these studies show that foundation models like SAM can outperform more conventional segmentation architectures when they are appropriately adapted to the target domain and prompting strategy. At the same time, they also show that U-Net remains a strong and efficient baseline for fully automated deployment. The dissertation therefore does not argue for foundation models as universal replacements, but rather for a nuanced use of architecture choice and prompting strategy depending on the operational setting.

Integrated Conclusion Across Studies

Taken together, the four studies span the full chain from climatic forcing through process-based modeling to automated monitoring. The dissertation demonstrates that the integration of advanced computational techniques—from spatiotemporal deep learning and foundation model adaptation to uncertainty-aware process modeling—can substantially improve the assessment of soil erosion and related hazards in agricultural landscapes. It also shows that methodological progress in one domain can directly strengthen another: updated rainfall erosivity maps improve climatic forcing for erosion models, while automated segmentation of erosion features can help generate validation data for those same models.

A further contribution of the dissertation is conceptual rather than purely technical. It argues that soil erosion research should not be organized as a sequence of isolated modules—climate input, erosion model, field survey, image analysis—but as a connected analytical system in which each component informs and constrains the others. This systems perspective is essential for producing robust knowledge under conditions of climatic change, conservation-oriented management, and rapidly evolving data technologies.

6.2 Broader Implications

The findings of this dissertation carry implications that extend beyond individual studies.

First, the century-scale rainfall erosivity product for Germany provides an updated, physically informed approach, offering deeper insights into the non-stationarity of precipitation regimes and their direct impact on soil loss trends under a changing climate. The documented increase in mean erosivity and the dramatic shortening of the return period for extreme erosivity years have direct consequences for soil conservation planning, infrastructure design, and agricultural risk management. Land managers, policymakers, and insurance providers can no longer rely on erosivity estimates calibrated under mid-20th century conditions. The new dataset offers a spatially explicit and temporally continuous foundation for reassessing erosion risk under contemporary and near-future climate conditions.

Second, the WaTEM/SEDEM evaluation provides clarity on the fitness for purpose of a widely used erosion model under soil conservation conditions. The finding that the model is suitable for long-term, landscape-scale conservation planning but not for annual micro-watershed predictions has important practical implications. Erosion models should be matched to the spatiotemporal scale at which management decisions are made, and model evaluation must account for the scale-dependent nature of both model performance and measurement uncertainty. The identification of structural limitations—particularly the inability to represent re-infiltration in grassed waterways—points toward specific model development priorities rather than a generic dismissal of the model.

Third, the deep learning segmentation studies demonstrate that foundation models such as SAM, when appropriately fine-tuned, can bridge a critical bottleneck in environmental monitoring. Erosion-SAM creates the opportunity to generate validation data for physically based erosion models and large-scale training datasets for data-driven erosion assessment. Similarly, the flood segmentation models enable rapid, automated damage assessment that can support emergency response teams and insurance applications. The finding that prompting strategy dramatically affects SAM's performance provides practical guidance for environmental segmentation tasks: point prompts should be preferred over bounding boxes when the target class occupies a large and spatially variable share of the image.

Fourth, the dissertation as a whole demonstrates the value of integrating multiple computational approaches within a unified framework for addressing interconnected environmental challenges. The rainfall erosivity maps from Chapter 3 can serve as direct climatic input to the erosion models evaluated in Chapter 4, while the automated erosion detection capabilities developed in Chapter 5 can provide the spatially distributed validation data that Chapter 4 identified as a critical gap. This integration points toward a future in which erosion assessment becomes more dynamic, data-rich, and operationally relevant.

6.3 Limitations

Several limitations should be acknowledged.

The rainfall erosivity reconstruction and forecasts inherit uncertainties from the training data period, the choice of predictors, and the extrapolation beyond historically observed climatic conditions. The ConvLSTM projections are conditional on the RCP4.5 emission scenario and on bias-adjusted CORDEX-EUR11 climate model outputs; higher-end emission scenarios would likely yield substantially larger erosivity increases. The iterative multi-stage forecasting approach, in which the ConvLSTM uses its own predictions as input for subsequent forecast windows, introduces error propagation that may accumulate over longer forecast horizons.

The WaTEM/SEDEM evaluation was conducted in six micro-scale watersheds at a single experimental farm site, which limits the generalizability of the findings to other landscapes, soil types, and conservation configurations. The model's static parameterization of connectivity and transport capacity cannot capture the temporal dynamics of grassed waterway effectiveness, which varies with seasonal vegetation development and event-specific runoff intensity. More broadly, the absence of spatially distributed observational erosion data constrains the ability to test simulated deposition and routing patterns directly.

The deep learning segmentation studies relied on datasets of moderate size, and the erosion training data were annotated by a single expert, introducing potential subjectivity into the ground truth. The aerial imagery used for erosion detection is subject to license restrictions, limiting reproducibility. In addition, the models were evaluated on high-resolution aerial imagery, and their transferability to satellite data of lower spatial resolution has not yet been tested. Finally, although the segmentation results are strong, the operational deployment of such models in routine workflows would require further benchmarking across diverse acquisition conditions, sensors, and landscapes.

6.4 Future Perspectives

The work presented in this dissertation opens several promising avenues for future research.

A first priority is the further development of explainable and ensemble-based erosivity modeling. Future iterations of the rainfall erosivity framework should incorporate explainable artificial intelligence techniques to clarify the relationships learned between climatic predictors and erosivity patterns. Understanding which variables and spatial features drive the models' predictions would enhance both scientific interpretability and user trust. In parallel, ensemble approaches combining multiple regional climate model runs and emission scenarios would allow probabilistic erosivity projections that better capture the full range of future uncertainty.

A second priority is the development of more dynamic parameterization strategies for erosion and sediment transport models. The structural limitations of WaTEM/SEDEM identified in this dissertation, particularly the inability to represent re-infiltration in grassed waterways and the static nature of connectivity parameters, motivate the integration of time-varying transport capacity formulations and improved infiltration representation. Future work could explore couplings between WaTEM/SEDEM-type routing schemes and process-based hydrological components that respond to seasonal vegetation development and event-specific runoff intensity. More broadly, updated, intensity-aware erosivity products should be integrated systematically into erosion modeling so that climatic forcing reflects contemporary and projected conditions rather than outdated empirical standards.

A third priority is the scaling up of automated erosion monitoring. Erosion-SAM represents an important first step toward automated, large-scale erosion detection, but future studies should expand the training dataset through multiple expert annotations to reduce subjectivity and improve robustness. Extending Erosion-SAM to very high-resolution satellite imagery from platforms such as WorldView-3, GeoEye-1, and KOMPSAT-3 would enable broader spatial coverage and higher temporal frequency, potentially opening the path toward near-real-time erosion surveillance over large agricultural regions.

A fourth priority is the advancement of flood segmentation for operational deployment. The flood mapping models developed in this dissertation could be improved by incorporating multi-temporal imagery to track flood progression, integrating digital elevation models for improved boundary delineation, and developing ensemble approaches that combine SAM and U-Net predictions to exploit the complementary strengths of both architectures. Standardized benchmark datasets for aerial flood imagery across diverse geographic and climatic contexts would facilitate systematic comparison and accelerate progress toward operational readiness.

Perhaps the most compelling future direction lies in the integration of the individual components developed in this dissertation into a coherent, end-to-end erosion assessment pipeline. In such a system, updated rainfall erosivity maps would provide the climatic input to spatially distributed erosion models, while Erosion-SAM would generate the spatially distributed erosion pattern data needed for model calibration and validation at scales that are currently difficult to access through field surveys alone. This coupling of climatic forcing, process-based modeling, and automated image-based monitoring could transform erosion assessment from a largely static and retrospective exercise into a dynamic, forward-looking capability for climate adaptation and soil conservation planning.

Appendix

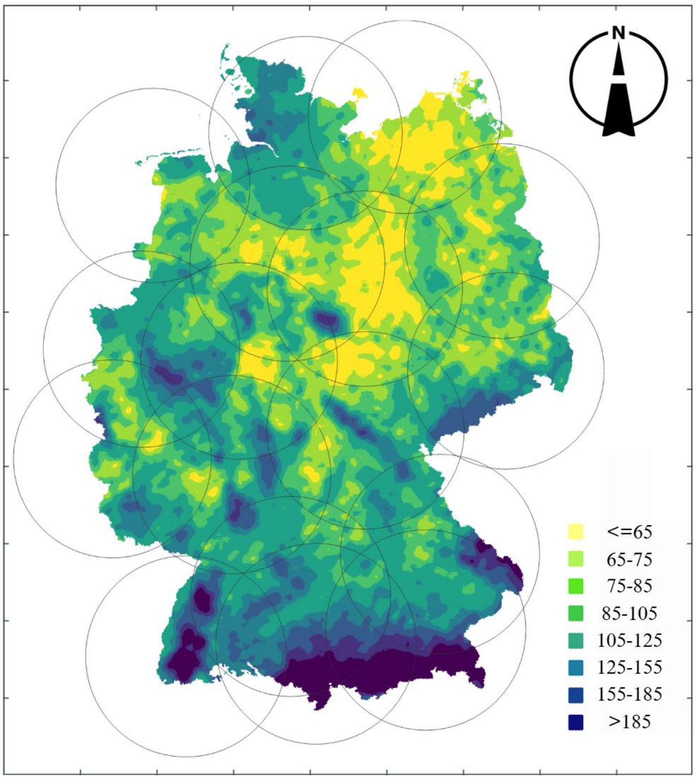


Figure 3.S1. Mean annual rainfall erosivity (R-factor, N h^{-1} yr^{-1}) for Germany, derived from 17 years of radar-based precipitation data (Auerswald et al., 2019). Black circles indicate the 128 km effective coverage radius of each of the 17 German weather radar sites. Axis tick marks correspond to 100 km distance intervals.

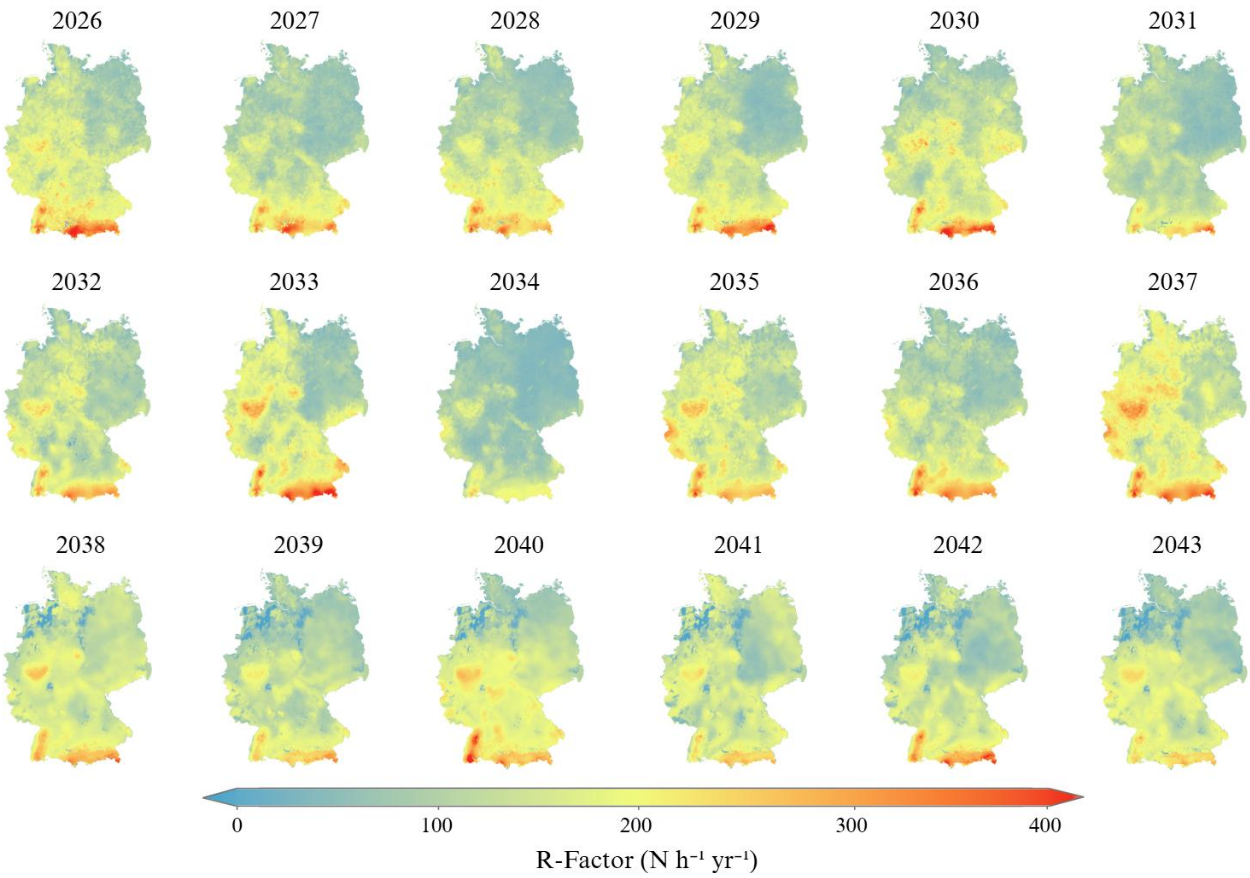


Figure 3.S2. ConvLSTM-projected annual rainfall erosivity (R-factor, N h^{-1} yr^{-1}) for Germany for each year from 2026 to 2043 under the RCP4.5 scenario.

Bibliography

Abaci, O., Dermisis, D., & Papanicolaou, A. N. (2010). Event based efficiency of grassed waterways in reducing sediment yield within an agricultural watershed. In *Watershed Management 2010: Innovations in Watershed Management under Land Use and Climate Change* (pp. 1202–1213).

Aghabeygi, M., Strauss, V., Paul, C., & Helming, K. (2024). Barriers of adopting sustainable soil management practices for organic and conventional farming systems. *Discover Soil*, *1*(1), 11. https://doi.org/10.1007/s44378-024-00008-1

Agnihotri, D., Kumar, T., & Jhariya, D. (2021). Intelligent vulnerability prediction of soil erosion hazard in semi-arid and humid region. *Environment, Development and Sustainability*, *23*(2), 2524–2551. https://doi.org/10.1007/s10668-020-00685-2

Ait El Asri, S., Negabi, I., El Adib, S., & Raissouni, N. (2023). Enhancing building extraction from remote sensing images through UNet and transfer learning. *International Journal of Computers and Applications*, *45*(5), 413–419. https://doi.org/10.1080/1206212X.2023.2219117

Alatorre, L. C., Beguería, S., & García-Ruiz, J. M. (2010). Regional scale modeling of hillslope sediment delivery: A case study in the Barasona Reservoir watershed (Spain) using WATEM/SEDEM. *Journal of Hydrology*, *391*(1), 109–123. https://doi.org/10.1016/j.jhydrol.2010.07.010

Andersson, J. A., & D'Souza, S. (2014). From adoption claims to understanding farmers and contexts: A literature review of Conservation Agriculture (CA) adoption among smallholder farmers in southern Africa. *AGR ECOSYST ENVIRON*, *187*, 116-132,. https://doi.org/10.1016/j.agee.2013.08.008

Angulo-Martínez, M., & Beguería, S. (2012). Trends in rainfall erosivity in NE Spain at annual, seasonal and daily scales, 1955–2006. *Hydrology and Earth System Sciences*, *16*(10), 3551–3559. https://doi.org/10.5194/hess-16-3551-2012

Angulo-Martínez, M., López-Vicente, M., Vicente-Serrano, S. M., & Beguería, S. (2009). Mapping rainfall erosivity at a regional scale: A comparison of interpolation methods in the Ebro Basin (NE Spain). *Hydrology and Earth System Sciences*, *13*(10), 1907–1920. https://doi.org/10.5194/hess-13-1907-2009

Anthropogenic Influence on Extreme Precipitation over Global Land Areas Seen in Multiple Observational Datasets, 12 Nature Communications 3944 (Nature Publishing Group UK London 2021). https://www.nature.com/articles/s41467-021-24262-x

Arnold, J. G., & Fohrer, N. (2005). SWAT2000: Current capabilities and research opportunities in applied watershed modelling. *Hydrological Processes*, *19*(3), 563–572. https://doi.org/10.1002/hyp.5611

Arslan, N., & Sekertekin, A. (2019). Application of Long Short-Term Memory neural network model for the reconstruction of MODIS Land Surface Temperature images. *Journal of Atmospheric and Solar-Terrestrial Physics*, *194*, 105100. https://doi.org/10.1016/j.jastp.2019.105100

Asadieh, B., & Krakauer, N. Y. (2015). Global trends in extreme precipitation: Climate models versus observations. *Hydrology and Earth System Sciences*, *19*(2), 877–891. https://doi.org/10.5194/hess-19-877-2015

Auerswald, K., & Fiener, P. (2019). Soil organic carbon storage following conversion from cropland to grassland on sites differing in soil drainage and erosion history. *SCI TOTAL ENVIRON*, *661*, 481-491,. https://doi.org/10.1016/j.scitotenv.2019.01.200

Auerswald, K., & Fiener, P. (2024). Assessing the impact of climate change on soil erosion by water. *Burleigh Dodds Series in Agricultural Science*, 51–76.

Auerswald, K., Fiener, P., & Dikau, R. (2009). Rates of sheet and rill erosion in Germany—A meta-analysis. *Geomorphology*, *111*, 182–193. https://doi.org/10.1016/j.geomorph.2009.04.018

Auerswald, K., Fiener, P., Gerl, G., & Wilken, F. (2019b). *Land use and land management data from the Scheyern experimental farm covering 14 small adjacent watersheds and their surroundings* [Dataset]. https://doi.org/10.13140/RG.2.2.26172.49285

Auerswald, K., Fiener, P., Martin, W., & Elhaus, D. (2014). Use and misuse of the K factor equation in soil erosion modeling: An alternative equation for determining USLE nomograph soil erodibility values. *CATENA*, *118*, 220-

225,. https://doi.org/10.1016/j.catena.2014.01.008

Auerswald, K., Fischer, F. K., Winterrath, T., & Brandhuber, R. (2019). Rain erosivity map for Germany derived from contiguous radar rain data. *Hydrology and Earth System Sciences*, *23*(4), 1819–1832. https://doi.org/10.5194/hess-23-1819-2019

Auerswald, K., Kainz, M., Scheinost, A., & Sinowski, W. (2001). *The Scheyern Experimental Farm: Research methods, the farming system and definition of the framework of site properties and characteristics* (Vol. In, 183–194, p. 10 1007 978-3-662-04504-6 10,).

Auerswald, K., Wilken, F., & Fiener, P. (2019a). *Soil properties at the Scheyern experimental farm covering 14 small adjacent watersheds and their surroundings* [Dataset]. https://doi.org/10.13140/RG.2.2.14231.83365

Baartman, J. E., Nunes, J. P., Masselink, R., Darboux, F., Bielders, C., Degré, A., & Wainwright, J. (2020). What do models tell us about water and sediment connectivity? *Geomorphology*, *367*, 107300.

Ballabio, C., Borrelli, P., Spinoni, J., Meusburger, K., Michaelides, S., Beguería, S., & Panagos, P. (2017). Mapping monthly rainfall erosivity in Europe. *Science of the Total Environment*, *579*, 1298–1315.

Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., & Gerlach, N. (2004). *Projekt RADOLAN - Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer): Zusammenfassender Abschlussbericht für die Projektlaufzeit von 1997 bis 2004*. https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/68177

Batista, P. V. G., Fiener, P., Scheper, S., & Alewell, C. (2022). A conceptual-model-based sediment connectivity assessment for patchy agricultural catchments. *HYDROL EARTH SYST SC*, *26*, 3753–3770, 10 5194-26-3753-2022,.

Beguería, S., Serrano-Notivoli, R., & Tomas-Burguera, M. (2018). Computation of rainfall erosivity from daily precipitation amounts. *Science of The Total Environment*, *637–638*, 359–373. https://doi.org/10.1016/j.scitotenv.2018.04.400

Behrens, T., & Scholten, T. (2006). A comparison of data-mining techniques in predictive soil mapping. *Developments in Soil Science*, *31*, 353–617.

Benavidez, R., Jackson, B., Maxwell, D., & Norton, K. (2018). A review of the (Revised) Universal Soil Loss Equation ((R) USLE): With a view to increasing its global applicability and improving soil loss estimates. *Hydrology and Earth System Sciences*, *22*(11), 6059–6086.

Beven, K. (2006). A manifesto for the equifinality thesis. *J HYDROL*, *320*, 18-36,. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *HYDROL PROCESS*, *6*, 279-298,. https://doi.org/10.1002/hyp.3360060305

Beven, K., & Lane, S. (2022). On (in)validating environmental models. 1. Principles for formulating a Turing-like Test for determining when a model is fit-for purpose. *HYDROL PROCESS*, *36, e14704*, 10 1002 14704,.

Bezak, N., Ballabio, C., Mikoš, M., Petan, S., Borrelli, P., & Panagos, P. (2020). Reconstruction of past rainfall erosivity and trend detection based on the REDES database and reanalysis rainfall. *Journal of Hydrology*, *590*, 125372.

Borrelli, P., Alewell, C., Alvarez, P., Anache, J. A. A., Baartman, J., Ballabio, C., Bezak, N., Biddoccu, M., Cerdà, A., Chalise, D., Chen, S., Chen, W., De Girolamo, A. M., Gessesse, G. D., Deumlich, D., Diodato, N., Efthimiou, N., Erpul, G., Fiener, P., … Panagos, P. (2021). Soil erosion modelling: A global review and statistical analysis. *Science of The Total Environment*, *780*, 146494. https://doi.org/10.1016/j.scitotenv.2021.146494

Borrelli, P., Diodato, N., & Panagos, P. (2016). Rainfall erosivity in Italy: A national scale spatio-temporal assessment. *International Journal of Digital Earth*, *9*, 835–850. https://doi.org/10.1080/17538947.2016.1148203

Brazier, R. E., Beven, K. J., Freer, J., & Rowan, J. S. (2000). Equifinality and uncertainty in physically based soil erosion models: Application of the GLUE methodology to WEPP–the Water Erosion Prediction Project–for sites in the UK and USA. *Earth Surface Processes and Landforms*, *25*(8), 825–845. https://doi.org/10.1002/1096-9837(200008)25:8%3C825::AID-ESP101%3E3.0.CO;2-3

Brus, D. J., & Akker, J. J. H. (2018). How serious a problem is subsoil compaction in the Netherlands? A survey based on probability sampling. *SOIL*, *4*, 37–45, 10 5194-4-37-2018,.

Bibliography

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of The Total Environment*, *721*, 137612. https://doi.org/10.1016/j.scitotenv.2020.137612

Cândido, B. M., Quinton, J. N., James, M. R., Silva, M. L. N., de Carvalho, T. S., de Lima, W., Beniaich, A., & Eltner, A. (2020). High-resolution monitoring of diffuse (sheet or interrill) erosion using structure-from-motion. *Geoderma*, *375*, 114477. https://doi.org/10.1016/j.geoderma.2020.114477

Carter, C. E., & Parsons, D. (1967). A.: Field tests on the coshocton-type wheel runoff sampler. *T ASAE*, *10*, 133-135,. https://doi.org/10.13031/2013.39613

Character, L., Ortiz JR, A., Beach, T., & Luzzadder-Beach, S. (2021). Archaeologic Machine Learning for Shipwreck Detection Using Lidar and Sonar. *Remote Sensing*, *13*(9), Article 9. https://doi.org/10.3390/rs13091759

Chen, S., & Zha, X. (2018). Effects of the ENSO on rainfall erosivity in the Fujian Province of southeast China. *Science of The Total Environment*, *621*, 1378–1388. https://doi.org/10.1016/j.scitotenv.2017.10.090

Chen, W., Lei, X., Chakrabortty, R., Pal, S. C., Sahana, M., & Janizadeh, S. (2021). Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility. *Journal of Environmental Management*, *284*, 112015.

Cheng, D., Qin, Z., Jiang, Z., Zhang, S., Lao, Q., & Li, K. (2023). *SAM on Medical Images: A Comprehensive Study on Three Prompt Modes* (arXiv:2305.00035). arXiv. https://doi.org/10.48550/arXiv.2305.00035

Dai, Q., Zhu, J., Lv, G., Kalin, L., Yao, Y., Zhang, J., & Han, D. (2023). Radar remote sensing reveals potential underestimation of rainfall erosivity at the global scale. *Science Advances*, *9*(32), eadg5551. https://doi.org/10.1126/sciadv.adg5551

De Leeuw, J., Vrieling, A., Shee, A., Atzberger, C., Hadgu, K. M., Biradar, C. M., Keah, H., & Turvey, C. (2014). The Potential and Uptake of Remote Sensing in Insurance: A Review. *Remote Sensing*, *6*(11), Article 11. https://doi.org/10.3390/rs61110888

de Vente, J., & Poesen, J. (2005). Predicting soil erosion and sediment yield at the basin scale: Scale issues and semi-quantitative models. *Earth-Science Reviews*, *71*(1), 95–125. https://doi.org/10.1016/j.earscirev.2005.02.002

Deelstra, J., Øygarden, L., Blankenberg, A. B., & Olav Eggestad, H. (2011). Climate change and runoff from agricultural catchments in Norway. *International Journal of Climate Change Strategies and Management*, *3*(4), 345–360. https://doi.org/10.1108/17568691111175641

Dermisis, D., Abaci, O., (Thanos) Papanicolaou, A. N., & Wilson, C. G. (2010). Evaluating grassed waterway efficiency in southeastern Iowa using WEPP. *Soil Use and Management*, *26*(2), 183–192. https://doi.org/10.1111/j.1475-2743.2010.00257.x

Desmet, P. J. J., & Govers, G. (1996). A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. *J SOIL WATER CONSERV*, *51*, 427-433,. https://doi.org/10.1016/j.cageo.2012.09.027

Diodato, N., Borrelli, P., Fiener, P., Bellocchi, G., & Romano, N. (2017). Discovering historical rainfall erosivity with a parsimonious approach: A case study in Western Germany. *Journal of Hydrology*, *544*, 1–9. https://doi.org/10.1016/j.jhydrol.2016.11.023

Ebrahimi-Khusfi, Z., Taghizadeh-Mehrjardi, R., Roustaei, F., Ebrahimi-Khusfi, M., Mosavi, A. H., Heung, B., Soleimani-Sardo, M., & Scholten, T. (2021). Determining the contribution of environmental factors in controlling dust pollution during cold and warm months of western Iran using different data mining algorithms and game theory. *Ecological Indicators*, *132*, 108287. https://doi.org/10.1016/j.ecolind.2021.108287

Ehmele, F., Kautz, L.-A., Feldmann, H., & Pinto, J. G. (2020). Long-term variance of heavy precipitation across central Europe using a large ensemble of regional climate model simulations. *Earth System Dynamics*, *11*(2), 469–490. https://doi.org/10.5194/esd-11-469-2020

Elhage, N., Hume, T., Olsson, C., Schiefer, N., Henighan, T., Kravec, S., Hatfield-Dodds, Z., Lasenby, R., Drain, D., Chen, C., Grosse, R., McCandlish, S., Kaplan, J., Amodei, D., Wattenberg, M., & Olah, C. (2022). *Toy Models of Superposition* (arXiv:2209.10652). arXiv. https://doi.org/10.48550/arXiv.2209.10652

Eltner, A., Baumgart, P., Maas, H.-G., & Faust, D. (2015). Multi-temporal UAV data for automatic measurement of rill and interrill erosion on loess soil. *Earth Surface Processes and Landforms*, *40*(6), 741–755. https://doi.org/10.1002/esp.3673

Bibliography

Ezzatvar, Y., & López-Gil, J. F. (2024). Urgent call for enhanced flood preparedness and response in Spain. *The Lancet*, *404*(10470), 2419–2420. https://doi.org/10.1016/S0140-6736(24)02506-6

Feng, X., Wang, Y., Chen, L., Fu, B., & Bai, G. (2010). Modeling soil erosion and its response to land-use change in hilly catchments of the Chinese Loess Plateau. *Geomorphology*, *118*(3), 239–248. https://doi.org/10.1016/j.geomorph.2010.01.004

Fiener, P., & Auerswald, K. (2003). Effectiveness of grassed waterways in reducing runoff and sediment delivery from agricultural watersheds. *J ENVIRON QUAL*, *32*, 927-936,. https://doi.org/10.2134/jeq2003.9270

Fiener, P., & Auerswald, K. (2018). *Grassed waterways, in, American Society of Agronomy*. Crop Science Society of America, Soil Science Society of America.

Fiener, P., Auerswald, K., & Weigand, S. (2005). Managing erosion and water quality in agricultural watersheds by small detention ponds. *AGR ECOSYST ENVIRON*, *110*, 132-142,. https://doi.org/10.1016/j.agee.2005.03.012

Fiener, P., Dostal, T., Krasa, J., Schmaltz, E., Strauss, P., & Wilken, F. (2020). Operational USLE-based modeling of soil erosion in Czech Republic, Austria, and Bavaria—Differences in model adaptation, parametrization, and data availability. *Appl. Science*, *10*, 1–18.

Fiener, P., Neuhaus, P., & Botschek, J. (2013). Long-term trends in rainfall erosivity–analysis of high resolution precipitation time series (1937–2007) from Western Germany. *Agricultural and Forest Meteorology*, *171–172*, 115–123. https://doi.org/10.1016/j.agrformet.2012.11.011

Fiener, P., Wilken, F., & Auerswald, K. (2019a). Filling the gap between plot and landscape scale – eight years of soil erosion monitoring in 14 adjacent watersheds under soil conservation at Scheyern, Southern Germany. *Advances in Geosciences*, *48*, 31–48, 10 5194-48-31-2019,.

Fiener, P., Wilken, F., & Auerswald, K. (2019b). *Runoff and sediment delivery data at the Scheyern experimental farm covering 14 small adjacent watersheds* [Dataset]. https://doi.org/10.13140/RG.2.2.30786.22729

Fischer, F. K., Kistler, M., Brandhuber, R., Maier, H., Treisch, M., & Auerswald, K. (2018). Validation of official erosion modelling based on high-resolution radar rain data by aerial photo erosion classification. *Earth Surface Processes and Landforms*, *43*(1), 187–194. https://doi.org/10.1002/esp.4216

Fischer, F. K., Winterrath, T., & Auerswald, K. (2018). Temporal- and spatial-scale and positional effects on rain erosivity derived from point-scale and contiguous rain data. *Hydrology and Earth System Sciences*, *22*(12), 6505–6518. https://doi.org/10.5194/hess-22-6505-2018

Foucher, A., Salvador-Blanes, S., Evrard, O., Simonneau, A., Chapron, E., Courp, T., Cerdan, O., Lefèvre, I., Adriaensen, H., Lecompte, F., & Desmet, M. (2014). Increase in soil erosion after agricultural intensification: Evidence from a lowland basin in France. *Anthropocene*, *7*, 10 1016 2015 02 001,.

Fully Differentiable, Fully Distributed Rainfall-Runoff Modeling, 29 Hydrology and Earth System Sciences 6257 (Copernicus Publications Göttingen, Germany 2025). https://hess.copernicus.org/articles/29/6257/2025/

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*, *7*(6), 953–961.

Garbrecht, J. D., & Starks, P. J. (2009). Watershed sediment yield reduction through soil conservation in a West-Central Oklahoma watershed. *Ecohydrology: Ecosystems, Land and Water Process Interactions, Ecohydrogeomorphology*, *2*(3), 313–320.

Gaus, Y. F. A., Bhowmik, N., Isaac-Medina, B. K. S., & Breckon, T. P. (2024). *Performance Evaluation of Segment Anything Model with Variational Prompting for Application to Non-Visible Spectrum Imagery*. 3142–3152. https://openaccess.thecvf.com/content/CVPR2024W/PBVS/html/Gaus_Performance_Evaluation_of_Segment_Anything_Model_with_Variational_Prompting_for_CVPRW_2024_paper.html

Gautam, P., Kubota, T., Sapkota, L. M., & Shinohara, Y. (2021). Landslide susceptibility mapping with GIS in high mountain area of Nepal: A comparison of four methods. *Environmental Earth Sciences*, *80*(9), 359. https://doi.org/10.1007/s12665-021-09650-2

Gayen, A., Pourghasemi, H. R., Saha, S., Keesstra, S., & Bai, S. (2019). Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Science of the Total Environment*, *668*, 124–138.

Gebrehiwot, A., Hashemi-Beni, L., Thompson, G., Kordjamshidi, P., & Langan, T. E. (2019). Deep convolutional neural network for flood extent mapping using unmanned aerial vehicles data. *Sensors*, *19*(7), 1486.

Ghaznavi, A., Saberioon, M., Brom, J., & Itzerott, S. (2024). Comparative performance analysis of simple U-Net, residual attention U-Net, and VGG16-U-Net for inventory inland water bodies. *Applied Computing and Geosciences*, *21*, 100150. https://doi.org/10.1016/j.acags.2023.100150

Gocke, M. I., Don, A., Heidkamp, A., Schneider, F., & Amelung, W. (2021). The phosphorus status of German cropland—An inventory of top- and subsoils. *Journal of Plant Nutrition and Soil Science*, *184*(1), 51–64. https://doi.org/10.1002/jpln.202000127

Gonzalez-Hidalgo, J. C., Batalla, R. J., Cerda, A., & Luis. (2012). M.: A regional analysis of the effects of largest events on soil erosion. *CATENA*, *95*, 85-90,. https://doi.org/10.1016/j.catena.2012.03.006

Hadlich, M., Marinov, Z., & Stiefelhagen, R. (2023). *AutoPET Challenge 2023: Sliding Window-based Optimization of U-Net* (arXiv:2309.12114). arXiv. https://doi.org/10.48550/arXiv.2309.12114

Hamming, R. W. (1956). *Digital Filters Prentice Hall*. NJ.

Hanel, M., Pavlásková, A., & Kyselỳ, J. (2016). Trends in characteristics of sub-daily heavy precipitation and rainfall erosivity in the Czech Republic. *International Journal of Climatology*.

Hänsel, S., Hoy, A., Brendel, C., & Maugeri, M. (2022). Record summers in Europe: Variations in drought and heavy precipitation during 1901–2018. *International Journal of Climatology*, *42*(12), 6235–6257. https://doi.org/10.1002/joc.7587

Hashemi-Beni, L., Jones, J., Thompson, G., Johnson, C., & Gebrehiwot, A. (2018). Challenges and Opportunities for UAV-Based Digital Elevation Model Generation for Flood-Risk Management: A Case of Princeville, North Carolina. *Sensors*, *18*(11), Article 11. https://doi.org/10.3390/s18113843

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. http://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Hlavčová, K., Kohnová, S., Velísková, Y., Studvová, Z., Sočuvka, V., & Ivan, P. (2018). Comparison of two concepts for assessment of sediment transport in small agricultural catchments. *J HYDROL HYDROMECH*, *66*, 404–415, 10 2478-2018-0032,.

Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory, Neural computation*. MIT press.

Hong, S., Kim, S., Joh, M., & Song, S. (2017). *PSIque: Next Sequence Prediction of Satellite Images using a Convolutional Sequence-to-Sequence Network* (arXiv:1711.10644). arXiv. https://doi.org/10.48550/arXiv.1711.10644

Hoomehr, S., Schwartz, J. S., & Yoder, D. C. (2016). *Potential changes in rainfall erosivity under GCM climate change scenarios for the southern Appalachian region*. Elsevier.

Hosseinzadehtalaei, P., Tabari, H., & Willems, P. (2020). Climate change impact on short-duration extreme precipitation and intensity–duration–frequency curves over Europe. *J HYDROL*, *590, 125249*, 10 1016 2020 125249,.

Hou, Y., Ren, H., Lv, Q., Wu, L., Yang, X., & Quan, Y. (2022). Radar-Jamming Classification in the Event of Insufficient Samples Using Transfer Learning. *Symmetry*, *14*(11), Article 11. https://doi.org/10.3390/sym14112318

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, *35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

Huber, P. J. (1992). Robust Estimation of a Location Parameter. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in Statistics* (pp. 492–518). Springer New York. https://doi.org/10.1007/978-1-4612-4380-9_35

Ishikawa, Y., Kasai, S., Aoki, Y., & Kataoka, H. (2021). *Alleviating Over-Segmentation Errors by Detecting Action Boundaries*. 2322–2331. https://openaccess.thecvf.com/content/WACV2021/html/Ishikawa_Alleviating_Over-Segmentation_Errors_by_Detecting_Action_Boundaries_WACV_2021_paper.html

Iuss, I., & Fao, I. (1999). *World reference base for soil resources*.

J., A. (2012). DWA Merkblatt M910—Stoffeinträge in Gewässer durch Bodenerosion. *Dt. Verein. Wasserwirt*.

Jackson, J., Yussif, S. B., Patamia, R. A., Sarpong, K., & Qin, Z. (2023). Flood or Non-Flooded: A Comparative Study of State-of-the-Art Models for Flood Image Classification Using the FloodNet Dataset with Uncertainty Offset Analysis. *Water*, *15*(5), Article 5. https://doi.org/10.3390/w15050875

Janeček, M., Květoň, V., Kubátová, E., Kobzová, D., Vošmerová, M., & Chlupsová, J. (2013). Values of rainfall erosivity factor for the Czech Republic. *Journal of Hydrology and Hydromechanics*, *61*(2), 97–102. https://doi.org/10.2478/johh-2013-0013

Jetten, V. G., & Favis-Mortlock, D. (2006). Modelling soil erosion in Europe. In *Soil erosion in Europe* (pp. 695–716). Wiley. https://research.utwente.nl/files/353470353/jetten_mod.pdf

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). *Disaster Monitoring using Unmanned Aerial Vehicles and Deep Learning* (arXiv:1807.11805). arXiv. https://doi.org/10.48550/arXiv.1807.11805

Karim, F., Sharma, K., & Barman, N. R. (2022). *Flood area segmentation*. Kaggle.

Karydas, C. G., Panagos, P., & Gitas, I. Z. (2014). A classification of water erosion models according to their geospatial characteristics. *International Journal of Digital Earth*, *7*(3), 229–250. https://doi.org/10.1080/17538947.2012.671380

Kathuria, A. (2018). Intro to optimization in deep learning: Momentum, RMSProp and Adam. *Paperspace Blog*.

Keller, T., Sandin, M., Colombi, T., Horn, R., & Or, D. (2019). Historical increase in agricultural machinery weights enhanced soil stress levels and adversely affected soil functioning. *Soil and Tillage Research*, *194, 104293*, 10 1016 2019 104293,.

Kelly, M., Schwarz, I., Ziegelaar, M., Watkins, A. B., & Kuleshov, Y. (2023). Flood Risk Assessment and Mapping: A Case Study from Australia's Hawkesbury-Nepean Catchment. *Hydrology*, *10*(2), Article 2. https://doi.org/10.3390/hydrology10020026

Khosravi, K., Rezaie, F., Cooper, J. R., Kalantari, Z., Abolfathi, S., & Hatamiafkoueieh, J. (2023). Soil water erosion susceptibility assessment using deep learning algorithms. *Journal of Hydrology*, *618*, 129229. https://doi.org/10.1016/j.jhydrol.2023.129229

Kim, S., Hong, S., Joh, M., & Song, S. (2017). *DeepRain: ConvLSTM Network for Precipitation Prediction using Multichannel Radar Data* (arXiv:1711.02316). arXiv. https://doi.org/10.48550/arXiv.1711.02316

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (arXiv:1412.6980). arXiv. https://doi.org/10.48550/arXiv.1412.6980

Kinnell, P. I. (2007). A.: Runoff dependent erosivity and slope length factors suitable for modelling annual erosion using the Universal Soil Loss Equation. *HYDROL PROCESS*, *21*, 2681-2689,. https://doi.org/10.1002/hyp.6493

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollar, P., & Girshick, R. (2023). *Segment Anything*. 4015–4026. https://openaccess.thecvf.com/content/ICCV2023/html/Kirillov_Segment_Anything_ICCV_2023_paper.html

Klemas, V. V. (2015). Coastal and Environmental Remote Sensing from Unmanned Aerial Vehicles: An Overview. *Journal of Coastal Research*, *31*(5), 1260–1267. https://doi.org/10.2112/JCOASTRES-D-15-00005.1

Kohrell, G. J., Mulla, D. J., & Gelder, B. (2023). Calibration and validation of hillslope runoff and soil loss outputs from the Water Erosion Prediction Project model in Minnesota agricultural watersheds. *JAWRA Journal of the American Water Resources Association*, *59*(6), 1529–1548. https://doi.org/10.1111/1752-1688.13148

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Kuenzer, C., Guo, H., Huth, J., Leinenkugel, P., Li, X., & Dech, S. (2013). Flood Mapping and Flood Dynamics of the Mekong Delta: ENVISAT-ASAR-WSM Based Time Series Analyses. *Remote Sensing*, *5*(2), Article 2. https://doi.org/10.3390/rs5020687

Kumar, B. V., Goel, M. K., & Singh, G. (2025). Flood Area Segmentation using Using U-Net: A Deep Learning Approach with UAV and Satellite Imagery. *2025 6th International Conference for Emerging Technology (INCET*, 1–5.

Laflen, J. M., Lane, L. J., & Foster, G. R. (1991). WEPP: A new generation of erosion prediction technology. *Journal of Soil and Water Conservation*, *46*(1), 34–38.

Bibliography

Lai, C., Chen, X., Wang, Z., Wu, X., Zhao, S., Wu, X., & Bai, W. (2016). *Spatio-temporal variation in rainfall erosivity during 1960–2012 in the Pearl River Basin*. Elsevier.

Lamprecht, S., Hill, A., Stoffels, J., & Udelhoven, T. (2017). A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data. *Remote Sensing*, *9*(5). https://doi.org/10.3390/rs9050505

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers, Special Issue: Progress of Machine Learning in Geosciences*, *7*(1), 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. *Nature*, *521*(7553), 436–444.

Lee, J., Lee, S., Hong, J., Lee, D., Bae, J. H., Yang, J. E., Kim, J., & Lim, K. J. (2021). Evaluation of Rainfall Erosivity Factor Estimation Using Machine and Deep Learning Models. *Water*, *13*(3). https://doi.org/10.3390/w13030382

Lehmiani, Y., Belattar, S., & Mahmoudi, A. (2023). A Comparative Study of SAM's Performance Against Three U-Net Architectures on Retinal Vessels Segmentation Task. *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, 1–7. https://doi.org/10.1109/SITA60746.2023.10373700

Lehmkuhl, F., Schüttrumpf, H., Schwarzbauer, J., Brüll, C., Dietze, M., Letmathe, P., Völker, C., & Hollert, H. (2022). Assessment of the 2021 summer flood in Central Europe. *Environmental Sciences Europe*, *34*(1), 107. https://doi.org/10.1186/s12302-022-00685-1

Lin, S.-S., Zhang, N., Xu, Y.-S., & Hino, T. (2020). Lesson Learned from Catastrophic Floods in Western Japan in 2018: Sustainable Perspective Analysis. *Water*, *12*(9), Article 9. https://doi.org/10.3390/w12092489

Liu, C., Li, H., Xu, J., Gao, W., Shen, X., & Miao, S. (2023). Applying Convolutional Neural Network to Predict Soil Erosion: A Case Study of Coastal Areas. *International Journal of Environmental Research and Public Health*, *20*(3), Article 3. https://doi.org/10.3390/ijerph20032513

Liu, W., Wang, Y., Zhong, D., Xie, S., & Xu, J. (2022). *ConvLSTM network-based rainfall nowcasting method with combined reflectance and radar-retrieved wind field as inputs, Atmosphere*. MDPI.

Lv, Q., Quan, Y., Feng, W., Sha, M., Dong, S., & Xing, M. (2022). Radar Deception Jamming Recognition Based on Weighted Ensemble CNN With Transfer Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–11. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2021.3129645

Mallah, S., Delsouz Khaki, B., Davatgar, N., Scholten, T., Amirian-Chakan, A., Emadi, M., Kerry, R., Mosavi, A. H., & Taghizadeh-Mehrjardi, R. (2022). Predicting Soil Textural Classes Using Random Forest Models: Learning from Imbalanced Dataset. *Agronomy*, *12*(11), Article 11. https://doi.org/10.3390/agronomy12112613

Manos, E., Witharana, C., Udawalpola, M. R., Hasan, A., & Liljedahl, A. K. (2022). Convolutional Neural Networks for Automated Built Infrastructure Detection in the Arctic Using Sub-Meter Spatial Resolution Satellite Imagery. *Remote Sensing*, *14*(11), Article 11. https://doi.org/10.3390/rs14112719

Matthews, F., Borrelli, P., Panagos, P., & Bezak, N. (2025). Dynamic assessment of rainfall erosivity in Europe: Evaluation of EURADCLIM ground-radar data. *Hydrology and Earth System Sciences*, *29*(20), 5299–5313.

Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., & Zhang, Y. (2023). Segment anything model for medical image analysis: An experimental study. *Medical Image Analysis*, *89*, 102918. https://doi.org/10.1016/j.media.2023.102918

MCBRATNEY, A. (2003). On digital soil mapping. *Geoderma*, *117*, 3–52.

Menz, C. (2023). *Bias Adjusted CORDEX-EUR11 Simulations*. https://www.wdc-climate.de/ui/entry?acronym=BAdCORD-EUR11

Mina, M., Rezaei, M., Sameni, A., Ostovari, Y., & Ritsema, C. (2022). Predicting wind erosion rate using portable wind tunnel combined with machine learning algorithms in calcareous soils, southern Iran. *Journal of Environmental Management*, *304*, 114171.

Mitasova, H., & Mitas, L. (2001). Multiscale Soil Erosion Simulations For Land Use Management. In R. S. Harmon & W. W. Doe (Eds.), *Landscape Erosion and Evolution Modeling* (pp. 321–347). Springer US. https://doi.org/10.1007/978-1-4615-0575-4_11

Moghimi, A., Welzel, M., Celik, T., & Schlurmann, T. (2024). A comparative performance analysis of popular deep learning models and segment anything model (SAM) for river water segmentation in close-range remote sensing imagery. *IEEE Access*, *12*, 52067-52085,. https://doi.org/10.1109/ACCESS.2024.3385425

Montanarella, L., Pennock, D. J., McKenzie, N., Badraoui, M., Chude, V., Baptista, I., Mamo, T., Yemefack, M., Singh Aulakh, M., & Yagi, K. (2016). World's soils are under threat. *SOIL*, *2*, 79–82, 10 5194-2-79-2016,.

Morgan, R. P. C. (1996). *Verificaton of the European Soil Erosion Model (EOROSEM) for varying slope and vegetation conditions*.

Mosavi, A., Sajedi-Hosseini, F., Choubin, B., Taromideh, F., Rahi, G., & Dineva, A. A. (2020). Susceptibility Mapping of Soil Water Erosion Using Machine Learning Models. *Water*, *12*(7), Article 7. https://doi.org/10.3390/w12071995

Moskolaï, W., Abdou, W., Dipanda, A., & Kolyang, D. T. (2020). *Application of LSTM architectures for next frame forecasting in Sentinel-1 images time series* (arXiv:2009.00841). arXiv. https://doi.org/10.48550/arXiv.2009.00841

Munawar, H. S., Ullah, F., Qayyum, S., & Heravi, A. (2021). Application of deep learning on uav-based aerial images for flood detection. *Smart Cities*, *4*(3), 1220–1242.

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., Sillmann, J., Schaller, N., Fischer, E., Schulz, M., & Stohl, A. (2019). Frequency of extreme precipitation increases extensively with event rareness under global warming. *SCI REP-UK*, *9, 16063*, 10 1038 41598-019-52277-4,.

Nadeu, E., Gobin, A., Fiener, P., Wesemael, B., & Oost, K. (2015). Modelling the impact of agricultural management on soil carbon stocks at the regional scale: The role of lateral fluxes. *Global Change Biology*, *21*(8), 3181–3192.

Nearing, M. (1998). Why soil erosion models over-predict small soil losses and under-predict large soil losses. *CATENA*, *32*, 15–22, 10 1016 0341-8162 97 00052-0,.

Nearing, M. (2000). Evaluating soil erosion models using measured plot data: Accounting for variability in the data. *EARTH SURF PROCESSES*, *25*(9), 1035-1043,.

Nguyen, K. A., Chen, W., Lin, B. S., & Seeboonruang, U. (2021). Comparison of ensemble machine learning methods for soil erosion pin measurements. *ISPRS International Journal of Geo-Information*, *10*(1), 42.

Oliveira, P. T. S., Wendland, E., & Nearing, M. A. (2013). *Rainfall erosivity in Brazil: A review, Catena*. Elsevier.

Onnen, N., Heckrath, G., Stevens, A., Olsen, P., Greve, M. B., Pullens, J. W. M., Kronvang, B., & Van Oost, K. (2019). Distributed water erosion modelling at fine spatial resolution across Denmark. *Geomorphology*, *342*, 150–162. https://doi.org/10.1016/j.geomorph.2019.06.011

Onojeghuo, A. O., Miao, Y., & Blackburn, G. A. (2023). Deep ResU-Net Convolutional Neural Networks Segmentation for Smallholder Paddy Rice Mapping Using Sentinel 1 SAR and Sentinel 2 Optical Imagery. *Remote Sensing*, *15*(6), Article 6. https://doi.org/10.3390/rs15061517

Oost, K., Govers, G., & Desmet, P. (2000). Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. *LANDSCAPE ECOL*, *15*, 577-589,. https://doi.org/10.1023/A:1008198215674

Osco, L. P., Wu, Q., Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., Li, J., & Junior, J. M. (2023). The segment anything model (sam) for remote sensing applications: From zero to one shot. *International Journal of Applied Earth Observation and Geoinformation*, *124*, 103540.

Pally, R. J., & Samadi, S. (2022). Application of image processing and convolutional neural networks for flood image classification and semantic segmentation. *Environmental Modelling & Software*, *148*, 105285. https://doi.org/10.1016/j.envsoft.2021.105285

Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving Precipitation Estimation Using Convolutional Neural Network. *Water Resources Research*, *55*(3), 2301–2321. https://doi.org/10.1029/2018WR024090

Panagos, P., Ballabio, C., Borrelli, P., & Meusburger, K. (2016). *Spatio-temporal analysis of rainfall erosivity and erosivity density in Greece, Catena*. Elsevier.

Panagos, P., Ballabio, C., Borrelli, P., Meusburger, K., Klik, A., Rousseva, S., Tadić, M. P., Michaelides, S., Hrabalíková, M., & Olsen, P. (2015). *Rainfall erosivity in Europe, Science of the Total Environment*. Elsevier.

Panagos, P., Ballabio, C., Meusburger, K., Spinoni, J., Alewell, C., & Borrelli, P. (2017). *Towards estimates of future rainfall erosivity in Europe based on REDES and WorldClim datasets, Journal of Hydrology*. Elsevier.

Panagos, P., Meusburger, K., Van Liedekerke, M., Alewell, C., Hiederer, R., & Montanarella, L. (2014). Assessing soil erosion in Europe based on data collected through a European network. *Soil Science and Plant Nutrition*, *60*(1), 15–29. https://doi.org/10.1080/00380768.2013.835701

Peeters, I., Van Oost, K., Govers, G., Verstraeten, G., Rommens, T., & Poesen, J. (2008). The compatibility of erosion data at different temporal scales. *Earth and Planetary Science Letters*, *265*(1), 138–152. https://doi.org/10.1016/j.epsl.2007.09.040

Petan, S., Rusjan, S., Vidmar, A., & Mikoš, M. (2010). *The rainfall kinetic energy–intensity relationship for rainfall erosivity estimation in the mediterranean part of Slovenia, Journal of Hydrology*. Elsevier.

Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics*, *28*(2), 126. https://doi.org/10.2307/2346729

Pimentel, D., Harvey, C., Resosudarmo, P., Sinclair, K., Kurz, D., McNair, M., Crist, S., Shpritz, L., Fitton, L., Saffouri, R., & Blair, R. (1995). Environmental and Economic Costs of Soil Erosion and Conservation Benefits. *Science*, *267*(5201), 1117–1123. https://doi.org/10.1126/science.267.5201.1117

Plate, E. J. (2002). Flood risk and flood management. *Journal of Hydrology, Advances in Flood Research*, *267*(1), 2–11. https://doi.org/10.1016/S0022-1694(02)00135-X

Polovina, S., Radić, B., Ristić, R., & Milčanović, V. (2024). Application of remote sensing for identifying soil erosion processes on a regional scale: An innovative approach to enhance the erosion potential model. *Remote Sensing*, *16*(13), 2390.

Portalés-Julià, E., Mateo-García, G., Purcell, C., & Gómez-Chova, L. (2023). Global flood extent segmentation in optical satellite images. *Scientific Reports*, *13*(1), 20316. https://doi.org/10.1038/s41598-023-47595-7

Prasad, D. V. V., Venkataramana, L. Y., Kumar, P. S., Prasannamedha, G., Harshana, S., Srividya, S. J., Harrinei, K., & Indraganti, S. (2022). Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Science of The Total Environment*, *821*, 153311. https://doi.org/10.1016/j.scitotenv.2022.153311

Pruski, F. F., & Nearing, M. A. (2002). Climate-induced changes in erosion during the 21st century for eight U.S. locations. *Water Resources Research*, *38*(12). https://doi.org/10.1029/2001WR000493

Pu, X., Jia, H., Zheng, L., Wang, F., & Xu, F. (2024). *ClassWise-SAM-Adapter: Parameter Efficient Fine-tuning Adapts Segment Anything to SAR Domain for Semantic Segmentation* (arXiv:2401.02326). arXiv. http://arxiv.org/abs/2401.02326

Qiao, X., Roy, S. K., & Huang, W. (2024). 3-D Sharpened Cosine Similarity Operation for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 1114–1125. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. https://doi.org/10.1109/JSTARS.2023.3337112

Qin, W., Guo, Q., Zuo, C., Shan, Z., Ma, L., & Sun, G. (2016). *Spatial distribution and temporal trends of rainfall erosivity in mainland China for 1951–2010, Catena*. Elsevier.

Quinn, P., Beven, K., Chevallier, P., & Planchon, O. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, *5*(1), 59–79. https://doi.org/10.1002/hyp.3360050106

Quinton, J. N., & Fiener, P. (2024). Soil erosion on arable land: An unresolved global environmental threat. *Progress in Physical Geography: Earth and Environment*, *48*(1), 136–161. https://doi.org/10.1177/03091333231216595

Rainfall Erosivity in Europe, 511 Science of the Total Environment 801 (Elsevier 2015). https://www.sciencedirect.com/science/article/pii/S004896971500011X

Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, *29*(9), 2352–2449. Neural Computation. https://doi.org/10.1162/neco_a_00990

Rehm, R., & Fiener, P. (2024). Model-based analysis of erosion-induced microplastic delivery from arable land to the stream network of a mesoscale catchment. *SOIL, 10*, 211–230, 10 5194-10-211-2024,.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Renard, K. G. (1997). *Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE)*. U.S. Department of Agriculture, Agricultural Research Service.

Rickson, R. J., Deeks, L. K., Graves, A., Harris, J. A. H., Kibblewhite, M. G., & Sakrabani, R. (2015). Input constraints to food production: The impact of soil degradation. *FOOD SECUR*, *7*, 351–364, 10 1007 12571-015-0437-,.

Risse, M., Nearing, M. A., Laflen, J. M., & Nicks, A. (1993). D.: Error assessment in the Universal Soil Loss Equation. *SOIL SCI SOC AM J*, *57*, 825-833,. https://doi.org/10.2136/sssaj1993.03615995005700030032x

Rompaey, A. J. J., Verstraeten, G., Oost, K., Govers, G., & Poesen, J. (2001). Modelling mean annual sediment yield using a distributed approach. *EARTH SURF PROCESSES*, *26*, 1221-1236,. https://doi.org/10.1002/esp.275

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Rudolf, B., & Rapp, J. (2003). The century flood of the River Elbe in August 2002: Synoptic weather development and climatological aspects. *Quarterly Rep*, *2*, 8–23.

Saadeldin, M., O'Hara, R., Zimmermann, J., Mac Namee, B., & Green, S. (2022). Using deep learning to classify grassland management intensity in ground-level photographs for more automated production of satellite land use maps. *Remote Sensing Applications: Society and Environment*, *26*, 100741. https://doi.org/10.1016/j.rsase.2022.100741

Safavi, F., & Rahnemoonfar, M. (2023). Comparative Study of Real-Time Semantic Segmentation Networks in Aerial Images During Flooding Events. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 4–20. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. https://doi.org/10.1109/JSTARS.2022.3219724

Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., & Ryo, M. (2023). Ten deep learning techniques to address small data problems with remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *125*, 103569. https://doi.org/10.1016/j.jag.2023.103569

Saha, S., Roy, J., Pradhan, B., & Hembram, T. K. (2021). Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at East Sikkim Himalayan, India. *Advances in Space Research*, *68*(7), 2819–2840. https://doi.org/10.1016/j.asr.2021.05.018

Sahay, R., & Savakis, A. (2024). On aligning SAM to remote sensing data. In *Geospatial Informatics: XIV (Vol. 13037* (pp. 10–18). SPIE.

Schindewolf, M., & Schmidt, J. (2012). Parameterization of the EROSION 2D/3D soil erosion model using a small-scale rainfall simulator and upstream runoff simulation. *CATENA, Experiments in Earth Surface Process Research*, *91*, 47–55. https://doi.org/10.1016/j.catena.2011.01.007

Schmidt, J., Werner, M. V., & Michael, A. (1999). Application of the EROSION 3D model to the CATSOP watershed, The Netherlands. *Catena*, *37*(3–4), 449–456.

Scholten, T., & Seitz, S. (2019). Soil Erosion and Land Degradation. *Soil Systems*, *3*(4), Article 4. https://doi.org/10.3390/soilsystems3040068

Scholz, F., Traub, M., Zarfl, C., Scholten, T., & Butz, M. V. (2025). *Fully differentiable, fully distributed rainfall-runoff modeling, Hydrology and Earth System Sciences*. Copernicus Publications Göttingen.

Schwertmann, U., Vogl, W., & Kainz, M. (1987). *Bodenerosion durch wasser*. Ulmer Verlag.

Senanayake, S., Pradhan, B., Alamri, A., & Park, H.-J. (2022). A new application of deep neural network (LSTM) and RUSLE models in soil erosion prediction. *Science of The Total Environment*, *845*, 157220. https://doi.org/10.1016/j.scitotenv.2022.157220

Shamsolmoali, P., Zareapoor, M., Wang, R., Zhou, H., & Yang, J. (2019). A Novel Deep Structure U-Net for Sea-Land Segmentation in Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(9), 3219–3232. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. https://doi.org/10.1109/JSTARS.2019.2925841

Sharma, S., & Mujumdar, P. P. (2019). On the relationship of daily rainfall extremes and local mean temperature. *Journal of Hydrology*, *572*, 179–191. https://doi.org/10.1016/j.jhydrol.2019.02.048

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, *28*.

https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html

Shokati, H., Engelhardt, A., Seufferheld, K., Taghizadeh-Mehrjardi, R., Fiener, P., Lensch, H. P. A., & Scholten, T. (2025). Erosion-SAM: Semantic segmentation of soil erosion by water. *CATENA*, *254*, 108954. https://doi.org/10.1016/j.catena.2025.108954

Shokati, H., Mashal, M., Noroozi, A., Abkar, A. A., Mirzaei, S., Mohammadi-Doqozloo, Z., Taghizadeh-Mehrjardi, R., Khosravani, P., Nabiollahi, K., & Scholten, T. (2024). Random Forest-Based Soil Moisture Estimation Using Sentinel-2, Landsat-8/9, and UAV-Based Hyperspectral Data. *Remote Sensing*, *16*(11), Article 11. https://doi.org/10.3390/rs16111962

Shokati, H., Mashal, M., Noroozi, A., Mirzaei, S., & Mohammadi-Doqozloo, Z. (2023). Assessing soil moisture levels using visible UAV imagery and machine learning models. *Remote Sensing Applications: Society and Environment*, *32*, 101076. https://doi.org/10.1016/j.rsase.2023.101076

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Simantiris, G., & Panagiotakis, C. (2024). Unsupervised Color-Based Flood Segmentation in UAV Imagery. *Remote Sensing*, *16*(12), Article 12. https://doi.org/10.3390/rs16122126

Sinha, R. K., Kushwaha, M., Choudhary, J., Singh, D. P., & Pandey, M. (2024). Flood image segmentation of uav aerial images using deep learning. *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS*, 1–6.

Smith, M. W., Carrivick, J. L., Hooke, J., & Kirkby, M. J. (2014). Reconstructing flash flood magnitudes using 'Structure-from-Motion': A rapid assessment tool. *Journal of Hydrology*, *519*, 1914–1927. https://doi.org/10.1016/j.jhydrol.2014.09.078

Song, B., Yang, H., Wu, Y., Zhang, P., Wang, B., & Han, G. (2024). A Multispectral Remote Sensing Crop Segmentation Method Based on Segment Anything Model Using Multistage Adaptation Fine-Tuning. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–18. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2024.3411398

Srikanthan, R., & McMahon, T. A. (2001). Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, *5*(4), 653–670. https://doi.org/10.5194/hess-5-653-2001

Steegen, A., Govers, G., Nachtergaele, J., Takken, I., Beuselinck, L., & Poesen, J. (2000). Sediment export by water from an agricultural catchment in the Loam Belt of central Belgium. *Geomorphology*, *33*(1), 25–36. https://doi.org/10.1016/S0169-555X(99)00108-7

Sugiura, R., Noguchi, N., & Ishii, K. (2005). Remote-sensing Technology for Vegetation Monitoring using an Unmanned Helicopter. *Biosystems Engineering*, *90*(4), 369–379. https://doi.org/10.1016/j.biosystemseng.2004.12.011

Sun, J., Yan, S., Yao, X., Gao, B., & Yang, J. (2024). A Segment Anything Model based weakly supervised learning method for crop mapping using Sentinel-2 time series images. *International Journal of Applied Earth Observation and Geoinformation*, *133*, 104085. https://doi.org/10.1016/j.jag.2024.104085

Taghizadeh-Mehrjardi, R., Bawa, A., Kumar, S., Zeraatpisheh, M., Amirian-Chakan, A., & Akbarzadeh, A. (2019). Soil erosion spatial prediction using digital soil mapping and RUSLE methods for Big Sioux River watershed. *Soil Systems*, *3*(3), 43.

Taguchi, R., Tanoue, M., Yamazaki, D., & Hirabayashi, Y. (2022). Global-Scale Assessment of Economic Losses Caused by Flood-Related Business Interruption. *Water*, *14*(6), Article 6. https://doi.org/10.3390/w14060967

Tanim, A. H., McRae, C. B., Tavakol-Davani, H., & Goharian, E. (2022). Flood Detection in Urban Areas Using Satellite Imagery and Machine Learning. *Water*, *14*(7), Article 7. https://doi.org/10.3390/w14071140

Tingsanchali, T. (2012). Urban flood disaster management. *Procedia Engineering, ISEEC*, *32*, 25–37. https://doi.org/10.1016/j.proeng.2012.01.1233

Tong, H. (2023). Nonasymptotic analysis of robust regression with modified Huber's loss. *Journal of Complexity*, *76*, 101744. https://doi.org/10.1016/j.jco.2023.101744

Uber, M., Haller, M., Brendel, C., Hillebrand, G., & Hoffmann, T. (2024). *Past, present and future rainfall erosivity in central Europe based on convection-permitting climate simulations, Hydrology and Earth System Sciences*.

Copernicus GmbH.

Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., & Pinto, J. G. (2003). The central European floods of August 2002: Part 1 - Rainfall periods and flood development. *Weather*, *58*(10), 371–377.

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, *14*(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

Verstraeten, G., Oost, K., Rompaey, A., Poesen, J., & Govers, G. (2002). Evaluating an integrated approach to catchment management to reduce soil loss and sediment pollution through modelling. *SOIL USE MANAGE*, *18*, 386–394, 10 1111 1475-2743 2002 00257 ,.

Verstraeten, G., Poesen, J., Demarée, G., & Salles, C. (2006). Long-term (105 years) variability in rain erosivity as derived from 10-min rainfall depth data for Ukkel (Brussels, Belgium): Implications for assessing soil erosion rates. *Journal of Geophysical Research: Atmospheres*, *111*(D22), 2006JD007169. https://doi.org/10.1029/2006JD007169

Vinayahalingam, S., Berends, B., Baan, F., Moin, D. A., van Luijn, R., Bergé, S., & Xi, T. (2023). Deep learning for automated segmentation of the temporomandibular joint. *Journal of Dentistry*, *132*, 104475. https://doi.org/10.1016/j.jdent.2023.104475

Wagner, F., Eltner, A., & Maas, H.-G. (2023). River water segmentation in surveillance camera images: A comparative study of offline and online augmentation using 32 CNNs. *International Journal of Applied Earth Observation and Geoinformation*, *119*, 103305. https://doi.org/10.1016/j.jag.2023.103305

Walling, D. E., He, Q., & Whelan, P. A. (2003). Using 137Cs<math><msup><mi></mi><mn>137</mn></msup><mtext>Cs</mtext></math> measurements to validate the application of the AGNPS and ANSWERS erosion and sediment yield models in two small Devon catchments. *Soil and Tillage Research, Field Application of the Cs-137 Technique in Soil Erosion and Sedimentation Studies*, *69*(1), 27–43. https://doi.org/10.1016/S0167-1987(02)00126-5

Wang, F., Sahana, M., Pahlevanzadeh, B., Chandra Pal, S., Kumar Shit, P., Piran, Md. J., Janizadeh, S., Band, S. S., & Mosavi, A. (2021). Applying different resampling strategies in machine learning models to predict head-cut gully erosion susceptibility. *Alexandria Engineering Journal*, *60*(6), 5813–5829. https://doi.org/10.1016/j.aej.2021.04.026

Wang, Q., Li, J., Jin, T., Chang, X., Zhu, Y., Li, Y., Sun, J., & Li, D. (2020). Comparative Analysis of Landsat-8, Sentinel-2, and GF-1 Data for Retrieving Soil Moisture over Wheat Farmlands. *Remote Sensing*, *12*(17), Article 17. https://doi.org/10.3390/rs12172708

Wang, S., Quan, D., Liang, X., Ning, M., Guo, Y., & Jiao, L. (2018). A deep learning framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing, Deep Learning RS Data*, *145*, 148–164. https://doi.org/10.1016/j.isprsjprs.2017.12.012

Wang, Y., Cheng, C., Xie, Y., Liu, B., Yin, S., Liu, Y., & Hao, Y. (2017). *Increasing trends in rainfall-runoff erosivity in the Source Region of the Three Rivers, 1961–2012, Science of the Total Environment*. Elsevier.

Wendt, R. C., Alberts, E. E., & Hjelmfelt Jr., A. T. (1986). Variability of Runoff and Soil Loss from Fallow Experimental Plots. *Soil Science Society of America Journal*, *50*(3), 730–736. https://doi.org/10.2136/sssaj1986.03615995005000030035x

Wieland, M., Martinis, S., Kiefl, R., & Gstaiger, V. (2023). Semantic segmentation of water bodies in very high-resolution satellite and aerial images. *Remote Sensing of Environment*, *287*, 113452. https://doi.org/10.1016/j.rse.2023.113452

Wilken, F., Fiener, P., & Auerswald, K. (2019a). *Meteorological data at the Scheyern experimental farm covering 14 small adjacent watersheds and their surroundings* [Dataset]. https://doi.org/10.13140/RG.2.2.34561.10088

Wilken, F., Fiener, P., & Auerswald, K. (2019b). *Topography at the Scheyern experimental farm covering 14 small adjacent watersheds and their surroundings* [Dataset]. https://doi.org/10.13140/RG.2.2.32044.51845

Wilken, F., Ketterer, M., Koszinski, S., Sommer, M., & Fiener, P. (2020). Understanding the role of water and tillage erosion from $^{239+240}$Pu tracer measurements using inverse modelling. *SOIL*, *6*(2), 549–564. https://doi.org/10.5194/soil-6-549-2020

Wilken, F., Sommer, M., Oost, K., Bens, O., & Fiener, P. (2017). Process-oriented modelling to identify main drivers of erosion-induced carbon fluxes. *Soil*, *3*(2), 83–94.

Bibliography

Wischmeier, W. H., & Smith, D. D. (1978). *Predicting Rainfall Erosion Losses: A Guide to Conservation Planning*. Department of Agriculture, Science and Education Administration.

Wu, Y., Ouyang, W., Hao, Z., Lin, C., Liu, H., & Wang, Y. (2018). *Assessment of soil erosion characteristics in response to temperature and precipitation in a freeze-thaw watershed, Geoderma*. Elsevier.

Xie, W., Willems, N., Patil, S., Li, Y., & Kumar, M. (2024). *SAM Fewshot Finetuning for Anatomical Segmentation in Medical Images*. 3253–3261. https://openaccess.thecvf.com/content/WACV2024/html/Xie_SAM_Fewshot_Finetuning_for_Anatomical_Segmentation_in_Medical_Images_WACV_2024_paper.html

Xu, Z., Wang, S., Stanislawski, L. V., Jiang, Z., Jaroenchai, N., Sainju, A. M., Shavers, E., Usery, E. L., Chen, L., Li, Z., & Su, B. (2021). An attention U-Net model for detection of fine-scale hydrologic streamlines. *Environmental Modelling & Software*, *140*, 104992. https://doi.org/10.1016/j.envsoft.2021.104992

Yang, D., Kanae, S., Oki, T., Koike, T., & Musiake, K. (2003). Global potential soil erosion with reference to land use and climate changes. *Hydrological Processes*, *17*(14), 2913–2928. https://doi.org/10.1002/hyp.1441

Yang, Q., Shi, L., Han, J., Zha, Y., & Zhu, P. (2019). Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research*, *235*, 142–153. https://doi.org/10.1016/j.fcr.2019.02.022

Yao, H., Qin, R., & Chen, X. (2019). Unmanned Aerial Vehicle for Remote Sensing Applications—A Review. *Remote Sensing*, *11*(12), Article 12. https://doi.org/10.3390/rs11121443

Ye, T., Qin, W., Zhao, Z., Gao, X., Deng, X., & Ouyang, Y. (2023). Real-Time Object Detection Network in UAV-Vision Based on CNN and Transformer. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–13. IEEE Transactions on Instrumentation and Measurement. https://doi.org/10.1109/TIM.2023.3241825

Yin, S., Nearing, M. A., Borrelli, P., & Xue, X. (2017). Rainfall erosivity: An overview of methodologies and applications. *Vadose Zone Journal*, *16*(12), 1–16.

Yin, S., Xie, Y., Liu, B., & Nearing, M. A. (2015). *Rainfall erosivity estimation based on rainfall data collected over a range of temporal resolutions, Hydrology and Earth System Sciences*. Copernicus GmbH Göttingen.

Yin, S., Xie, Y., Nearing, M. A., & Wang, C. (2007). *Estimation of rainfall erosivity using 5-to 60-minute fixed-interval rainfall data from China, Catena*. Elsevier.

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*(2), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*, 111716. https://doi.org/10.1016/j.rse.2020.111716

Zamboni, P. A. P., Blanch, X., Marcato Junior, J., Gonçalves, W. N., & Eltner, A. (2025). Do we need to label large datasets for river water segmentation? Benchmark and stage estimation with minimum to non-labeled image time series. *International Journal of Remote Sensing*, *46*, 2719-2747,. https://doi.org/10.1080/01431161.2025.2457131

Zeder, J., & Fischer, E. M. (2020). Observed extreme precipitation trends and scaling in Central Europe. *Weather and Climate Extremes*, *29*, 100266. https://doi.org/10.1016/j.wace.2020.100266

Zhang, H., Li, Y., Jiang, Y., Wang, P., Shen, Q., & Shen, C. (2019). Hyperspectral Classification Based on Lightweight 3-D-CNN With Transfer Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(8), 5813–5828. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2019.2902568

Zhang, Y., Liu, P., Chen, L., Xu, M., Guo, X., & Zhao, L. (2023). A new multi-source remote sensing image sample dataset with high resolution for flood area extraction: GF-FloodNet. *International Journal of Digital Earth*, *16*(1), 2522–2554. https://doi.org/10.1080/17538947.2023.2230978

Zhou, X., Liang, F., Chen, L., Liu, H., Song, Q., Vivone, G., & Chanussot, J. (2024). MeSAM: Multiscale enhanced segment anything model for optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–15.