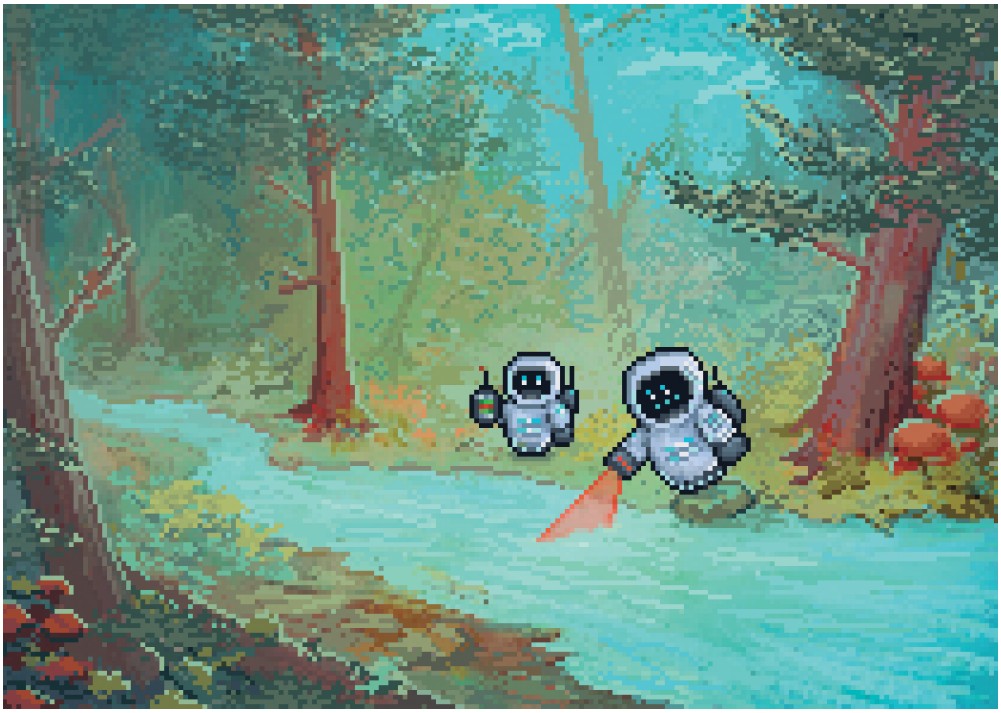
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2025:81
Faculty of Forest Sciences

Mapping small water channels using machine learning

Mariana Dos Santos Toledo Busarello



Mapping small water channels using machine learning

Mariana Dos Santos Toledo Busarello
Faculty of Forest Sciences
Department of Forest Ecology and Management
Umeå

SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Umeå 2025

Acta Universitatis Agriculturae Sueciae
2025:81

Cover: Two robots classifying a channel, by Eybraian Lopes Bonifácio (2025)

ISSN 1652-6880
ISBN (print version) 978-91-8124-065-8
ISBN (electronic version) 978-91-8124-111-2
https://doi.org/10.54612/a.1vuvm11qn6

Swedish University of Agricultural Sciences, Department of Forest Ecology and Management, Umeå, Sweden

Print: SLU Grafisk service, Uppsala 2025

Mapping small water channels using machine learning

Abstract

Boreal landscapes are shaped by a dense network of natural streams, modified streams, and ditches. Together, they regulate hydrology, nutrient transport, and ecosystem function. Historically, streams were modified to accommodate log transportation, and drainage ditches were dug to improve food and timber production. Although new ditching has mostly stopped, historical changes to the drainage network still affect forestry and water management. Small streams and ditches are the landscape's capillaries, but they remain poorly mapped despite their vital hydrological and ecological roles. This thesis addresses this gap in knowledge by developing a novel, national-scale framework for mapping small streams and ditches using high-resolution topographic data and machine learning techniques. Combining convolutional neural networks, XGBoost classification, uncertainty quantification, and drainage analyses, this work identifies geomorphological and hydrological indices that distinguish streams from ditches across the landscape. The highest-performing model shows that integrating digital elevation models with terrain indices and machine learning delineates the channel networks successfully for ditches (recall=76%, precision=88%) and moderately for natural streams (recall=58%, precision=56%). Furthermore, the produced uncertainty maps highlight low-certainty pixels from the background that can be used to potentially improve the mapping of streams in the future. To the best of our knowledge, this is the first study that can separate streams and ditches on maps across an entire nation. By providing consistent, scalable maps of small channels, this research supports restoration prioritization, sustainable forestry planning, and national reporting under EU and UN environmental frameworks. The methodology also offers a reproducible approach for characterizing coupled natural-artificial drainage systems in boreal and temperate regions worldwide.

Keywords: machine learning, deep learning, neural network, ditch, drainage, water channels, XGBoost, U-Net

Kartläggning av små vattendrag och diken med hjälp av maskininlärning

Abstract

Det boreala landskapet präglas av ett omfattande nätverk av naturliga vattendrag, modifierade vattendrag och diken som tillsammans reglerar hydrologi, näringsflöden och ekosystemfunktioner. Historiskt har vattendrag rätats och rensats för timmerflottning medan diken har grävts för att öka jordbruks- och skogsproduktionen. Trots att nydikning numera är ovanlig påverkar dessa historiska ingrepp fortfarande vattenförvaltning och skogsbruk. Små vattendrag och diken fungerar som landskapets kapillärer men är fortfarande bristfälligt kartlagda, trots deras centrala hydrologiska och ekologiska betydelse. Denna avhandling utvecklar en metod för att kartlägga små vattendrag och diken på nationell skala med högupplösta topografiska data och maskininlärning. Genom att kombinera konvolutionella neurala nätverk, XGBoost-klassificering, osäkerhetsanalys och data på flödesackumulering identifieras geomorfologiska och hydrologiska indicier som skiljer diken från naturliga vattendrag. Den bästa metoden visar att digitala höjdmodeller och terrängindicier kan användas för att effektivt avgränsa vattendrag och diken. Metoden hade hög precision för diken (recall=76%, precision=88%) och mer måttliga resultat för naturliga vattendrag (recall=58%, precision=56%). Osäkerhetskartor visar dessutom var framtida förbättringar av kartläggningen bör riktas.

Detta är den första studien som framgångsrikt särskiljer vattendrag och diken för ett helt land. Genom att skapa konsekventa, skalbara kartor över små vattendrag och diken bidrar forskningen till restaureringsprioritering, hållbar skogsförvaltning och nationell miljörapportering inom EU:s och FN:s ramar. Metoden erbjuder även ett reproducerbart sätt att beskriva sammankopplade naturliga och artificiella dräneringssystem i boreala och tempererade regioner.

Keywords: maskininlärning, djupinlärning, neurala nätverk, diken, dränering, vattendrag, XGBoost, U-Net

Mapeando cursos d’água estreitos usando aprendizado de máquina

Resumo

As paisagens boreais são moldadas por uma rede densa de cursos d’água naturais, cursos d’água modificados e valas de drenagem. Juntos, eles regulam a hidrologia, o transporte de nutrientes e o funcionamento dos ecossistemas. Historicamente, na Suécia os cursos d’água foram modificados para o transporte de madeira, e valas de drenagem foram escavadas para melhorar a agricultura e produção florestal. Embora a abertura de novas valas não seja mais permitida, as alterações históricas na drenagem ainda afetam o manejo florestal e a gestão da água. Pequenos cursos d’água e valas funcionam como as veias do terreno, mas continuam sendo pouco mapeados, apesar do seu papel essencial em hidrologia e ecologia.

Esta tese foca nessa lacuna no conhecimento ao desenvolver um fluxo de trabalho inédito, em escala nacional, para o mapeamento de pequenos cursos d’água e valas, utilizando dados topográficos de alta resolução e técnicas de aprendizado de máquina. Combinando redes neurais convolucionais, classificação com XGBoost, quantificação de incertezas e análises de drenagem, o estudo identifica índices geomorfológicos e hidrológicos que distinguem cursos d’água de valas em toda a paisagem. O modelo com melhor desempenho demonstra que a combinação de modelos digitais de elevação com índices de terreno e aprendizado de máquina mapeia com sucesso as valas (recall=76%, precisão=88%) e de forma moderada os cursos d’água naturais (recall=58%, precisão=56%). Além disso, os mapas de incerteza produzidos destacam pixels de baixa confiabilidade do background, que podem ser usados para aprimorar o mapeamento de cursos d’água no futuro.

Até onde sabemos, este é o primeiro estudo capaz de identificar separadamente cursos d’água naturais e valas de todo um país. Ao fornecer mapas consistentes e reprodutíveis de cursos d’água estreitos, esta pesquisa ajuda a determinar a prioridade para ações de restauração, o planejamento florestal sustentável e os relatórios nacionais sob os marcos ambientais da UE e da ONU. A metodologia também oferece uma abordagem reprodutível para caracterizar sistemas de drenagem naturais e artificiais interconectados em regiões boreais e temperadas em todo o mundo.

Keywords: aprendizado de máquina, aprendizagem profunda, redes neurais, valas, drenagem, rios, cursos d’água, XGBoost, U-Net

Credits: Léa Tessier

Dedication

In memory of my grandmother Aracy Matheus dos Santos (☥05/02/2021). I wonder what she’s cooking.

“Science isn’t about why. It’s about WHY NOT.”’

– Cave Johnson

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Busarello, M.D.S.T., Ågren, A.M., Westphal, F. & Lidberg, W. (2025). Automatic detection of ditches and natural streams from digital elevation models using deep learning. Computers & Geosciences, 196, 105875. https://doi.org/10.1016/j.cageo.2025.105875

II. Busarello, M.D.S.T., Ågren, A.M., Westphal, F., Hasselquist, E.M. & Lidberg, W. (2025). Classifying stream and ditch channels mapped from High-Resolution Digital Elevation Models Using Machine Learning. *Manuscript.*

III. Westphal, F., Ågren, A.M., Busarello, M.D.S.T. & Lidberg, W. Uncertainty quantification for LiDAR-based maps of ditches and natural streams. (2025). Environmental Modelling & Software, 191, 106488. https://doi.org/10.1016/j.envsoft.2025.106488

IV. Busarello, M.D.S.T., Ågren, A.M., Westphal, F. & Lidberg, W. (2025) Upscaling small channel classification from High-Resolution Digital Elevation Models using Deep Learning. *Manuscript.*

“Sometimes scientists change their minds. New developments cause a rethink. If this bothers you, consider how much damage is being done to the world by people for whom new developments do not cause a rethink.”

– Terry Pratchett, *The Science of Discworld*

List of tables

“I want to know why. Why *everything*. I don't know the answers, but a few days ago I didn't know there were questions.”

– Terry Pratchett, *Nation*

List of figures

Abbreviations

ALS	Aerial Laser Scanning
CNN	Convolutional Neural Network
DEM	Digital Elevation Model
DL	Deep Learning
FCP	Feature Conformal Prediction
FN	False Negative
FP	False Positive
HPMF	High-Pass Median Filter
LiDAR	Light Detection and Ranging
MCC	Matthew's Correlation Coefficient
ML	Machine Learning
TN	True Negative
TP	True Positive
XGBoost	eXtreme Gradient Boosting

Credits: Fernanda Schäfer Hackenhaar

1. Introduction

Boreal forests are the largest biome on Earth (Sanderson et al. 2012), spanning northern Europe, Asia, and North America (Figure 1). This global forest contains dense networks of small streams, wetlands, and peatlands that regulate hydrology and nutrient cycling (Pomeroy et al. 1998). Over centuries, these water systems have been modified in Sweden, Finland, and Canada (Lavoie et al. 2005), where ditches have drained the soil to improve forest productivity and expand agricultural land (Jacks 2019)(Figure 2), altering stream morphology, disrupting habitats, and impacting ecosystem processes. Therefore, understanding how natural and artificial channels interact is key to effectively managing and restoring water systems in these countries.

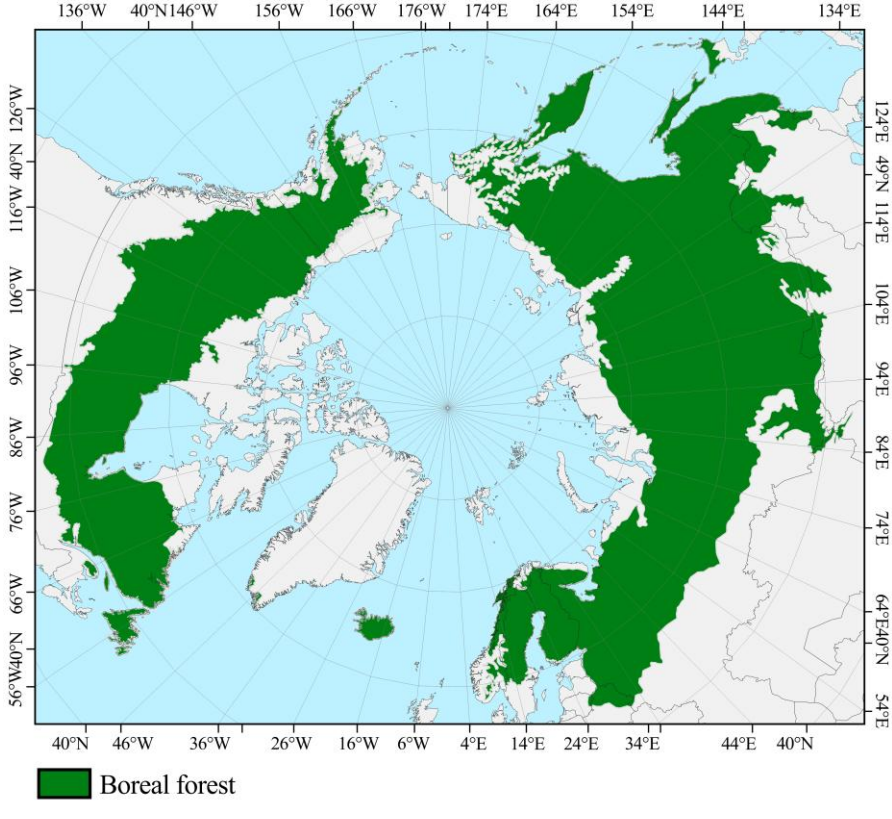


Figure 1. The coverage of boreal forests across the higher latitudes of the globe. Plotted with data from Boucher et al. (2024).

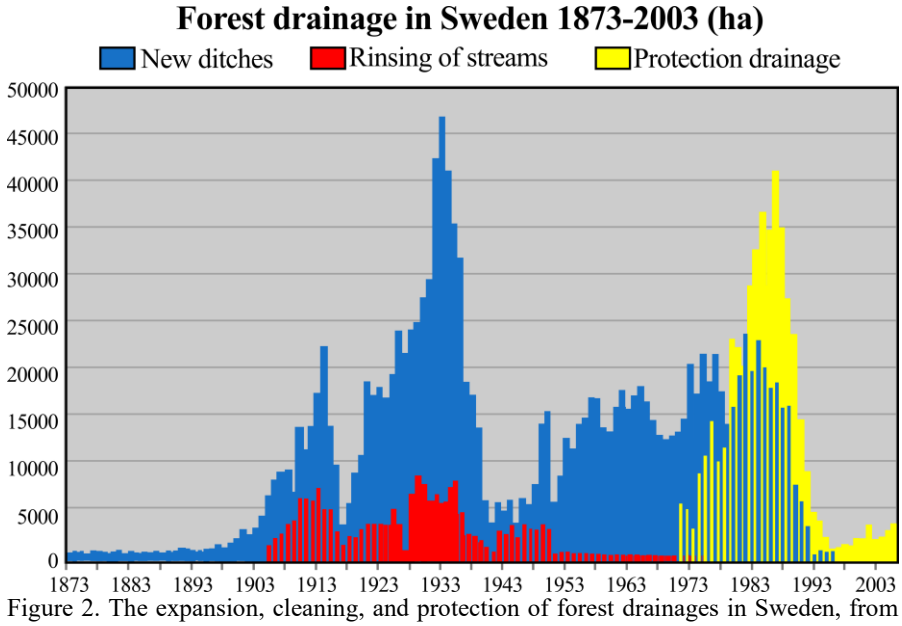


Figure 2. The expansion, cleaning, and protection of forest drainages in Sweden, from 1873 to 2003. Adapted from (Jacks 2019).

In Sweden, the historical development of the forest industry further altered watercourses. To transport logs downstream, streams and rivers were modified: boulders were blasted or removed, watercourses were straightened or redirected, and dams were built to retain water during low-flow periods (Törnlund & Östlund 2006). These interventions changed stream morphology and ecology, disrupting habitats, reducing biodiversity, altering flow regimes, and affecting nutrient cycling (Törnlund & Östlund 2002). When timber floating ceased and road transport became dominant, these channel modifications remained, continuing to affect the local system.

Ditches are another form of human modification of the Swedish landscape. Drainage ditches have been constructed since medieval times (Jacks 2019) and today constitute about 67 % of the country's total channel network, which is currently estimated at 1.2 million km (Laudon et al. 2022a). This proportion corresponds to approximately 800 000 km of ditch channels, the majority of which are forest ditches (Paul et al. 2023). Now, ditches present a complex relationship with natural streams within the Swedish hydrological network. Distinguishing one from the other may appear straightforward: ditches are typically linear, straight channels, but not always (Figure 3B, Figure 3D, Figure 3F), while natural streams tend to

meander (Figure 3A, Figure 3C, Figure 3E), but not always. Furthermore, ditches may have undergone naturalization processes over decades since they were dug, developing their own ecosystems that support diverse fauna and flora (Armitage et al. 2003; Herzon & Helenius 2008) and creating important ecological functions. This creates challenges for restoration, management, and compliance with environmental regulations.

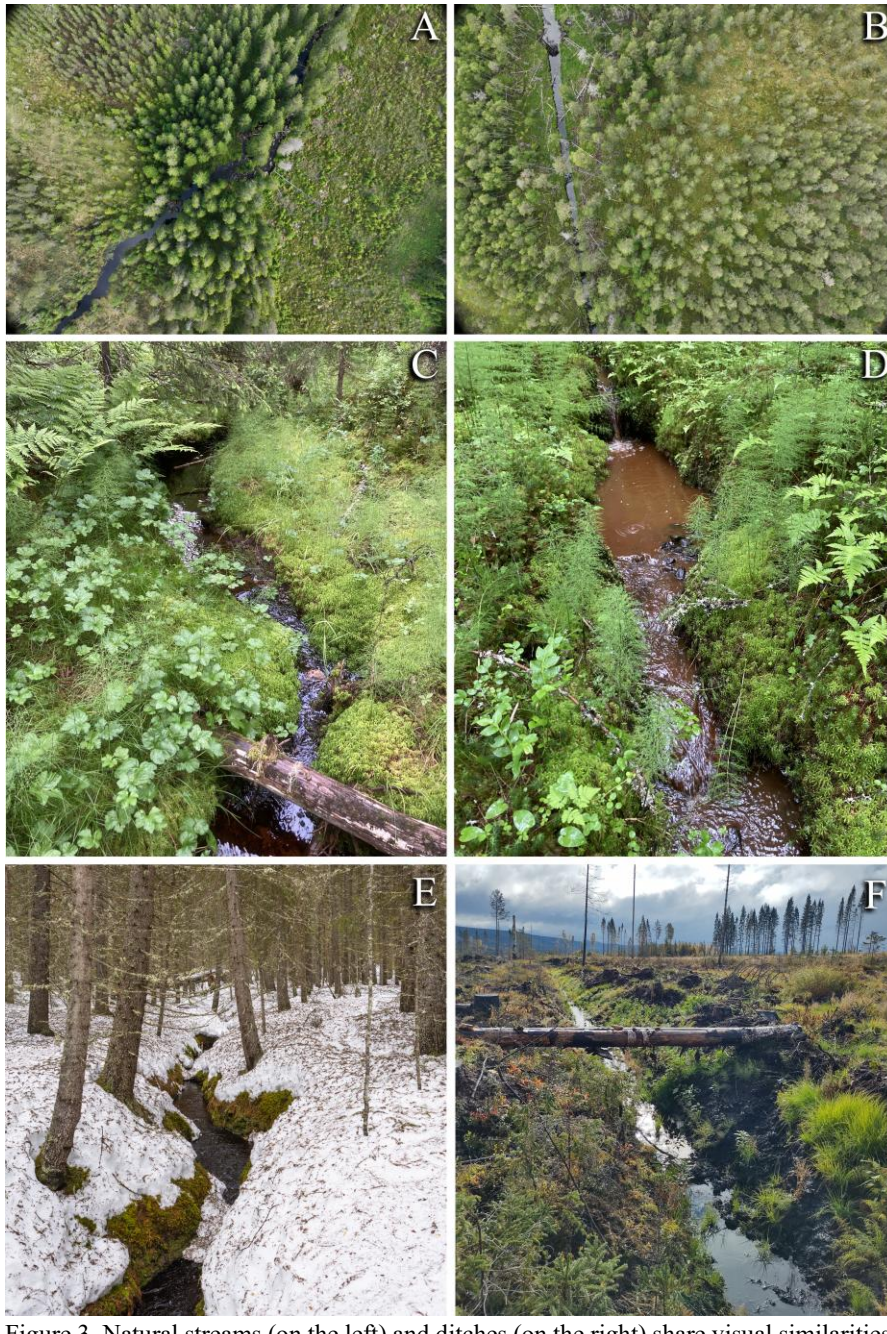


Figure 3. Natural streams (on the left) and ditches (on the right) share visual similarities. Image credits: (A,B) Alejandro Gandara, (C,D) Cedrik Åkermark, (E) Andreas Palmén.

Both ditches and natural streams are subject to hydrological processes shaped by climate change. Global warming is predicted to increase precipitation in boreal regions, raising discharge in streams and expanding groundwater-fed areas along channels (Nilsson et al. 2013) which may benefit plant species richness. Higher temperatures increase the occurrence of heatwaves (Coumou & Rahmstorf 2012) and droughts (Trenberth 2011), which affects the water quality of boreal streams (Gómez-Gener et al. 2020). For streams in peatland-dominated boreal catchments, the DOC exports shift due to changes in runoff and precipitation (Dore 2005; Gauthier et al. 2015; Prijac et al. 2023). Earlier snowmelt affects spring flood peaks (Falloon & Betts 2006), reducing the channel geomorphological activity because of weaker extreme flood events (Andréasson et al. 2004), impacting channel connectivity (Croke et al. 2013). This represents a challenge for forestry practices because, following an initial increase in forest productivity with warmer temperatures, the adverse consequences outweigh any benefit (World Bank 2014).

Increasing the soil carbon sequestration is an important climate change mitigation strategy (Minx et al. 2018). Peatlands, the predominant wetland type in boreal landscapes, are efficient natural long-term carbon sinks and help regulate water flow during dry periods (Karimi et al. 2025; Laudon et al. 2025). Unfortunately, extensive ditching for forestry in Sweden (940.000 ha; Hånell 1990) and Finland (13% of the country's area; Peltomaa 2007) has drained these wetlands, reducing their hydrological and ecological functions.

To prevent the expansion of ditched areas in Sweden, regulations were implemented starting in 1986 (Hasselquist et al. 2020), such as the requirement of permits to dig new ditches (Skogsstyrelsen 2022). However, after one consults the Swedish Forest Agency (Swedish PEFC 2017), ditches can be cleaned to improve drainage capability; that is, the accumulated sediments can be removed using an excavator. This harms the established ecosystem and releases more sediments and nutrients from the soil, impacting the watercourses that receive this water (Nieminen et al. 2018). Currently, natural streams are a target of protection at many levels. For example, the Swedish Forest Agency recommends the use of a 30 m riparian buffer along natural streams to protect the aquatic habitat from forestry and agricultural practices (Skogsstyrelsen 2022). Such forest buffers act as a nutrient sink on the channel margins and promote biodiversity by hosting

different species than the surrounding area (Gundersen et al. 2010). However, in practice, these buffers vary greatly in terms of width, enforcement, and effectiveness. Adhesion by forest owners is voluntary and, as estimated in Kuglerová et al. (2020), only 25% of small streams even have said buffers. Also, the average width is ±4 meters, far from the recommended.

Restoration efforts in small waterways mostly focus on ditch blocking. The aim is to reestablish wetland hydrology by blocking ditches to recover groundwater storage and support ecological functions (Maanavilja et al. 2014; Bring et al. 2022)(Figure 4). However, rewetting drained peatlands can conflict with forestry interests (Lõhmus et al. 2015). The widespread high drainage density across the Swedish landscape shows the impact of ditching (Laudon et al. 2022), peaking 15 km/km² in some areas. Around 53% of Sweden's peatlands were altered, i.e., 23% of the national landscape (Vasander et al. 2003). Since 21.6% of degraded peatlands are boreal, there is an opportunity for large-scale rewetting and carbon storage: rewetting 60% of today's degraded peatlands could turn the global land system into a net carbon sink by 2100 (Humpenöder et al. 2020). Reflecting this potential, studies on boreal peatland restoration have been conducted in Sweden (Elenius et al. 2025; Laudon et al. 2025; Zannella et al. 2025), Finland (Komulainen et al. 1999; Haapalehto et al. 2011), and Canada (Nugent et al. 2019).

Figure 4. A ditch being filled during the restoration of the Stormyran mire, in the Trollberget study site. Image credits: Andreas Palmén.

Restoration attempts of natural streams that were altered by timber floating have been made (Gardeström et al. 2013) with varying results (Helfield et al. 2007; Hasselquist et al. 2017; Frainer et al. 2018; Pilotto et al. 2018), but recovering pre-floating conditions is still a challenge (Nilsson et al. 2005). Meanwhile, the restoration of wetlands means removing the effect of ditches, by blocking water using a plug or filling them in completely with an excavator. The differences in legislation and management practices between natural streams or ditches show the importance of classifying them correctly to follow the goals of Agenda 2030.

The Nature Restoration Law sets a critical target for freshwater ecosystems: restoring at least 20% of degraded ecosystems by 2030 and 90% by 2050 (Council of the European Union 2023). On a larger scale, the Sustainable Development goals listed in the United Nations Agenda 2030 focus on the importance of protecting water resources from degradation and promoting their sustainable management (United Nations General Assembly 2015). This led the European Union to adopt the Water Framework Directive guidelines (European Commission 2000), which require member states to implement policies aimed at improving the ecological and chemical status of water bodies. Consequently, the monitoring and management of freshwater

ecosystems becomes a priority to achieve restoration goals, with mapping being key to this development.

Planning the management of small waterways (<6 m width) has been a challenge in Sweden because the majority of them were missing from topographical maps, where, as reported in Flyckt et al. (2022), only 9% of the ditches, 25% of the straightened watercourses, and 45% of the natural watercourses were present. In fact, Bishop et al. (2008) named them the *Aqua Incognita* - the unknown headwaters. While other studies have focused on understanding the hydrology, water quality, and ecology of small waterways, this thesis focuses on mapping the networks using novel technology. Common tools to map and classify these channels are field surveys (Brookes 1987), comparing current and historical maps to aerial imagery (Ruuska & Helenius 1996), and checking channel continuity (Zaharia et al. 2018). However, such detailed work on a national scale has a high cost and takes a long time to be completed: we estimated it would take 90 years to digitize all the small water channels in Sweden, which does not match the urgency for compliance with the Agenda 2030. Considering the limitations of large-scale surveys, digital methods rose as an affordable alternative.

Topography-based methods using Digital Elevation Models (DEMs) derived from Aerial Laser Scanning (ALS) can be scaled with robust results. DEMs capture fine-scale elevation data, enabling the modelling of hydrological features such as flow accumulation (Jenson & Domingue 1988; Moore et al. 1991). However, these models are not without limitations: they tend to misclassify depressions, often miss ditches due to their placement in wetlands or flat terrain, and may require extensive preprocessing (e.g., stream burning, breaching) that introduces further uncertainties, especially at road crossings (Lidberg et al. 2017).

The field of artificial intelligence has shown promise in overcoming these limitations. This is a broad term encompassing, among other things, efforts to make machines perform tasks that require cognitive capabilities, such as reasoning, learning, and problem-solving (Russell & Norvig 2021). Using it, complex, time-demanding tasks can be automated in many cases at a lower cost. While this alone is not sufficient to classify a machine as an autonomous intelligent agent in the same sense as humans (Korteling et al. 2021), the automation of such tasks has many useful applications in many research areas (Pham & Pham 1999; Hamet & Tremblay 2017; As et al. 2018; De Almeida et al. 2019).

Machine Learning (ML) is a subfield of artificial intelligence (Shinde & Shah 2018). It involves using an algorithm that learns directly from the data by applying statistical methods to address various tasks. Within the environmental and ecological context, it has applications in species distribution (Pasha & Reddy 2024), forest health assessment (Estrada et al. 2023), landslide susceptibility (Merghadi et al. 2020), and rainfall prediction (Barrera-Animas et al. 2022), among others. For water channels, it was used in river ice mapping (Han et al. 2024), streamflow (Szczepanek 2022), evaluation of water quality (Khoi et al. 2022), and reach classification (Guillon et al. 2020; Olusola et al. 2022).

Deep Learning (DL), a subset of ML, was inspired by the structure of the human brain, using artificial neural networks to learn hierarchical representations of data. It started with the creation of neuron models (McCulloch & Pitts 1943), until reaching multi-layer perceptrons trained by backpropagation (Rumelhart et al. 1986). Within it, convolutional neural networks (CNN; LeCun et al. 1989) have several applications in forestry, such as delineating tree crowns (Ball et al. 2023), mapping biomass (Fu et al. 2024), and species identification (Zhang et al. 2022). In Earth Sciences, it has been used for the analysis of mineral resource distribution (Li et al. 2024), mapping volcanic and glacial landforms (Kazemi Garajeh et al. 2022), and simultaneous earthquake detection (Mousavi et al. 2020).

Surface waters have been mapped with satellite data and DL before (Isikdogan et al. 2017; Jiang et al. 2018; Fei et al. 2022; Mazhar et al. 2022; Thirumalraj et al. 2023). However, detecting small channels remains challenging because satellite imagery typically has a spatial resolution of 10–50 m. Orthophotos (Lantmäteriet 2021) provide much higher resolution, but in Sweden, dense tree cover often hides the terrain and the channels underneath it. High-resolution digital elevation models (DEMs) derived from ALS (Lantmäteriet 2022) address this limitation: with resolutions of 0.10–2 m, they make it possible to filter out vegetation and focus on the underlying terrain.

Small-scale channels have been extracted from topographic indices and remote sensing data before (Koski et al. 2023; Du et al. 2024) using U-Net (Ronneberger et al. 2015), a CNN architecture. It was used in Sweden, too, where Lidberg et al. (2023) increased the mapping of ditches from 9% to 86% using ALS-derived data. Given these developments, there is a growing opportunity to integrate machine learning tools to not only detect channels

but also to classify them based on geomorphological and hydrological attributes. Factors such as stream slope, length, and catchment area could aid in distinguishing ditches from natural streams, which is crucial for both regulatory compliance and ecological restoration.

By advancing the methodology to map and classify small water channels in forested landscapes, this study aims to support better integration of scientific data into environmental policy and land management, contributing to the restoration and protection goals of both national and international environmental agendas.

2. Research objectives

The main goal of this thesis was to improve the automatic mapping and classification of small water channels using data derived from high-resolution DEMs.

- Identifying the best settings to map channels using U-Net based on individual and combined topographic indices, with different dataset setups (using only ditches, only streams, ditches and streams combined as "channels", and ditches and streams separated) (Study I)
- Improve the classification of the detected channels from Study I with eXtreme Gradient Boosting (XGBoost), removing false positives (FPs) and increasing the amount of stream channels correctly classified (Study II)
- Comparing the performance of different methods to measure uncertainty when classifying pixels with concrete dropout (Study III)
- Determining if there is any location-specific variability in the U-Net performance across Sweden (Study IV)

"'They're pretty high mountains,' said Azhural, his voice now edged with doubt.
'Slopes go up, slopes go down,' said M'bu gnomically.
'That's true,' said Azhural. 'Like, on *average*, it's flat all the way.'"

– Terry Pratchett, *Moving Pictures*

3. Materials and Methods

3.1 Study areas

The twelve study areas used to train and test the models are distributed across Sweden (Figure 5), a Scandinavian country with an area of 450.295 km^2. 55% of its territory is covered in forests, mostly boreal in the northern and central areas (Diekmann 1999). On a smaller scale, deciduous forests are found in the south, where the fertile plains are also located, while peatlands and wetlands are found in the central and northern areas (Sjörs 1999). The sites were selected to be 1) mainly forested areas (86-99% forest cover (Busarello et al. 2025)); 2) as diverse as possible when it came to runoff conditions, soil type, topographic variation, and tree species; and 3) within the constraints of areas with available higher resolution LiDAR cover, which, at the time, was limited to a few areas of the country.

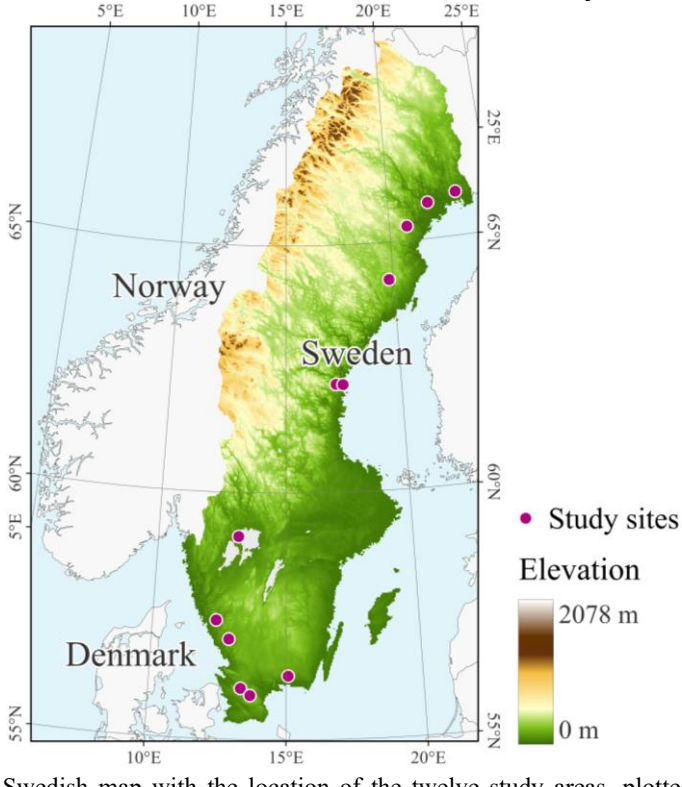


Figure 5. Swedish map with the location of the twelve study areas, plotted over the elevation.

3.2 Data collection

Using orthophotos and terrain visualization techniques, experts digitally mapped small water channels (<6 m width) across the study areas' landscape to create the data used to train the models. Ditches were visually identified from terrain indices and high-resolution (0.17-0.5 m) orthophotos (Lantmäteriet 2021)(Figure 6A) and traced as vector data. After the channel heads (> 2 ha) were located and the connections between the stream and ditch network marked, their downstream stream paths were also manually edited. In total, the dataset had 2235 km of ditches and 335 km of natural streams.

In all Studies, the indices that aided the mapping and classification were derived from the 0.5 m resolution DEM, obtained from the ALS with 1-2 points per square meter (Lantmäteriet 2022)(Figure 6B). All indices and their applications are listed in Section 3.2.1.

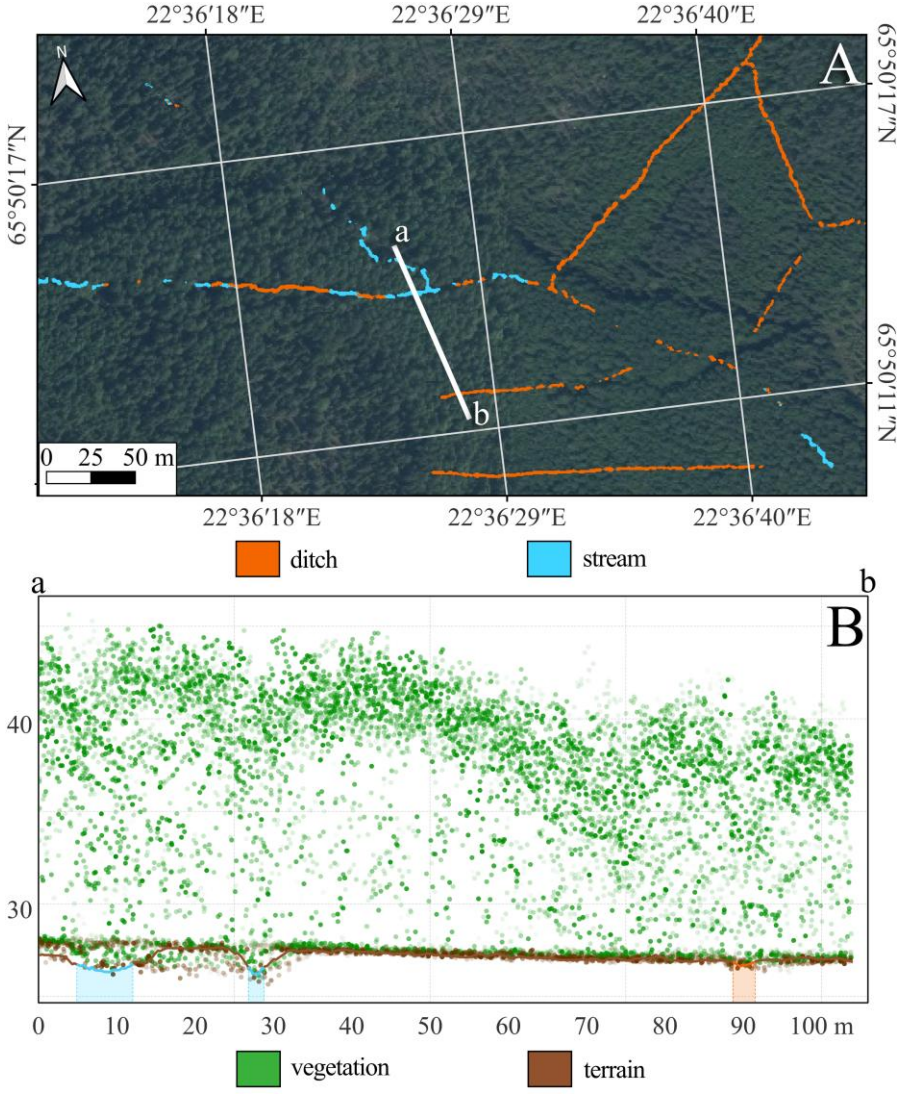


Figure 6. (A) An example of an orthophoto from Lantmäteriet (2021) and the channel prediction from the inference output of Study IV, both plotted in 0.5 m resolution, and with a cross-section a-b. (B) LiDAR point cloud from the cross-section a-b, plotted with data from Lantmäteriet (2022). The cross-section cuts through two streams (in turquoise) and a ditch (in orange), highlighted on the elevation surface. The opacity of the data points represents the distance from the cross-section (maximum 20 m), with distant points having a higher transparency.

The laser data from Lantmäteriet (2022) were organized as square tiles with a side of 2500 m (Figure 7). These tiles were further divided into chips with

dimensions of 250 m x 250 m (Figure 7B) to be used by the ML models. In Study I, 80% of the chips were randomly selected for training, while the 20% remaining were used for testing. In Study II, chips from eight of the study areas randomly selected were used to retrain a U-Net model, with chips from the four remaining areas used to train the XGBoost model. In Study III and IV, the chips were divided into nine folds for training, plus one for testing and one for calibration, applying stratified sampling to preserve the representativeness of all study areas.

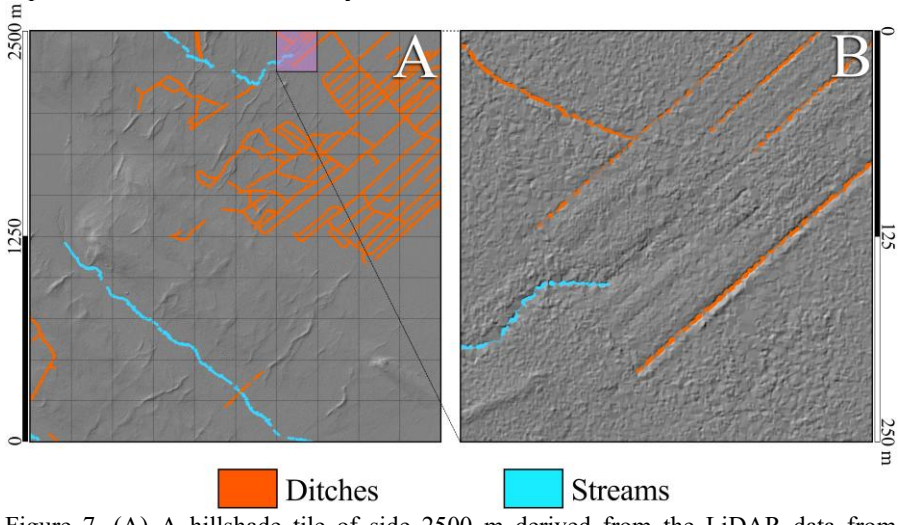


Figure 7. (A) A hillshade tile of side 2500 m derived from the LiDAR data from Lantmäteriet (2022). The tile is further divided into 100 chips, each with a side of 250 m, as shown in (B).

In Study IV, the U-Net model was used for a national-scale prediction. To capture the Swedish landscape variability on such a large area, the independent National Inventory of Landscapes in Sweden (NILS; Ståhl et al. 2011)(Figure 8) was used for evaluation. In it, 631 line transects are distributed across Sweden's different land cover areas, including forests and wetlands, providing reliable national estimates. In the study area, 6 576 channels were analyzed.

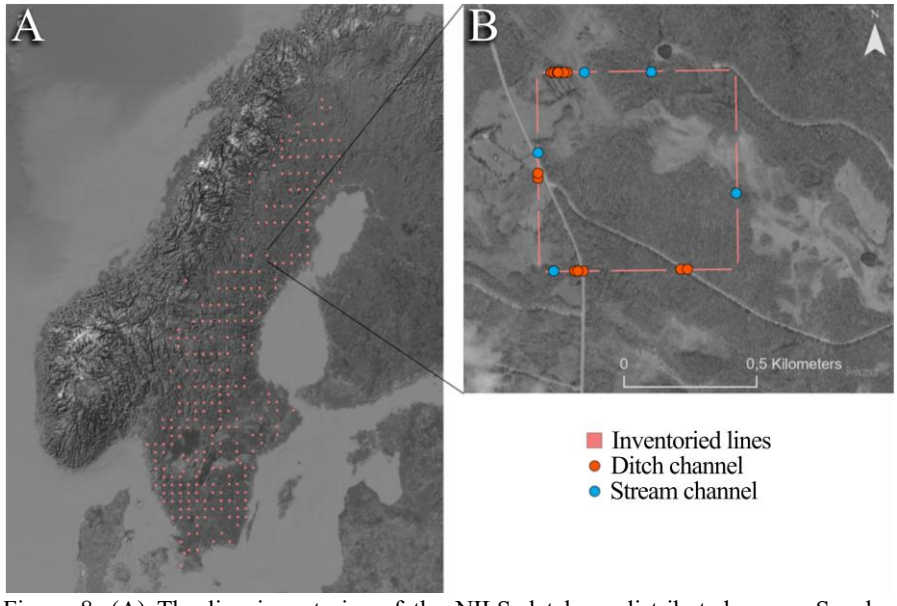


Figure 8. (A) The line inventories of the NILS database distributed across Sweden (n=631). (B) Example of a single line inventory and its small ditches and streams observed.

The channels from the U-Net did not overlap perfectly with the NILS database due to GPS uncertainties, so we used the snap function with a 25 m radius to relocate channel observations to predicted channels.

3.2.1 Topographic and hydrological indices

Topographic indices were derived from DEMs to describe terrain characteristics and explain how the topography influences landforms, erosion, and water flow, among others. Hydrology indices are a part of them, but focus on describing hydrological processes by encompassing water dynamics to characterize different components of the flow regime (Olden & Poff 2003). In this work, most of the indices were calculated using Whitebox Tools (Lindsay 2016), an open-source geospatial analysis library for python used for GIS and remote sensing applications. The only exception is the sky-view factor, obtained using the Relief Visualization Toolbox (Zakšek et al. 2011), a relief visualization package for geospatial analysis.

In this thesis, we detected the water channels using only topographic indices to train the U-Net models. Furthermore, in Study II, we built on those results by combining the detected channel network with hydrological indices

and morphology to improve the classification between ditches and streams. Below, we list the most relevant indices used in our work (Figure 9).

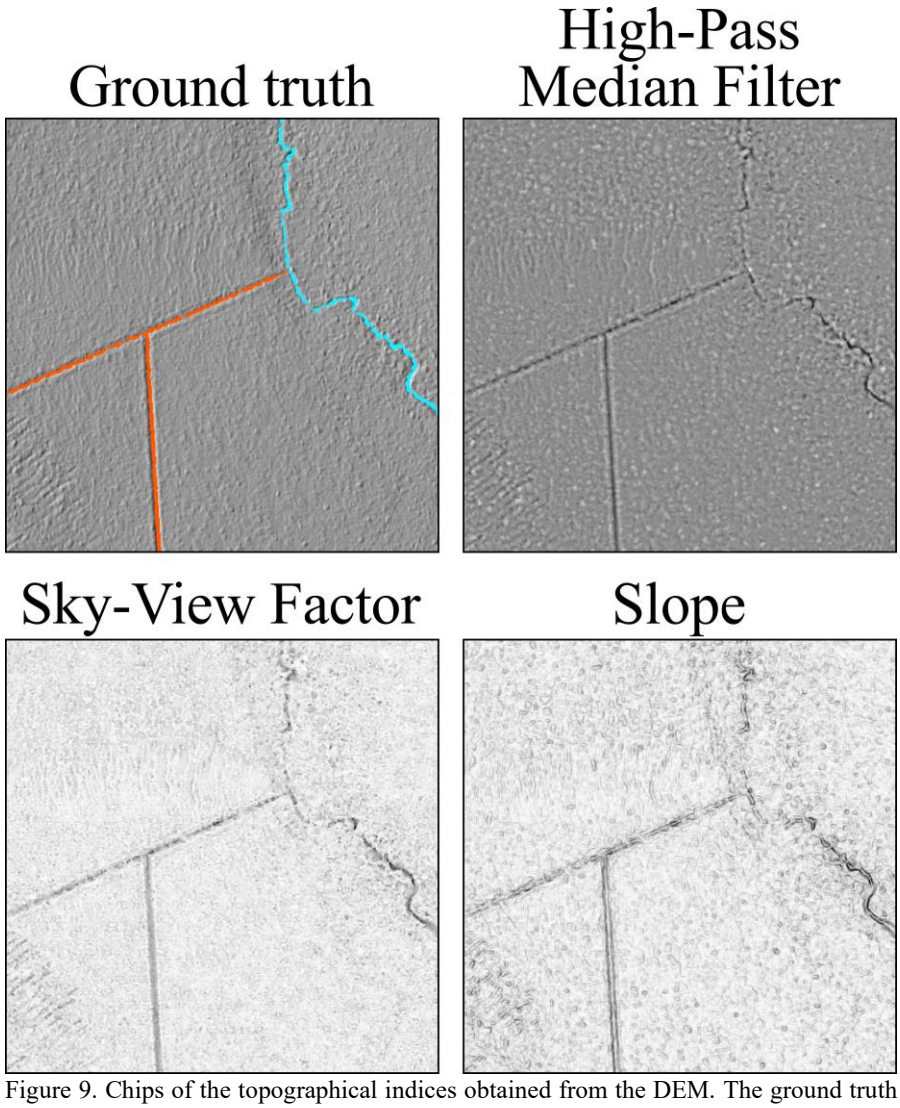


Figure 9. Chips of the topographical indices obtained from the DEM. The ground truth is plotted over the hillshade at 90°.

High-Pass Median Filter

The High-Pass Median Filter (HPMF) highlights short-range variability in the elevation by a moving window that returns the difference between the

pixel value and the median value of the pixels around it within the window. The window size used in this study is 11x11 pixels, which in our resolution means 5.5 m x 5.5 m. This index was combined with the vector lines from the ground truth to create the labels to train the U-Net model used in our studies.

Sky-View Factor

The sky-view factor is a terrain visualization technique that shows the portion of the sky visible from each DEM cell, accounting for obstructions caused by the surrounding terrain. Used in Studies I, II, and IV, combined with the slope.

Slope

Slope is a terrain analysis that calculates the rate of maximum change in elevation for each cell of a DEM, the steepness. It ranges from 0° to 90°, where 0° is a completely flat terrain, and 90° is a vertical cliff. It was calculated by a polynomial fit of the elevation using a 5x5 cell window (in this case, 2.5 m x 2.5 m) for a more robust result. Used in Study III by itself and combined with the sky-view factor in Studies I, II, and IV.

Flow Accumulation

The flow accumulation grid was calculated using the single-flow-direction method with the D8 algorithm (O'Callaghan & Mark 1984), without flow divergence. In this approach, water from each cell flows entirely into a single downslope neighbouring cell.

Average Flowpath Slope

This tool calculates the slope steepness of the flowpaths passing through each cell of the DEM.

Average Upslope Flowpath Length

This index is the calculation of the average length of the upslope flowpaths that pass through each cell of the DEM.

Maximum Upslope Flowpath Length

This index is the same as the above, but instead it calculates the maximum length of the upslope flowpaths that pass through each DEM cell.

Upslope Depression Storage

Returns the average upslope depression depth, with smooth terrain having lower values than rough terrain. It first calculates the upslope depth of the depression storage, then divides it by the number of upslope cells. Uses the FD8 flow algorithm (Freeman 1991) for the calculation.

3.2.2 Additional data

Other data that were not derived from the DEM were added to train the models in Study II. Sinuosity is an indicator of how straight a water channel is (Lazarus & Constantine 2013), being calculated by dividing the length of the channel by the distance between its start and end points. Data was also extracted from the Study I inference by counting the frequency of each channel class pixel within the channel buffer and obtaining which was the majority class.

3.2.3 Drainage index

For Study IV, knowing that deeper ditches have stronger drainage effects that decrease with distance, we calculated the drainage index. The ditch influence was modelled by calculating the logarithmic decay of the regression function described by Bring et al. (2022), resulting in:

$$Index = D - \left(\frac{D}{\ln(M+1)} \times \ln(d+1)\right)$$

D is the estimated ditch depth in meters, M is the maximum influence distance (150 m), and d is the distance from a pixel to the nearest ditch in meters.

3.3 Machine Learning approaches

3.3.1 Convolutional Neural Networks

CNNs (LeCun et al. 2015) are DL methods commonly used in computer vision tasks for spatial data, such as image segmentation, classification, and object detection. With one of their first applications being the recognition of handwritten numbers (LeCun et al. 1989), CNNs have remained relevant in

the fields of medicine (Polsinelli et al. 2020; Zuluaga-Gomez et al. 2021), climate change (Kim et al. 2022; Elshewey et al. 2025), chemistry (Derry et al. 2023), traffic flow forecasting (Sun et al. 2020), and several others.

These models offer an advantage by automatically extracting hierarchical features from images and requiring fewer training instances than standard vision transformers. Architecturally, CNNs consist of a sequence of convolutional and pooling layers, which transform the input representation several times before reaching one or more fully connected layers (Figure 10). Within a convolutional layer (Figure 11), learnable filters (kernels) are applied across the input image to detect basic visual features, such as edges and corners. The resulting feature maps encode the spatial presence of these patterns, where a non-linear activation function is subsequently applied, allowing the network to model complex, non-linear relationships within the data. Pooling layers then perform downsampling operations, reducing the spatial dimensions of the feature maps while retaining important information. This step enhances computational efficiency and improves robustness to local variations and distortions.

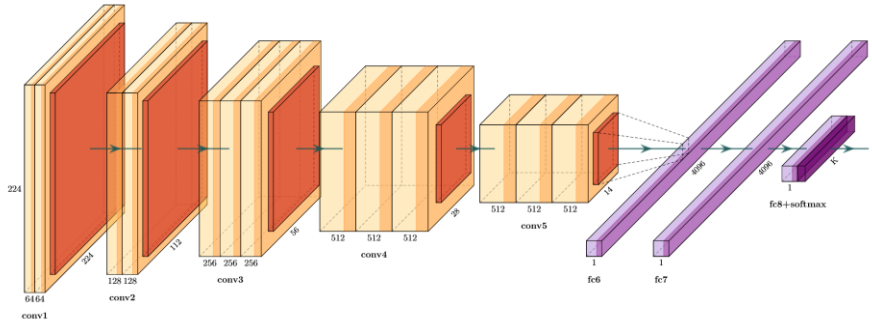


Figure 10. Example of a VGG-16 convolutional neural network (Simonyan & Zisserman 2015), with the convolutional layers in orange, and the three fully-connected layers in purple. The third fully connected layer performs the classification, with a final softmax layer computing the class probabilities. Image from https://github.com/HarisIqbal88/PlotNeuralNet.

Through repeated convolution and pooling operations, CNNs progressively capture higher-level, abstract features. Ultimately, the multidimensional feature maps are flattened and passed into fully connected layers, where the extracted features are combined to classify previously unseen data.

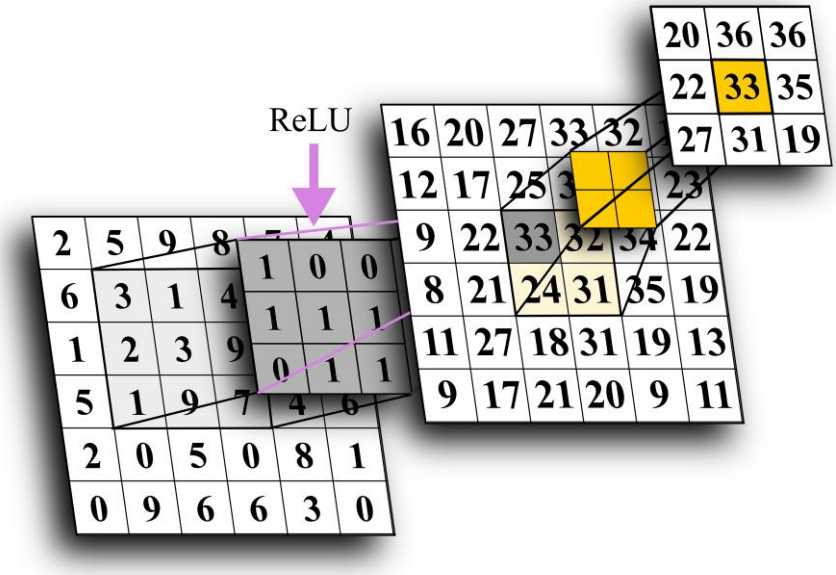


Figure 11. A convolutional layer. In the example, we assume zeroes fill the cells surrounding the input layer (i.e., zero padding). The 3x3 learning filter, in grey, is composed of trainable weights, the values within the cells. It moves through the input layer, cell by cell, multiplying the elements that overlap and adding their values. Then, an activation function, in purple, is applied to produce values for the feature map created by this filter. In this example, we use ReLU (Rectified Linear Unit), which does not change positive values and converts negative ones to zero. After this, a 2x2 maxpooling window, in yellow, passes through the cells, outputting the maximum value from within the window to the final output, which has reduced dimensions.

3.3.2 U-Net

U-Net (Ronneberger et al. 2015) is a CNN model known for its U-shape (Figure 12) that comes from its downsampling encoding path and the upsampling decoding path. The encoding path works similarly to the CNN described before, pooling after each convolution and extracting relevant features, increasing the number of feature channels. After that, the decoding path performs transposed convolutions, reconstructing the original spatial resolution and classifying pixels. At each level, skip connections link the encoder and decoder paths, transferring feature maps to preserve spatial details that may be lost during downsampling. In Study I, standard dropout is used after the first convolutions in each block, randomly setting a fraction of activations to zero during training, as specified by the dropout rate. In

Study III, some changes were made in the U-Net architecture following Teng et al. (2023), together with choosing specific angles for the data augmentation translation (0°, 90°, 180°, and 270°). The model was retrained using only the slope for a shorter processing time. In Study IV, the model with the architecture from Study III was retrained too, this time using sky-view factor and the slope.

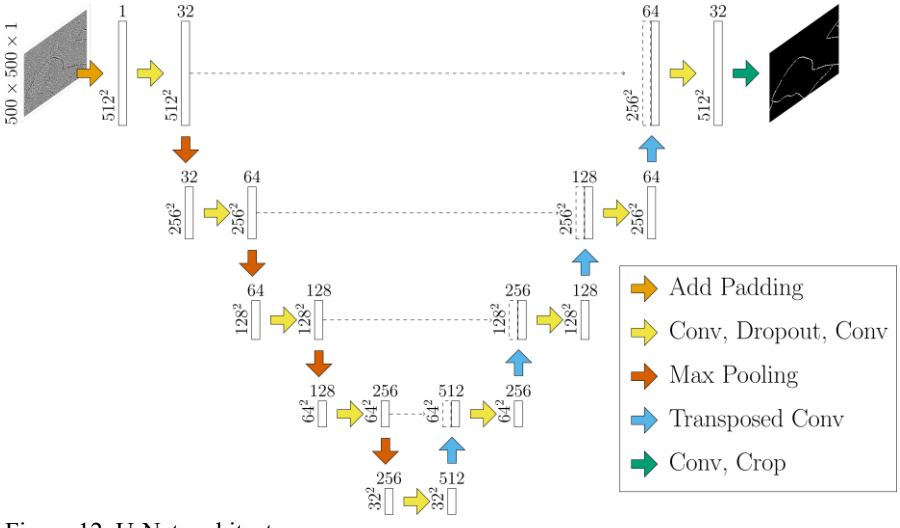


Figure 12. U-Net architecture.

3.3.3 XGBoost

XGBoost (Chen & Guestrin 2016) works differently from CNNs, not relying on spatial structure. It is an ensemble learning method that implements the boosting framework developed by Friedman (2001) for gradient-boosted decision trees. Decision trees (Figure 13) are models that make predictions by splitting data at nodes into branches based on feature values. The splitting criteria vary depending on the algorithm: in standard decision trees, common measures include entropy or Gini impurity, while XGBoost uses a regularized gain function based on second-order gradient statistics. Splitting stops when the maximum tree depth is reached, when each leaf contains fewer than a minimum number of samples, or, in the case of XGBoost, when the gain from a potential split falls below a predefined threshold. Boosting, which is one ensemble learning approach, combines multiple models named

"weak learners" to create a stronger one. The models are developed in sequence, with each new one correcting the mistakes of the previous ones.

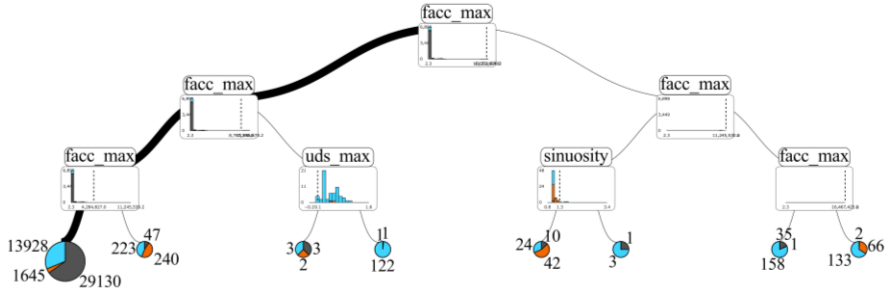


Figure 13. The final decision tree from the XGBoost training model. In the example, the maximum flow accumulation (facc_max) was used for the first decision split (node) by a high threshold, with most samples split to the left side (branch) and only a few to the right. Next, different lower thresholds are set for maximum flow accumulation again, further separating the samples. In the final split layer, new thresholds of maximum flow accumulation, maximum upslope depression storage (uds_max), and sinuosity are used for the final classification (leaves). Ideally, most of the channels in a leaf after the final division would be of the same type, something that can be observed in some leaves that contain mostly streams. However, some leaves still show an even split between channel types.

3.3.4 Feature Conformal Prediction

In Study III, several analyses were conducted to evaluate the model uncertainty. Instead of using the standard dropout in the U-Net model, concrete dropout was used, and the dropout rate was learned during training. During the inference, dropout was kept active with the adoption of Monte Carlo dropout (Gal & Ghahramani 2016): multiple stochastic forward passes were run, with different units dropped at each run. At the end, the predictions were aggregated by computing the mean and variance across the runs, building uncertainty maps to produce more robust predictions.

The Feature Conformal Prediction (FCP; Teng et al. 2023) quantifies the uncertainty of a neural network by estimating the range of output values that would include the correct value with a specific probability, such as 90%. It works by recording the feature representations produced by a chosen layer of the network for all instances in a calibration set, which can be the convolutional layer just before the output in a U-Net, for example. For each instance, FCP measures the distance between the recorded feature

representation and the minimal feature representation that would give the correct output. These distances are then sorted, and a percentile (e.g., the 90th) is selected to determine the maximum allowable deviation from the original feature representation that still guarantees the correct output. When evaluating a new instance, FCP varies its feature representation within this calculated distance and records the resulting changes in the network's output. This process identifies the highest and lowest possible outputs consistent with the calibrated distance. The difference between these values serves as an indication of uncertainty: larger differences indicate higher uncertainty, while small differences indicate greater confidence in the prediction.

3.4 Evaluation

The U-Net evaluation of model performance was made on the pixel level, i.e., how many pixels were correctly classified when comparing the inference to the ground truth. The metrics used to verify the performance are derived from the confusion matrix, a table that assesses not only how many instances were correctly classified by the model, but also how many were incorrect, and which class these instances were predicted to have instead. The class that we want to evaluate is the positive one, while the other is the negative one. Instances that are correctly predicted as positive are named true positives (TPs). Those that are correctly predicted as being negative are named true negatives (TNs). Those that are incorrectly predicted to be positive, false positives (FPs). And those that were incorrectly predicted to be negative, false negatives (FNs). These values are then used to calculate other relevant metrics that better illustrate the model's performance, such as recall, precision, F1-score, and the Matthew's Correlation Coefficient (Matthews 1975).

Recall estimates the number of true positive instances predicted by the model from all ground truth positive instances (TP and FN combined), calculated by the ratio:

$$recall = \frac{TP}{TP + FN}$$

Precision is the number of true positive instances predicted by the model from all those predicted to be positive (TP and FP combined), calculated with:

$$precision = \frac{TP}{TP + FP}$$

The F-1 score is the harmonic mean of precision and recall, returning a number that shows the balance between them, obtained with:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

The MCC is also known as the phi-coefficient, a metric that measures the correlation between the predicted labels and ground truth, and is considered more reliable for very imbalanced datasets (Chicco & Jurman 2020).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

In this thesis, we have a multiclass dataset with three classes (background, ditch, and stream), which required a more complex 3x3 confusion matrix (Figure 14). With a multiclass dataset, the metric needs to be calculated for each class separately, which is done by considering the two other classes as negative and combining their values.

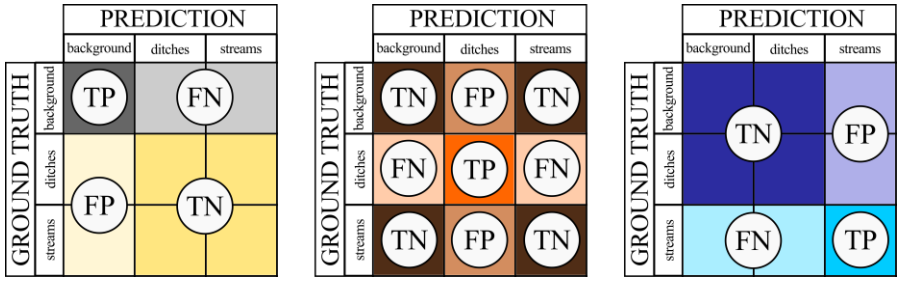


Figure 14. Confusion matrix structure for each channel class. In grey and yellow, "background" is the positive class. In brown and orange, the positive class is "ditch", and in blue and purple, the positive class is "stream".

For XGBoost in Study II, we applied the same metrics but evaluated channel segments instead of pixels, counting the number of instances correctly classified when comparing the model inference to the ground truth polylines. To quantify the impact of each index on the model prediction, we have used

SHAP plots (Zhang et al. 2023). These are plots that show how each attribute impacted the model classification, favoring one class or another, and how much. This way, we can track which features contributed the most to the model's decision.

For Study IV, we also compared the classes observed in the NILS database survey with the channel classes predicted by the U-Net retrained model, then calculated the evaluation metrics.

3.5 Workflow

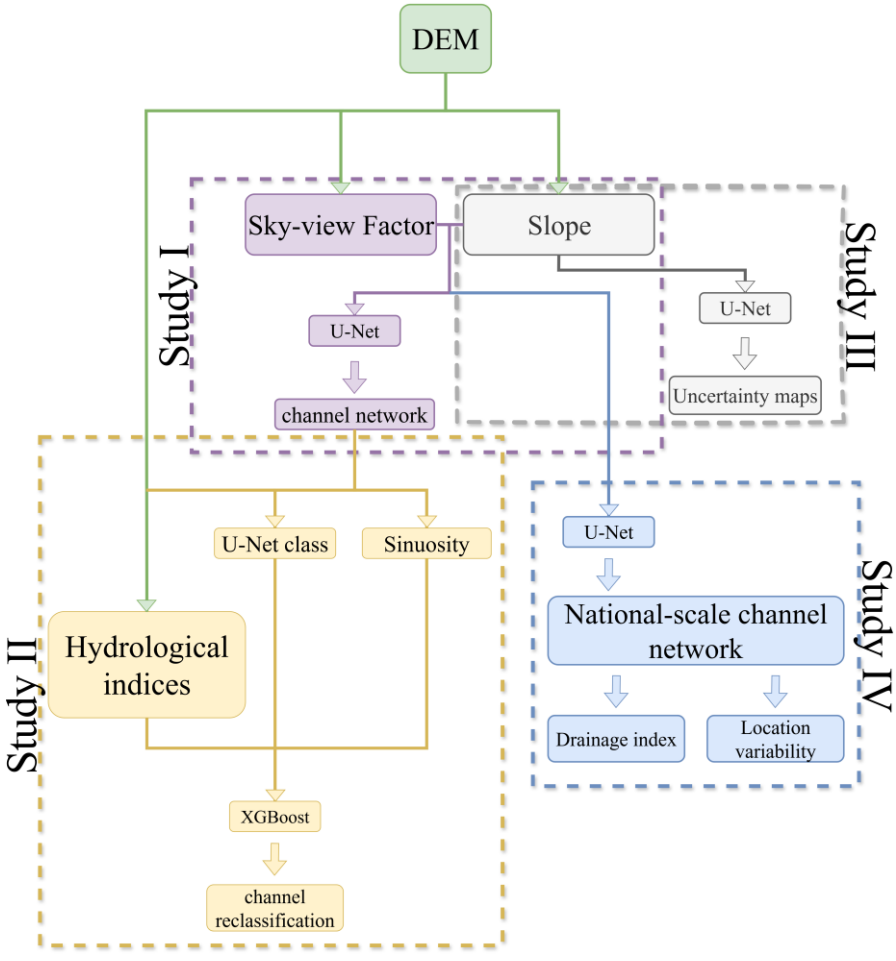


Figure 15. Combined workflow for the four articles. In purple, the steps involved in Study I; in yellow, the steps involved in Study II; in light grey, the steps involved in Study III; and in blue, the steps involved in Study IV. The dashed boxes highlight the steps corresponding to each study.

All studies used topographic indices to train a DL (U-Net) model for mapping water channels at the pixel level. Study II combined the output from the highest-ranking model from Study I with hydrological indices to train an ML (XGBoost) model, improving channel classification. Study III extended the methodology with a new U-Net architecture trained with only slope to evaluate different uncertainty quantification approaches. Study IV retrained

the model from Study III using the highest-ranking combination (sky-view factor and slope) and obtained a national-scale prediction of channels. A national drainage index was calculated, and the location-variability was assessed.

“May whatever tests await you on the other side either support or disprove your hypotheses.”

– GLaDOS

4. Summary of results and discussion

4.1 Channel detection and classification

In Study I, the highest-ranking model was the one combining the sky-view factor and slope data in the training process. 89.7% of the ditches were detected using it, but 6.2% were incorrectly classified as streams. 75.5% of the stream channels were detected by the model too, with 15.8% of the streams being incorrectly classified as ditches. The MCC was 0.74 for ditches and 0.31 for streams.

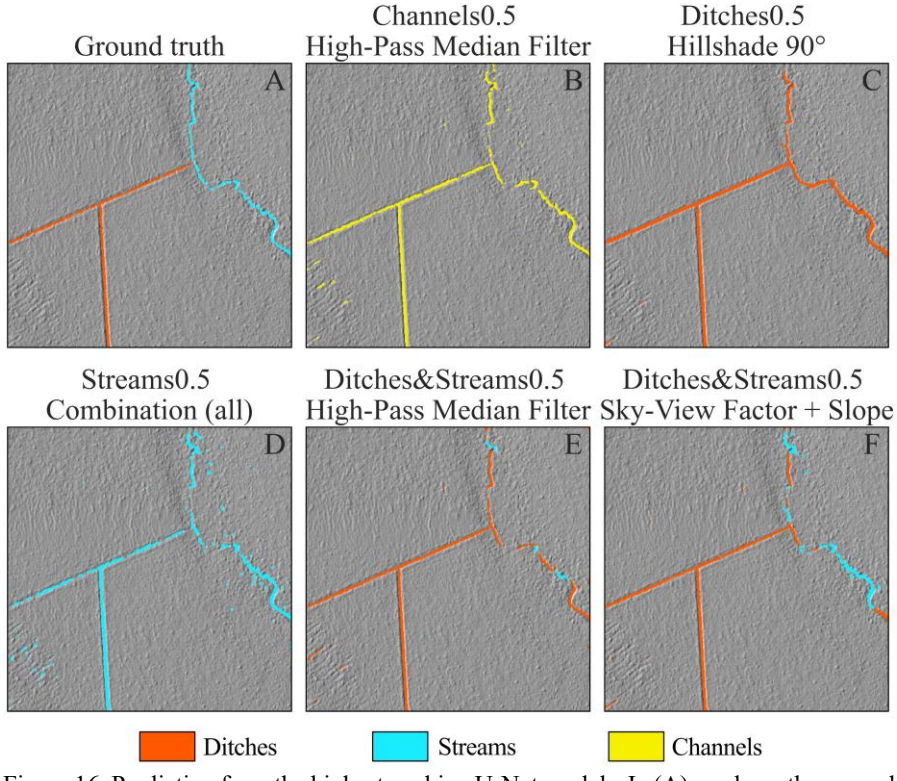


Figure 16. Prediction from the highest-ranking U-Net models. In (A), we have the ground truth; (B) is the output of a model trained without a channel type specified; (C) is the output of using only ditches in the training data; (D) is the output of using only streams in the training data; (E) is the output of using ditches and streams for training with the

HPMF; and (F) is the output of the model trained using ditches and streams with sky-view factor and slope.

When converting these detected channels from raster to vector data and reclassifying them with the hybrid model and hydrological data (Study II), a new evaluation was made on the U-Net performance for channel segments of several lengths (Figure 17). The U-Net model demonstrated a high recall rate for ditch channels (79%) but with low precision (5%). In contrast, it showed both low recall (8%) and precision (8%) for stream channels. When XGBoost was applied to reclassify the channels, the precision for ditches improved substantially to 50%, although recall decreased to 63%. For stream channels, both recall (71%) and precision (52%) increased. The background class was reported only in the hybrid plot, as it represents FPs from the U-Net model that were reclassified by XGBoost. This class showed strong performance, with a recall of 79% and a precision of 88%.

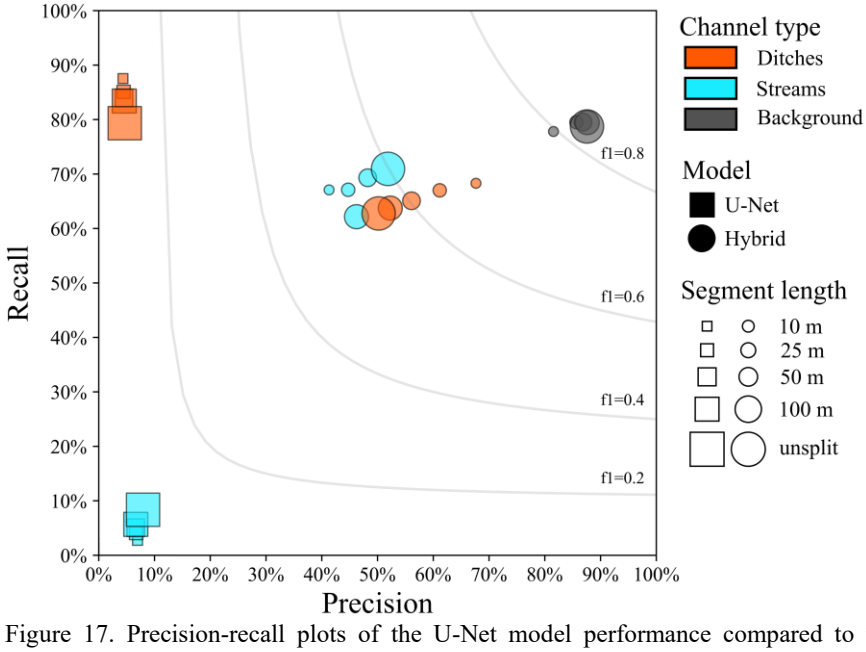


Figure 17. Precision-recall plots of the U-Net model performance compared to the ground truth (squares), and the XGBoost model performance compared to the ground

truth (circles) across the different segment lengths. The grey lines are iso-F1 lines, placing the trained models in a performance area according to their F1-scores.

A visual analysis showed that a substantial number of FPs were now properly classified as background and could be removed from the maps (Figure 18), illustrating that the channels detected in Study I were successfully reclassified as stream TPs and FPs (background TPs).

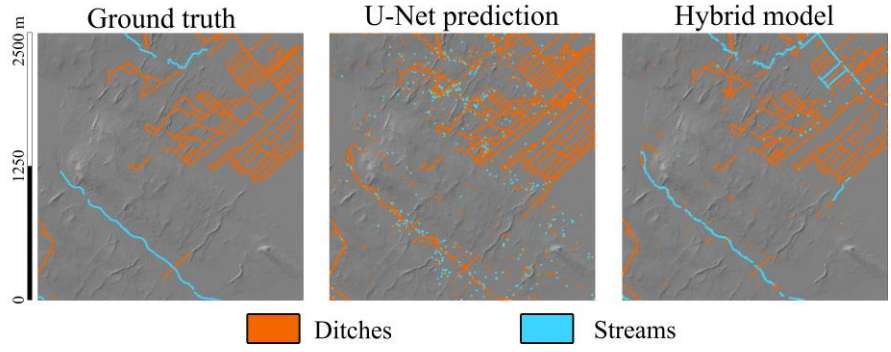


Figure 18. Prediction of the hybrid model compared to the ground truth and U-Net. On the left is the ground truth data, which was manually labeled. The center shows the predicted channels from the original U-Net model, and on the right are the predicted channels after post-processing by the hybrid model.

In Study III, we found that FCP provided the most reliable uncertainty estimates despite the higher execution time, and network probability was the best for correcting misclassified pixels. Visual inspection (Figure 19) shows that the top 5% most uncertain background pixels still outline a potential channel, even though the model did not classify them as such.

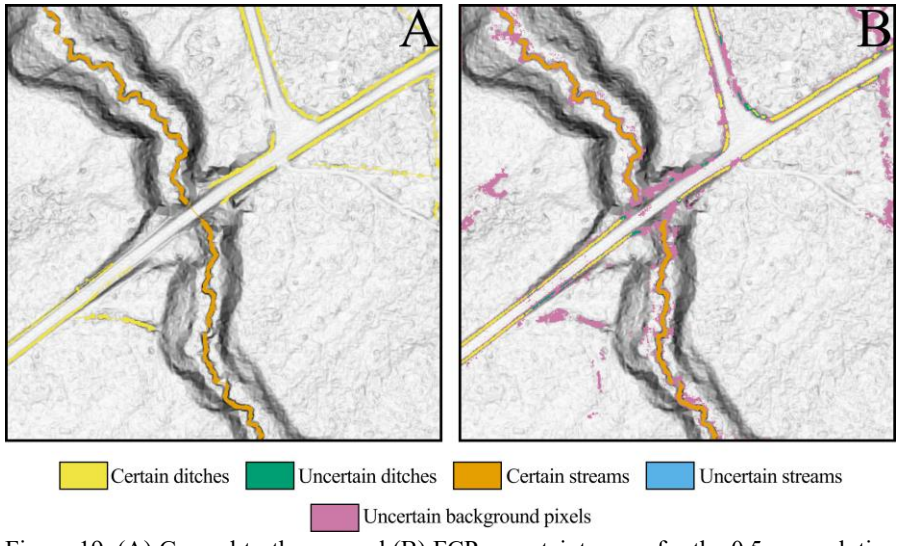


Figure 19. (A) Ground truth map and (B) FCP uncertainty map for the 0.5 m resolution over the slope index. The 5% most uncertain pixels from the uncertainty quantification approaches are plotted. Wherever the slope is visible, it represents certain background pixels. Image credits: Westphal et al. (2025).

In Study IV, the model predicted 1 153 749 km of ditch channels. With a precision of 88%, we estimate that approximately 1 015 299 km of the predicted ditches represent actual ditch channels. This is about a 20% increase compared to the previous estimate by Laudon et al. (2022), who estimated a total of 1.2 million kilometres of channels, of which 67% (about 800,000 km) were considered ditch channels. Further, 145 646 km of streams were mapped, of which 56% (81 562 km) were estimated to be actual stream channels. The model performance in Study IV improved compared to the overprediction of the ditch channels in Study I. The recall for the ditch label remained high (76%), while precision increased from moderate to high (57.6% to 88%). For streams, performance also improved: the low precision (16.4%) from Study I increased to moderate values (56%) despite a small reduction in recall (59.7% to 58%). We have evaluated the new U-Net model with an independent test dataset and the NILS database (Table 1), illustrating the overall performance across the Swedish landscape. Ditches performed better than streams, yet both channel types still surpassed the Swedish topographical maps.

Table 1. Summary of the different models when evaluated on test data and topographical maps. The lines from the topographical maps were evaluated against both stream points and ditch points in the NILS database separately.

Evaluation method	Class	Recall	Precision	F1-score
Model evaluation on test data	Ditch	76%	88%	0.81
	Stream	58%	56%	0.56
NILS vs U-Net	Ditch	83%	88%	0.86
	Stream	26%	93%	0.41
NILS vs topographical maps	Ditch	11%	27%	0.16
	Stream	36%	21%	0.27

The confusion matrix for the NILS dataset compared to the U-Net model (Figure 20. Confusion matrix for Study IV's model compared to NILS channel observations. confirms the findings from the table, showing that ditches have the largest number of TPs and only a few of them were undetected (16%). Most of the streams, however, were still undetected, and part of them (20%) were incorrectly classified as ditches.

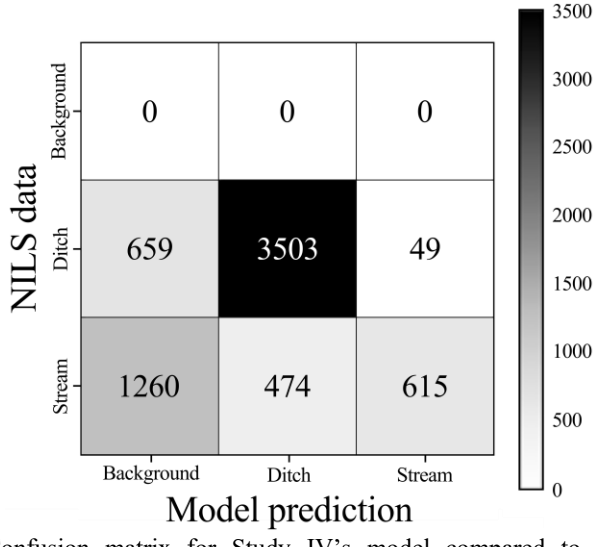


Figure 20. Confusion matrix for Study IV's model compared to NILS channel observations.

The distribution patterns of streams and ditches were different depending on the channel type. The northwestern region had the highest concentration of natural streams (Figure 21A), while fewer streams were found in the south.

Ditch channels were very few in the mountainous areas and were instead concentrated in the south and along the east coast (Figure 21B).

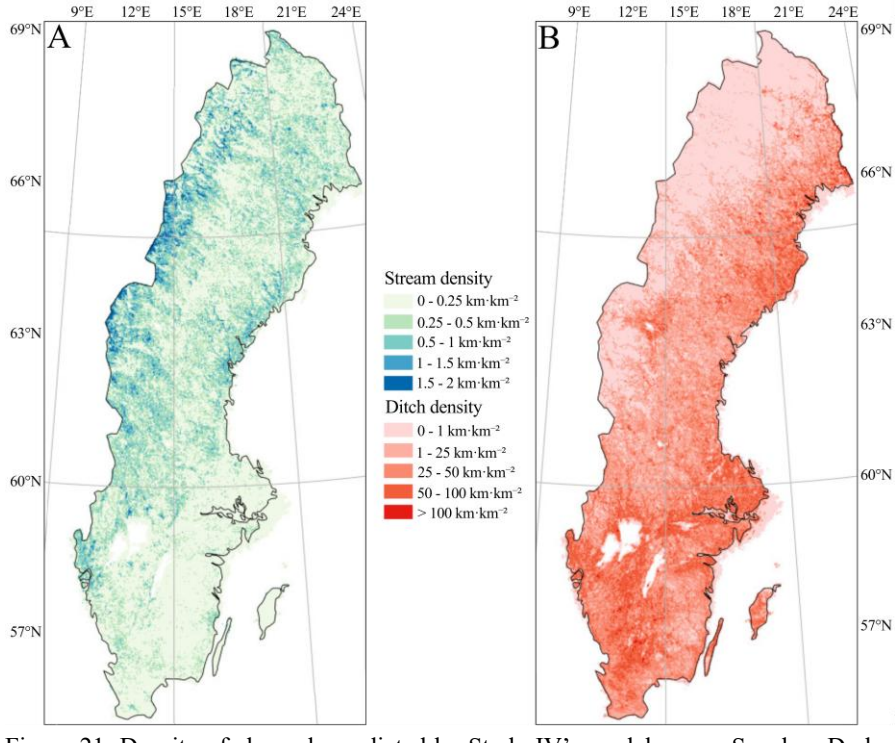


Figure 21. Density of channels predicted by Study IV's model across Sweden. Darker tones represent higher channel density. (A) Streams have been detected more often in the northwest. (B) Ditches were often detected around the east coast and southern Sweden.

Evaluating the drainage index, we could verify that deeper ditches exert a greater influence, which decreases logarithmically with distance (Figure 22). With this, we developed a depth-weighted drainage index to map the spatial influence of artificial drainage across Sweden. The index integrates estimated ditch depths from high-resolution DEMs with a logarithmic distance-decay function (maximum influence 150 m), producing a continuous surface of drainage impact. The results suggest that drainage effects on soil, greenhouse gas emissions, and vegetation are likely more widespread than previously recognized.

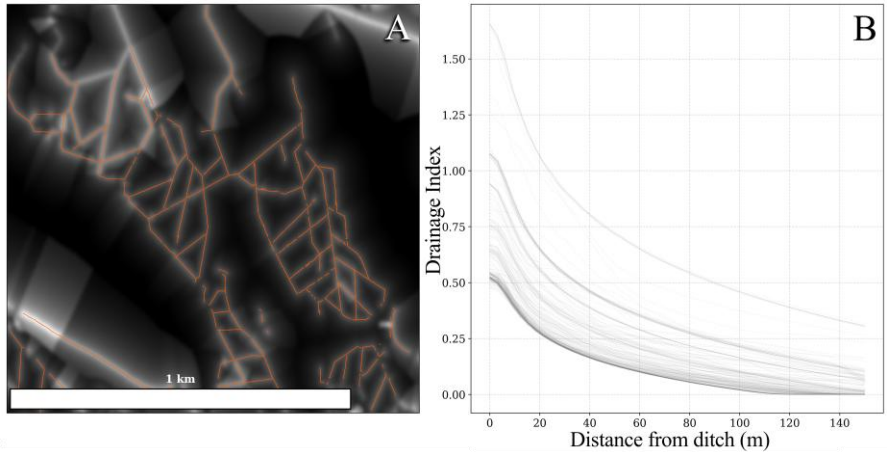


Figure 22. (A) Predicted ditch channels used to calculate the drainage index. (B) Deeper ditches exert a greater influence that decreases logarithmically with distance.

4.2 Methodology reflection

Ditches were successfully mapped using the U-Net across all studies in this thesis. Our work represents the first effort to map ditches and streams separately using high-resolution LiDAR data and ML on a national scale. Previous studies used a single index with DL to detect channels (Koski et al. 2023; Du et al. 2024), but the channel types were either not classified or were exclusively ditches. The highest performing model was trained on a combination of topographic indices (sky-view factor and slope), which agrees with some other studies where a combination of indices had a higher performance (Du et al. 2019; Kazimi et al. 2020), however, channels were not the only thing they were detecting, and they used a coarser resolution. Du et al. (2024) reported similar precision (88%) and higher recall (89%) compared to Study III; however, our model additionally detected small streams as separate features, making it more functional for management applications.

We have also assessed the number of "channel" pixels that were detected, i.e., how many pixels were not labelled as background. This was done to compare the performance of the DL methodology with traditional methods and available maps, since the channel network from flow accumulation would not differentiate between ditches and streams. We then compared these pixels with the ground truth to determine how many corresponded to

ditches and how many to streams, thereby establishing a benchmark (Table 2). We observed that many stream pixels were detected using flow accumulation alone; however, the ditches were not detected to the same extent. This was expected because flow accumulation represents the points in the landscape where water would converge, accumulating, meaning that there's an increased likelihood of it matching the location of natural streams. Ditches were mostly placed where soil drainage was required, which is why their locations do not necessarily match natural channels or the highest flow accumulation. Using only flow accumulation would result in an acceptable performance for detecting natural streams, but only with DL would one be able to detect ditches in the same output.

Table 2. Comparison between the performance of different models from Study I, the Swedish property map, and the traditional flow accumulation methodology. The percentage of stream pixels refers to how many stream pixels were detected, even if mislabeled as ditches. The Swedish property map and flow accumulation do not differentiate by channel type; these numbers were obtained by comparing these channels with our ground truth.

Method	Detected ditch pixels	Detected stream pixels
Swedish property map	8.1%	27.5%
Flow accumulation (2 ha)	33.8%	76%
Study I	89.7%	75.5%

Our exploration of alternative ML architectures led to improved channel reclassification. Using the hybrid model, the channels detected in Study I were reclassified into ditches and streams, resulting in more accurate classification. The stream morphology was initially considered one of the most relevant characteristics to classify channels, but the measures of sinuosity did not reflect this expectation. Most of the values were close to 1.0 for any channel type, meaning that they would be little meandering. With this, the contribution of sinuosity was heavily reduced, as can be seen in the absolute SHAP plot (Figure 23). Instead, the maximum flow accumulation was the most important index for classifying both ditches and streams, followed by the maximum average flowpath slope. This shows that catchment-level hydrological dynamics had a more important role in the channel classification.

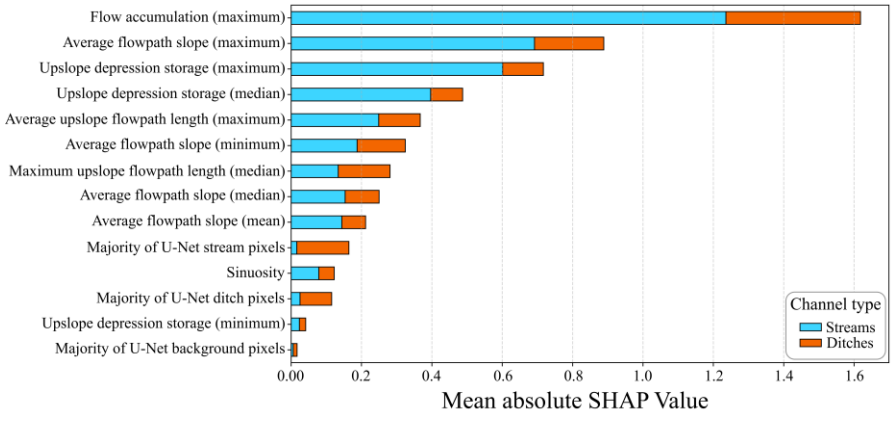


Figure 23. SHAP plot of mean absolute values for the data used in Study II, showing that the sinuosity had a low impact on the classification of channels.

In Study IV, we have enhanced the understanding of the extent and distribution of the Swedish drainage network. The location of ditches in agricultural areas and managed forests across the landscape follows their historical implementation for this purpose, with a density that surpassed 100 km/km^{-2}. The natural streams were concentrated in the northern boreal area, where their density peaked 2 km/km^{-2}.

4.3 Ecology, management, and policy

Because boreal forests are major carbon stores that are increasingly becoming carbon sources (Bradshaw & Warkentin 2015), their channel networks play an important role in regulating climate-related processes. Misclassifying streams as ditches (Figure 20) could lead to under-protection, despite ditches supporting their own distinct biodiversity, while failing to detect a stream may result in its disturbance during ditch cleaning or by forestry machinery. Such errors can impact flow patterns, reduce organic matter retention (Muotka et al. 2002), and increase downstream sedimentation (Bishop et al. 2009), with cumulative effects across the channel network.

In Study II, we addressed this issue by improving the initial channel classification produced by the model from Study I through a hybrid approach. One relevant challenge was selecting an appropriate segment length, as that could affect both classification accuracy and practical management decisions. Although uniform segment lengths simplify model

evaluation, natural channels rarely follow such regular patterns. We therefore tested multiple segment sizes and identified 50 m segments as the most balanced option, offering a compromise between model performance and operational feasibility for management planning.

Our models not only improved channel mapping but also provided the spatial and geomorphic context needed for catchment-scale restoration planning. Creating high-resolution, reliable maps is valuable not only for identifying drainage patterns and supporting ecological applications such as stream and riparian restoration (Baker et al. 2007; Gergel et al. 2007). With more accurately delineated channels, our results enable representative assessments of the condition and extent of ditches and natural streams across Sweden, guiding restoration priorities.

Stream restoration measures vary depending on project objectives and may include reconnecting floodplains, modifying flow regimes, or reconfiguring channels (Wohl et al. 2015). Some restoration initiatives have already been tested in Sweden, such as the demonstration project described by (Gardeström et al. 2013), where several methods were evaluated in a previously channelized wide stream. Similarly, Negishi & Richardson (2003) restored narrow boreal streams (<6 m) by increasing habitat heterogeneity with in-stream boulders. This intervention led to short-term improvements in detritivore productivity, but long-term monitoring is still required to evaluate long-term ecological restoration.

In Study IV, we quantified the extensive ditch network and high-resolution drainage density across the Swedish landscape, increasing the information available to support future analyses. In peatlands, the ditch-induced lowering of the groundwater table increases peat decomposition, leading to emissions of up to 7 Mtonnes CO_2eq per year in fertile forested soils (He et al. 2016). With our model outputs, these emissions can be better assessed to coordinate climate change mitigation plans. Rewetting after clear-cutting restores the water table (Bring et al. 2022), but an effective implementation relies on policymakers. Restoration effects also require post-restoration monitoring and management. Bring et al. (2022) showed that ditch blocking raised the groundwater levels near the blocked ditch, though the effect was halved after a 9 m distance, while the drainage effect persisted until 21 m before being halved. Moreover, these interventions on nutrient-rich peatlands can increase the export of DOC and nutrients, impacting the water quality (Koskinen et al. 2017). These findings highlight the need for

careful monitoring of interventions to ensure that rewetting achieves its intended ecological and climate benefits.

Laine et al. (2024) suggested converting forestry-drained, nutrient-rich peatlands into tree-covered pine or spruce mires with a sub-surface water level. Our fine-scale assessment of drainage density can guide this type of implementation, ensuring that restoration targets are both realistic and compatible with land-use priorities. This balance is particularly relevant given that these areas are also highly suitable for forestry, where climate mitigation and forest production goals often compete.

Achievable goals must be defined for restoration, however. While Agenda 2030 calls for the restoration of water systems, it leaves the methodology to policymakers. The overlap and misclassification between streams and ditches from our outputs reflect how these systems have converged morphologically and ecologically, emphasizing that full restoration to pre-alteration conditions may not be realistic. Most channels adapted to the changes over time: ditches developed their own communities (Williams et al. 2004; Verdonschot et al. 2011) and altered the assemblage composition of non-aquatic biodiversity in forests (Remm et al. 2013). Also, the knowledge about natural processes in channels across Fennoscandia before the disturbances is limited (Nilsson et al. 2015; Mason & Polvi 2023). Wohl et al. (2015) point out that restoration focused on channel physical connectivity can be highly detailed, even though it may successfully restore ecological function. Instead of focusing on resetting the channel to a historical classification, a more sustainable alternative would be to restore community function and foster a system robust to perturbation, reaching a more dynamic and less degraded ecological state (Palmer et al. 1997; 2005).

4.4 Limitations and future research

The national LiDAR dataset (Lantmäteriet 2022), which forms the foundation of our approach to mapping streams and ditches, was collected at a density of 1–2 points per square meter, from which a digital elevation model (DEM) with a 0.5 m resolution was derived. Future LiDAR acquisitions with potentially higher point densities and DEMs at decimeter-scale resolution are likely to improve the detection and differentiation of stream channels and ditches (Roelens et al. 2018; Song & Jung 2023). One limitation we faced with DL was the need for large, high-quality datasets

with ground truth data to train the models. It is common for published benchmarking datasets such as ImageNet (Deng et al. 2009) or Cityscapes (Cordts et al. 2016) to be massive, whereas creating custom labeled data is time-consuming. There is a possibility that, with more manually labeled data, our models could achieve even higher performance; however, Wang & Perez (2017) emphasize that increasing the dataset size does not necessarily improve a model. Also, the real-world data cannot be changed: only 14% of the channel length is natural streams. Studies III and IV demonstrated that adjustments to the architecture resulted in higher performance despite using the same dataset, suggesting that architecture design can play a greater role than data volume in some cases. Different weights during training could help improve the results at the cost of lower precision, with aggressive weights used for ditches and streams in the U-Net, and post-processing the results with a decision tree model, similar to what we have done in Study II, with improved performance.

Our models did not determine whether the channels contained water or not. High-resolution (0.8 m) multispectral remote sensing imagery has been used to map streams before (Leckie et al. 2005) with an average 80% accuracy. This methodology could be combined with our data to improve channel classification, although dissolved organic carbon also needs to be taken into consideration. For this, one discerning method for verifying water presence in the channel could be data gathering during different seasons (Islam et al. 2022). However, this would not be a helpful practice in Swedish forest streams because of the dominance of evergreen conifers, which would keep the forest cover on these channels all year round.

For Study II, the use of zonal statistics based on a fixed 3 m buffer provided valuable near-channel information but may have limited our ability to capture riparian zone characteristics. Expanding buffer analyses in future studies could provide additional ecological context and improve differentiation between channel types. Analyses that were initially performed on a smaller scale, for example, could now be applied to a larger extent. Take the buffer width estimation around natural streams. Kuglerová et al. (2020) used 111 Swedish small streams in their analysis. With our maps as input, the size of protective buffers could be automatically estimated using other remote sensing vegetation data (such as satellite images or LiDAR point clouds), creating a large-scale verification of whether the recommended buffer size is being followed or not.

With the uncertain FCP maps from Study III, we could verify that the presence of uncertain background pixels indirectly indicated undetected channels (Figure 19B, in pink). This illustrates the potential for improving the mapping of natural streams in the future using the highest uncertain background pixels as a proxy. In all our models, it was not uncommon to have interrupted stream segments in the prediction instead of a fully connected channel network. This could be caused by the LiDAR signal not reaching the ground in some areas with dense vegetation. A connection across the gap could be created to resemble traditional stream networks, but further studies are needed to quantify whether this would bring an improvement or new errors to the maps.

Another limitation is that the classification presented here is strictly related to the channel's origin, either artificial or natural. An automatic classification between stream types (pools, riffles, rapids, etc.) would shift the act of mapping from a functional task to an ecologically and geomorphologically based classification. More specific in-channel information could be added, such as soil type, channel bed, and sediment granulometry, providing more data about the water ecosystem.

While we produce maps of natural streams and ditches, we advise that this classification should not be trusted implicitly because errors are still present. For example, the model from Study IV still misclassified 20% of the stream observations as ditches, despite only 1% of the ditches being misclassified as streams (Figure 20). The streams predicted by the model in Study IV had a high F1-score compared to the topographical maps (Table 1). However, 74% were missing when compared to the NILS database. This highlights the challenges of mapping natural streams at a national scale and raises awareness when using our maps for management applications. From a legal perspective, to be certain that a channel is a ditch or a natural stream, the historical documentation (“vattenverksamhet”) needs to be consulted. However, these archives are from different agencies, regions, and landowners, making it unlikely to compile them into a national database. Our maps are useful to indicate “likely” ditches and streams. The drainage index is a valuable tool for restoration, carbon budgeting, and nutrient assessment, but it remains a model-based estimate unadjusted for soil and unvalidated in the field. Future work should include water table and biogeochemical measurements to improve its reliability across landscapes.

"Hiding the self through a faithful mapping of the universe is the only path to eternity."

– Cixin Liu, *The Dark Forest*

4.5 Conclusion

Despite certain limitations and opportunities for further improvement, this study demonstrates that the novel approach of using deep learning to map small channels in the landscape has been successful. While existing topographical maps fail to distinguish between ditches and natural streams, this research represents, to our knowledge, the first attempt globally to map and classify these channel types separately. We showed that this can be achieved by training separate models for ditches and streams. However, using a deep learning model to first detect all channels and subsequently classify them into ditches and streams based on channel characteristics with machine learning further improved performance.

The predicted ditch channels in Study IV showed both high precision and recall, achieving a higher F1-score (0.86) than the model implemented by Laudon et al. (2022)(0.71). This highlights the robustness of our model for large-scale ditch mapping and further analyses, such as the drainage index, a valuable tool to support future hydrological assessments, ecological studies, and landscape management decisions. However, 54% of the streams were still unmapped, and 20% were classified as ditches; hence, future research needs to focus on natural stream channels. The hybrid approach with XGBoost has yet to be implemented nationally, but so far, its use has resulted in the highest F1-score for streams (0.60). Combining it with the uncertainty analysis of background pixels could increase the number of mapped streams.

Furthermore, employing a deep learning model significantly reduces the time required compared to manual digitization. To train these machine learning models, 315 km of natural streams and 2 235 km of ditches were manually digitized across 12 study areas. Had we continued with manual digitization at the same pace and staffing levels, mapping the entirety of Sweden would have taken approximately 90 years, a task that could now be achieved within this four-year PhD project.

This work follows an operational mapping framework that can be continuously improved as new data becomes available. Beyond its scientific contribution, the resulting datasets and tools provide a guide for hydrological restoration, sustainable forest management, and national reporting under the goals of Agenda 2030 and the EU Nature Restoration Law. In doing so, this research connects environmental monitoring and decision support, offering a reproducible model for other boreal and temperate areas seeking to balance productivity with ecosystem resilience.

Credits: Mariana Dos Santos Toledo Busarello

References

Andréasson, J., Bergström, S., Carlsson, B., Graham, L.P. & Lindström, G. (2004). Hydrological Change – Climate Change Impact Simulations for Sweden. *AMBIO: A Journal of the Human Environment*, 33 (4), 228–234. https://doi.org/10.1579/0044-7447-33.4.228

Armitage, P.D., Szoszkiewicz, K., Blackburn, J.H. & Nesbitt, I. (2003). Ditch communities: a major contributor to floodplain biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 13 (2), 165–185. https://doi.org/10.1002/aqc.549

As, I., Pal, S. & Basu, P. (2018). Artificial intelligence in architecture: Generating conceptual design via deep learning. *International Journal of Architectural Computing*, 16 (4), 306–327. https://doi.org/10.1177/1478077118800982

Baker, M.E., Weller, D.E. & Jordan, T.E. (2007). Effects of stream map resolution on measures of riparian buffer distribution and nutrient retention potential. *Landscape Ecology*, 22 (7), 973–992. https://doi.org/10.1007/s10980-007-9080-z

Ball, J.G.C., Hickman, S.H.M., Jackson, T.D., Koay, X.J., Hirst, J., Jay, W., Archer, M., Aubry-Kientz, M., Vincent, G. & Coomes, D.A. (2023). Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN. *Remote Sensing in Ecology and Conservation*, 9 (5), 641–655. https://doi.org/10.1002/rse2.332

Barrera-Animas, A.Y., Oyedele, L.O., Bilal, M., Akinosho, T.D., Delgado, J.M.D. & Akanbi, L.A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, 7, 100204. https://doi.org/10.1016/j.mlwa.2021.100204

Bishop, K., Allan, C., Bringmark, L., Garcia, E., Hellsten, S., Högbom, L., Johansson, K., Lomander, A., Meili, M., Munthe, J., Nilsson, M., Porvari, P., Skyllberg, U., Sørensen, R., Zetterberg, T. & Åkerblom, S. (2009). The Effects of Forestry on Hg Bioaccumulation in Nemoral/Boreal Waters and Recommendations for Good Silvicultural Practice. *AMBIO: A Journal of the Human Environment*, 38 (7), 373–380. https://doi.org/10.1579/0044-7447-38.7.373

Boucher, D., Schepaschenko, D.G., Gauthier, S., Bernier, P., Kuuluvainen, T. & Shvidenko, A.Z. (2024). World boreal forest and managed boreal forest

extent. Natural Resources Canada. https://doi.org/10.23687/88D70716-2600-4995-8D5F-86F96E383ABF
Bradshaw, C.J.A. & Warkentin, I.G. (2015). Global estimates of boreal forest carbon stocks and flux. *Global and Planetary Change*, 128, 24–30. https://doi.org/10.1016/j.gloplacha.2015.02.004
Bring, A., Thorslund, J., Rosén, L., Tonderski, K., Åberg, C., Envall, I. & Laudon, H. (2022). Effects on groundwater storage of restoring, constructing or draining wetlands in temperate and boreal climates: a systematic review. *Environmental Evidence*, 11 (1), 38. https://doi.org/10.1186/s13750-022-00289-5
Brookes, A. (1987). The distribution and management of channelized streams in Denmark. *Regulated Rivers: Research & Management*, 1 (1), 3–16. https://doi.org/10.1002/rrr.3450010103
Busarello, M.D.S.T., Ågren, A.M., Westphal, F. & Lidberg, W. (2025). Automatic detection of ditches and natural streams from digital elevation models using deep learning. *Computers & Geosciences*, 196, 105875. https://doi.org/10.1016/j.cageo.2025.105875
Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA, August 13 2016. 785–794. ACM. https://doi.org/10.1145/2939672.2939785
Chicco, D. & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21 (1), 6. https://doi.org/10.1186/s12864-019-6413-7
Coumou, D. & Rahmstorf, S. (2012). A decade of weather extremes. *Nature Climate Change*, 2 (7), 491–496. https://doi.org/10.1038/nclimate1452
Council of the European Union (2023). *Interinstitutional File: 2022/0195 (COD)*. (15907/23)
Croke, J., Fryirs, K. & Thompson, C. (2013). Channel–floodplain connectivity during an extreme flood event: implications for sediment erosion, deposition, and delivery. *Earth Surface Processes and Landforms*, 38 (12), 1444–1456. https://doi.org/10.1002/esp.3430
De Almeida, A.F., Moreira, R. & Rodrigues, T. (2019). Synthetic organic chemistry driven by artificial intelligence. *Nature Reviews Chemistry*, 3 (10), 589–604. https://doi.org/10.1038/s41570-019-0124-0

Derry, A., Krzywinski, M. & Altman, N. (2023). Convolutional neural networks. *Nature Methods*, 20 (9), 1269–1270. https://doi.org/10.1038/s41592-023-01973-1
Diekmann, M. (1999). Southern deciduous forests. In: *Swedish plant geography*. (Acta Phytogeographica Suecica; 84). Svenska växtgeografiska sällskapet : Opulus Press [distributör]. 33–53.
Dore, M.H.I. (2005). Climate change and changes in global precipitation patterns: What do we know? *Environment International*, 31 (8), 1167–1181. https://doi.org/10.1016/j.envint.2005.03.004
Du, L., McCarty, G.W., Li, X., Zhang, X., Rabenhorst, M.C., Lang, M.W., Zou, Z., Zhang, X. & Hinson, A.L. (2024). Drainage ditch network extraction from lidar data using deep convolutional neural networks in a low relief landscape. *Journal of Hydrology*, 628. https://doi.org/10.1016/j.jhydrol.2023.130591
Du, L., You, X., Li, K., Meng, L., Cheng, G., Xiong, L. & Wang, G. (2019). Multi-modal deep learning for landform recognition. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, 63–75. https://doi.org/10.1016/j.isprsjprs.2019.09.018
Elenius, M., Pers, C., Schützer, S., Lindström, G. & Arheimer, B. (2025). Where can rewetting of forested peatland reduce extreme flows? Model experiment on the hydrology of Sweden. *Hydrology and Earth System Sciences*, 29 (17), 4307–4325. https://doi.org/10.5194/hess-29-4307-2025
Elshewey, A.M., Jamjoom, M.M. & Alkhammash, E.H. (2025). An enhanced CNN with ResNet50 and LSTM deep learning forecasting model for climate change decision making. *Scientific Reports*, 15 (1), 14372. https://doi.org/10.1038/s41598-025-97401-9
Estrada, J.S., Fuentes, A., Reszka, P. & Auat Cheein, F. (2023). Machine learning assisted remote forestry health assessment: a comprehensive state of the art review. *Frontiers in Plant Science*, 14, 1139232. https://doi.org/10.3389/fpls.2023.1139232
European Commission (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02000L0060-20141120 [2025-05-19]

Falloon, P.D. & Betts, R.A. (2006). The impact of climate change on global river flow in HadGEM1 simulations. *Atmospheric Science Letters*, 7 (3), 62–68. https://doi.org/10.1002/asl.133
Fei, J., Liu, J., Ke, L., Wang, W., Wu, P. & Zhou, Y. (2022). A deep learning-based method for mapping alpine intermittent rivers and ephemeral streams of the Tibetan Plateau from Sentinel-1 time series and DEMs. *Remote Sensing of Environment*, 282, 113271. https://doi.org/10.1016/j.rse.2022.113271
Flyckt, J., Andersson, F., Lavesson, N., Nilsson, L. & Å gren, A.M. (2022). Detecting ditches using supervised learning on high-resolution digital elevation models. *Expert Systems with Applications*, 201, 116961. https://doi.org/10.1016/j.eswa.2022.116961
Frainer, A., Polvi, L.E., Jansson, R. & McKie, B.G. (2018). Enhanced ecosystem functioning following stream restoration: The roles of habitat heterogeneity and invertebrate species traits. Cao, Y. (ed.) (Cao, Y., ed.) *Journal of Applied Ecology*, 55 (1), 377–385. https://doi.org/10.1111/1365-2664.12932
Freeman, T.G. (1991). Calculating catchment area with divergent flow based on a regular grid. *Computers & Geosciences*, 17 (3), 413–422. https://doi.org/10.1016/0098-3004(91)90048-I
Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29 (5). https://doi.org/10.1214/aos/1013203451
Fu, H., Zhao, H., Jiang, J., Zhang, Y., Liu, G., Xiao, W., Du, S., Guo, W. & Liu, X. (2024). Automatic detection tree crown and height using Mask R-CNN based on unmanned aerial vehicles images for biomass mapping. *Forest Ecology and Management*, 555, 121712. https://doi.org/10.1016/j.foreco.2024.121712
Gal, Y. & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *Proceedings of 33rd International Conference on Machine Learning*, New York, NY, USA, 2016. JMLR: W&CP
Gardeström, J., Holmqvist, D., Polvi, L.E. & Nilsson, C. (2013). Demonstration Restoration Measures in Tributaries of the Vindel River Catchment. *Ecology and Society*, 18 (3), art8. https://doi.org/10.5751/ES-05609-180308

Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A.Z. & Schepaschenko, D.G. (2015). Boreal forest health and global change. *Science*, 349 (6250), 819–822. https://doi.org/10.1126/science.aaa9092
Gergel, S.E., Stange, Y., Coops, N.C., Johansen, K. & Kirby, K.R. (2007). What is the Value of a Good Map? An Example Using High Spatial Resolution Imagery to Aid Riparian Restoration. *Ecosystems*, 10 (5), 688–702. https://doi.org/10.1007/s10021-007-9040-0
Gómez-Gener, L., Lupon, A., Laudon, H. & Sponseller, R.A. (2020). Drought alters the biogeochemistry of boreal stream networks. *Nature Communications*, 11 (1), 1795. https://doi.org/10.1038/s41467-020-15496-2
Guillon, H., Byrne, C.F., Lane, B.A., Sandoval Solis, S. & Pasternack, G.B. (2020). Machine Learning Predicts Reach-Scale Channel Types From Coarse-Scale Geospatial Data in a Large River Basin. *Water Resources Research*, 56 (3), e2019WR026691. https://doi.org/10.1029/2019WR026691
Gundersen, P., Laurén, A., Finér, L., Ring, E., Koivusalo, H., Sætersdal, M., Weslien, J.-O., Sigurdsson, B.D., Högbom, L., Laine, J. & Hansen, K. (2010). Environmental Services Provided from Riparian Forests in the Nordic Countries. *AMBIO*, 39 (8), 555–566. https://doi.org/10.1007/s13280-010-0073-9
Haapalehto, T.O., Vasander, H., Jauhiainen, S., Tahvanainen, T. & Kotiaho, J.S. (2011). The Effects of Peatland Restoration on Water-Table Depth, Elemental Concentrations, and Vegetation: 10 Years of Changes. *Restoration Ecology*, 19 (5), 587–598. https://doi.org/10.1111/j.1526-100X.2010.00704.x
Hamet, P. & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, S36–S40. https://doi.org/10.1016/j.metabol.2017.01.011
Han, H., Kim, T. & Kim, S. (2024). River Ice Mapping from Landsat-8 OLI Top of Atmosphere Reflectance Data by Addressing Atmospheric Influences with Random Forest: A Case Study on the Han River in South Korea. *Remote Sensing*, 16 (17), 3187. https://doi.org/10.3390/rs16173187
Hånell, B. (1990). Torvtäckta Marker, Dikning Och Sumpskogar i Sverige. In: *Skogsfakta*. SLU.
Hasselquist, E.M., Hasselquist, N.J., Sparks, J.P. & Nilsson, C. (2017). Recovery of nitrogen cycling in riparian zones after stream restoration using δ 15N along a 25-year chronosequence in northern Sweden. *Plant and Soil*, 410 (1–2), 423–436. https://doi.org/10.1007/s11104-016-3038-3

Hasselquist, E.M., Mancheva, I., Eckerberg, K. & Laudon, H. (2020). Policy change implications for forest water protection in Sweden over the last 50 years. *Ambio*, 49 (7), 1341–1351. https://doi.org/10.1007/s13280-019-01274-y

He, H., Jansson, P.-E., Svensson, M., Björklund, J., Tarvainen, L., Klemedtsson, L. & Kasimir, Å. (2016). Forests on drained agricultural peatland are potentially large sources of greenhouse gases – insights from a full rotation period simulation. *Biogeosciences*, 13 (8), 2305–2318. https://doi.org/10.5194/bg-13-2305-2016

Helfield, J.M., Capon, S.J., Nilsson, C., Jansson, R. & Palm, D. (2007). Restoration of Rivers Used for Timber Floating: Effects on Riparian Plant Diversity. *Ecological Applications*, 17 (3), 840–851. https://doi.org/10.1890/06-0343

Herzon, I. & Helenius, J. (2008). Agricultural drainage ditches, their biological importance and functioning. *Biological Conservation*, 141 (5), 1171–1183. https://doi.org/10.1016/j.biocon.2008.03.005

Humpenöder, F., Karstens, K., Lotze-Campen, H., Leifeld, J., Menichetti, L., Barthelmes, A. & Popp, A. (2020). Peatland protection and restoration are key for climate change mitigation. *Environmental Research Letters*, 15 (10), 104093. https://doi.org/10.1088/1748-9326/abae2a

Isikdogan, F., Bovik, A.C. & Passalacqua, P. (2017). Surface Water Mapping by Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10 (11), 4909–4918. https://doi.org/10.1109/JSTARS.2017.2735443

Islam, Md.T., Yoshida, K., Nishiyama, S., Sakai, K. & Tsuda, T. (2022). Characterizing vegetated rivers using novel unmanned aerial vehicle-borne topo-bathymetric green lidar: Seasonal applications and challenges. *River Research and Applications*, 38 (1), 44–58. https://doi.org/10.1002/rra.3875

Jacks, G. (2019). Drainage in Sweden -the past and new developments. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 69 (5), 405–410. https://doi.org/10.1080/09064710.2019.1586991

Jenson, S.K. & Domingue, J.O. (1988). Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. *PHOTOGRAMMETRIC ENGINEERING*,

Jiang, W., He, G., Long, T., Ni, Y., Liu, H., Peng, Y., Lv, K. & Wang, G. (2018). Multilayer Perceptron Neural Network for Surface Water Extraction

in Landsat 8 OLI Satellite Images. *Remote Sensing*, 10 (5), 755. https://doi.org/10.3390/rs10050755
Karimi, S., Mosquera, V., Maher Hasselquist, E., Järveoja, J. & Laudon, H. (2025). Does peatland rewetting mitigate flooding from extreme rainfall events? *Hydrology and Earth System Sciences*, 29 (12), 2599–2614. https://doi.org/10.5194/hess-29-2599-2025
Kazemi Garajeh, M., Li, Z., Hasanlu, S., Zare Naghadehi, S. & Hossein Haghi, V. (2022). Developing an integrated approach based on geographic object-based image analysis and convolutional neural network for volcanic and glacial landforms mapping. *Scientific Reports*, 12 (1), 21396. https://doi.org/10.1038/s41598-022-26026-z
Kazimi, B., Thiemann, F. & Sester, M. (2020). DETECTION OF TERRAIN STRUCTURES IN AIRBORNE LASER SCANNING DATA USING DEEP LEARNING. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2–2020, 493–500. https://doi.org/10.5194/isprs-annals-V-2-2020-493-2020
Khoi, D.N., Quan, N.T., Linh, D.Q., Nhi, P.T.T. & Thuy, N.T.D. (2022). Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam. *Water*, 14 (10), 1552. https://doi.org/10.3390/w14101552
Kim, J., Lee, M., Han, H., Kim, D., Bae, Y. & Kim, H.S. (2022). Case Study: Development of the CNN Model Considering Teleconnection for Spatial Downscaling of Precipitation in a Climate Change Scenario. *Sustainability*, 14 (8), 4719. https://doi.org/10.3390/su14084719
Komulainen, V., Tuittila, E., Vasander, H. & Laine, J. (1999). Restoration of drained peatlands in southern Finland: initial effects on vegetation change and CO_2 balance. *Journal of Applied Ecology*, 36 (5), 634–648. https://doi.org/10.1046/j.1365-2664.1999.00430.x
Korteling, J.E. (Hans)., Van De Boer-Visschedijk, G.C., Blankendaal, R.A.M., Boonekamp, R.C. & Eikelboom, A.R. (2021). Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4, 622364. https://doi.org/10.3389/frai.2021.622364
Koski, C., Kettunen, P., Poutanen, J., Zhu, L. & Oksanen, J. (2023). Mapping Small Watercourses from DEMs with Deep Learning—Exploring the Causes of False Predictions. *Remote Sensing*, 15 (11), 2776. https://doi.org/10.3390/rs15112776

Koskinen, M., Tahvanainen, T., Sarkkola, S., Menberu, M.W., Laurén, A., Sallantaus, T., Marttila, H., Ronkanen, A.-K., Parviainen, M., Tolvanen, A., Koivusalo, H. & Nieminen, M. (2017). Restoration of nutrient-rich forestry-drained peatlands poses a risk for high exports of dissolved organic carbon, nitrogen, and phosphorus. *Science of The Total Environment*, 586, 858–869. https://doi.org/10.1016/j.scitotenv.2017.02.065

Kuglerová, L., Jyväsjärvi, J., Ruffing, C., Muotka, T., Jonsson, A., Andersson, E. & Richardson, J.S. (2020). Cutting Edge: A Comparison of Contemporary Practices of Riparian Buffer Retention Around Small Streams in Canada, Finland, and Sweden. *Water Resources Research*, 56 (9), e2019WR026381. https://doi.org/10.1029/2019WR026381

Laine, A.M., Ojanen, P., Lindroos, T., Koponen, K., Maanavilja, L., Lampela, M., Turunen, J., Minkkinen, K. & Tolvanen, A. (2024). Climate change mitigation potential of restoration of boreal peatlands drained for forestry can be adjusted by site selection and restoration measures. *Restoration Ecology*, 32 (7), e14213. https://doi.org/10.1111/rec.14213

Lantmäteriet (2021). Ortophoto

Lantmäteriet (2022). Laser data. https://www.lantmateriet.se/globalassets/geodata/geodataprodukter/hojddata/e_pb_laserdata_nedladdning_nh.pdf [2025-01-31]

Laudon, H., Järveoja, J., Ågren, A., Peichl, M. & Lindgren, A. (2025). Rewetting drained forested peatlands: A cornerstone of Sweden's climate change mitigation strategy. *Ambio*,. https://doi.org/10.1007/s13280-025-02220-x

Laudon, H., Lidberg, W., Sponseller, R.A., Maher Hasselquist, E., Westphal, F., Östlund, L., Sandström, C., Järveoja, J., Peichl, M. & Ågren, A.M. (2022). Emerging technology can guide ecosystem restoration for future water security. *Hydrological Processes*, 36 (10), e14729. https://doi.org/10.1002/hyp.14729

Lavoie, M., Paré, D., Fenton, N., Groot, A. & Taylor, K. (2005). Paludification and management of forested peatlands in Canada: a literature review. *Environmental Reviews*, 13 (2), 21–50. https://doi.org/10.1139/a05-006

Lazarus, E.D. & Constantine, J.A. (2013). Generic theory for channel sinuosity. *Proceedings of the National Academy of Sciences*, 110 (21), 8447–8452. https://doi.org/10.1073/pnas.1214074110

Leckie, D.G., Cloney, E., Jay, C. & Paradine, D. (2005). Automated Mapping of Stream Features with High-Resolution Multispectral Imagery. *Photogrammetric Engineering & Remote Sensing*, 71 (2), 145–155. https://doi.org/10.14358/PERS.71.2.145

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, 521 (7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. & Jackel, L.D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1 (4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Li, C., Xiao, K., Sun, L., Tang, R., Dong, X., Qiao, B. & Xu, D. (2024). CNN-Transformers for mineral prospectivity mapping in the Maodeng–Baiyinchagan area, Southern Great Xing'an Range. *Ore Geology Reviews*, 167, 106007. https://doi.org/10.1016/j.oregeorev.2024.106007

Lidberg, W., Nilsson, M., Lundmark, T. & Ågren, A.M. (2017). Evaluating preprocessing methods of digital elevation models for hydrological modelling. *Hydrological Processes*, 31 (26), 4660–4668. https://doi.org/10.1002/hyp.11385

Lidberg, W., Paul, S.S., Westphal, F., Richter, K.F., Lavesson, N., Melniks, R., Ivanovs, J., Ciesielski, M., Leinonen, A. & Ågren, A.M. (2023). Mapping Drainage Ditches in Forested Landscapes Using Deep Learning and Aerial Laser Scanning. *Journal of Irrigation and Drainage Engineering*, 149 (3), 04022051. https://doi.org/10.1061/JIDEDH.IRENG-9796

Lindsay, J.B. (2016). Whitebox GAT: A case study in geomorphometric analysis. *Computers & Geosciences*, 95, 75–84. https://doi.org/10.1016/j.cageo.2016.07.003

Lõhmus, A., Remm, L. & Rannap, R. (2015). Just a Ditch in Forest? Reconsidering Draining in the Context of Sustainable Forest Management. *BioScience*, 65 (11), 1066–1076. https://doi.org/10.1093/biosci/biv136

Maanavilja, L., Aapala, K., Haapalehto, T., Kotiaho, J.S. & Tuittila, E.-S. (2014). Impact of drainage and hydrological restoration on vegetation structure in boreal spruce swamp forests. *Forest Ecology and Management*, 330, 115–125. https://doi.org/10.1016/j.foreco.2014.07.004

Mason, R.J. & Polvi, L.E. (2023). Unravelling fluvial versus glacial legacy controls on boulder-bed river geomorphology for semi-alluvial rivers in Fennoscandia. *Earth Surface Processes and Landforms*, 48 (14), 2900–2919. https://doi.org/10.1002/esp.5666

Matthews, B.W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405 (2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Mazhar, S., Sun, G., Bilal, A., Hassan, B., Li, Y., Zhang, J., Lin, Y., Khan, A., Ahmed, R. & Hassan, T. (2022). AUnet: A Deep Learning Framework for Surface Water Channel Mapping Using Large-Coverage Remote Sensing Images and Sparse Scribble Annotations from OSM Data. *Remote Sensing*, 14 (14), 3283. https://doi.org/10.3390/rs14143283

McCulloch, W.S. & Pitts, W.H. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5, 115–133

Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R. & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207, 103225. https://doi.org/10.1016/j.earscirev.2020.103225

Minx, J.C., Lamb, W.F., Callaghan, M.W., Fuss, S., Hilaire, J., Creutzig, F., Amann, T., Beringer, T., De Oliveira Garcia, W., Hartmann, J., Khanna, T., Lenzi, D., Luderer, G., Nemet, G.F., Rogelj, J., Smith, P., Vicente Vicente, J.L., Wilcox, J. & Del Mar Zamora Dominguez, M. (2018). Negative emissions—Part 1: Research landscape and synthesis. *Environmental Research Letters*, 13 (6), 063001. https://doi.org/10.1088/1748-9326/aabf9b

Moore, I.D., Grayson, R.B. & Ladson, A.R. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5 (1), 3–30. https://doi.org/10.1002/hyp.3360050103

Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y. & Beroza, G.C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11 (1), 3952. https://doi.org/10.1038/s41467-020-17591-w

Muotka, T., Paavola, R., Haapala, A., Novikmec, M. & Laasonen, P. (2002). Long-term recovery of stream habitat structure and benthic invertebrate communities from in-stream restoration. *Biological Conservation*, 105 (2), 243–253. https://doi.org/10.1016/S0006-3207(01)00202-6

Negishi, J.N. & Richardson, J.S. (2003). Responses of organic matter and macroinvertebrates to placements of boulder clusters in a small stream of southwestern British Columbia, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, 60 (3), 247–258. https://doi.org/10.1139/f03-013
Nieminen, M., Palviainen, M., Sarkkola, S., Laurén, A., Marttila, H. & Finér, L. (2018). A synthesis of the impacts of ditch network maintenance on the quantity and quality of runoff from drained boreal peatland forests. *Ambio*, 47 (5), 523–534. https://doi.org/10.1007/s13280-017-0966-y
Nilsson, C., Jansson, R., Kuglerová, L., Lind, L. & Ström, L. (2013). Boreal Riparian Vegetation Under Climate Change. *Ecosystems*, 16 (3), 401–410. https://doi.org/10.1007/s10021-012-9622-3
Nilsson, C., Lepori, F., Malmqvist, B., Törnlund, E., Hjerdt, N., Helfield, J.M., Palm, D., Östergren, J., Jansson, R., Brännäs, E. & Lundqvist, H. (2005). Forecasting Environmental Responses to Restoration of Rivers Used as Log Floatways: An Interdisciplinary Challenge. *Ecosystems*, 8 (7), 779–800. https://doi.org/10.1007/s10021-005-0030-9
Nilsson, C., Polvi, L.E., Gardeström, J., Hasselquist, E.M., Lind, L. & Sarneel, J.M. (2015). Riparian and in-stream restoration of boreal streams and rivers: success or failure? *Ecohydrology*, 8 (5), 753–764. https://doi.org/10.1002/eco.1480
Nugent, K.A., Strachan, I.B., Roulet, N.T., Strack, M., Frolking, S. & Helbig, M. (2019). Prompt active restoration of peatlands substantially reduces climate impact. *Environmental Research Letters*, 14 (12), 124030. https://doi.org/10.1088/1748-9326/ab56e6
O'Callaghan, J.F. & Mark, D.M. (1984). The Extraction of Drainage Networks from Digital Elevation Data. *Computer Vision, Graphics, And Image Processing*, 28, 323–344
Olden, J.D. & Poff, N.L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19 (2), 101–121. https://doi.org/10.1002/rra.700
Olusola, A.O., Olumide, O., Fashae, O.A. & Adelabu, S. (2022). River sensing: the inclusion of red band in predicting reach-scale types using machine learning algorithms. *Hydrological Sciences Journal*, 67 (11), 1740–1754. https://doi.org/10.1080/02626667.2022.2098752
Palmer, M.A., Ambrose, R.F. & Poff, N.L. (1997). Ecological Theory and Community Restoration Ecology. *Restoration Ecology*, 5 (4), 291–300. https://doi.org/10.1046/j.1526-100X.1997.00543.x

Palmer, M.A., Bernhardt, E.S., Allan, J.D., Lake, P.S., Alexander, G., Brooks, S., Carr, J., Clayton, S., Dahm, C.N., Follstad Shah, J., Galat, D.L., Loss, S.G., Goodwin, P., Hart, D.D., Hassett, B., Jenkinson, R., Kondolf, G.M., Lave, R., Meyer, J.L., O'Donnell, T.K., Pagano, L. & Sudduth, E. (2005). Standards for ecologically successful river restoration. *Journal of Applied Ecology*, 42 (2), 208–217. https://doi.org/10.1111/j.1365-2664.2005.01004.x

Pasha, S.V. & Reddy, C.S. (2024). Global spatial distribution of Prosopis juliflora - one of the world's worst 100 invasive alien species under changing climate using multiple machine learning models. *Environmental Monitoring and Assessment*, 196 (2), 196. https://doi.org/10.1007/s10661-024-12347-1

Paul, S.S., Hasselquist, E.M., Jarefjäll, A. & Ågren, A.M. (2023). Virtual landscape-scale restoration of altered channels helps us understand the extent of impacts to guide future ecosystem management. *Ambio*, 52 (1), 182–194. https://doi.org/10.1007/s13280-022-01770-8

Peltomaa, R. (2007). Drainage of forests in Finland. *Irrigation and Drainage*, 56 (S1), S151–S159. https://doi.org/10.1002/ird.334

Pham, D.T. & Pham, P.T.N. (1999). Artificial intelligence in engineering. *International Journal of Machine Tools and Manufacture*, 39 (6), 937–949. https://doi.org/10.1016/S0890-6955(98)00076-5

Pilotto, F., Nilsson, C., Polvi, L.E. & McKie, B.G. (2018). First signs of macroinvertebrate recovery following enhanced restoration of boreal streams used for timber floating. *Ecological Applications*, 28 (2), 587–597. https://doi.org/10.1002/eap.1672

Polsinelli, M., Cinque, L. & Placidi, G. (2020). A light CNN for detecting COVID-19 from CT scans of the chest. *Pattern Recognition Letters*, 140, 95–100. https://doi.org/10.1016/j.patrec.2020.10.001

Pomeroy, J., Granger, R., Pietroniro, A., Elliott, J., Toth, B., & Hedstrom, N. (1998). Classification of the Boreal Forest for Hydrological Processes. *Proceedings of Ninth international boreal forest research association conference*, 1998. 49–59

Prijac, A., Gandois, L., Taillardat, P., Bourgault, M.-A., Riahi, K., Ponçot, A., Tremblay, A. & Garneau, M. (2023). Hydrological connectivity controls dissolved organic carbon exports in a peatland-dominated boreal catchment stream. *Hydrology and Earth System Sciences*, 27 (21), 3935–3955. https://doi.org/10.5194/hess-27-3935-2023

Remm, L., Lõhmus, P., Leis, M. & Lõhmus, A. (2013). Long-Term Impacts of Forest Ditching on Non-Aquatic Biodiversity: Conservation Perspectives for a Novel Ecosystem. Chen, H.Y.H. (ed.) (Chen, H. Y. H., ed.) *PLoS ONE*, 8 (4), e63086. https://doi.org/10.1371/journal.pone.0063086
Roelens, J., Höfle, B., Dondeyne, S., Van Orshoven, J. & Diels, J. (2018). Drainage ditch extraction from airborne LiDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 409–420. https://doi.org/10.1016/j.isprsjprs.2018.10.014
Ronneberger, O., Fischer, P. & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv. http://arxiv.org/abs/1505.04597 [2023-03-14]
Rumelhart, D.E., Hintont, G.E. & Williams, R.J. (1986). Learning representations by back-propagating errors.
Russell, S.J. & Norvig, P. (2021). *Artificial intelligence: a modern approach*. Fourth Edition. Pearson. (Pearson Series in Artificial Intelligence)
Ruuska, R. & Helenius, J. (1996). GIS analysis of change in an agriculture landscape in Central Finland. *Agricultural and Food Science*, 5 (6), 567–576. https://doi.org/10.23986/afsci.72770
Sanderson, L.A., Mclaughlin, J.A. & Antunes, P.M. (2012). The last great forest: a review of the status of invasive species in the North American boreal forest. *Forestry*, 85 (3), 329–340. https://doi.org/10.1093/forestry/cps033
Shinde, P.P. & Shah, S. (2018). A Review of Machine Learning and Deep Learning Applications. *Proceedings of 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, August 2018. 1–6. IEEE. https://doi.org/10.1109/ICCUBEA.2018.8697857
Simonyan, K. & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Proceedings of 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA, April 10 2015. https://doi.org/10.48550/arXiv.1409.1556
Sjörs, H. (1999). The background: geology, climate and zonation. In: *Swedish plant geography*. (Acta Phytogeographica Suecica; 84). Svenska växtgeografiska sällskapet : Opulus Press [distributör]. 5–14.
Skogsstyrelsen (2022). Skogsvårdslagstiftningen
Song, H. & Jung, J. (2023). Scalable Surface Water Mapping up to Fine-scale using Geometric Features of Water from Topographic Airborne LiDAR Data. arXiv. https://doi.org/10.48550/arXiv.2301.06567

Ståhl, G., Allard, A., Esseen, P.-A., Glimskär, A., Ringvall, A., Svensson, J., Sundquist, S., Christensen, P., Torell, Å.G., Högström, M., Lagerqvist, K., Marklund, L., Nilsson, B. & Inghe, O. (2011). National Inventory of Landscapes in Sweden (NILS)—scope, design, and experiences from establishing a multiscale biodiversity monitoring system. *Environmental Monitoring and Assessment*, 173 (1), 579–595. https://doi.org/10.1007/s10661-010-1406-7

Sun, S., Wu, H. & Xiang, L. (2020). City-Wide Traffic Flow Forecasting Using a Deep Convolutional Neural Network. *Sensors*, 20 (2), 421. https://doi.org/10.3390/s20020421

Swedish PEFC (2017). PEFC Sweden Forest Standard 002:4. https://pefc.se/vara-standarder/svenska-pefc-standarden/swedish-pefc-standard-in-english [2024-01-21]

Szczepanek, R. (2022). Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost. *Hydrology*, 9 (12), 226. https://doi.org/10.3390/hydrology9120226

Teng, J., Wen, C., Zhang, D., Bengio, Y., Gao, Y. & Yuan, Y. (2023). Predictive Inference with Feature Conformal Prediction. *Proceedings of The Eleventh International Conference on Learning Representations*, Rwanda, 2023. https://openreview.net/forum?id=0uRm1YmFTu

Thirumalraj, A., Anusuya, V.S. & Manjunatha, B. (2023). Detection of Ephemeral Sand River Flow Using Hybrid Sandpiper Optimization-Based CNN Model: In: Kumar, A., Srivastav, A.L., Dubey, A.K., Dutt, V., & Vyas, N. (eds) *Advances in Civil and Industrial Engineering*. IGI Global. 195–214. https://doi.org/10.4018/979-8-3693-1194-3.ch010

Törnlund, E. & Östlund, L. (2002). Floating Timber in Northern Sweden: The Construction of Floatways and Transformation of Rivers. *Environment and History*, 8 (1), 85–106. https://doi.org/10.3197/096734002129342611

Törnlund, E. & Östlund, L. (2006). Mobility without Wheels: The Economy and Ecology of Timber Floating in Sweden, 1850–1980. *The Journal of Transport History*, 27 (1), 48–70. https://doi.org/10.7227/TJTH.27.1.5

Trenberth, K. (2011). Changes in precipitation with climate change. *Climate Research*, 47 (1), 123–138. https://doi.org/10.3354/cr00953

United Nations General Assembly (2015). Transforming Our World: The 2030 Agenda for Sustainable Development

Vasander, H., Tuittila, E.-S., Lode, E., Lundin, L., Ilomets, M., Sallantaus, T., Heikkilä, R., Pitkänen, M.-L. & Laine, J. (2003). Status and restoration

of peatlands in northern Europe. *Wetlands Ecology and Management*, 11, 51–63
Verdonschot, R.C.M., Keizer-vlek, H.E. & Verdonschot, P.F.M. (2011). Biodiversity value of agricultural drainage ditches: a comparative analysis of the aquatic invertebrate fauna of ditches and small lakes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 21 (7), 715–727. https://doi.org/10.1002/aqc.1220
Wang, J. & Perez, L. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. 11
Westphal, F., Lidberg, W., Busarello, M.D.S.T. & Ågren, A.M. (2025). Uncertainty quantification for LiDAR-based maps of ditches and natural streams. *Environmental Modelling & Software*, 191, 106488. https://doi.org/10.1016/j.envsoft.2025.106488
Williams, P., Whitfield, M., Biggs, J., Bray, S., Fox, G., Nicolet, P. & Sear, D. (2004). Comparative biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in Southern England. *Biological Conservation*, 115 (2), 329–341. https://doi.org/10.1016/S0006-3207(03)00153-8
Wohl, E., Lane, S.N. & Wilcox, A.C. (2015). The science and practice of river restoration. *Water Resources Research*, 51 (8), 5974–5997. https://doi.org/10.1002/2014WR016874
World Bank (2014). *Turn Down the Heat: Confronting the New Climate Normal*. The World Bank. (World Bank E-Library Archive). https://doi.org/10.1596/978-1-4648-0437-3
Zaharia, L., Ioana-Toroimac, G., Moroşanu, G.-A., Gălie, A.-C., Moldoveanu, M., Čanjevac, I., Belleudy, P., Plantak, M., Buzjak, N., Bočić, N., Legout, C., Bigot, S. & Ciobotaru, N. (2018). Review of national methodologies for rivers' hydromorphological assessment: A comparative approach in France, Romania, and Croatia. *Journal of Environmental Management*, 217, 735–746. https://doi.org/10.1016/j.jenvman.2018.04.017
Zakšek, K., Oštir, K. & Kokalj, Ž. (2011). Sky-View Factor as a Relief Visualization Technique. *Remote Sensing*, 3 (2), 398–415. https://doi.org/10.3390/rs3020398
Zannella, A., Eklöf, K., Hasselquist, E.M., Laudon, H., Garnett, M.H. & Wallin, M.B. (2025). Changes in Aquatic Carbon Following Rewetting of a Nutrient-Poor Northern Peatland. *Journal of Geophysical Research: Biogeosciences*, 130 (4), e2024JG008565. https://doi.org/10.1029/2024JG008565

Zhang, C., Zhou, J., Wang, H., Tan, T., Cui, M., Huang, Z., Wang, P. & Zhang, L. (2022). Multi-Species Individual Tree Segmentation and Identification Based on Improved Mask R-CNN and UAV Imagery in Mixed Forests. *Remote Sensing*, 14 (4), 874. https://doi.org/10.3390/rs14040874

Zhang, J., Xianglong, M., Zhang, J., Sun, D., Zhou, X., Mi, C. & Wen, H. (2023). Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *Journal of Environmental Management*, 332. https://doi.org/10.1016/j.jenvman.2023.117357

Zuluaga-Gomez, J., Al Masry, Z., Benaggoune, K., Meraghni, S. & Zerhouni, N. (2021). A CNN-based methodology for breast cancer diagnosis using thermal images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 9 (2), 131–145. https://doi.org/10.1080/21681163.2020.1824685

Popular science summary

In the northern forests, the land is covered by a dense network of small streams and man-made ditches. Together, they control how water moves through the landscape, carry nutrients, and support a wide range of plants and animals. In the past, people altered many of these streams to float timber and dug ditches to dry out land for farming and forestry purposes. Even though digging new ditches has mostly stopped, the old ones are still draining today's forests and affecting water systems. These small channels are like the capillaries of the landscape, but most of them are still missing from maps. This thesis focuses on that problem by exploring a new way to map small streams and ditches nationwide. The method uses detailed elevation data from laser scanning (LiDAR) together with artificial intelligence (AI) to find and classify channels automatically. By training AI models to tell the difference between natural streams and man-made ditches, the study shows that this can be done successfully on a national scale. The best model correctly found most ditches and improved the mapping of natural streams in comparison to topographic maps. Furthermore, we show that future mapping of natural streams can use our "uncertainty maps". These are maps that indicate places where the models are less certain of a channel presence. This is the first time that streams and ditches have been separated on maps for an entire country. The results can help improve forest and water management, guide stream restoration, and support national environmental reporting. The method can also be used in other countries with high-resolution LiDAR data and similar forested landscapes.

“Science is not about building a body of known ‘facts’. It is a method for asking awkward questions and subjecting them to a reality-check, thus avoiding the human tendency to believe whatever makes us feel good.”

– Terry Pratchett, *The Science of Discworld*

Populärvetenskaplig sammanfattning

I de boreala skogarna i norr finns ett tätt nätverk av små vattendrag och konstgjorda diken. Tillsammans styr de hur vattnet rör sig genom landskapet, transporterar näringsämnen och utgör habitat för många växter och djur. Tidigare förändrade människor majoriteten av dessa vattendrag för att flotta timmer, grävde diken för att dränera mark för jordbruk och skogsbruk. Även om ny dikning i stort har upphört, dränerar de gamla dikena fortfarande dagens skogar och påverkar vattensystemen. Dessa små kanaler är som landskapets kapillärer där marken står i kontakt med vattnet. Men de flesta små vattendragen saknas fortfarande på dagens kartor. De här avhandlingen fokuserar på det problemet genom att skapa en ny metod att kartlägga små vattendrag och diken på nationell skala. Metoden använder detaljerade höjddata från laserskanningar (LiDAR) tillsammans med artificiell intelligens (AI) för att automatiskt hitta och klassificera vattendrag som bäckar och diken. Genom att träna AI-modeller att skilja mellan naturliga vattendrag och konstgjorda diken visar studien att detta kan göras framgångsrikt på nationell nivå. Den bästa modellen identifierade majoriteten av alla diken korrekt och förbättrade kartläggningen av naturliga vattendrag jämfört med topografiska kartor. Dessutom visar vi att framtida kartläggning av naturliga vattendrag kan dra nytta av våra ”osäkerhetskartor”, som visar var modellerna är mindre pålitlig. Detta är första gången som vi kan särskilja naturliga vattendrag från diken på nationell nivå. Resultaten från den här avhandlingen kan bidra till bättre skogs- och vattenförvaltning, vägleda restaurering av vattendrag och stödja nationell miljörapportering. Metoden kan även användas i andra länder med liknande högupplöst laserdata och skogstäckta landskap.

Credits: Luciano Censoni

Acknowledgements

I want to start by thanking **William**, my supervisor, whom I am forever thankful to. Thank you not only for following the petty bureaucratic constraints of being my supervisor, but also for all the support and motivation that you shared with me. You always went out of your way to make things smoother for me and made sure that I was able to get to the end of this. Before things got dramatically demanding with the thesis, I remember asking, "Can I survive this?" to which you answered, "Yeah, sure, don't worry". I was skeptical, but here we are. *Anu belore dala'na*. I really hope you were not playing Alliance, but I fear you were. Well, no one is perfect.

Secondly, thank you so much, **Anneli**, for being my supervisor too. You never agreed with something just because I made it sound like a good idea (a geologist's passive skill), but convincing you made me a better professional who must take their time to think before speaking, even though I still speak a lot. Thank you very much for being available constantly, and I hope some day we can dance *forró*. Thank you, **Florian**, my third supervisor all the way from Jönköping (which severely limited the amount of time I could talk about random things with him), for your patience in explaining things to me more than twice, and for making sure that our hardcore coding part was working properly.

A special thanks to **Eliza**, who helped me tremendously in understanding artificial and natural channels in Sweden, their policies, legislation, and management. There is no way I would have been able to do it without you. I have seen how much you care about your students, which is why I know you cared a lot about helping me, too. One more honorable thanks to **Lenka**, who got us a win in a *pepparkakashustavling* from the department in 2023 due to her awesome argumentation in favor of our construction. One day, she encouraged me to join her in field work to help an intern out, and, despite that not being part of my personal disposition, I had a great time doing so and learned a lot from her, too.

From work, there are many people to thank. I'll start by saying that **Koffi**'s thesis organization inspired me when doing my own, since it looked super neat. Also, since he defended before me, I was asking questions to him constantly about the process. I also thank you for all the lunches we shared and the funny conversations we had, and all the spex we planned. And I'm sorry I'm such a terrible singer!! I also need to thank **Shirin**, with whom I've

spent such fun times together, and who recommended me the best rice of all. Thanks to **Clydecia**, who, during desperate times, allowed me to watch one of her lectures for a course, to which I paid back with Melitta coffee. **Betty** had the absolute best laugh from the whole department, and definitely one of the best defense parties of all time. Kudos to her and her family for making it so: you have set a standard. Another killer defense party was **Arvid**'s, who, by the way, has abandoned me reading Chainsaw Man by myself and moved on to better manga. But he also introduced me to Hooja, so I guess that evens out. Please, always remember: you are Kenough.

Thanks, **Sijia,** for being the only one in the department who understands why K-pop choreography is so cool and fun, and for making me the best birthday noodles I have ever had. I'll always remember us trying to make boba tea. Thanks, **Antonia**, for showing me that gluten-free food is super tasty. I hope I never made you sick with anything I've cooked for you. Thank you, **Barry**, who had endured bravely the department loneliness of December until we met for the first time on the day I started working. We did so many fun things together, I just hope to never step on wetlands with hiking shoes ever again. Thank you, **Kohsuke**, for cooking the first ochazuke I've ever had, not to mention the best dumplings and mapo tofu. I think no one works as hard as you, and I hope some of that brushed off on me. Thank you, **RuiRui**, for having the patience to listen to my music, including Faye Wong, which I have always loved. Thank you, **Lei**, for always asking questions during our conversations, which caused you to be the last one to finish eating. More thanks to **Ilse**, **Moritz**, **Ash**, **Lorenzo**, **Johannes**, **Martin**, **Anne**, **Joss**, **Francesco**, **Eli**, and **Lena**.

Special thanks to **Alejandro** and **Vicky**, who know what it feels like to be Latin American in Sweden and with whom I share a craving for sunny days. When **Olivia** started working, and I noticed we shared a lot of things in common, I was super happy! It was awesome having someone to talk to about video games and anime, I've only watched Oshi no Ko because she told me to, otherwise, I would have skipped one of the best anime ever. Thank you for being part of our office. Now **Lin** and I shared more than an office, we also shared supervisors and commentary on society. I remember when I was told they had chosen you, I was already trying to find out when you would move to Umeå and if we would be friends. You were the one who introduced me to climbing, so I am glad they selected you and that we were able to become closer during this period. I hope I was as helpful to you as

you have been to me. I admire you a lot for speaking your mind without committing “sincericide”, as we say it in Brazil.

From WASP-HS, I want to thank **Eva**, who was worried about me when I was sick, and who was always quick to help out with administrative issues. I also want to thank **Hannah** for the insightful gossip, awesome pictures, and funny blog posts. Thank you, Irene, for all the time we spent climbing together, eating delicious food, and laughing out loud. Thank you, **Felix**, who was as crazy as me to volunteer when the student council needed leadership, who always reaches out to check how I’m doing, who never leaves anyone behind. For all the *kompisar* I made there, extra thanks: **Igor**, for always being chill and knowing all the board games; **Lux**, for the unhinged conversations and even more unhinged climbing; **Sarah**, for hanging out with me during weird dinners and long hikes; **Denitsa**, who helped me during my worst (COVID); **Bijona**, for sharing the same opinions about a *insopportabile* person when I thought I was losing my mind; **Karim**, for being someone to whom I can always complain about very specific things.

Triple thanks to **Joakim**, who I told everyone was my project partner even when we were a bit lacking in the joint project thing, who lent me *The Three-Body Problem,* consequently ruining my life forever, who sometimes knows more about Brazil than I do, and who has the funniest voice for cat dubbing. Thank you, **Sarah,** for giving Joakim a chance and, therefore, giving Umeå a chance because we wouldn’t have met if that didn’t happen. I admire you very much, and you deserve every good thing that comes your way. Thank you two for the Eurovision and På Spåret nights, if I am integrated in Europe at all, it’s because of you.

Climbing has become one of the things I can’t shut up about, so I couldn’t not mention all the amazing people I’ve met there (but not only there): **Kai**, **Magda**, and extra especially, **Meredith**, **Amanda** and **Léa**. Sometimes, when cheering for me to go higher, you were all motivating me in more ways than you knew. Sometimes I only need to remember to breathe. And maybe brush a little.

Thank you, **Fernanda**, my best Brazilian friend, who is also my favorite Brazilian. I appreciate all our shopping sprees, fancy dinners, and humble gossip. I love you soooo much, I am super thankful for having you in my life, and I’ll fight anyone who gets in our way. I want to thank **Ting**, who made me so happy by messaging me out of the blue to hang out when we were in Komvux. I always thought you were the coolest gal in the class!

Having you in my life is the best, and I'll always be here to cheer you on and help you when you need it, and I'm already in place to be an Auntie to Loke. Thank you, **Ana**. I love you and, every day, I miss living in the same building as you. Not in the same apartment, though, as we have tried that before, with funny results now that we look back on it. I am always cheering for your success and happiness because you are the one who deserves it the most. And on this note, I couldn't do this without **Daniel**'s support. Every time I needed to vent, you were there to listen to me and agree about how bad everything was. Not denying my feelings, just validating them. We have always been on the same wavelength, and even if sometimes it feels like we are out of phase, I know we'll always go back in sync. I love you very much. I also want to thank **Bia**, the kindest soul I have ever met, who inspires me and never lets me forget to blow cinnamon at my door on the first day of the month, every month.

For the eternal college friendships, I'd like to start saying thanks to those who have been part of all my journeys for a long time now. We have lived countless adventures together, we got drunk, we tested weird products, we recorded videos, we danced, and we partied hard. But most importantly, we were always there for each other, supporting each other, celebrating every victory, every achievement, building our family, and sharing memes. This has not changed to this day, and I hope we can still keep it going for a long time. Thank you, **Jorge**, **Marcelo**, **Thomás**, **Seiji**, **Akira**, **Lucas**, **Mikhael**, **Fábio**, and **Arthur** (gosh, it feels very weird to write these names).

Now for the blood ties, I want to thank my brother **Cairo**, who has the coolest name ever. I love you and I am very happy to be your sister. Mostly because I'm older than you, but also because we don't have to share a house anymore. I wish every day that your dreams come true, and that you are constantly happy! Thank you, my stepfather **Alexandre**, who has always been a dad to us, worrying, caring, and stepping up. We love you, and we are thankful to have you in our lives. I also want to thank my mom, **Mirian**, for always being supportive of me and cheering for my happiness. I remember all those mornings you were listening to progressive rock, and I thought it was the most annoying thing ever, just to end up going to those same concerts when I grew up. You are a human lie detector, and I love you very much. Thanks to my mother-in-law, **Marcia**, who is not a blood tie, but who treats me as her own daughter and always supports me and Luciano.

Thank you, **Luciano**, my match, my boyfriend, my fiancé, and my husband. You have absorbed my way of speaking random things, not like a sponge, but more like a catchment, storing it and discharging it in increasingly unpredictable times (specialists will disagree with this statement, though), surprising me and making me laugh twice as hard. Thank you for the immeasurable help you have given me since the beginning. Thank you for bringing us this far. I would not have achieved half of what I have without you. Which is why I am sure we are going to achieve even more. I love you. And thank you for agreeing to take in these two furballs that fill our life with happiness and destruction: our cats **Monet** and **Mondrian**.

To my grandmother, **Aracy**, it is not easy to find the words. You have built me this way, for good and for bad, but I have chosen differently from the path that was set for me, over and over. I have learned to rely on friends. I have learned that we can trust people. I have reframed frustrations. And yet, I still hope that I made you a little bit proud. Maybe this is the thing that does it. Would you be telling those you knew that I was going to become a doctor (of Philosophy)? I will never know. But I hope one day it won't matter as much. Today is not that day. It still matters, which is why you are remembered.

I type this as I ride the train during a quick trip to southern Sweden, I look out the window, and I can't feel anything else but gratitude. It is a privilege to be sitting on a train with such great scenery to look at. The autumn colors paint the trees that still hold their leaves against the increasingly harsh winds, and out of every tunnel, you are welcomed by the potential sight of the Baltic Sea in the early morning. Some bright houses might pop up across the landscape, just to make sure you're paying attention. I am addicted to being alive and experiencing the many small joys that life brings, and making my thesis did not ruin this feeling, which I hope I can hold on to forever. I am here. I did this. Thank you, past me.

I

Computers & Geosciences 196 (2025) 105875

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Automatic detection of ditches and natural streams from digital elevation models using deep learning☆

Mariana Dos Santos Toledo Busarello [a,*], Anneli M. Ågren [a], Florian Westphal [b], William Lidberg [a]

[a] Swedish University of Agricultural Sciences, Skogsmarksgränd, 901 83, Umeå, Sweden
[b] Jönköping University, Gjuterigatan 5, 551 11 Jönköping, Sweden

ARTICLE INFO

Keywords:
Streams
Ditches
Deep learning
LiDAR
Semantic segmentation

ABSTRACT

Policies focused on waterbody protection and restoration have been suggested to European Union member countries for some time, but to adopt these policies on a large scale the quality of small water channel maps needs considerable improvement. We developed methods to detect and classify small stream and ditch channels using airborne laser scanning and deep learning. The research questions covered the influence of the resolution of the digital elevation model on channel extraction, the efficacy of different terrain indices to identify channels, the potential advantages of combining indices, and the performance of a U-net model in mapping both ditches and stream channels. Models trained in finer resolutions were more accurate than models trained with coarser resolutions. No single terrain index consistently outperformed all others, but some combinations of indices had higher MCC values. Natural stream channels were not classified to the same extent as ditches. The model trained on the 0.5 m resolution had the most balanced performance using a combination of indices trained using the dataset with both types of channel separately. The deep learning model outperformed traditional mapping methods for ditches, increasing the recall from less than 10% to over 92%, while the recall for natural channels was around 71%. However, despite the successful detection of ditches, the models frequently misclassified streams as ditches. This poses a challenge, as natural channels are protected under land use management practices, while ditches are not.

1. Introduction

The primary objective of the United Nations Agenda 2030 for Sustainable Development is the protection of the planet from further environmental degradation (United Nations General Assembly, 2015), highlighting the importance of protecting and restoring water-related ecosystems. A similar goal is present in the European Water Framework Directive (where policy changes implemented in 2000 brought an integrated approach to the management and protection of aquatic environments) adopted throughout the European Union. Furthermore, a proposal for new targets of nature restoration is currently being drawn up by the European Commission, aiming at successful restoration of 20% of the target area by 2030, and 90% by 2050 (Council of the European Union, 2023). However, the management strategies for applying these initiatives differ among countries.

Most countries use different sizes of riparian buffer zones to protect surface waters during land-use operations, but these policies vary when it comes to small streams. In Finland, for example, stream channels are protected through a forest buffer of minimum width (Ring et al., 2018). In Sweden, the Swedish Forest Act (Skogsstyrelsen, 2013) also prescribes forest water protection through riparian buffers of variable width (Hasselquist et al., 2020). This is a necessary measure because over 75% of the total river network is estimated to be small streams (Bishop et al., 2008), and therefore even small changes in the network can impact downstream channels dramatically. Even so, the data shows that after 2004 as few as 25% of the small streams in Sweden were protected in such a manner, and when a buffer is present it usually has a width of 4 ± 0.4 m (Kuglerová et al., 2020), despite the recommended 5–30 m width of no-harvesting zones.

Some laws only address watercourses in general and do not

☆ Link to the code: https://github.com/mbusarello/Automatic-Detection-of-Ditches-and-Natural-Streams-from-Digital-Elevation-Models-Using-Deep-Learning.

* Corresponding author.
E-mail addresses: mariana.busarello@slu.se, mariana.busarello@gmail.com (M.D.S.T. Busarello), anneli.agren@slu.se (A.M. Ågren), florian.westphal@ju.se (F. Westphal), william.lidberg@slu.se (W. Lidberg).

https://doi.org/10.1016/j.cageo.2025.105875
Received 18 March 2024; Received in revised form 21 January 2025; Accepted 23 January 2025
Available online 27 January 2025

differentiate between natural channels and those altered or made by humans, while other laws go into more depth on different types of watercourses. For example, according to the Swedish Forestry Act (September 1, 2022) ditches are divided into two categories: "ditches" and "protective ditches". Simple "ditches" are dug for permanent soil drainage to change the land use of an area. "Protective ditches," on the other hand, are temporarily dug to mitigate groundwater level rise following clear-cutting. Protective ditches must not be cleaned, as they are temporary, and should not be more than 50 cm deep. No permit is needed to clean ditches, while digging new ones does requires official permission (Swedish PEFC, 2023). The idea is that ditches should gradually fill in with sediments and vegetation, eventually disappearing with time. The management of ditches can also include damming/plugging them to restore wetlands (Nieminen et al., 2018). Because of this variability in the practices which are allowed by law, knowing if a channel is natural or altered by man determines the best management choice.

Within the context of environmental impact, forest ditches can be strong anthropogenic emitters of greenhouse gases (Peacock et al., 2021b), with methane offsetting the uptake from terrestrial CH_4; they also transport suspended solids, which impacts water quality (Nieminen et al., 2018). Even though the differences between ditches and small natural streams are not always clear, factors such as morphology and hydrology do stress the distinction between channel types. Some of these attributes can also influence the quantity of methane being emitted (Peacock et al., 2021a), resulting in an annual flux slightly higher for ditches than for streams.

There is wide recognition of the importance of hydrological variability to the ecology of small streams (Huryn and Wallace, 1987; Lanka et al., 1987; Wohl, 2017), after all, the characteristics of meandering, pools, and rapids can define habitats (Beschta and Platts, 1986; Wiens, 2002; Martínez et al., 2013), nutrient cycling (Alexander et al., 2007; Claessens et al., 2010), and water quality (Cox et al., 2023). Yet, the mapping of small water channels (<6 m wide) on Sweden's traditional digital maps was poor: 55% of the natural streams and 91% of ditches were not detected in the Swedish property map (Flyckt et al., 2022). Plus, the simplified digitized line from this dataset (Lantmäteriet, 2014) has limited usefulness for research in ecology when working across the landscape scale with geographic information system methods. Still, the number of mapped ditches was increased from 9% to 86% by Lidberg et al. (2023) using deep learning (LeCun et al., 2015) and remote sensing, turning the once laborious manual task with a substantial investment of cost and time into an automated process. Many countries have already been scanned with airborne laser scanning (ALS), and, using the latest return data, digital elevation models (DEMs) can be constructed, revealing small-scale channels (Raber et al., 2002).

Deep learning approaches have been used to map stream channels based on satellite images and Digital Elevation Models (Mazhar et al., 2022; Fei et al., 2022; Isikdogan et al., 2017). However, the main focus of these studies has been on larger rivers, while deep learning applications in small streams is limited. Koski et al. (2023) mapped small channels but did not separate between ditches and natural streams, while others have focused only on ditches based on ALS data (Du et al., 2024; Lidberg et al., 2023), or aerial photos (Robb et al., 2023). Despite these efforts, a research gap remains for small natural streams – the headwaters. Headwater streams are like the capillary system in the body – just as the health of the whole organism depends on a functioning capillary system, the health of larger streams and rivers depend upon an intact headwater stream network (Kuglerová et al., 2017), hence there is a large societal need for improving the mapping of the headwaters. Traditionally, headwaters are mapped from DEMs by calculating flow accumulation and applying a threshold to determine where streams begin (Ågren et al., 2015). However, the high natural variability in stream initiation thresholds makes these networks unreliable (Paul et al., 2023). Additionally, channel networks derived from flow accumulation are subject to further uncertainties because flow accumulation requires extensive preprocessing to of the DEM which introduces more uncertainties especially at stream/road crossings (Lidberg et al., 2017). Therefore, the goal of this study was to develop a method for mapping channels in the landscape without including upstream areas or considering the presence of water. Instead, the focus was on detecting the physical structure of the channel, specifically the elongated depression visible in the DEM.

Building on the successful use of deep learning to map ditches in Lidberg et al. (2023), this article extends the methodology by incorporating the digitization of small natural stream channels into a dataset that was previously limited to ditches and adding one more study area. Topographic indices derived from ALS data and the manually mapped channels were used to train a U-net model to detect small-scale channels (both ditches and natural streams). Here, we explore for the first time if deep learning can be used to detect small streams from the high-resolution DEM considering not only the channels' location, but also their variable width instead of just buffering them. The following research questions were answered:

1) How important is the resolution of the DEM for detecting ditches and natural channels? Here we explore two resolutions: 0.5 m and 1 m.
2) When highlighting the channels using digital terrain indices, is there a best one? Is the same index best for natural channels and ditches, or do they differ?
3) When detecting channels, is it better to work with just one terrain index, or to combine the information from many indices?
4) Can a U-net model be used to detect natural channels as well as ditches? Is it better to include ditches and natural channels in the same model, or to make separate models?

2. Methodology

Digital terrain indices were extracted from the DEM obtained from the high-resolution ALS data. These terrain indices were combined to form a database of manually mapped water channels, this then being used to train a deep neural network to detect and classify small-scale channels.

2.1. Study areas

We used remote sensing data and field data from the 12 regions described by Lidberg et al. (2023). The original dataset was exclusively composed of ditches; smaller (<6 m width) natural streams were added later by Paul et al. (2023). This data were revised and updated by comparing the location of the channels directly to orthophotos with a resolution ranging from 0.17 to 0.5 m (Lantmäteriet, 2021a) and the High-Pass Median Filter (HPMF) terrain analysis, increasing the length of channels to 2235 km of ditches and 315 km of natural streams.

Following Paul et al. (2023), these sites illustrated the diversity of the country's landscape properties, with land use mainly represented by forests covering 86–99% of the area, and agriculture ranging from 0 to 13.2% coverage among sites. Variability in characteristics such as soil type, tree species, runoff, and topography were considered in the site selection process. Overall, the Swedish landscape has been heavily ditched, tripling the originally unaltered channel length density, with the majority of the channels built being forest ditches. Most of the natural channel heads can be found in the northern areas, but transition points (i.e., the connection between a natural channel and an upstream ditch network) happened more often in the south. Small natural channels in Sweden are meandering and blend with the surrounding terrain, as boulders in their course minimize stark contrasts (Fig. 2B). Ditches are instead straight and smooth-looking, with generally well-defined borders resulting from the removal of boulders during the digging process. Most of the ditches in the dataset were forest ditches (56%), with road ditches in second (25%), and agricultural ditches last (6%, Paul et al. (2023)).

2.2. Training data

2.2.1. Topographic indices

The ALS data (Lantmäteriet, 2021b) were collected by an aircraft flying at a height of 2888–3000 m with a compact laser-based system onboard (Leica ALS80-HP-8236) generating point clouds with a density of 1–2 points per square meter. LiDAR Tin Gridding from Whitebox Tools was used to create DEMs with 0.5 m and 1 m resolutions over the study areas, totaling 430 km^2. We selected seven topographic indices that could visually highlight small-scale channels present in the DEMs (Fig. 2) as a proxy for the differences in elevation. Many indices could have been experimented on, but there is a limitation in the number of variables that could be used in the study considering the amount of time and effort involved in calculating new indices and preparing them as input for training the models. It was also observed that larger moving windows provided excessive smoothing, blending small channels in the landscape, while small scales introduced a high amount of noise. This is why the choice in scale relied on the visual evaluation for the cases where the size of the moving window was not arbitrarily defined by the tool in use.

The topographic indices were normalized between zero and one before being divided into chips of 500 × 500 pixels for input to the deep learning algorithm (Fig. 1B and C). Whitebox Tools was used to calculate all topographic indices, except for the Sky-view Factor, which was obtained using the Relief Visualization Toolbox v. 2.2.0 (Kokalj et al., 2016).

2.2.2. High-Pass Median Filter

The HPMF (Lindsay, 2016) emphasizes short-range variability, subtracting the pixel value from the median value of the other pixels inside a window. The window size kernel is user-defined; this study used 11 in both X and Y directions. The data were normalized by applying the Min–Max Normalization. Negative values indicate depressions and can be used to highlight channels, i.e. elongated depressions in the soil. This index is similar to the topographic position index, which is obtained through the subtraction of the mean value of the area covered by a moving window, however, HPMF was chosen due to the previously successful application in Lidberg et al. (2023), and because the median is more resistant to extreme values in the data.

2.2.3. Hillshade

The shaded relief (Wilson and Gallant, 2000) makes it possible to visualize a three-dimensional surface considering its slope and aspect, with shadows distributed according to the illumination source position (altitude and azimuth). This study has used the fixed altitude of 30° and the azimuths 0°, 45°, 90°, and 135°. The values were normalized

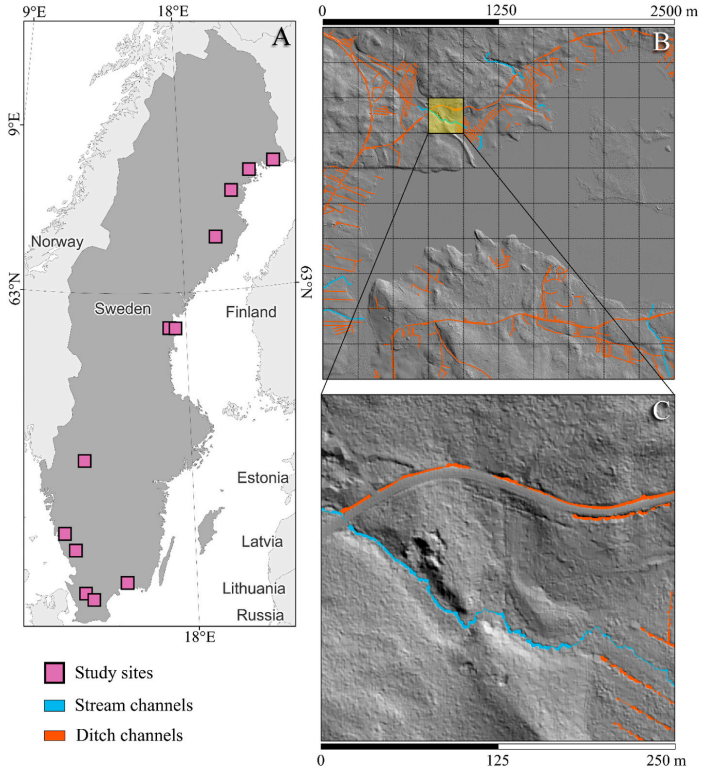


Fig. 1. Study areas. (A) 12 regions spread across Sweden where all ditches and streams were manually digitized; (B) Study regions split into 2.5 km× 2.5 km tiles. Locations of manually mapped water channels were separated by type, with ditches in orange and natural channels in turquoise, drawn over the hillshaded elevation model. Each grid cell represents chips with sides of 500 × 500 pixels. (C) An example of a 0.5 m resolution image chip obtained after splitting the tile. These chips are the images that the deep learning models will use as training data. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

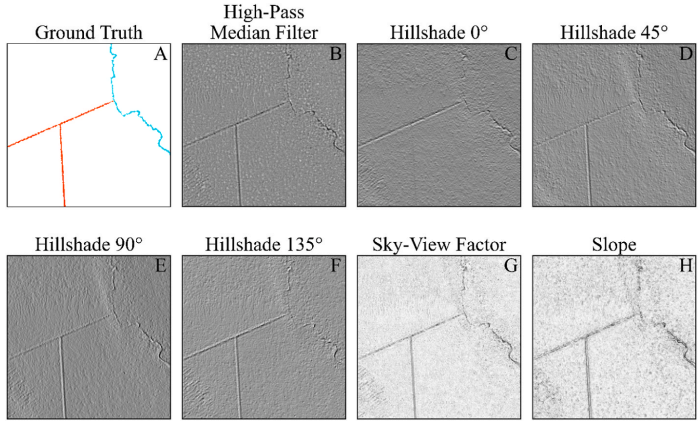


Fig. 2. Examples of the ground truth and topographic indices. Orange represents ditches and turquoise represents natural streams. Images displayed represent an area of 250 m × 250 m. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

afterward through their division by the maximum value. The bottom of a channel would be shaded unless it was hit by sunlight along the direction of the channel. To address this issue, we included hillshades from four different angles.

2.2.4. Sky-view factor

This index is defined by the ratio between the radiation received at a specific grid cell and the one emitted through the whole hemispheric environment around it (Zakšek et al., 2011). Considering a visual observation of the channels, the chosen radius was 5 m with 16 directions.

2.2.5. Slope

This topographic index represents the change in elevation between every pixel in the DEM with a moving window sized 5 x 5 for increased accuracy and stronger reduction of high-frequency noise (Florinsky, 2016), with the inclination displayed in degrees. To perform the normalization, all values were divided by the theoretical maximum value of 90°.

2.2.6. Labels

When the word "channel" is used in this article, it includes both ditches and natural streams. Using Whitebox Tools, we started by obtaining the flow accumulation (O'Callaghan and Mark, 1984). First, we filled the single cell depressions in the DEM (*FillDepressions*), then burning streams at roads using data from the Swedish Property map (Lantmäteriet, 2014) to ensure stream continuity across roads (*BurnStreamsAtRoads*). Remaining larger depressions were breached (*BreachDepressionsLeastCost*) to keep the flow continuity, using this as the input to calculate the D8 flow accumulation (*D8FlowAccumulation*). Streams were extracted (*ExtractStreams*) using the lowest stream initiation threshold from the distribution observed for natural channel heads in Paul et al. (2023): 2 ha.

Following this methodology, the channel heads and connections to the ditch network were identified, and downstream stream paths manually marked and edited. Ditches were visually identified from HPMF and ortophotos, being manually mapped as vector lines by a team of experts. We have utilized the HPMF values within the channels to give these lines a variable width, creating structures that more closely resemble the actual shape of the channels. Based on the method described in Lidberg et al. (2023), the HPMF analysis had its pixels reclassified based on the threshold of −0.075 (determined through visual inspections), receiving the label 0 when they are above it, and 1 when below. A 3 m buffer surrounding the vector lines was generated, later overlapping the relabeled data and extracting the non-null pixels within it. Finally, we applied the majority filter to these selected pixels to remove strays, preserving the continuity of the channels (Fig. 2A).

Eight different datasets were created (Fig. 3), initially separated by how the channels were represented:

- Channels: all channels, merged to a combined dataset with no separation of ditches and streams. Two class labels; channel and background (Fig. 3A and E)
- Ditches: a separate dataset of only ditches. Two class labels; ditch and background (Fig. 3B and F).
- Streams: a separate dataset of only streams. Two class labels; streams and background (Fig. 3C and G)
- Ditches&Streams: a combined dataset with three class labels; ditches, streams, and background (Fig. 3D and H)

Each type of representation was calculated for both 0.5 m and 1 m resolution to analyze how this impacted the results; each one is noted as an added "0.5" or "1" the dataset names.

The datasets exhibited significant class imbalance. To compensate for that, only the chips containing more than 250 pixels with the positive label were selected for the analysis, resulting in 4615 chips in total. From these, 1.1% of the total pixels were ditches and 0.1% were streams. Not all chips contained both types of channels, so datasets with only streams or ditches had fewer chips (Busarello et al., 2024).

2.3. Semantic segmentation

The convolutional neural network (CNN) U-net (Ronneberger et al., 2015) (Fig. 4) was chosen for having successful real-world applications in different scientific fields such as medicine (Siddique et al., 2021), geology (Gao et al., 2022), and forestry (Korznikov et al., 2021), being both robust and versatile. It also has the advantage of concatenating the feature maps of the downsampling path to the upsampling path, preventing the loss of information during downsampling. A limitation of the study was the amount of chips available in the datasets: CNN models usually require thousands of training data examples, and for this reason, acquiring training data is the most challenging part of the process. The use of data augmentation (Tanner and Wong, 1987) increased the number and diversity of training images by adding slightly different

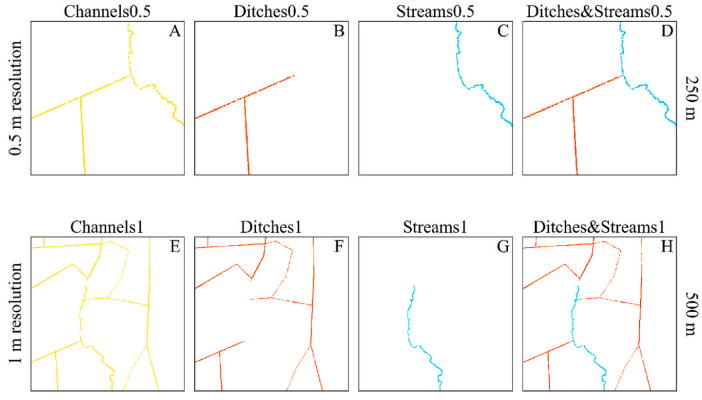


Fig. 3. Training data chip examples of both resolutions. Top row represents 0.5 m, and bottom row shows 1 m resolution. Chip size is 250 m × 250 m for 0.5 m resolution and 500 m × 500 m for 1 m resolution. Yellow lines in dataset Channels represent channels, without distinction between stream channels and ditch channels. Ditch channels are represented in orange in the datasets Ditches and Ditches&Streams. Turquoise represents stream channels in datasets Streams and Ditches&Streams. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

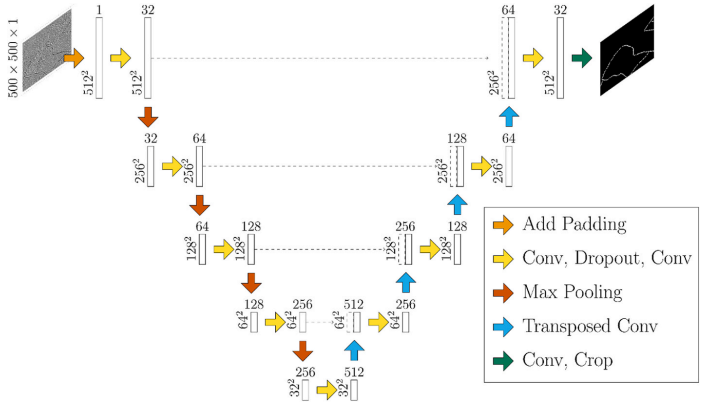


Fig. 4. U-net architecture. The left side shows the encoding/down-sampling process, where the main features are extracted while the input is compacted. On the right side is the decoding/up-sampling path, which upscales the features until it reaches the same size as the input.

copies of them to the dataset, obtained through transformations. The geometric transformations used in this work were the random rotation and random flips (horizontal and vertical). Random rotation rotated images in a random angle within the specified range of 0°–360°, helping improve the model's generalization by increasing the pattern recognition regardless of the object orientation in the image. The horizontal random flipping rotated the image along its vertical axis, swapping left and right, while the vertical random flipping flipped across the horizontal axis, swapping the top and bottom of the image instead.

Considering that the proportion between the classes of pixels showed considerable imbalance, median frequency balancing (Eigen and Fergus, 2015) was used to establish the class weights used for training. Adam (Kingma and Ba, 2015) was used as the optimization algorithm, and the chosen batch size was 16. In the beginning, the topographical indices were used individually as input to train the first models, while the last model combines all indices, resulting in 64 different models. Later, all possible combinations were used as training data for the dataset Ditches&Streams to determine if combining indices is a better option than using them individually.

The general architecture of U-net incorporates two paths: encoding and decoding. During the encoding phase, hierarchical features are extracted by a combination of convolutions and the pooling of feature maps, down-sampling the data resulting in a compact representation of the input, with an increased number of channels. Subsequently, in the decoding phase, transposed convolutions are applied to upscale the spatial dimension until the output matches the input original size. After each transposed convolution, a skip connection happens between corresponding layers in both paths. This allows the network to keep fine-grained details in the up-sampling process. The final convolution reduces the number of channels, producing the final segmentation map. In it, each pixel is assigned a probability of belonging to a class.

The processing time for calculating the topographic indices was tracked, as well as the inference time, being further extrapolated for the whole area of Sweden to estimate how long it would take to detect the location of channels throughout the entire country. Training and inference were done using an NVIDIA RTX A6000 GPU and AMD Ryzen

Threadripper 3990X Processor.

2.4. Evaluation

The data were split into two parts, 80% for training and 20% to evaluate the performance of all models, comparing the ground truth pixels with the detected pixels. Precision, Recall, F-score, and Matthews correlation coefficient (MCC; Matthews, 1975) were the key metrics used to evaluate the models, along with information retrieval tables. Precision is the metric that accounts for the accuracy of positive predictions from a model, being affected by the number of false positives. It assesses how much of the detection and classification made by the model was right. Recall, on the other hand, accounts for how much of the ground truth was correctly detected. F-score is the harmonic average of precision and recall, and MCC is a special case of the phi coefficient. The F-score was calculated to easily compare the performance of this study with other publications, but MCC reports the overall quality of the classification performed by the model, being more reliable for imbalanced datasets (Chicco and Jurman, 2020). The Precision-Recall curves were plotted to display the tradeoff between recall and precision in the highest-ranking models. In addition to these metrics, we also used models with the highest MCC values from each dataset to illustrate the location of detected channels. For the final evaluation, the inference of the best-performing models was compared to the ground truth in order to account for how much of each type of channel was detected by them.

2.5. Benchmark

We have used the traditional flow accumulation method of the 0.5m resolution as a benchmark to compare with our deep learning approach and our manually labeled dataset. The process to obtain the flow accumulation has been described in section 2.2.2, but now we have included the other two stream initiation thresholds of 6 ha and 10 ha, also observed in Paul et al. (2023). To make the comparison fair, the extracted streams went through the same described process to create the labels with natural contours: buffering, multiplying the buffer with the reclassified HPMF data, majority filtering, and combination with rasterlines. Additionally, the Swedish property map (1:12 500) was also used for comparison. It was rasterized (*VectorLinesToRaster*) and underwent the same process described in section 2.2 to create natural contours. All of this data was compared pixel by pixel to the labeled dataset, counting how many pixels labeled as ditch or natural stream were identified as channel by the flow accumulation.

Furthermore, the inference results from the deep learning model from Lidberg et al. (2023) was also compared to our ground truth data. Despite their model being trained exclusively on ditches, it indirectly detected some natural channels, allowing for a relevant comparison. To ensure we did not evaluate on data that the previous model might have been trained on, we used data from the newly added study site for this process, as it was not included in the previous model's training data.

3. Results and discussion

3.1. Importance of DEM scale for the modeling of channels using deep learning

The precision and recall values were higher for datasets with a 0.5 m resolution than for the 1 m counterpart. This was the case for all datasets and topographic indices (Fig. 5). Despite this, some models displayed higher values at either metric individually, and some overlap between the resolutions has been observed. This is partially in line with previous research on mapping terrain features with deep learning and DEM data, where higher resolution had better results (Chowdhuri et al., 2021) but also showed that the difference in performance between resolutions was not very strong (Robson et al., 2020).

The recall had different values for all datasets at 1 m resolution, with small differences between ditches and channels. The precision was similar for either resolution, with a variation of around 10%. We can assume that the performance of the models with 1 m resolution was impacted by the topographic index used in the training process. This impact was also observed in the 0.5 m resolution but to a lesser extent, which could indicate that models trained on a higher resolution were stable. The stability was not present on channels labeled as streams: in both scales and with any dataset, as seen in the black crosses in Fig. 5, the recall values were different while the precision was similar, not going over 25%.

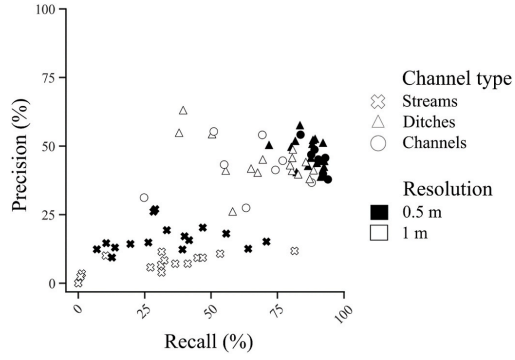


Fig. 5. Precision by Recall plot of the trained models, grouped by resolution and channel type. Black represents the 0.5 m resolution, while the 1 m resolution is represented by the white color. When referring to dataset Ditches&Streams0.5, the channel types were analyzed separately. The "Ditches" identified in the legend refers to this class in dataset Ditches and Ditches&Streams, while "Streams" addresses this class in dataset Streams and Ditches&Streams. "Channels" describes the models trained with the dataset Channels, combining ditch and stream channels in the same class.

The estimated processing time required to both extract the topographical indices and apply the model differed substantially between the DEM resolutions (Table 1). The Sky-view Factor in particular was computationally demanding compared to the other topographical indices, regardless of resolution. This happens because the source-code for the RVT library was written in python, which is an interpreted language. The tools from WBT, on the other hand, were coded in Rust – a compiled language. Compiled programs are faster than those that have to be interpreted (Kwame et al., 2017), and one way to have similar processing times would be to have all the processing steps written in a compiled language. Furthermore, parallelizing the codes for execution on the GPU could potentially mean a considerable speed improvement. The inference time of the deep learning model was about the same for

Table 1
Time spent to calculate each topographic index individually and in combinations, and the time spent to apply a deep learning model on new data (inference): both in two resolutions and measured in seconds by square kilometers. It was also estimated how long it would take (in days) to calculate the topographic index(es) and apply the model to the processed data for the whole surface area of Sweden (447 425 km^2). Hillshade had the same processing time regardless of the angle.

Topographic Index	Processing time (s/km^2)		Inference time (s/km^2)		Estimated time for Sweden (days)	
	0.5 m	1 m	0.5 m	1 m	0.5 m	1 m
HPMF	0.30	0.09	6.71	1.68	36	9
Hillshade	0.25	0.07	6.69	1.67	36	9
Slope	0.27	0.08	6.69	1.67	36	9
Sky-view Factor	3.01	0.68	6.64	1.67	50	12
Combination (all)	4.58	1.13	6.81	1.70	59	15

models trained with one index or several combined.

As models trained on the 0.5 m resolution datasets had the highest recall, the rest of this work focused on the models trained on topographical indices with a 0.5 m resolution. The analyses for 1 m resolution are in Appendices A.1, A.2, A.3, and A.4.

3.2. Impact of different terrain indices for detecting ditch and stream channels

We did not find a particular topographical index that consistently outperformed the others in this study. Models trained on Hillshades had the highest recall, while models trained on HPMF and Hillshade 0° had the highest precision using the datasets Channels0.5 (Fig. 6A) and Ditches0.5 (Fig. 6B). The model trained on the dataset Streams0.5 had the highest recall when trained on a combination of all topographical indices (Fig. 6C). That model had a recall of 70%, but the precision was still low at 20%. The highest recall for ditches with dataset Ditches&Streams0.5 was from the combination of all indices with 92% and 7% for streams using Hillshade 90° (Fig. 6D). The precision for the model trained on this dataset was highest with the HPMF for ditches, and Slope for streams. We believe that MCC gives the most balanced measure of the overall model performance, but there was no clear winner among models trained on different digital terrain indices (Table 2).

Indices not used in our work were listed as the most effective ones in studies focused on channels and fluvial features using the DEM (Du et al., 2019; Koski et al., 2023), or topographic positive openness for ditches (Du et al., 2024). Koski et al. (2023) detected channels using deep learning and several terrain indices besides the DEM, finding recall and precision values ranging 16–77% and 43–86%, respectively, while the F-score varied 0.23–0.81. The best terrain indices in our study for this type of dataset scored higher recall (83–93%), but lower precision (range 42–52%, Fig. 6A) and lower F-score (0.54–0.63, Table 3). The reasons for the differences are analyzed in section 3.4. Similarly, Du et al. (2024) detected ditches with deep learning, combining topographic and other features. Recall and precision were in the range of 73–76% and 63–69%, respectively, and F-score 0.69–0.71. Meanwhile, our similar dataset had higher recall (72–92%), lower precision 42–52% (Fig. 6B), and lower F-score 0.57–0.66 (Table 3). This difference could be because of the U.S. study having a higher resolution (0.3 m against our 0.5 m). Lidberg et al. (2023), however, obtained a higher MCC value than this study using the HPMF (0.78), which could be due to the different deep learning architecture.

The variation in the performance of the hillshade indices could be explained by the variation in channel orientation. In Fig. 2C, for example, part of the stream and the vertical ditch do not show because they were parallel to 0°, while the channels oriented perpendicularly were highlighted. Therefore, no matter the amount of data acquired and data augmentation performed, when using an index there is a chance that the channels might not be visible at all. This further motivated our choice to combining them.

3.3. Combining topographic indices

Combining all of the topographic indices did not result in a higher MCC compared to using them individually as input training data for most datasets, except Streams0.5 (Table 2). This dataset (Fig. 6C) and Ditches&Streams0.5 (Fig. 6D) had higher recall values.

However, when all of the possible combinations between the indices with dataset Ditches&Streams0.5 were analyzed (Appendix B) we observed that, for ditches, the HPMF was surpassed by the combination of Sky-view Factor + Slope in the ditches class (MCC = 0.69 (Table 3) against 0.74 (Fig. 7)) and the streams class (MCC = 0.09 against 0.31). Furthermore, for streams the Slope was surpassed by the combination of Hillshade 45° + Hillshade 90° + Hillshade 135°, not in the ditch class (MCC = 0.63 against 0.63) but in the stream class (MCC = 0.28 against 0.36). These results are in line with Kazimi et al. (2020) and Du et al. (2019), where a combination outperformed the single index, even though both studies used a coarser resolution (50 m) to detect fluvial structures (among others). We believe that the resolution did not influence this difference between combining indices or not, since these results also matched the coarser one analyzed by us (Appendix A.4).

Additionally to the observed trend that no single index was better than a combination of indices, we noted that the best performing combinations are those that combined two or three indices (Fig. 7). This appears reasonable since each index extracted different information from the DEM and as such may not contain all necessary information.

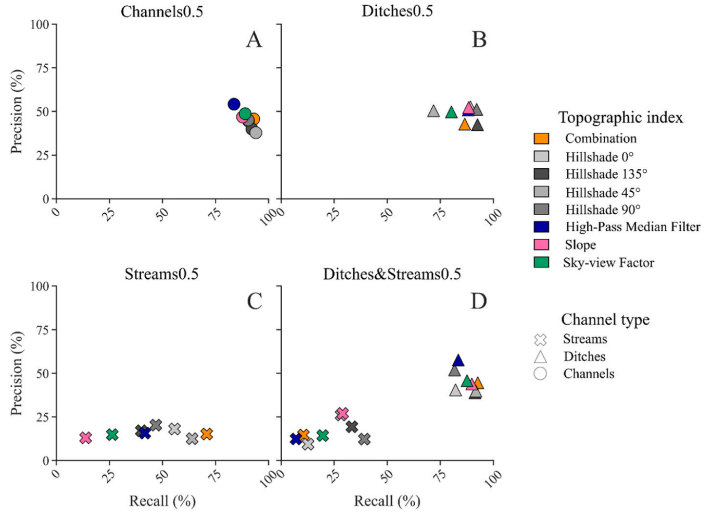


Fig. 6. Precision by Recall plots separated by dataset with 0.5 m resolution. Each color represents a topographic index, and each symbol represents a channel type. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
MCC values for all datasets with the 0.5 m resolution. The terrain indices with the highest MCC are highlighted in bold.

Topographic Indices	Channels0.5	Ditches0.5	Streams0.5	Ditches&Streams0.5 (ditches)	Ditches&Streams0.5 (streams)
Combination	0.65	0.61	**0.32**	0.64	0.12
Hillshade 0°	0.63	0.68	0.31	0.57	0.11
Hillshade 45°	0.59	0.60	0.28	0.60	0.27
Hillshade 90°	0.64	**0.69**	0.30	0.65	0.22
Hillshade 135°	0.60	0.63	0.26	0.59	0.25
HPMF	**0.67**	0.67	0.25	**0.69**	0.09
Slope	0.64	0.68	0.13	0.63	**0.28**
Sky-view Factor	0.66	0.63	0.19	0.63	0.17

Table 3
Evaluation metrics for each model, dataset, and its highest-performing topographic index. The recall, precision, F-score, and MCC values are also presented.

Model	TP	FP	TN	FN	Recall	Precision	F-score	MCC
Channels0.5 High-Pass Median Filter	2395624	2028771	224358343	467262	83.7%	54.1%	0.66	0.67
Ditches0.5 Hillshade 90°	2396057	2282997	203616321	204625	92.1%	51.2%	0.66	0.69
Streams0.5 Combination	185753	1040722	54697074	76451	70.8%	15.1%	0.25	0.32
Ditches&Streams0.5 High-Pass Median Filter (ditches)	2170656	1597868	224812682	430026	83.4%	57.6%	0.68	0.69
Ditches&Streams0.5 High-Pass Median Filter (streams)	18318	130158	226965020	243886	6.9%	12.3%	0.12	0.09

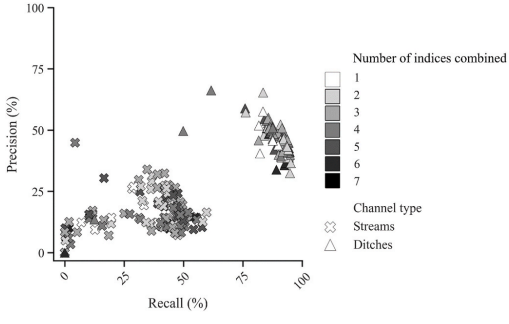


Fig. 7. Precision and Recall plots for all of the possible combinations of topographic indices. The color indicates the number of combined indices, and the shape represents the type of channel.

For example, ditches running from north to south were difficult to see in Hillshade 0° (Fig. 2C). However, adding indices to the considered combination, which introduce only slight variations of the information already provided by the considered indices, harmed performance, since it made the learning problem more difficult. This issue has been observed by others, for example by Yang et al. (2023) and Koski et al. (2023), who trained models directly on the DEM. These models performed similar or better than models trained on the DEM combined with indices derived from it, since all required information was already included in the DEM. Still, we argue that it is reasonable to assume that a model trained on topographic indices can generalize better due to the more uniform representation of the relevant topographic features.

Processing time could affect the decision to use multiple topographical indices, considering that it can increase greatly with a higher resolution. It seems that combining multiple topographical indices derived from the same LiDAR data could be beneficial, and so, including aerial photographs in the topographical data is something that might be worth exploring. Robb et al. (2023) obtained a higher F-score than our study (0.79 against 0.66) using orthophotos with a 0.25 m resolution to detect ditches, but this was not observed by Koski et al. (2023), where combining the orthophotos had the worst performance detecting channels. The aerial imagery data used by the Finnish study had a coarser resolution (0.5 m; NLS (2023)) which could be creating this difference. Koski et al. (2023) also points out that the extent of tree coverage hindered the performance of this input data to some extent, something that seems not to have happened in the UK publication, judging by the fact that the study area was less forested.

3.4. Evaluating model performance with different datasets

The models with the highest MCC values were selected for further evaluation under section 3.4. By "datasets" we mean if the model was trained to identify channels, streams, and/or ditches. The models trained with dataset Ditches0.5 with the highest MCC had a recall of 92.1%, while models trained on the dataset Channels0.5 had a recall of 83.7% (Table 3).). The same observation was made in the Precision-Recall curves, with AP = 0.76 for Channels0.5 (Fig. 8A) versus AP = 0.82 for Ditches0.5 (Fig. 8B).A Finnish dataset similar to Channels0.5 was used by Koski et al. (2023), with lower recall values (77.3%) but notably greater precision (85.6% against our 54%, Table 3). Starting in the 1950s, the ditching process in peatlands that was conducted in Finland altered the shape of most small natural channels (Muotka et al., 2002), with a low number of unaltered small streams left. This could mean that the uncertainty brought in by natural channels was smaller, as unaltered streams might be rarer in Finland, resulting in a higher precision. This could be an indication that when streams and ditches had the same label, uncertainty was introduced in the training process, blurring the detection and classification of channels. With the streams labeled as background, the separation became clearer and more channels were detected (despite the number of false positives also increasing).

The precision-recall curves strengthen the observations from Table 3. The average precision values reported were higher than the ones seen in the table because this metric is an approximation of the area under the precision-recall curve (Aslam et al., 2005), i.e., a summary of the precision-recall performance across all thresholds. However, we could still see similarities in the overall poor performance of the stream class in the Streams0.5 dataset (AP = 0.22, Fig. 8C), Ditches&Streams0.5 trained with HPMF (AP = 0.06, Fig. 8D), and the improvement brought to it by combining Sky-View Factor and Slope (AP = 0.28, Fig. 8E). Overall, the ditch label performed better across all datasets, showing that whichever high-ranking model was chosen, their detection would be similar. The differences, though, could be seen in the inferences (Fig. 9), where the interruption in channels happened more often within Channels0.5 (Fig. 9B) than Ditches0.5 (Fig. 9C).

For the model where the channels were trained with three labels (ditches, streams, and background (Fig. 3D)) we evaluated the ditches and streams separately. Ditch channels were correctly classified

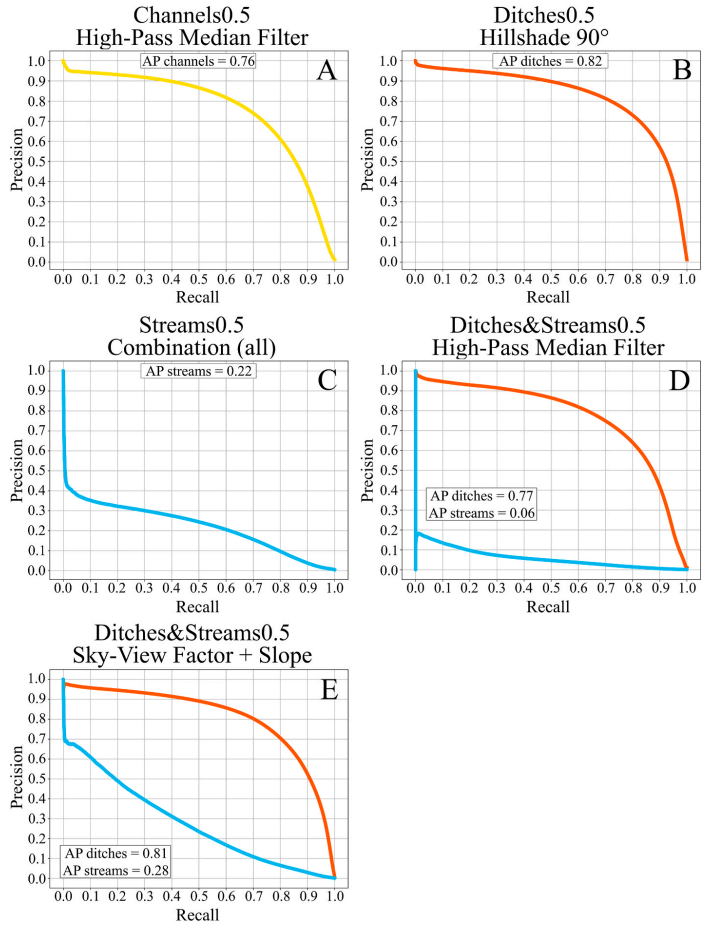


Fig. 8. Precision-Recall curves of the highest-ranking trained models and their average precision (AP).

frequently, which could mean that these channels had morphological attributes that made them more easily recognized by the neural network, while streams did not. Comparing this dataset (3-class) to dataset Ditches0.5 (binary), the recall was lower (83.4% against 92.1%), a result similar to Phinzi et al. (2020) when comparing the performance of a binary and a multiclass dataset to detect gullies with machine learning.

Models trained with the binary datasets had false positives more often, meaning that labeling streams and ditches separately in the training process could have helped distinguish both from the background data. A visual analysis of the detection (Fig. 9E) demonstrates that the models were not able to separate ditches and streams, but the number of false positives for the stream channels and ditches was low (0.06% and 0.7%, respectively; Table 3). For the dataset Ditches0.5 (Fig. 9C), stream channels were mainly misclassified as ditches despite being detected, while in the dataset Streams0.5 (Fig. 9D) the opposite happened, with frequent channel interruptions. This discontinuity was also observed in dataset Channels0.5 (Fig. 9B).

The channel interruptions were observed in small sections where the width was narrower than the average 3 m. In the ground truth data, these gaps were absent because the original polyline shapefile was converted to raster format and merged with HPMF-extracted values. This provided channel continuity, but limited their width to a single pixel. Gaps in the channel network are not unusual due to not only natural processes like sedimentation, falling trees and logs, but also to anthropogenic modifications such as culverts, bridges, road embankments (Lindsay and Dhun, 2015), which would explain why parts of the channel would be absent in ground truth. With that, they would not be detected in the inference either.

The highest-ranking models (Table 3) detected channels but were not as effective when classifying them, so we have calculated how much of each channel type was detected by each model regardless of the model's classification (Table 4). For the binary datasets, "detection" was the same as recall (TP/TP + FN), while "classification" was the same as precision (TP/TP + FP). However, we also used the multilabel ground truth (with pixels labeled 0, 1, or 2) to evaluate the performance of the models on channels, calculating how much of each channel type was detected. For the multilabel dataset (Ditches&Streams0.5), "detection"

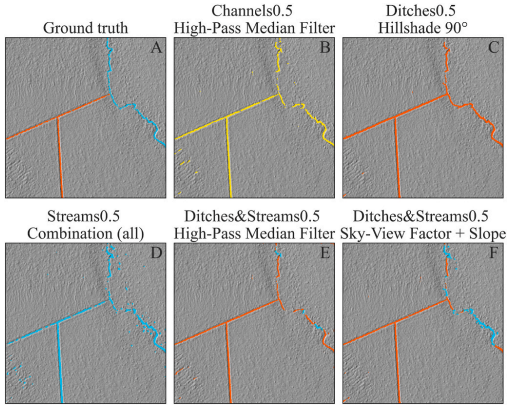


Fig. 9. Detected channels by the highest performing model from every dataset using the 0.5 m resolution, plotted over the hillshade. The colors represent channel type: ditch channels are orange, stream channels are turquoise, and combined channels are yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 4
Amount of channel pixels detected by each model, separated by channel type. The last two columns are only relevant to the multilabel dataset and describe the quantity of detected channels that were correctly classified by the model as their ground truth channel type.

Dataset used to train the model	Detected ditches	Detected streams	Classified as ditches	Classified as streams
Channels0.5	85.9%	61.4%	–	–
Ditches0.5	92.1%	55.9%	–	–
Streams0.5	81.5%	70.8%	–	–
Ditches&Streams0.5	83.8%	54.9%	99.5%	12.7%

meant not being predicted as background (label 0). At the same time, "classification" verified how many of the channel type predictions were correct, i.e., ditch pixels predicted to be ditches and stream pixels predicted to be streams. This was done because, despite a pixel being classified incorrectly as either ditch or stream, as long as it was not classified as "background" (label 0), it was still counted as a channel per the definition we use in this work: the combination of ditches and streams.

When both channel types had the same label (Channels0.5), the detection was higher than when they were labeled separately in the same dataset (Ditches&Streams0.5). Models with only one channel type (Ditches0.5 and Streams0.5) detected the other class, and in the case of Streams0.5 more ditches were detected than stream channels. Streams can be characterized by relative depth, continuity, and high sinuosity. Ditches are also characterized by relative depth and continuity, and low sinuosity (more straight). However, not all streams are meandering and not all ditches are straight. These similarities make it challenging to distinguish between both channel types, while the straight aspect tends to simplify the recognition of ditches. Furthermore, we employed median frequency balancing (Eigen and Fergus, 2015), which assigns larger class weights to less frequent classes, leading to larger errors when pixels of these classes are mislabeled. With this in mind, we observed how different labeling strategies affect the tradeoff different models made between precision and recall (Fig. 8).

In the three binary classification datasets (Channels0.5, Ditches05, and Streams0.5), misclassifying background as the respective positive class is comparatively inexpensive, due to the small class weight for the background class. Thus, the models favored higher recall despite an increase in false positives. In the 3-class dataset (Ditches&Streams0.5), labeling uncertain pixels as a minority class was costly due to the large class weights assigned to the ditch and the stream class. Mislabeling stream pixels as ditches incurred a significant penalty, while correctly identifying a small number of additional ditch pixels had limited benefits given their rarity. Conversely, background pixels offered the lowest relative cost, as they outnumbered the other two classes significantly. This led to higher precision but lower recall for ditches and streams. Additionally, the different recall values for ditches and streams in the three binary classification datasets were presumably due to the difficulty of identifying streams compared to ditches. When only streams were labeled (Stream0.5), the model needed to account for the meandering, sometimes nearly interrupted pattern of streams (Fig. 9A), which appeared to push the model toward recognizing other features in the landscape which have a similar pattern (Fig. 9D). This did not happen when only ditches were labeled (Fig. 9C), presumably because the model exploits the linear aspect of ditches, which allowed it to ignore other landscape features. When ditches and streams were labeled as channels (Channels0.5), the model needed to find a tradeoff between only focusing on the linear aspect, to allow it to find more streams than the ditch model, and recognizing too many landscape features, to achieve a better precision than the stream model. It appears to find this tradeoff by detecting more meandering interrupted features of the landscape as channels, while labeling more uncertain pixels as background, leading to more interrupted ditches (Fig. 9B).

Furthermore, because Ditches&Streams0.5 was a multilabel dataset we could verify how much of one label is classified as the other. In this case, from the number of ditches detected (83.8%), 99.5% were ground truth ditches. Meanwhile, only 12.7% of the streams detected by the model (54.9%) were correctly classified as streams. The difference in performance between stream and ditch channels in this dataset could be partially explained by the imbalance in the datasets. While the number of pixels with ditch labels was around 1.11% of the data, the stream pixels were underrepresented, with 0.01%. Contrasting class prior probabilities is a common occurrence in real-world data, and some techniques could be used to overcome it (Kotsiantis et al., 2006). In this work, the use of median frequency balancing (Eigen and Fergus, 2015) was motivated by its successful application in other studies such as Xu et al. (2022) and Kampffmeyer et al. (2016). However, despite the positive impact it had on the ditch class, an increase in performance of the stream class was not observed to the same extent. This represents a model limitation because the incorrect classification of streams as ditches is a regular occurrence. Adding more training data containing small natural streams would be an option to try to reduce the data imbalance, while an alternative would have been to perform a chip-based sampling, choosing chips that have more stream than ditch pixels in it. This would require further manual labor, though, where choices to reduce the costs of data acquisition could be explored, such as the use of semi-automated methods for labeling (Desmond et al., 2021) and crowdsourcing, despite the limitations that may arise regarding those who are not domain-specific experts (Clough et al., 2013).

3.5. Comparison to the benchmark

Our model (Streams0.5) had a recall of 70.8% of stream pixels, while the flow accumulation had the highest recall rate of stream pixels for 2 ha and 6 ha of initiation threshold (Table 5). 76.0% of the natural stream network was detected by the flow accumulation of 2 ha of the catchment area, 71.3% by 6 ha, and 70.2% by 10 ha. The Swedish Property map had a recall of 27.5% of pixels from the same channel type, which could be explained by the fact that the stream headwaters have been digitized from grainy black-and-white orthophotos in this data, being often obscured by canopy cover, which impacted its performance. Meanwhile, Lidberg et al. (2023) had an indirect recall (i.e., how much of the label "stream" was detected despite the model being trained with only

Table 5
Comparison between the recall performance of different methods of channel detection separated by type of channel pixels. "Recall of ditch pixels" refers to how many ditch pixels could be detected when compared to the ground truth. "Recall of stream pixels" refers to how many stream pixels were detected. All methods were evaluated on the same study areas, except Lidberg et al. (2023), which was evaluated on the study area that was not included in its training data. The MCC values listed were calculated with only the streams as the positive class to make a fair comparison between the methods.

Method	Recall of ditch pixels	Recall of stream pixels	MCC of ditches	MCC of natural streams
Swedish property map	8.1%	27.5%	0.16	0.28
Flow accumulation (2 ha)	33.8%	76.0%	0.32	0.21
Flow accumulation (6 ha)	21.5%	71.3%	0.26	0.26
Flow accumulation (10 ha)	17.3%	70.2%	0.24	0.29
Deep learning (Lidberg et al., 2023)	82.1%	25.7%	0.63	0.09
Deep learning (Ditches0.5)	92.1%	55.9%	0.68	0.29
Deep learning (Streams0.5)	81.5%	70.8%	0.59	0.32

ditches) of 16.9%.

For ditches, our model Ditches0.5 had the highest recall: 92.1% against the reported 86.0% of ditch pixels from Lidberg et al. (2023); 33.8% (2 ha), 21.5% (6 ha), and 17.3% (10 ha) from the flow accumulation; and 27.5% from the Swedish property map. We believe that the differences between our deep learning model and the one from Lidberg et al. (2023), for either channel type, comes from the resolution: their model used 1 m, whereas our data was at a finer 0.5 m one. The lower recall rates of ditch pixels from the flow accumulation and Swedish property map could be explained by the absence of the ditch network, reported to be 91% missing from Swedish maps (Flyckt et al., 2022) before the use of deep learning.

Despite having a high recall rate for stream pixels, the MCC values had a low performance in both the baseline data and deep learning models, showing that there could be a bias towards finding positives at the expense of accuracy. In conclusion, our deep learning-based method for detecting channels outperformed traditional methods regarding ditches, where the recall reached 92.1%, but did not outperform the detection of natural streams. However, while one might argue that missing 29.2% of headwaters still requires further improvement, these results demonstrate that deep learning holds significant promise for improving automatic headwater mapping.

3.6. Limitations and future research

We believe that more studies are needed to improve the performance of class separation. Extracting additional features to the channels and training a separate model with them might improve the classification, especially with attributes related to drainage. The use of hydrological features in the future might answer whether the channel contains water or not and improve the network connectivity, avoiding the interruption of channels in the inference (Fig. 9). However, defining the banks of low relief channels can be particularly challenging if there are wetlands along the river course (Wohl, 2017), something that was observed in the study areas, causing the interruption of visible channels in the HPMF. Adding future information about culverts (Lidberg, 2025) and bridges might impact the inference connectivity as well. To deal with these occurrences, traditional topographic modeling could be applied, and with techniques such as burning and breaching, it might be possible to create the missing connectivity in the ground truth.

The dense canopy cover could have impacted the classification of small streams, potentially affecting the comparison of resolution performance too. The number of laser points is directly related to the resolution of the calculated DEM, however, as the canopy coverage becomes more dense in forested areas, the number of laser points that are able to penetrate it decreases (Chasmer et al., 2004). This could result in wrong terrain elevation estimates for densely covered areas, lowering the performance of the classification of small natural streams. With a higher amount of training data, it would be possible to separate the forested areas from the open ones to train the models, evaluating how much the tree tops were impacting the resolution performance. However, doing so with these datasets would result in a lower performance overall.

At the same time, while adding more data for this type of channel might seem like a solution, Yang et al. (2022) showed that this might not necessarily improve the models. Not only that, but the most time-consuming and expensive part of training a model with machine learning is acquiring the ground truth data, which in this study is due to the manual labeling and classification of channels relying on the terrain data and ortophotos. However, in dense vegetation covered sites, the ortophotos were not helpful, requiring an expert to visit the location and evaluate the channel type, which in turn increased the costs and time of the process. Despite these difficulties, the inclusion of aerial photographs and other data sources combined with ALS might be beneficial to the models, adding new characteristics to the channels.

Forwarding ruts were not observed in our dataset, but we acknowledge that this could be a cause for false positives. Some publications have focused on their identification using image data from drones (Bhatnagar et al., 2022) or conventional cameras (Pierzchała et al., 2016) unlike our study, which was based on the DEM. Another issue is that the vegetation can hinder the visual identification of these structures, making it hard to remove them from the data.

3.7. Water channel management and policies

Knowing the ambitious scope of the suggested actions by Agenda (2030) regarding water ecosystems, the management of both types of channels needs to be addressed. The measures allowed depend on the type of channel: riparian buffers are prescribed around streams, while ditch channels can be cleaned without permits (Swedish PEFC, 2023). Most ditches were detected in this study; however, streams were often misclassified as ditches. This is a cause for concern as streams have stronger protection policies than ditches during forest management. For example, crossing streams with heavy forest machinery should be avoided according to best management practices (Skogsstyrelsen, 2016) to avoid disturbing soils near and in the stream; such disturbance causes downstream sedimentation (Bishop et al., 2009). Meanwhile, ditches are not protected, and the full length of the ditch can be dug out and cleaned, also causing downstream sedimentation (Bishop et al., 2009); management procedures applied on natural channels would negatively change their characteristics, such as flow patterns and retention potential of detritus input (Muotka et al., 2002). Therefore, streams misclassified as ditches on maps could lead to the deterioration of both local and downstream environments if these maps were unquestioningly trusted by practitioners.

We suggest caution then when implementing models trained on just ditches: our model trained on this dataset misclassified 50% of the stream channels as ditch channels. This advice also concerns the ditch map developed by Lidberg et al. (2023). We are confident that further studies on how to separate ditches and streams on maps are needed.

A restoration process is currently underway to turn some of the Finnish channelized streams back to their natural status, thus improving sport fisheries (Erkinaro et al., 2011), while demonstration restorations have also been done in a number of Swedish rivers (Gardeström et al., 2013). However, studies focused on the restoration of small stream channels (<6 m) of the sort that we investigated are still missing. A

better classification of natural streams can benefit these studies and practices, further helping us to reach the water goals set by the Agenda 2030.

4. Conclusion

With this work, we have identified several key findings:

1) Resolution impact: The 0.5 m resolution significantly improved the detection of both ditches and natural stream channels, leading to higher overall performance. However, the finer resolution also required more computing power for processing the training data, training and testing the model, and running inference. highlighting the need for parallelizing the code and executing it on the GPU.
2) Topographic Index Performance: The highest-scoring topographic index varied depending on the dataset. The High-Pass Median Filter performed best for Channels0.5 and Ditches&Streams0.5 (ditch label), while the Hillshade 90° was the top-ranking for Ditches0.5. For Streams0.5, Hillshade 0° ranked higher.
3) Combining indices: Using a combination of indices resulted in higher values of MCC than single indices, with the combination of Sky-view Factor and Slope having the highest value for the stream label.
4) U-net performance: Our deep learning model Ditches0.5 was able to detect ditches better than any previous method (Table 5). In comparison with traditional mapping methods, the detection for ditches increased from less than 40% to over 92%, while Streams0.5 could map 70.8% of stream pixels.

Hence, our study shows great potential for using deep learning for mapping small headwaters, whether natural or man-made. However, the detection of natural streams still needs improving as close to 30% of them are still missing on the resulting maps. Future research should focus on identifying shared morphological features between ditch and stream channels, exploring methods to reduce class imbalance, and incorporating additional data such as information on soils, catchment area, and channel morphology. Improving automatic channel detection and classification of natural and man-made channels can provide valuable support for future improved management decisions for surface waters and optimize resource allocation for landscape planning.

CRediT authorship contribution statement

Mariana Dos Santos Toledo Busarello: Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis, Data curation. **Anneli M. Ågren:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization. **Florian Westphal:** Writing – review & editing, Visualization, Software, Methodology, Investigation. **William Lidberg:** Writing – review & editing, Supervision, Software, Resources, Methodology, Funding acquisition, Conceptualization.

Code availability section

Contact: Mariana Dos Santos Toledo Busarello, Skogsmarksgränd 901 83 Umeå, Sweden, mariana.busarello@slu.se.

The hardware requirements are an x86 CPU with at least 16 GB of RAM. All code was written in Python, and the software requirements are Python and Docker.

The open source codes (MIT) are available for download at: https://github.com/mbusarello/Automatic-Detection-of-Ditches-and-Natural-Streams-from-Digital-Elevation-Models-Using-Deep-Learning.

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.

Acknowledgements

We thank Liselott Nilsson, Eliza Maher Hasselquist, Elijah Ourth, Siddhartho Shekhar Paul, Amanda Jarefjäll, Gudrun Norstedt, Lars Strand, Anders Hejnebo, Björn Lehto, Magnus Martinsson, Marcus Björsell, Tobias Johansson, Andrew Landström, and Catarina Welin for digitizing the ditches and streams used to train the model in this study. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation, the Marcus and Amalia Wallenberg Foundation, Kempestiftelserna, the Swedish Forest Agency, Formas (Proj no. 2021-00115), and Future Silviculture (2018.0259) financed by the Knut and Alice Wallenberg Foundation.

Data availability

The data used in this project can be found at the link: https://doi.org/10.5878/jrex-z325.

References

Ågren, A.M., Lidberg, W., Ring, E., 2015. Mapping temporal dynamics in a forest stream network – implications for riparian forest management. Forests 6, 2982–3001. https://doi.org/10.3390/f6092982.

Alexander, R.B., Boyer, E.W., Smith, R.A., Schwarz, G.E., Moore, R.B., 2007. The role of headwater streams in downstream water quality 1. JAWRA Journal of the American Water Resources Association 43, 41–59. https://doi.org/10.1111/j.1752-1688.2007.00005.x.

Aslam, J.A., Yilmaz, E., Pavlu, V., 2005. A geometric interpretation of r-precision and its correlation with average precision. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Presented at the SIGIR05: the 28th ACM/SIGIR International Symposium on Information Retrieval 2005. Salvador, Brazil, pp. 573–574. https://doi.org/10.1145/1076034.1076134. ACM.

Beschta, R.L., Platts, W.S., 1986. Morphological features of small streams: significance and function 1. JAWRA Journal of the American Water Resources Association 22 (3), 369–379.

Bhatnagar, S., Puliti, S., Talbot, B., Heppelmann, J.B., Breidenbach, J., Astrup, R., 2022. Mapping wheel-ruts from timber harvesting operations using deep learning techniques in drone imagery. Forestry 95 (5), 698–710.

Bishop, K., Buffam, I., Erlandsson, M., Fölster, J., Laudon, H., Seibert, J., Temnerud, J., 2008. *Aqua Incognita* : the unknown headwaters. Hydrol. Process. 22, 1239–1242. https://doi.org/10.1002/hyp.7049.

Bishop, K., Allan, C., Bringmark, L., Garcia, E., Hellsten, S., Högbom, L., Johansson, K., Lomander, A., Meili, M., Munthe, J., Nilsson, M., Porvari, P., Skyllberg, U., Sørensen, R., Zetterberg, T., Åkerblom, S., 2009. The effects of forestry on Hg bioaccumulation in nemoral/boreal waters and recommendations for good silvicultural practice. AMBIO A J. Hum. Environ. 38, 373–380. https://doi.org/10.1579/0044-7447-38.7.373.

Busarello, M.d.S.T., Lidberg, W., Ågren, A., Westphal, F., 2024. Automatic Detection of Ditches and Natural Streams from Digital Elevation Models Using Deep Learning (Version 1). Swedish University of Agricultural Sciences. https://doi.org/10.5878/jrex-z325 [Data set].

Chasmer, L., Hopkinson, C., Treitz, P., 2004. Assessing the three-dimensional frequency distribution of airborne and ground-based LiDAR data for red pine and mixed deciduous forest plots. International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences 36, 66–69.

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 21, 6. https://doi.org/10.1186/s12864-019-6413-7.

Chowdhuri, I., Pal, S.C., Saha, A., Chakrabortty, R., Roy, P., 2021. Evaluation of different DEMs for gully erosion susceptibility mapping using in-situ field measurement and validation. Ecol. Inf. 65, 101425. https://doi.org/10.1016/j.ecoinf.2021.101425.

Claessens, L., Tague, C.L., Groffman, P.M., Melack, J.M., 2010. Longitudinal and seasonal variation of stream N uptake in an urbanizing watershed: effect of organic matter, stream size, transient storage and debris dams. Biogeochemistry 98, 45–62.

Clough, P., Sanderson, M., Tang, J., Gollins, T., Warner, A., 2013. Examining the limits of crowdsourcing for relevance assessment. IEEE Internet Computing 17, 32–38. https://doi.org/10.1109/MIC.2012.95.

Council of the European Union, 2023. Interinstitutional File: 2022/0195 (COD). No. 15907/23). Brussels.

Cox, E.J., Gurnell, A.M., Bowes, M.J., Bruen, M., Hogan, S.C., O'Sullivan, J.J., Kelly-Quinn, M., 2023. A multi-scale analysis and classification of the hydrogeomorphological characteristics of Irish headwater streams. Hydrobiologia 850 (15), 3391–3418.

Desmond, M., Duesterwald, E., Brimijoin, K., Brachman, M., Pan, Q., 2021. Semi-automated data labeling. J. Mach. Learn. Res. 133, 159–169.

Du, L., You, X., Li, K., Meng, L., Cheng, G., Xiong, L., Wang, G., 2019. Multi-modal deep learning for landform recognition. ISPRS J. Photogrammetry Remote Sens. 158, 63–75. https://doi.org/10.1016/j.isprsjprs.2019.09.018.

Du, L., McCarty, G.W., Li, X., Zhang, Xin, Rabenhorst, M.C., Lang, M.W., Zou, Z., Zhang, Xuesong, Hinson, A.L., 2024. Drainage ditch network extraction from lidar data using deep convolutional neural networks in a low relief landscape. J. Hydrol. 628, 130591. https://doi.org/10.1016/j.jhydrol.2023.130591.

Eigen, D., Fergus, R., 2015. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: 2015 IEEE International Conference on Computer Vision (ICCV). Presented at the 2015. IEEE International Conference on Computer Vision (ICCV), IEEE, Santiago, Chile, pp. 2650–2658. https://doi.org/10.1109/ICCV.2015.304.

Erkinaro, J., Laine, A., Mäki-Petäys, A., Karjalainen, T.P., Laajala, E., Hirvonen, A., Orell, P., Yrjänä, T., 2011. Restoring migratory salmonid populations in regulated rivers in the northernmost Baltic Sea area, Northern Finland - biological, technical and social challenges: restoring migratory salmonid populations in regulated rivers. J. Appl. Ichthyol. 27, 45–52. https://doi.org/10.1111/j.1439-0426.2011.01851.x.

Fei, J., Liu, J., Ke, L., Wang, W., Wu, P., Zhou, Y., 2022. A deep learning-based method for mapping alpine intermittent rivers and ephemeral streams of the Tibetan Plateau from Sentinel-1 time series and DEMs. Rem. Sens. Environ. 282, 113271.

Florinsky, I.V., 2016. Digital terrain modeling. In: Digital Terrain Analysis in Soil Science and Geology. Elsevier. https://doi.org/10.1016/B978-0-12-804632-6.00001-8.

Flyckt, J., Andersson, F., Lavesson, N., Nilsson, L., Å gren, A.M., 2022. Detecting ditches using supervised learning on high-resolution digital elevation models. Expert Syst. Appl. 201, 116961. https://doi.org/10.1016/j.eswa.2022.116961.

Gao, K., Huang, L., Zheng, Y., 2022. Fault detection on seismic structural images using a nested residual U-net. IEEE Trans. Geosci. Rem. Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2021.3073840.

Gardeström, J., Holmqvist, D., Polvi, L.E., Nilsson, C., 2013. Demonstration restoration measures in tributaries of the vindel river catchment. Ecol. Soc. 18, art8. https://doi.org/10.5751/ES-05609-180308.

Hasselquist, E.M., Mancheva, I., Eckerberg, K., Laudon, H., 2020. Policy change implications for forest water protection in Sweden over the last 50 years. Ambio 49, 1341–1351. https://doi.org/10.1007/s13280-019-01274-y.

Huryn, A.D., Wallace, J.B., 1987. Local geomorphology as a determinant of macrofaunal production in a mountain stream. Ecology 68 (6), 1932–1942.

Isikdogan, F., Bovik, A.C., Passalacqua, P., 2017. Surface water mapping by deep learning. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 10 (11), 4909–4918.

Kampffmeyer, M., Salberg, A.-B., Jenssen, R., 2016. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Las Vegas, NV, USA, pp. 680–688. https://doi.org/10.1109/CVPRW.2016.90.

Kazimi, B., Thiemann, F., Sester, M., 2020. Detection of terrain structures in Airborne Laser Scanning data using deep learning. ISPRS Annals of the Photogrammetry. Remote Sensing and Spatial Information Sciences 493–500. https://doi.org/10.5194/isprs-annals-V-2-2020-493-2020. V-2–2020.

Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization. Conference Track Proceedings, San Diego, CA, USA.

Kokalj, Ž., Zakšek, K., Oštir, K., Pehani, P., Čotar, K., Somrak, M., 2016. Relief Visualization Toolbox, Ver. 2.2.1 Manual.

Korznikov, K.A., Kislov, D.E., Altman, J., Doležal, J., Vozmishcheva, A.S., Krestov, P.V., 2021. Using U-Net-Like deep convolutional neural networks for precise tree recognition in very high resolution RGB (red, green, blue) satellite images. Forests 12, 66. https://doi.org/10.3390/f12010066.

Koski, C., Kettunen, P., Poutanen, J., Zhu, L., Oksanen, J., 2023. Mapping small watercourses from DEMs with deep learning—exploring the causes of false predictions. Rem. Sens. 15, 2776. https://doi.org/10.3390/rs15112776.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P., 2006. Handling imbalanced datasets: a review. GESTS international transactions on computer science and engineering 30 (1), 25–36.

Kuglerová, L., Hasselquist, E.M., Richardson, J.S., Sponseller, R.A., Kreutzweiser, D.P., Laudon, H., 2017. Management perspectives on *Aqua incognita* : connectivity and cumulative effects of small natural and artificial streams in boreal forests. Hydrol. Process. 31, 4238–4244. https://doi.org/10.1002/hyp.11281.

Kuglerová, L., Jyväsjärvi, J., Ruffing, C., Muotka, T., Jonsson, A., Andersson, E., Richardson, J.S., 2020. Cutting edge: a comparison of contemporary practices of riparian buffer retention around small streams in Canada, Finland, and Sweden. Water Resour. Res. 56, e2019WR026381. https://doi.org/10.1029/2019WR026381.

Kwame, A.E., Martey, E.M., Chris, A.G., 2017. Qualitative assessment of compiled, interpreted and hybrid programming languages. Communications on Applied Electronics 7 (7), 8–13.

Lanka, R.P., Hubert, W.A., Wesche, T.A., 1987. Relations of geomorphology to stream habitat and trout standing stock in small Rocky Mountain streams. Trans. Am. Fish. Soc. 116 (1), 21–28.

Lantmäteriet, 2014. GSD-fastighetskartan, Vektor.

Lantmäteriet, 2021a. Ortophoto.

Lantmäteriet, 2021b. Laser Data.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lidberg, W., 2025. Deep learning-enhanced detection of road culverts in high-resolution digital elevation models: improving stream network accuracy in Sweden. J. Hydrol.: Reg. Stud. 57, 102148. https://doi.org/10.1016/j.ejrh.2024.102148.

Lidberg, W., Nilsson, M., Lundmark, T., Ågren, A.M., 2017. Evaluating preprocessing methods of digital elevation models for hydrological modelling. Hydrol. Process. 31, 4660–4668. https://doi.org/10.1002/hyp.11385.

Lidberg, W., Paul, S.S., Westphal, F., Richter, K.F., Lavesson, N., Melniks, R., Ivanovs, J., Ciesielski, M., Leinonen, A., Ågren, A.M., 2023. Mapping drainage ditches in forested landscapes using deep learning and aerial laser scanning. J. Irrigat. Drain. Eng. 149, 04022051. https://doi.org/10.1061/JIDEDH.IRENG-9796.

Lindsay, J.B., 2016. Whitebox GAT: a case study in geomorphometric analysis. Comput. Geosci. 95, 75–84. https://doi.org/10.1016/j.cageo.2016.07.003.

Lindsay, J.B., Dhun, K., 2015. Modelling surface drainage patterns in altered landscapes using LiDAR. Int. J. Geogr. Inf. Sci. 29, 397–411. https://doi.org/10.1080/13658816.2014.975715.

Martínez, A., Larranaga, A., Basaguren, A., Pérez, J., Mendoza-Lera, C., Pozo, J., 2013. Stream regulation by small dams affects benthic macroinvertebrate communities: from structural changes to functional implications. Hydrobiologia 711, 31–42.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta Protein Struct. 405, 442–451. https://doi.org/10.1016/0005-2795(75)90109-9.

Mazhar, S., Sun, G., Bilal, A., Hassan, B., Li, Y., Zhang, J., Lin, Y., Khan, A., Ahmed, R., Hassan, T., 2022. AUnet: a deep learning framework for surface water channel mapping using large-coverage remote sensing images and sparse scribble annotations from OSM data. Rem. Sens. 14 (14), 3283.

Muotka, T., Paavola, R., Haapala, A., Novikmec, M., Laasonen, P., 2002. Long-term recovery of stream habitat structure and benthic invertebrate communities from in-stream restoration. Biol. Conserv. 105, 243–253. https://doi.org/10.1016/S0006-3207(01)00202-6.

Nieminen, M., Piirainen, S., Sikström, U., Löfgren, S., Marttila, H., Sarkkola, S., Laurén, A., Finér, L., 2018. Ditch network maintenance in peat-dominated boreal forests: review and analysis of water quality management options. Ambio 47, 535–545. https://doi.org/10.1007/s13280-018-1047-6.

NLS, 2023. NLS orthophotos [WWW Document]. URL. https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/orthophotos.

O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Comput. Vis. Graph Image Process 28, 323–344.

Paul, S.S., Hasselquist, E.M., Jarefjäll, A., Ågren, A.M., 2023. Virtual landscape-scale restoration of altered channels helps us understand the extent of impacts to guide future ecosystem management. Ambio 52, 182–194. https://doi.org/10.1007/s13280-022-01770-8.

Peacock, M., Audet, J., Bastviken, D., Futter, M.N., Gauci, V., Grinham, A., Harrison, J. A., Kent, M.S., Kosten, S., Lovelock, C.E., Veraart, A.J., Evans, C.D., 2021a. Global importance of methane emissions from drainage ditches and canals. Environ. Res. Lett. 16, 044010. https://doi.org/10.1088/1748-9326/abeb36.

Peacock, M., Granath, G., Wallin, M.B., Högbom, L., Futter, M.N., 2021b. Significant emissions from forest drainage ditches—an unaccounted term in anthropogenic greenhouse gas inventories? J. Geophys. Res.: Biogeosciences 126. https://doi.org/10.1029/2021JG006478.

Phinzi, K., Abriha, D., Bertalan, L., Holb, I., Szabó, S., 2020. Machine learning for gully feature extraction based on a pan-sharpened multispectral image: multiclass vs. Binary approach. International Journal of Geo-Information 9, 252. https://doi.org/10.3390/ijgi9040252.

Pierzchała, M., Talbot, B., Astrup, R., 2016. Measuring wheel ruts with close-range photogrammetry. Forestry: Int. J. Financ. Res. 89 (4), 383–391.

Raber, G.T., Jensen, J.R., Schill, S.R., Schuckman, K., 2002. Creation of digital terrain models using an adaptive lidar vegetation point removal process. Photogramm. Eng. Rem. Sens. 68, 1307–1315.

Ring, E., Lode, P.E., Libiete, Z., Oksanen, E., Gil, W., Šimkevičius, K., 2018. Good Practices for Forest Buffers to Improve Surface Water Quality in the Baltic Sea Region. Arbetsrapport No. 995–2018).

Robb, C., Pickard, A., Williamson, J.L., Fitch, A., Evans, C., 2023. Peat drainage ditch mapping from aerial imagery using a convolutional neural network. Rem. Sens. 15, 499. https://doi.org/10.3390/rs15020499.

Robson, B.A., Bolch, T., MacDonell, S., Hölbling, D., Rastner, P., Schaffer, N., 2020. Automated detection of rock glaciers using deep learning and object-based image analysis. Rem. Sens. Environ. 250, 112033. https://doi.org/10.1016/j.rse.2020.112033.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III. Springer International Publishing, pp. 234–241, 18.

Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V., 2021. U-net and its variants for medical image segmentation: a review of theory and applications. IEEE Access 9, 82031–82057.

Skogsstyrelsen, 2013. SKSFS 2013:2.

Skogsstyrelsen, 2016. Nya Och Reviderade Målbilder För God Miljöhänsyn, No. 2016, p. 12.

Swedish PEFC, 2023. Forest use standard. Technical Report PEFC SWE 002:5. Swedish PEFC.

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. J. Am. Stat. Assoc. 82 (398), 528–540.

United Nations General Assembly, 2015. Transforming Our World: the 2030 Agenda for Sustainable Development.

Wiens, J.A., 2002. Riverine landscapes: taking landscape ecology into the water. Freshw. Biol. 47 (4), 501–515.

Wilson, J.P., Gallant, J.C. (Eds.), 2000. Terrain Analysis: Principles and Applications. Wiley, New York.
Wohl, E., 2017. The significance of small streams. Front. Earth Sci. 11, 447–456.
Xu, S., Song, Y., Hao, X., 2022. A comparative study of shallow machine learning models and deep learning models for landslide susceptibility assessment based on imbalanced data. Forests 13, 1908. https://doi.org/10.3390/f13111908.
Yang, J., Zhang, Z., Gong, Y., Ma, S., Guo, X., Yang, Y., Xiao, S., Wen, J., Li, Y., Gao, X., Lu, W., Meng, Q., 2022. Do deep neural networks always perform better when eating more data? arXiv 2022. arXiv preprint arXiv:2205.15187.
Yang, J., Xu, J., Lv, Y., Zhou, C., Zhu, Y., Cheng, W., 2023. Deep learning-based automated terrain classification using high-resolution DEM data. Int. J. Appl. Earth Obs. Geoinf. 118, 103249. https://doi.org/10.1016/j.jag.2023.103249.
Zakšek, K., Oštir, K., Kokalj, Ž., 2011. Sky-view factor as a relief visualization technique. Rem. Sens. 3, 398–415. https://doi.org/10.3390/rs3020398.

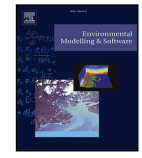
III

Environmental Modelling and Software 191 (2025) 106488

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Uncertainty quantification for LiDAR-based maps of ditches and natural streams

Florian Westphal [a,*], William Lidberg [b], Mariana Dos Santos Toledo Busarello [b], Anneli M. Ågren [b]

[a] Jönköping AI Lab, School of Engineering, Jönköping University, Gjuterigatan 5, Jönköping, 551 11, Sweden
[b] Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Skogsmarksgränd 17, Umeå, 901 83, Sweden

ARTICLE INFO

Dataset link: Uncertainty Quantification for LiDAR-based Maps of Ditches and Natural Streams (Original data), Automatic Detection of Ditches and Natural Streams from Digital Elevation Models Using Deep Learning (Reference data)

Keywords:
Semantic segmentation
Uncertainty quantification
Monte Carlo dropout
Conformal prediction
Small-scale hydrology
LiDAR

ABSTRACT

This article compares novel and existing uncertainty quantification approaches for semantic segmentation used in remote sensing applications. We compare the probability estimates produced by a neural network with Monte Carlo dropout-based approaches, including predictive entropy and mutual information, and conformal prediction-based approaches, including feature conformal prediction (FCP) and a novel approach based on conformal regression. The chosen task focuses on identifying ditches and natural streams based on LiDAR derived digital elevation models. We found that FCP's uncertainty estimates aligned best with the neural network's prediction performance, leading to the lowest Area Under the Sparsification Error curve of 0.09. For finding misclassified instances, the network probability was most suitable, requiring a correction of only 3% of the test instances to achieve a Matthews Correlation Coefficient (MCC) of 0.95. Conformal regression produced the best confident maps, which, at 90% confidence, covered 60% of the area and achieved an MCC of 0.82.

1. Introduction

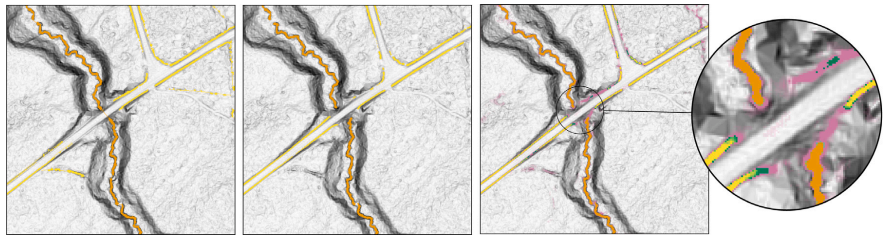
Having accurate maps of a landscape is crucial for supporting informed decisions in various applications, including sustainable land-use management (Pagella and Sinclair, 2014). Creating large-scale maps, such as those covering an entire country, is a labor-intensive process that requires significant human effort. Consequently, the automated analysis of remote sensing data has become a common solution (Blaschke, 2010). This involves the analysis of data from sources such as optical images, synthetic aperture radar, hyperspectral imaging, and Light Detection and Ranging (LiDAR) (Toth and Jóźków, 2016). Historically, traditional computer graphics-based approaches have been used for remote sensing applications (Savelonas et al., 2022), but more recently, deep learning-based methods have been used successfully in these applications (Yuan et al., 2020). Deep learning-based approaches tend to convert the remote sensing data into images, and apply semantic segmentation to assign one of the classes of interest to every pixel of the image. For example, O'Neil et al. (2020) have mapped wetlands based on aerial images and topographic indices calculated based on a LiDAR derived digital elevation model (DEM). Similarly, Busarello et al. (2025) have investigated the use of different topographic indices as representation of a DEM derived from LiDAR data. Based on these rasterized representations, they trained a neural network to detect ditches and natural streams.

One challenge when working with automatically generated maps is assessing their reliability. A common approach to estimating the quality of these maps is by comparing them with a representative portion of the actual landscape, which provides a good general estimate as long as the evaluated landscape is representative of the overall terrain. However, the actual quality can vary significantly depending on location, with some parts being more accurate and others less so (Kasraei et al., 2021). For decision-making purposes, it is important to have an estimate of reliability at specific locations, which can be achieved by quantifying the uncertainty of the used model at the point of interest (Xu et al., 2022).

Quantifying uncertainty in deep learning models initially appears straightforward, as they typically provide class-wise probabilities for each pixel. However, research has shown that these estimates tend to be overconfident, due to the training process rewarding overconfident predictions (Guo et al., 2017; Sensoy et al., 2018). In response, various methods have been developed to quantify neural network uncertainty, which Gawlikowski et al. (2023) categorize into four primary

* Corresponding author.
E-mail address: florian.westphal@ju.se (F. Westphal).

https://doi.org/10.1016/j.envsoft.2025.106488
Received 17 December 2024; Received in revised form 14 April 2025; Accepted 21 April 2025
Available online 2 May 2025



(a) Groundtruth Labels (b) Predicted Labels (c) Visualized Uncertainty

Fig. 1. Illustration of the semantic segmentation task. Ditches (yellow) and natural streams (orange) should be identified in a given chip based on the slope image derived from the digital elevation model at a 0.5 m resolution. The uncertainty of the 5% most uncertain pixels, as quantified by Feature Conformal Prediction is displayed using pink for background pixels, green for ditches and blue for streams. The strength of the color is determined by the uncertainty value.

directions: single network deterministic approaches, Bayesian methods, ensemble techniques, and test-time augmentation methods.

Deterministic methods, such as Dirichlet prior networks (Gawlikowski et al., 2022), have been used in remote sensing applications, as well as ensemble techniques, such as deep ensembles (Lakshminarayanan et al., 2017). For example, Chaudhary et al. (2022) utilized deep ensembles to quantify uncertainty in generated maximum water depth hazard maps, which aid in estimating the risk of flooding. Additionally, deep ensembles have been leveraged to estimate the uncertainty in wavelength bands from Sentinel-2 whose spatial resolution had been enhanced to a resolution of 10 m (Iagaru and Gottschling, 2023).

However, the primary focus has centered on Bayesian methods. The most prevalent approach among these Bayesian methods is Monte Carlo dropout (MC dropout) (Gal and Ghahramani, 2016), which has been used, for example, by Kampffmeyer et al. (2016) to quantify the uncertainty of their method on an urban object classification task based on a digital surface model (DSM). MC dropout has also been used by Martínez-Ferrer et al. (2022) for uncertainty quantification of their approach to retrieve different biophysical variables, such as leaf area index and canopy water content from surface reflectance data. Another notable Bayesian approach involves the application of Bayesian neural networks (Blundell et al., 2015; Goan and Fookes, 2020). Hertel et al. (2023) have conducted a comparative analysis of both methodologies and advocate for the use of Bayesian neural networks, as they tend to be less likely to indicate high confidence in incorrect predictions.

One other approach to uncertainty quantification is the conformal prediction framework (Vovk et al., 2005), which has been primarily applied to simple classification and regression tasks, but more recently was adapted to semantic segmentation. For example, Wieslander et al. (2021) have used conformal prediction for medical image segmentation, while Labuzzetta (2022) has applied subsample conformal prediction to the task of surface water and grassed waterway segmentation. Additionally, Singh et al. (2024) have demonstrated how conformal prediction can be applied to different tasks in earth observation, such as tree species mapping, land cover classification and canopy height estimation, and advocate for its more widespread use. While these works are based on more traditional formulations of conformal prediction, Teng et al. (2023) have proposed Feature Conformal Prediction (FCP), which is particularly adjusted to the use with deep neural networks, and has been shown to be more effective at quantifying the uncertainty of a neural network in general semantic segmentation tasks.

This article compares uncertainty estimates derived from the predictions of a neural network (network probability) with mutual information and predictive entropy — two uncertainty metrics calculated through MC dropout — to those obtained via conformal regression and FCP. We focus on these methods, in contrast to Bayesian neural networks (Blundell et al., 2015) or deep ensembles (Lakshminarayanan et al., 2017), since they can be integrated into existing network architectures for semantic segmentation tasks, and do not incur extensive training times, due to the need to train multiple models. Notably, conformal prediction-based methods enable the production of predictions with a specified confidence level. Ideally, this would result in a map featuring only confident predictions, such as those above a 90% confidence level. Therefore, we investigate the usefulness of those confident maps.

For our comparison, we select the remote sensing task of detecting ditches and natural streams from a DEM (Fig. 1), which has been derived from LiDAR data. In particular, we perform this detection task on data derived from a DEM at 1 m resolution, as well as at 0.5 m resolution. This task is especially challenging due to the narrowness of the objects of interest, requiring high detection precision. In contrast to other semantic segmentation problems, most pixels are background pixels, while only few represent ditches and even fewer represent natural streams, leading to a significant class imbalance. Additionally, distinguishing between streams and ditches in a DEM can be difficult, as they often appear similar. These challenges contribute to uncertainty in predictions, which we aim to estimate.

Uncertainty quantification is crucial in this context because it could help identify natural streams that have been erroneously predicted as ditches. This distinction is significant, as natural streams require distinct management strategies to preserve their ecological integrity (Swedish PEFC, 2023). For example, avoiding the crossing of these streams with heavy machinery can prevent soil disturbance, which otherwise can exacerbate sedimentation and disrupt ecological functions (Bishop et al., 2009). In contrast, ditches can be more easily cleaned or maintained without needing permits.

This article addresses the following research questions:

1. Which of the investigated uncertainty quantification approaches, i.e., network probability, mutual information, predictive entropy, conformal regression, and FCP produces the most reliable uncertainty estimates?
2. To what degree does the resolution of the DEM impact the uncertainty estimates?
3. To what extent is it possible to generate useful maps with a specific confidence level using conformal regression and FCP?

2. Methodology

2.1. Mapping ditches and streams: Network probability

For mapping ditches and streams, our approach employs a U-Net architecture (Ronneberger et al., 2015) similar to that used by Busarello et al. (2025) (Fig. 2), which has been demonstrated to be effective for this task. The U-Net takes as input a 500×500 pixels large chip of the landscape represented by the local slope derived from a DEM. This

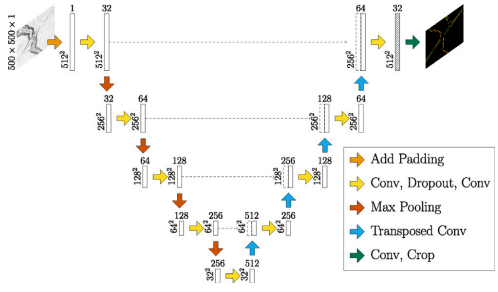


Fig. 2. U-Net architecture for mapping streams and ditches. The colored arrows show different processing steps, the dashed arrows indicate concatenation of feature maps, and the shaded feature maps indicate the ones being used for Feature Conformal Prediction.

input is then downsampled through a series of convolutional, dropout, and max pooling layers. Notably, our approach differs from Busarello et al. (2025) in that we utilize concrete dropout (Gal et al., 2017), which has been shown to improve uncertainty estimates obtained through MC dropout (Mukhoti and Gal, 2018).

After four downsampling steps, the extracted feature maps are upsampled using transposed convolutions, and processed by convolution and dropout layers to reach the original input size. At each upsampling step, the feature maps of the corresponding downsampling step are concatenated to ensure that no relevant information is lost. The final output is produced by applying a convolutional layer to the last feature maps (shaded feature maps in Fig. 2) . The output consists of three bands, each representing one of the considered classes: background, ditch, and natural stream.

In contrast to most U-Net architectures, our approach does not utilize a softmax layer, which would map the output at each pixel to a probability distribution over the three classes and be trained using cross entropy loss. Instead, we employ a linear activation function in the last convolutional layer and train the network using mean squared error, as proposed by Teng et al. (2023) to improve uncertainty estimates of FCP. Labels are mapped farther apart using a double log transform, resulting in large positive and negative values. Unlike Teng et al. (2023), who applied a Gaussian blur to the labels, we found this approach to be detrimental to performance, likely due to the narrow nature of our objects of interest, i.e., ditches and streams. To address class imbalance, we implement median frequency balancing (Eigen and Fergus, 2015) as suggested by Busarello et al. (2025).

Uncertainty estimates are derived from predicted network probabilities. This involves reversing the double log transform to obtain probabilities between 0 and 1 for each pixel and class. It should be noted that these probabilities are not calibrated in any way. The class with the highest probability is selected for each pixel. Uncertainty values are then calculated as the difference between the predicted probability and 1. This approach assumes that high confidence predictions yield probabilities close to 1, whereas low confidence predictions result in lower probabilities and thus higher uncertainty values.

2.2. MC dropout: Predictive entropy and mutual information

MC dropout has been proposed by Gal and Ghahramani (2016) as a method for estimating the uncertainty of a neural network. The main idea behind MC dropout is that if a neural network is certain about its prediction, introducing small random changes in its execution will not affect its prediction. Conversely, when a network is uncertain about its prediction, these small changes will lead to large variations in the predicted outcome. Thus, the network's uncertainty can be estimated by observing the variability in its predicted output when run multiple times. MC dropout introduces small random changes using dropout layers within the network architecture.

In a dropout layer (Srivastava et al., 2014), a randomly selected subset of neurons has its output set to zero. At each new input, a predefined probability determines which neurons are dropped. This probability is learned in concrete dropout (Gal et al., 2017), which we use in this study. Unlike the traditional use of dropout layers, which typically activates them only during training to promote robustness, MC dropout keeps those layers active during inference, resulting in varying outputs for identical inputs processed multiple times.

MC dropout estimates the uncertainty by using these varying outputs to compute two different metrics: predictive entropy and mutual information. These metrics measure different types of uncertainty, viz. aleatoric and epistemic uncertainty. Aleatoric uncertainty captures uncertainty caused by the data, such as ambiguity at the border between ditch and background, whereas epistemic uncertainty captures uncertainty caused by the model itself, for example, due to insufficient training data.

Predictive entropy captures both aleatoric and epistemic uncertainty and is approximated for a given input $\boldsymbol{x}$ and a given training set D_{train} as:

$$\hat{\mathbb{H}}[y|\boldsymbol{x}, D_{train}] = - \sum_{c \in C} \left(\frac{1}{T} \sum_{t=1}^{T} p\left(y = c|\boldsymbol{x}, \hat{w}_t\right) \right) \ln \left(\frac{1}{T} \sum_{t=1}^{T} p\left(y = c|\boldsymbol{x}, \hat{w}_t\right) \right) \quad (1)$$

Here, C is the set of classes, T is the number of outputs y to collect for variations of the neural network $\hat{w}_t$, which are produced by the dropout layers, and $p\left(y = c|\boldsymbol{x}, \hat{w}_t\right)$ is the probability of input $\boldsymbol{x}$ being in class c. In contrast, mutual information measures only the epistemic uncertainty and is approximated as:

$$\hat{\mathbb{I}}[y, w|\boldsymbol{x}, D_{train}] = \hat{\mathbb{H}}[y|\boldsymbol{x}, D_{train}] + \frac{1}{T} \sum_{c \in C} \sum_{t=1}^{T} \left(p\left(y = c|\boldsymbol{x}, \hat{w}_t\right) \ln p\left(y = c|\boldsymbol{x}, \hat{w}_t\right) \right) \quad (2)$$

This study computes predictive entropy and mutual information values for each pixel within every output chip, based on 1 000 outputs collected for each chip.

2.3. Conformal regression

Conformal regression is a part of the conformal prediction framework (Vovk et al., 2005), offering guarantees for machine learning model predictions. Unlike standard regression, conformal regression generates prediction intervals rather than single numerical values. The framework ensures that, for a pre-defined percentage of predictions (e.g., 90%), the true value lies within the provided interval. While this can be achieved easily by making this interval arbitrarily large, the challenge lies in finding a narrow yet guarantee-ensuring interval.

While there are two types of conformal regression, this article focuses on the inductive case, as it does not require frequent re-training. Inductive conformal regression estimates the size of the prediction interval based on a calibration set, which is separate from the training, validation, and test datasets. The interval is derived by measuring the difference between the predicted value and the true value for all instances of the calibration set, using a non-conformity function, such as mean absolute error (MAE), resulting in a non-conformity score. Based on a pre-defined confidence-level, e.g., 90%, the difference or non-conformity score of the 90th percentile is selected, and the interval is set as the value predicted by the machine learning model plus or minus the selected value. This ensures that the true value of 90% of instances in the calibration set lies within the produced interval, since their prediction errors were smaller than the one chosen. Because the calibration set is required to be exchangeable with the test set, i.e., they

both come from the same distribution, it can be expected that this guarantee will hold also for unseen instances from the test set.

One issue with the described approach is that it assigns the same interval to all instances, leading to overly large intervals for most of them. This can be addressed by normalizing non-conformity scores through instance difficulty estimation. For example, Cortés-Ciriano and Bender (2019) estimate instance difficulty using MC dropout, recording predicted outputs for the same instance i multiple times with enabled dropout layers and calculating mean μ_i and standard deviation σ_i over those outputs. The non-conformity score α_i is then computed based on the corresponding true value y_i over all instances in the calibration set D_{cal}, resulting in a list of non-conformity scores S, which is then sorted in ascending order.

$$\alpha_i = \frac{|y_i - \mu_i|}{e^{\sigma_i}}$$
$$S = \alpha_1, \ldots, \alpha_q, \text{ with } q = |D_{cal}| \tag{3}$$

Based on this list, the non-conformity score α_p is selected, which corresponds to the chosen confidence level $1 - \epsilon$ (e.g., 0.9 for $\epsilon = 0.1$). For a new instance j, the prediction interval around the mean of the MC dropout samples μ_j is then derived by multiplying the selected α_p with the instance's difficulty, as measured by the standard deviation over the MC dropout samples σ_j (Cortés-Ciriano and Bender, 2019).

$$p = \lceil (1 - \epsilon)(q + 1) \rceil, \text{ for } \alpha_p$$
$$\mu_j \pm \alpha_p \cdot e^{\sigma_j} \tag{4}$$

Another challenge in deriving regression intervals is that the distribution of non-conformity scores may vary depending on certain properties of the instances. For example, when dealing with instances having large true values, the error may be greater than for those with small true values. If this difference in distribution is not taken into account, the derived regression intervals will be larger than necessary for instances with small true values and possibly too narrow for instances with large true values, depending on their prevalence in the calibration set.

For classification problems, Mondrian conformal prediction (Vovk et al., 2005) addresses these issues by categorizing instances based on a Mondrian taxonomy that considers certain properties of each instance. A separate conformal predictor is then built for each category. Mondrian regression, proposed by Boström and Johansson (2020), follows a similar approach. It divides the calibration instances into different categories based on a Mondrian taxonomy, specifically an estimate of difficulty. The prediction interval within each category is derived from the non-conformity score at a specific percentile. This methodology allows for more tailored prediction intervals that are narrower for instances belonging to simpler categories and wider for those in harder categories. Since simpler categories typically have low errors and thus low non-conformity scores, their prediction intervals can be narrower. In contrast, harder categories will have higher non-conformity scores, leading to broader prediction intervals.

In our implementation, each pixel in an input chip is associated with three real values indicating to which of the three classes it belongs. After reverting the double log transform, we perform conformal regression to derive a prediction interval for the three class values of each pixel. Since the class values can be seen as the probability of the pixel to belong to each of the classes, the estimated intervals can be interpreted as probability ranges. The estimation of these intervals involves calculating non-conformity scores per class for every pixel in all calibration set chips, followed by normalization using 100 Monte Carlo samples as proposed by Cortés-Ciriano and Bender (2019).

While we record non-conformity scores per class, we also employ Mondrian conformal regression to obtain more tailored intervals. This approach differs from the original Mondrian taxonomy by Boström and Johansson (2020), which utilized estimated instance difficulty. In contrast, our taxonomy categorizes predictions for each class into two categories: pixels with predicted probabilities close to zero and those near one. This distinction is important because we observed in initial experiments the tendency of classes with few pixels to have most commonly a predicted probability value of zero with a low non-conformity score. Conversely, when the actual class is predicted (i.e., the predicted probability exceeds 0.5), the non-conformity scores tend to be substantially higher. Given this observation, it is reasonable to create categories based on the predicted values.

Thus, we group the non-conformity scores of instances from the calibration set D_{cal} for each class individually into two lists, one for which the predicted probability is lower than 0.5, $S^{<0.5}$, and one for which the predicted probability is larger or equal, $S^{\geq 0.5}$. Those lists are then sorted in ascending order, and the non-conformity scores corresponding to the chosen confidence-level $1 - \epsilon$ are selected as before.

$$S^{\geq 0.5} = \alpha_1^{\geq 0.5}, \ldots, \alpha_r^{\geq 0.5}$$
$$S^{<0.5} = \alpha_1^{<0.5}, \ldots, \alpha_s^{<0.5}, \text{ with } r + s = |D_{cal}|$$
$$t = \lceil (1 - \epsilon)(r + 1) \rceil, \text{ for } \alpha_t^{\geq 0.5}$$
$$u = \lceil (1 - \epsilon)(s + 1) \rceil, \text{ for } \alpha_u^{<0.5} \tag{5}$$

We then calculate intervals for each pixel j in a new chip by collecting 100 Monte Carlo samples of output predictions for the pixel and computing the respective mean μ_j and standard deviation σ_j. Given the selected non-conformity scores and the estimated means and standard deviations, the interval for one of the possible classes for pixel j is derived as follows:

$$\mu_j \pm \left(\mu_j \alpha_t^{\geq 0.5} + (1 - \mu_j) \alpha_u^{<0.5} \right) \cdot e^{\sigma_j} \tag{6}$$

By multiplying the selected non-conformity scores with the probability mean and its inverse respectively, the final interval is derived as combination of both scores depending on how much the pixel's prediction agrees with the respective categories. This way of assigning the corresponding non-conformity score to a pixel is computationally more efficient than having to find the applicable score based on some other feature of the pixel, such as difficulty, via a look-up, as it is the case in the Mondrian approaches by Boström and Johansson (2020), Wieslander et al. (2021), and Labuzzetta (2022).

The uncertainty value for each class is determined by the size of the interval, where a larger interval indicates greater uncertainty in the prediction. Unlike MC dropout, which produces uncertainty values per pixel, the conformal regression approach derives an uncertainty value per pixel per class.

2.4. Feature conformal prediction (FCP)

In contrast to conformal regression, which computes non-conformity scores based on the output of a machine learning model, FCP (Teng et al., 2023) calculates these scores based on an intermediate feature representation of a neural network. This feature representation can be, for example, the feature maps produced by a convolutional layer. These feature maps are then converted into a single vector by flattening the corresponding tensor, enabling FCP to obtain a predicted output for an input instance as a point in a high-dimensional vector space.

When applying conformal regression, it is clear what constitutes a true value for computing the non-conformity score, i.e., the target value of an instance. In contrast, identifying the true feature representation of an instance is not straightforward. FCP assumes this true representation to be the infimum, which corresponds to the feature representation with the smallest numerical values, which produces the correct output. However, finding this optimal representation is challenging. As a result, FCP approximates the infimum by optimizing the original feature representation for a given input instance to produce the correct output using gradient descent. It should be noted that this approach modifies the values of the feature representation rather than adjusting neural network weights. The non-conformity score is then computed using a norm distance, such as the infinity norm, between the vector of the

original representation and the one derived through gradient descent. This yields a single non-conformity score per instance, differing from the conformal regression case where multiple scores are generated corresponding to each output.

The base score is derived, similar to conformal regression, by computing the non-conformity scores for the calibration set and selecting, for example, the 90th percentile. Given a test instance, FCP derives its corresponding feature representation and applies perturbations to this representation, ensuring that the resulting new feature representations do not deviate beyond the distance indicated by the base score. These perturbations are achieved using Linear Relaxation based Perturbation Analysis (LiRPA) (Xu et al., 2020). Subsequently, FCP estimates the resulting output intervals by applying the neural network to the perturbed feature representations. In summary, FCP performs conformal regression in feature space and derives output prediction intervals through perturbation analysis. Mathematical proofs of the correctness and efficiency of the method have been derived by Teng et al. (2023).

Our implementation utilizes feature maps generated prior to the output layer (shaded feature maps in Fig. 2) for FCP. In contrast to Teng et al. (2023), who found that features can be extracted from various layers without altering the prediction intervals, our findings suggest that using feature maps from any other layer results in unreasonably large prediction intervals for our task and network architecture. This may be because the skip connections in our U-Net architecture interfered presumably with the perturbation step, as the perturbations were applied only to the feature maps of the upsampling path and not those of the downsampling path. We employ perturbation analysis to derive prediction intervals for every pixel and class. Similar to our conformal regression implementation, the size of the interval is interpreted as uncertainty, where larger intervals indicate higher uncertainty.

3. Experiments

3.1. Dataset

For this article, we used a dataset provided by Busarello et al. (2025)., consisting of LiDAR-derived DEMs for 12 distinct regions in Sweden, further described by Lidberg et al. (2023). The dataset is available in two resolutions, 0.5 m and 1 m, corresponding to input chips of 500 × 500 pixels representing areas of 250 m×250 m and 500 m×500 m, respectively. To address class imbalance, chips with less than 250 ditch or stream pixels were removed, resulting in a dataset where still only 1.1% and 0.1% of all pixels belong to the ditch and natural stream class, respectively (Busarello et al., 2025).

Topographic indices are utilized to provide a rasterized representation of the DEM. In our experiments, the local slope was used, which signifies the change in elevation between every pixel in the DEM, with inclination displayed in degrees (Florinsky, 2016). This index was chosen due to its superior performance in stream detection and satisfactory results for ditch detection (Busarello et al., 2025). To reduce execution time, we focused on a single index; however, all uncertainty quantification methods remain applicable when multiple indices are considered.

To evaluate the chosen uncertainty quantification methods, we employed 10-fold cross-validation to facilitate statistical analysis. However, since conformal regression and FCP require a calibration set, the dataset was divided into 11 folds: nine for training, one for calibration, and one for testing to ensure exchangeability between folds. Stratified sampling by region ensured that chips in each fold cover the 12 distinct regions similarly well, preserving representativeness throughout training, calibration, and test set.

Apart from ensuring exchangeability, we needed to prevent information about the test set from leaking into the training and calibration set to avoid biasing the evaluation and obtaining miscalibrated uncertainty estimates. This was achieved using the following partitioning strategy. The dataset was divided into chips without overlap, ensuring that no chip's information was shared between training, calibration and test set. Within each region, chips were grouped to minimize borders with adjacent chips in other folds. To optimize this grouping, a heuristic algorithm was used due to the NP-hard nature of the problem[1], yielding an approximate optimal solution for partitioning.

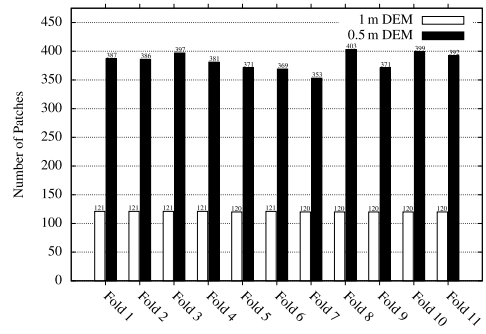


Fig. 3. Number of chips in each of the 11 folds for the digital elevation model (DEM) with resolution 1 m and 0.5 m.

After splitting the chips from the 1 m DEM into 11 folds, the corresponding chips were then selected for the 0.5 m DEM, ensuring that both resolutions contained the same ditches and streams within each fold. This design prevented differences in performance between the two resolutions being attributed to varying levels of complexity, rather than resolution itself. While the number of chips for the 1 m DEM was nearly the same for all folds, this number varied more for the 0.5 m DEM (Fig. 3). The reason for this variation was that a different number of chips was dropped in each fold, depending on the number of 0.5 m DEM chips containing at least 250 ditch or stream pixels.

3.2. Performance metrics

The neural network's performance in classifying pixels as background, ditch, or natural stream was evaluated using the Matthews Correlation Coefficient (MCC) (Matthews, 1975; Yule, 1912) and F_1 score. Given that there were more than two classes, we used the multiclass version of MCC proposed by Gorodkin (2004). MCC provides a balanced view of the classification performance across all classes, while F_1 score focuses on the performance for a specific class, making it particularly suitable for investigating the network's performance for one class of interest (Chicco et al., 2021).

To evaluate the performance of uncertainty quantification approaches, we utilized the Area Under the Sparsification Error Curve (AUSE) (Ilg et al., 2018). Unlike the commonly used Patch Accuracy vs. Patch Uncertainty (PAvPU) (Mukhoti and Gal, 2018), AUSE also considers the uncertainty estimates for accurate predictions and does not require parameter tuning (Dreissig et al., 2023). Furthermore, AUSE is more suitable than the Expected Calibration Error (ECE) (Pakdaman Naeini et al., 2015) because ECE tends to overestimate calibration performance on imbalanced datasets (Dreissig et al., 2023). In contrast, AUSE can be combined with a performance metric that is robust to imbalanced data, such as MCC (Chicco et al., 2021). The main idea behind AUSE is that network outputs should be correct when

[1] NP-hard problems are computational problems for which there is no known algorithm which finds a solution in a number of steps polynomial in its input (Garey and Johnson, 1979). There is no efficient algorithm to solve them.

estimated to have low uncertainty, but may be incorrect when their uncertainty is high.

The sparsification curve is obtained by sorting pixels by their uncertainty and removing a fraction of the most uncertain pixels. Then, classification performance is measured on the remaining pixels. Here, we used MCC for multi-class evaluation and F_1 score for single-class evaluation. This process is repeated for increasing fractions of pixels. The resulting performance curve should gradually increase if uncertainty aligns with correctness.

The sparsification error curve is obtained by subtracting the sparsification curve for one uncertainty quantification approach from the oracle curve, i.e., the sparsification curve derived by sorting and removing pixels by actual distance between predicted and true values. This optimal sorting removes the most incorrect predictions first and is thus the best an uncertainty quantification method can achieve. For a good uncertainty quantification method, there will be a small area under the sparsification error curve, which can be used as single measure to compare between uncertainty quantification approaches.

Furthermore, we evaluated the practical use of those approaches using a correction curve, which we propose for this evaluation. This curve illustrates the impact different uncertainty quantification methods would have when used for correcting uncertain pixels, rather than removing them as is done for the sparsification curve. This correction curve shows how many pixels would need manual investigation to achieve a specified MCC value or F_1 score, facilitating informed decision-making. The correction error curve can be obtained by subtracting the correction curve of a particular uncertainty quantification method from the oracle correction curve. Based on this, we define the Area Under the Correction Error Curve (AUCE) as a metric for evaluating how well an uncertainty quantification approach identifies pixels that require correction relative to the optimal solution.

3.3. Experiment design

In our experiments, 10 U-Net models were trained on different fold combinations using a unique calibration and test set for each model. The implementation utilized pytorch 2.0.1 (Ansel et al., 2024) with training performed on a computer equipped with approximately 1 TB of RAM, two Intel Xeon Platinum processor with 32 cores each, and one 40 GB partition of an NVIDIA A100 GPU. We performed training using the Adam optimizer (Kingma and Ba, 2015) and a batch size of 16. Furthermore, each model was trained for 300 epochs in case of the 1 m DEM, and for 165 epochs, in case of the 0.5 m DEM, as these values were determined to be optimal based on validation loss performance. Given the reduced instance count for the 1 m DEM, training for more epochs was reasonable since there were fewer weight update steps per epoch.

After training the models, their performance was evaluated using MCC and F_1 score on the respective test sets. A Bayesian t-test for correlated observations (Corani and Benavoli, 2015) was conducted to determine if there were significant differences between the models' performance on the 1 m and 0.5 m DEM data. This statistical test was chosen, since it avoids the shortcomings of more traditional null hypothesis significance tests (Benavoli et al., 2017). Basically, it computes the probability of the performance difference between two approaches to lie within or outside of a pre-defined region of practical equivalence (ROPE). In our evaluation, we chose the ROPE to be a difference in MCC value of 0.05, meaning that the performance difference of two methods would have to be at least 0.05, for us to consider one method significantly better or worse than the other. Given that this test is a paired test, we paired the MCC result on one test fold from the 1 m DEM with its corresponding test fold from the 0.5 m DEM, i.e., the fold which covers the same areas, just at a higher resolution.

Given the trained models, we calibrated the conformal regression and FCP approaches on the respective calibration sets. We then derived uncertainty estimates for the chips in the corresponding test sets using the investigated approaches, i.e., network probability, mutual information, predictive entropy, conformal regression, and FCP. The execution time was measured for each approach. We then calculated the AUSE for all approaches on each test fold and both resolutions. This allowed us to investigate whether a lower resolution lead to poorer uncertainty estimates by comparing the AUSE scores between resolutions using the Bayesian t-test. Specifically, we paired the scores for each test fold and method of one resolution with those of the other resolution to determine if there were significant differences in uncertainty estimation quality.

Furthermore, we compared the AUSE scores for different uncertainty quantification methods using the Bayesian t-test to determine which method performed best. This comparison involved pairing the AUSE score of each two methods based on the corresponding folds and resolution. When comparing the AUSE, we considered a ROPE of 0.05 sufficient to identify practically relevant differences in performance among the evaluated methods. To facilitate efficient comparison of methods, high-density intervals (HDIs) were derived using the Bayesian t-test. The HDI plot displays the 95% probability intervals in which performance differences between methods lie, as well as the ROPE. By focusing on intervals not overlapping with the ROPE, statistically significant differences can be identified between methods.

To illustrate the practicality of these methods, we derived correction curves considering all classes, as well as curves focusing solely on predicted ditch and stream pixels. This allowed us to investigate the effort required to correct errors where natural streams were mistakenly predicted to be ditches or vice versa. Since, for illustrative purposes only, sparsification and correction curves displaying the performance of a single model had to be selected, the model with AUSE and AUCE values closest to the mean performance at both 1 m and 0.5 m resolutions was selected. The chip used for illustration was chosen as the one containing the most ditch and stream pixels from the test set of this model.

Lastly, we explored the possibility of generating reliable prediction maps using conformal regression and FCP. To this end, we calibrated these methods for various confidence thresholds, spanning from 50% to 90%, and included only pixels for which the probability interval of the most probable class did not overlap with those of any other class. We then computed the recall for each class, as well as the average recall over all classes. The recall was derived by dividing the number of confidently and correctly predicted pixels of a class by the total number of pixels of that class in the test set. Thus, giving an indication of the percentage of classified pixels in those confident maps. We also evaluated the classification performance on only those pixels classified with high confidence, excluding the ground truth of all pixels to which no single class was assigned. This gave an indication of the correctness of those confident maps.

4. Results

4.1. Mapping performance

Our analysis of the mapping performance revealed that all trained models performed best on the background and second best on the ditch class, but struggled with natural streams (Table 1). Models trained on the 0.5 m DEM outperformed those on the 1 m DEM in terms of MCC. A Bayesian t-test confirmed a significant advantage for the 0.5 m DEM models, estimating that with a probability of 100% they yielded a 0.05 points higher MCC than their 1 m DEM counterparts. This result remained the same even when increasing the ROPE to 0.1.

Table 1
Mapping performance on the 1 m and 0.5 m resolution data as measured by the Matthews Correlation Coefficient (MCC) for all classes, and the F_1 score for the background ($F_1^{(b)}$), ditches ($F_1^{(d)}$), and natural streams ($F_1^{(s)}$). The reported values indicate the mean and standard deviation over 10 test folds. Best performance indicated in bold.

Resolution	$F_1^{(b)}$	$F_1^{(d)}$	$F_1^{(s)}$	MCC
1 m	**1.00 ± 0.00**	0.62 ± 0.02	0.39 ± 0.06	0.61 ± 0.02
0.5 m	**1.00 ± 0.00**	**0.77 ± 0.03**	**0.43 ± 0.08**	**0.76 ± 0.03**

Table 2
Area Under the Sparsification Error Curve (AUSE) for the 1 m and 0.5 m resolution data derived for the background ($AUSE^{(b)}$), ditch ($AUSE^{(d)}$), and natural stream ($AUSE^{(s)}$) class using F_1 score as performance metric, and the overall AUSE score using the Matthews Correlation Coefficient for network probability ($\mathcal{U}_{prob}$), predictive entropy ($\mathcal{U}_{pe}$), mutual information ($\mathcal{U}_{mi}$), conformal regression ($\mathcal{U}_{cr}$), and feature conformal prediction ($\mathcal{U}_{fcp}$). The reported values indicate the mean and standard deviation over 10 test folds. Best result indicated in bold.

	$AUSE^{(b)}$	$AUSE^{(d)}$	$AUSE^{(s)}$	AUSE
	1 m			
$\mathcal{U}_{prob}$	**0.00 ± 0.00**	0.46 ± 0.23	0.58 ± 0.22	0.42 ± 0.19
$\mathcal{U}_{pe}$	**0.00 ± 0.00**	0.96 ± 0.01	0.97 ± 0.03	0.95 ± 0.03
$\mathcal{U}_{mi}$	**0.00 ± 0.00**	0.95 ± 0.01	0.98 ± 0.00	0.95 ± 0.01
$\mathcal{U}_{cr}$	0.02 ± 0.00	0.33 ± 0.03	0.52 ± 0.07	0.35 ± 0.03
$\mathcal{U}_{fcp}$	**0.00 ± 0.00**	**0.20 ± 0.10**	**0.39 ± 0.11**	**0.20 ± 0.10**
	0.5 m			
$\mathcal{U}_{prob}$	**0.00 ± 0.00**	0.61 ± 0.28	0.64 ± 0.22	0.51 ± 0.21
$\mathcal{U}_{pe}$	**0.00 ± 0.00**	0.73 ± 0.16	0.84 ± 0.18	0.65 ± 0.14
$\mathcal{U}_{mi}$	**0.00 ± 0.00**	0.90 ± 0.07	0.96 ± 0.06	0.84 ± 0.09
$\mathcal{U}_{cr}$	0.02 ± 0.00	0.20 ± 0.03	0.51 ± 0.09	0.23 ± 0.03
$\mathcal{U}_{fcp}$	**0.00 ± 0.00**	**0.09 ± 0.04**	**0.34 ± 0.12**	**0.09 ± 0.04**

4.2. Uncertainty quantification performance

MCC values increased faster for models trained on the 0.5 m DEM compared to those on the 1 m DEM when removing the most uncertain pixels, as indicated by the sparsification curves (Figs. 4(a) and 4(b)). This suggests that uncertainty quantification methods are more effective in identifying misclassified pixels for the 0.5 m DEM than the 1 m DEM. Consequently, areas between sparsification curves and the oracle curve were smaller for the 0.5 m DEM (Table 2).

The trend of improved identification of incorrect pixels with higher resolution did not hold for network probability ($\mathcal{U}_{prob}$), where higher resolution resulted in worse identification. Nonetheless, the Bayesian t-test found that a higher resolution (0.5 m DEM) led to better uncertainty estimates than a lower resolution (1 m DEM) with a probability of 83% (ROPE=0.05). Excluding $\mathcal{U}_{prob}$ increased this probability to 99% (ROPE=0.05).

While the uncertainty quantification performance varied between resolutions for sparsification curves and AUSE, it showed mostly minor differences for correction curves (Figs. 4(c) and 4(d)) and AUCE scores (Table 3). The only exception was conformal regression ($\mathcal{U}_{cr}$) for which correction curves and AUCE scores improved with higher resolution. A Bayesian t-test revealed that, with a probability of 85% (ROPE=0.05), the performances at different resolutions were practically equivalent, i.e, the performance differences lay within the ROPE. Without $\mathcal{U}_{cr}$, this probability rose to 98% (ROPE=0.05).

Comparative analysis of uncertainty quantification methods revealed distinct differences in their sparsification curves (Figs. 4(a) and 4(b)). Notably, the MCC scores for methods, such as mutual information ($\mathcal{U}_{mi}$), predictive entropy ($\mathcal{U}_{pe}$), and network probability ($\mathcal{U}_{prob}$), decreased significantly, especially when the first 5% of uncertain pixels were removed (Fig. 4(b)). This drop in performance was caused by the fact that these methods assigned high uncertainty values to correctly classified pixels, particularly ditch and natural stream pixels (Fig. 5). This tendency is reflected in the higher AUSE scores for these classes (Table 2). The HDIs (Fig. 6), derived from the Bayesian t-test, confirmed that FCP ($\mathcal{U}_{fcp}$) outperformed MC Dropout based approaches, such as predictive entropy ($\mathcal{U}_{pe}$) and mutual information ($\mathcal{U}_{mi}$) with a 100% probability, even when assuming a ROPE of 0.4. Furthermore, $\mathcal{U}_{fcp}$ was estimated to perform better than network probability ($\mathcal{U}_{prob}$) with a probability of 99.4%, and better than conformal regression with a probability of 99.6% (ROPE=0.05).

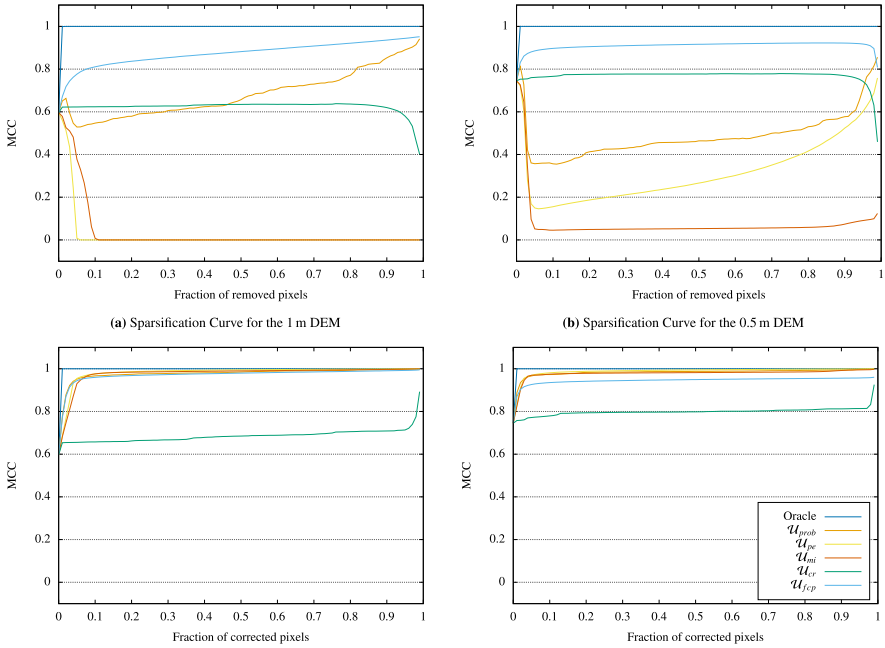
Table 3
Area Under the Correction Error Curve (AUCE) for the 1 m and 0.5 m resolution data derived for the background ($AUCE^{(b)}$), ditch ($AUCE^{(d)}$), and natural stream ($AUCE^{(s)}$) class using F_1 score as performance metric, and the overall AUCE score using the Matthews Correlation Coefficient for network probability ($\mathcal{U}_{prob}$), predictive entropy ($\mathcal{U}_{pe}$), mutual information ($\mathcal{U}_{mi}$), conformal regression ($\mathcal{U}_{cr}$), and feature conformal prediction ($\mathcal{U}_{fcp}$). The reported values indicate the mean and standard deviation over 10 test folds. Best result indicated in bold.

	$AUCE^{(b)}$	$AUCE^{(d)}$	$AUCE^{(s)}$	AUCE
	1 m			
$\mathcal{U}_{prob}$	**0.00 ± 0.00**	**0.01 ± 0.00**	0.04 ± 0.01	0.02 ± 0.00
$\mathcal{U}_{pe}$	**0.00 ± 0.00**	**0.01 ± 0.00**	**0.03 ± 0.01**	**0.01 ± 0.00**
$\mathcal{U}_{mi}$	**0.00 ± 0.00**	0.02 ± 0.00	0.04 ± 0.01	0.02 ± 0.00
$\mathcal{U}_{cr}$	**0.00 ± 0.00**	0.29 ± 0.04	0.38 ± 0.06	0.29 ± 0.04
$\mathcal{U}_{fcp}$	**0.00 ± 0.00**	0.10 ± 0.07	0.17 ± 0.12	0.10 ± 0.07
	0.5 m			
$\mathcal{U}_{prob}$	**0.00 ± 0.00**	**0.01 ± 0.00**	0.04 ± 0.01	0.01 ± 0.00
$\mathcal{U}_{pe}$	**0.00 ± 0.00**	**0.01 ± 0.00**	**0.03 ± 0.01**	**0.01 ± 0.00**
$\mathcal{U}_{mi}$	**0.00 ± 0.00**	0.01 ± 0.01	0.05 ± 0.02	0.02 ± 0.01
$\mathcal{U}_{cr}$	**0.00 ± 0.00**	0.19 ± 0.02	0.41 ± 0.07	0.20 ± 0.02
$\mathcal{U}_{fcp}$	**0.00 ± 0.00**	0.06 ± 0.04	0.16 ± 0.11	0.06 ± 0.04

The correction curves (Figs. 4(c) and 4(d)) revealed that $\mathcal{U}_{cr}$ and $\mathcal{U}_{fcp}$ exhibited inferior performance compared to $\mathcal{U}_{prob}$, $\mathcal{U}_{pe}$, and $\mathcal{U}_{mi}$. This indicates that correcting pixels identified by the latter enables faster achievement of higher performance. This is likely caused by their strong focus on ditches and natural streams (Fig. 5), which make up only a small portion of the dataset, but are frequently misclassified (Table 1). Specifically, an MCC of 0.95 was attainable with an average correction rate of 3% (approximately 2.87 million pixels) using $\mathcal{U}_{pe}$. Using the Bayesian t-test, we found that the probability of $\mathcal{U}_{pe}$, $\mathcal{U}_{mi}$, and $\mathcal{U}_{prob}$ being practically equivalent to be 100% (ROPE=0.05). Furthermore, the test suggested that $\mathcal{U}_{cr}$ performed significantly worse than all other methods with a probability of 100% (ROPE=0.05). $\mathcal{U}_{fcp}$ was found to perform significantly worse than $\mathcal{U}_{pe}$, $\mathcal{U}_{prob}$, and $\mathcal{U}_{mi}$ with a probability of 78.1%, 73.5%, and 69.1% (ROPE=0.05) respectively.

When focusing solely on pixel classifications predicted to be ditches or streams, overall $\mathcal{U}_{mi}$ was found to be most effective in identifying misclassified streams and ditches (Figs. 7(a) and 7(b)). A Bayesian t-test revealed that for ditch pixels incorrectly classified as stream pixels, $\mathcal{U}_{mi}$ had a significantly higher AUCE score with a probability greater than 95% (ROPE=0.05) when compared to $\mathcal{U}_{prob}$, $\mathcal{U}_{cr}$, and $\mathcal{U}_{fcp}$. Using $\mathcal{U}_{mi}$ to correct these errors, on average 70.6% of stream pixels (≈ 40 000) needed to be corrected to achieve an F_1 score of 0.95 for ditches. For correcting pixels classified as ditch, the Bayesian t-test revealed that $\mathcal{U}_{mi}$ had a significantly higher AUCE score than $\mathcal{U}_{cr}$ with a probability of 99.1% (ROPE=0.05). However, we found that $\mathcal{U}_{fcp}$ and $\mathcal{U}_{prob}$ lead to achieving an F_1 score of 0.95 for the stream pixels with fewer corrections than $\mathcal{U}_{mi}$. Both required on average the correction of 75% pixels (≈ 714 000). In contrast, $\mathcal{U}_{mi}$ required a correction of 79.7%. It should be noted that these F_1 scores were calculated not on all pixels, but only on those initially classified as ditch or natural stream.

$\mathcal{U}_{fcp}$ had significantly faster inference times compared to $\mathcal{U}_{cr}$ and the MC dropout-based $\mathcal{U}_{pe}$ and $\mathcal{U}_{mi}$ (Table 4). Specifically, processing the entire surface area of Sweden at a 0.5 m resolution using $\mathcal{U}_{fcp}$, producing both the actual prediction and the uncertainty estimates, would take approximately 80 h, whereas an MC dropout-based approach would require around 3 years on the same hardware. It should be noted that both MC dropout-based approaches have the same execution time,



(a) Sparsification Curve for the 1 m DEM

(b) Sparsification Curve for the 0.5 m DEM

(c) Correction Curve for the 1 m DEM

(d) Correction Curve for the 0.5 m DEM

Fig. 4. Sparsification and correction curves for the oracle, network probability ($\mathcal{U}_{prob}$), predictive entropy ($\mathcal{U}_{pe}$), mutual information ($\mathcal{U}_{mi}$), conformal regression ($\mathcal{U}_{cr}$), and feature conformal prediction ($\mathcal{U}_{fcp}$) computed on one test fold. The curves are shown for both resolutions of the digital elevation model (DEM), 1 m and 0.5 m. The classification performance was measured across all classes using the Matthews Correlation Coefficient (MCC).

Table 4
Execution times in seconds for predictive entropy ($\mathcal{U}_{pe}$), mutual information ($\mathcal{U}_{mi}$), conformal regression ($\mathcal{U}_{cr}$), and feature conformal prediction ($\mathcal{U}_{fcp}$) on one chip covering an area of 500 m ×500 m (1 m resolution) or 250 m ×250 m (0.5 m resolution). The reported values indicate the mean and standard deviation over all chips in the 10 test sets. Fastest execution time indicated in bold.

	$t_{1\,\mathrm{m}}$ (s)	$t_{0.5\,\mathrm{m}}$ (s)
$\mathcal{U}_{pe}/\mathcal{U}_{mi}$	14.13 ± 0.78	14.00 ± 0.72
$\mathcal{U}_{cr}$	1.75 ± 1.53	1.49 ± 0.22
$\mathcal{U}_{fcp}$	**0.06 ± 0.01**	**0.04 ± 0.01**

Table 5
Recall for the confident maps generated from the 0.5 m resolution data using conformal regression for different confidence levels, measured for the background ($Recall^{(b)}$), ditches ($Recall^{(d)}$), natural streams ($Recall^{(s)}$), and the class average ($Recall$). The reported values indicate the mean and standard deviation over 10 test folds.

Confidence	$Recall^{(b)}$	$Recall^{(d)}$	$Recall^{(s)}$	$Recall$
90.0%	1.00 ± 0.00	0.56 ± 0.04	0.24 ± 0.08	0.60 ± 0.03
80.0%	1.00 ± 0.00	0.59 ± 0.04	0.25 ± 0.08	0.61 ± 0.04
70.0%	1.00 ± 0.00	0.62 ± 0.05	0.27 ± 0.08	0.63 ± 0.04
60.0%	1.00 ± 0.00	0.65 ± 0.05	0.29 ± 0.09	0.64 ± 0.04
50.0%	1.00 ± 0.00	0.67 ± 0.04	0.30 ± 0.09	0.66 ± 0.04

since that time is dominated by the sampling process, which is the same for both approaches.

4.3. Conformal prediction performance

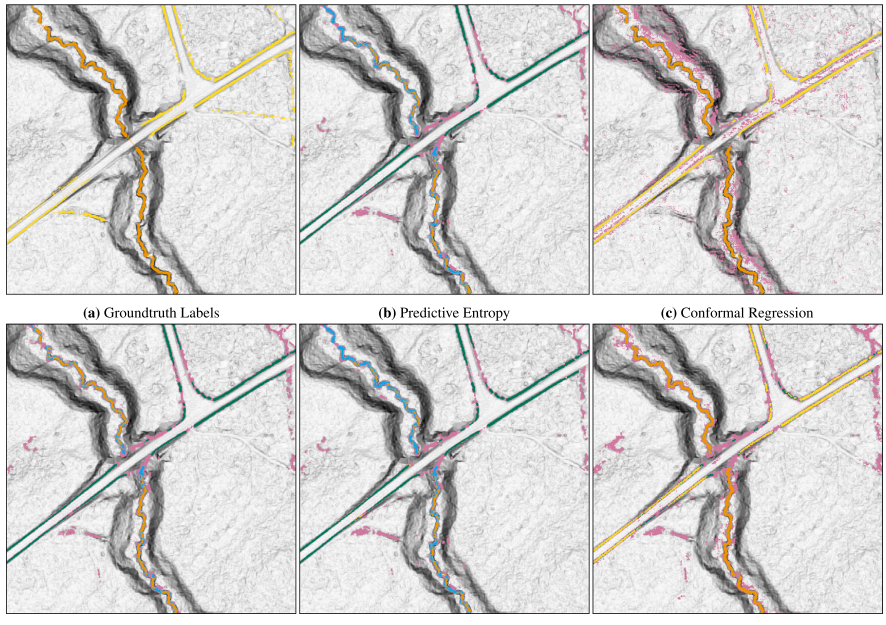
When generating confident maps using the conformal prediction approaches, FCP resulted in significantly lower recall for all confidence levels than conformal regression ($\mathcal{U}_{fcp}$: 0.12–0.13; $\mathcal{U}_{cr}$: 0.60–0.66), prompting a focus on maps generated using the latter. As expected, recall increased with decreasing confidence (Table 5). However, even highly confident maps covered a sizeable portion of background (100%), ditch (56%), and natural stream pixels (24%).

Similarly to expectation, classification performance degraded with decreasing confidence levels, with one notable exception being the background class, whose performance remained stable (Table 6). However, even at 50% confidence, the performance on confidently classified pixels, as measured by MCC, surpassed the overall performance on all pixels (Table 1).

5. Discussion

5.1. Choice of uncertainty quantification method

When comparing the evaluated uncertainty quantification approaches, FCP outperformed others in terms of AUSE (Table 2) but not in terms of AUCE (Table 3). This discrepancy stems from AUSE and AUCE addressing different questions. AUSE assesses alignment between predictions and uncertainty estimates (Dreissig et al., 2023), while AUCE evaluates the ability to identify misclassified pixels. The choice of method depends on the goal: AUSE is more informative for creating



(a) Groundtruth Labels **(b)** Predictive Entropy **(c)** Conformal Regression

(d) Network Probability **(e)** Mutual Information **(f)** Feature Conformal Prediction

Fig. 5. Illustration of the groundtruth map, as well as the uncertainty maps for the 0.5 m resolution showing the 5% most uncertain pixels as estimated by the evaluated uncertainty quantification approaches. The maps show the local slope image for certain background pixels and uncertain ones in **pink**. Furthermore, the maps show certain (**yellow**) and uncertain (**green**) ditches, as well as certain (**orange**) and uncertain (**blue**) streams.

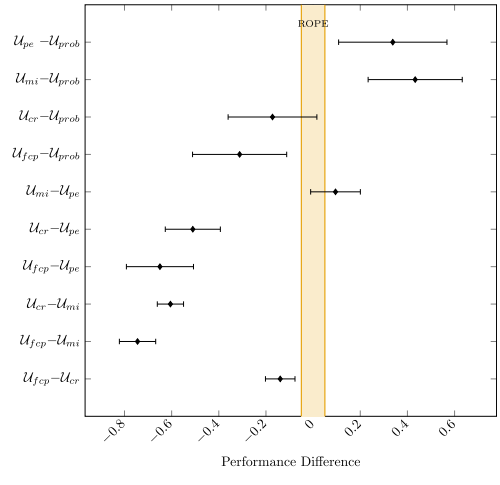


Fig. 6. High-density intervals derived using a Bayesian t-test for correlated observations indicating the intervals in which the performance differences between the compared methods, network probability ($\mathcal{U}_{prob}$), predictive entropy ($\mathcal{U}_{pe}$), mutual information ($\mathcal{U}_{mi}$), conformal regression ($\mathcal{U}_{cr}$), and feature conformal prediction ($\mathcal{U}_{fcp}$), lie with a probability of 95%. The performance is measured as area under the sparsification error curve for all classes, and the Region of Practical Equivalence (ROPE) indicates a performance difference of 0.05.

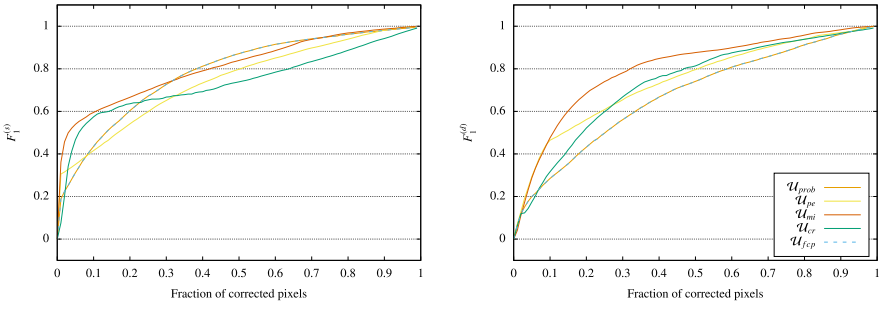
Table 6
Mapping performance for only the pixels included in the confident maps generated using conformal regression on the 0.5 m resolution data as measured by the Matthews Correlation Coefficient (MCC) for all classes, and the F_1 score for the background ($F_1^{(b)}$), ditches ($F_1^{(d)}$), and natural streams ($F_1^{(s)}$). The reported values indicate the mean and standard deviation over 10 test folds.

Confidence	$F_1^{(b)}$	$F_1^{(d)}$	$F_1^{(s)}$	MCC
90.0%	1.00 ± 0.00	0.83 ± 0.03	0.44 ± 0.10	0.82 ± 0.03
80.0%	1.00 ± 0.00	0.82 ± 0.03	0.44 ± 0.10	0.81 ± 0.03
70.0%	1.00 ± 0.00	0.81 ± 0.03	0.44 ± 0.09	0.80 ± 0.03
60.0%	1.00 ± 0.00	0.80 ± 0.03	0.43 ± 0.09	0.79 ± 0.03
50.0%	1.00 ± 0.00	0.79 ± 0.02	0.43 ± 0.09	0.78 ± 0.03

prediction uncertainty maps, whereas AUCE appears to be suitable for pixel-level correction.

Upon examining the uncertainty map generated by FCP for a broader area (Fig. 8(b)), it becomes clear that the model is generally confident in its ditch predictions, except in border regions or where ditches exhibit unusual bends. Additionally, while the two natural streams in the area (the orange lines in Fig. 8(a)) were not well identified by the model, it is relatively straightforward to trace their paths from the uncertainty maps due to the presence of uncertain background pixels on the map. This can help alert a human viewer to the presence of these streams, which would be imperceptible in the prediction map alone.

When examining the top-performing uncertainty quantification methods according to AUCE, we found that network probability, predictive entropy, and mutual information consistently identified predictions on ditch and natural stream pixels as the most uncertain ones, regardless of prediction correctness (Fig. 5). On the other hand, predictive entropy and mutual information tended to exhibit overconfidence in



(a) Correcting predicted ditch pixels

(b) Correcting predicted stream pixels

Fig. 7. Correction curves for the network probability ($\mathcal{U}_{prob}$), predictive entropy ($\mathcal{U}_{pe}$), mutual information ($\mathcal{U}_{mi}$), conformal regression ($\mathcal{U}_{cr}$), and FCP ($\mathcal{U}_{fcp}$) computed only on pixels from one test fold on the 0.5 m resolution data. The curves indicate the F_1 score for the stream class ($F_1^{(s)}$) and the ditch class ($F_1^{(d)}$) respectively, considering only pixels previously classified as ditch or stream.

incorrect predictions, as evidenced by their low AUSE scores. This tendency aligns with findings by Hertel et al. (2023), who also observed this characteristic of MC dropout-based approaches. Given that only about 1% of pixels belong to ditches, and even fewer to natural streams, it is likely that the methods' strong AUCE performance is an artifact of the highly skewed class distribution. This phenomenon arises because fixing a few pixels in classes with low instance counts and generally poorer performance can improve MCC scores more than correcting pixels from the mostly correct majority class (Table 1). As a result, one may find that for more balanced datasets, the AUCE scores of these methods may be lower compared to FCP. Additionally, the tendency to identify correctly classified pixels as uncertain can be problematic for their use in detecting incorrectly classified pixels, since the high false positive rate may lead people to dismiss detections of potentially misclassified pixels (Axelsson, 2000).

When specifically examining corrections of pixels misclassified as ditches or natural streams, we observed that mutual information outperformed other approaches in identifying ditch pixels mistakenly classified as streams. Conversely, FCP and network probability were more effective at identifying stream pixels incorrectly classified as ditches. This disparity may stem from the fact that most ditch pixels were accurately predicted, leaving only few natural stream pixels to be detected. In this scenario, overconfidence in incorrect predictions is more detrimental than when there is a larger number of misclassified pixels, as it was the case for the pixels classified as natural stream. Given that natural streams underlie stronger protections (Swedish PEFC, 2023), it is more important to identify stream pixels misclassified as ditch than vice versa.

One notable finding was that network probability achieved comparable AUCE scores to MC dropout-based approaches, while outperforming them in AUSE scores. The strong performance in identifying stream pixels among those classified as ditch is likely a consequence of that. Thus, it appears that network probability has effectively balanced high uncertainty values for ditches and streams with cautious avoidance of undue certainty in incorrectly classified pixels, at least for this dataset.

In Fig. 8(c), we observe the pixel corrections for pixels marked as most uncertain by network probability. It is evident that all predicted ditch and stream pixels were corrected due to their relatively high uncertainty. However, there are also instances where pixels were not corrected despite being wrongly predicted (stream pixels in Fig. 8(c), zoomed-in region), resulting from the model's undue confidence in its predictions. This confidence can be attributed to the fact that the natural stream is not visible in the DEM, as indicated by the one pixel wide line in the ground truth. Given that the figure showcases the correction of the 5% most uncertain pixels, a significant number of background pixels were also corrected, even though they were correctly predicted. One notable aspect of these corrected background pixels is that they appear to follow a specific pattern. Upon analyzing the slope values of those corrected background pixels, we found them to be significantly higher than average slope values. Furthermore, similar patterns have been observed in data from other regions, but not consistently across all areas, suggesting that these may be caused by minor differences in the data collection process.

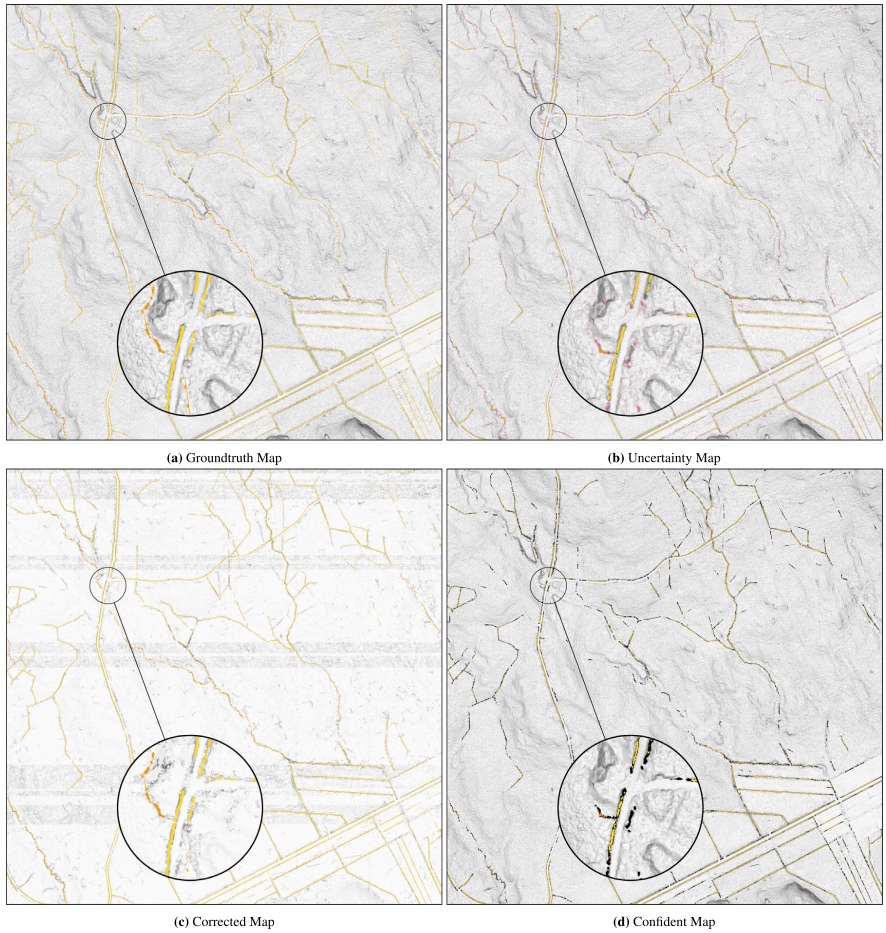
When evaluating execution performance, arguably, the fastest uncertainty estimates were derived using network probability, since it equals the model's inference speed of approximately 0.014 s per chip, resulting in an estimated processing time of 28 h for all of Sweden. While this was significantly shorter than the 80 h required for FCP, we deem FCP still feasible, especially when compared to the execution times for MC dropout-based approaches ($\approx$ 3 years) or conformal regression ($\approx$ 124 days). It is worth noting that these times can be significantly reduced by using fewer Monte Carlo samples. For example, utilizing just 10 samples, as Kampffmeyer et al. (2016), would reduce the time required for MC dropout and conformal regression to 280 h and 298 h, respectively. However, this may come at the cost of reduced uncertainty quantification performance.

In summary, our results show that FCP yielded the most accurate uncertainty estimates at a reasonable processing speed. Therefore, we believe it is well-suited as a method for generating uncertainty maps. However, when attempting to identify which pixels require correction in the generated ditch and stream maps, we found that using network probability was more effective. This approach identified the pixels that needed correction better and resulted in lower execution times.

5.2. Impact of resolution

The classification performance was improved when detecting ditches and streams on higher resolution data (Table 1). This is reasonable since landscape outlines were captured more accurately, which simplified the detection problem. This finding aligns with the findings by Busarello et al. (2025) on mapping ditches and streams, but also with findings on mapping other terrain features, such as ephemeral gullies (Chowdhuri et al., 2021), and rock glaciers (Robson et al., 2020).

Higher resolution DEMs also yielded more accurate uncertainty estimates as indicated by the obtained AUSE scores. While it is unsurprising, that a lower resolution leads to a higher uncertainty (Pogson and Smith, 2015; Wu et al., 2024), the observed reduced alignment between estimated model uncertainty and actual performance is likely due to the network's generally poorer performance on lower resolution data. In contrast to AUSE, the AUCE scores were mostly unaffected by the resolution, presumably since AUCE performance was largely



(a) Groundtruth Map **(b)** Uncertainty Map

(c) Corrected Map **(d)** Confident Map

Fig. 8. Illustration of the groundtruth, uncertainty, corrected, and confident map over an area of 1.5 km × 1.5 km at a 0.5 m resolution. In all maps, certain or correct background pixels are shown by the local slope image, while ditches are shown in **yellow**, and streams in **orange**. The uncertainty map was generated using feature conformal prediction and displays the 5% most uncertain background (**pink**), ditch (**green**), and stream (**blue**) pixels. The corrected map was derived by correcting the 5% most uncertain pixels as estimated by network probability. Corrected pixels are shown with full intensity, while not corrected pixels have low intensity. The confident map was derived using conformal regression at a 90% confidence level, and pixels where the model did not commit to one class are shown in **black**.

improved by ditch and stream detection rather than uncertainty quantification accuracy. Thus, as long as a method could identify most ditch and stream pixels it would get a high AUCE score, even if it marked many correctly classified pixels as uncertain.

Most methods showed increased uncertainty quantification performance with higher resolutions, except network probability, which decreased due to overconfidence in its predictions. This overconfidence was caused by the simplified learning problem, which allowed the model to assign more extreme probability estimates to pixels, as incentivized by the training process. As noted by Guo et al. (2017) and Sensoy et al. (2018), this leads to poorer uncertainty estimates.

There was no difference in processing time for a chip of 1 m resolution versus one with a 0.5 m resolution (Table 4), since both have the same number of pixels. However, four 0.5 m resolution chips are required to cover the same area as one 1 m resolution chip. This results in four times longer processing times for the 0.5 m resolution. As such, it is important to consider whether the gained performance improvements justify the increased processing costs.

In summary, there is a motivation for conducting high-resolution LiDAR scans to improve ditch and stream detection and to obtain more accurate uncertainty estimates. However, this may decrease the accuracy of uncertainty estimates obtained by network probability as performance improves.

5.3. Confident segmentation maps

When generating confident segmentation maps, we found that only $\mathcal{U}_{cr}$ consistently produced a reasonable number of single-class predictions for various confidence levels, ruling out $\mathcal{U}_{fcp}$ from further evaluation. This appears contradictory to the findings by Teng et al. (2023), who showed that FCP produced shorter confidence bands than a baseline conformal prediction approach. It is reasonable to assume

that shorter confidence bands also would lead to a higher number of single-class predictions. However, it should be noted that the conformal prediction approach used by Teng et al. (2023) differs from $\mathcal{U}_{cr}$ used in this article, which is the likely reason for the observed differences.

For $\mathcal{U}_{cr}$, recall improved as the confidence level decreased (Table 5). This was expected since lower confidence thresholds allow $\mathcal{U}_{cr}$ to make more errors and thus commit to single-class predictions for more pixels. Similarly in line with expectations was the observed decrease in precision, indicated by lower MCC and F_1 scores (Table 6). This decrease is caused by $\mathcal{U}_{cr}$ actually making more errors at lower confidence levels.

Compared to the models' results on all pixels (Table 1), we observed improved classification performance for predictions with high confidence levels (Table 6). Specifically, we achieved an MCC of 0.82 for 90% confident predictions, surpassing the MCC of 0.76 obtained on all pixel predictions. This performance difference was largely due to clear improvement in the ditch class, which was attained through $\mathcal{U}_{cr}$ not assigning a class in border regions where it is challenging to determine where the ditch ends and the background begins, or areas where the ditch was not clearly visible in the DEM (Fig. 8(d), zoomed-in region). These observations align well with the findings by Koski et al. (2023), who found that the main causes of error in detecting small watercourses with deep learning were boundary issues and unclear visual expression in the DEM.

Despite committing to a single class with high confidence, it is possible for $\mathcal{U}_{cr}$ to make errors. For example, many natural stream pixels were confidently predicted as background (Fig. 8(d), zoomed-in region), which was not unexpected. This outcome is consistent with the fact that $\mathcal{U}_{cr}$ allows for 10% errors at a 90% confidence level. It is important to note that the guarantees provided by this method apply to probability intervals rather than the classes themselves. A model that consistently missed to predict the natural stream class, would make significantly fewer errors than 10%, due to its low occurrence rate (less than 1%). Instead, it would in over 99% of the cases be correct in predicting the probability for the stream class to be close to 0%. Consequently, $\mathcal{U}_{cr}$ primarily prevented overprediction in minority classes, such as ditch and stream, as observed in Fig. 8(d) and reflected in their low recall values (Table 5).

Our analysis revealed that neither $\mathcal{U}_{cr}$ nor $\mathcal{U}_{fcp}$ are particularly suitable for generating confident maps of ditches and natural streams. Although $\mathcal{U}_{cr}$ produced more confident predictions than $\mathcal{U}_{fcp}$, the generated maps only covered around 60% of all pixels, particularly omitting ditch and stream pixels. This means that the prediction sets for pixels of these classes frequently contained more than one possible prediction. This observation is in line with the findings by Ghosh et al. (2023), who show that conformal prediction tends to result in large prediction sets for challenging datasets, while obtaining narrower sets for simple ones. Apart from this issue, it also took a considerable amount of time to generate the confident maps (Table 4).

5.4. Limitations and future work

This article's evaluation of uncertainty quantification methods is limited to one specific remote sensing task with an extreme class distribution. This may have skewed results, as MC dropout-based solutions likely perform differently in terms of AUCE on tasks with more balanced distributions. Although investigating extreme cases is valuable, given that classes with relatively few instances are not uncommon in remote sensing (Kossmann et al., 2021), it would be interesting to investigate if MC dropout's AUCE performance would decrease when applied to tasks with more balanced distributions.

Furthermore, the dataset used in this study is limited by its two-resolution format (1 m and 0.5 m). As demonstrated, classification and uncertainty quantification performance improve with increasing resolution. However, it is plausible that returns diminish at some point, warranting investigation into the optimal resolution threshold. Additionally, the uncertainty quantification performance of $\mathcal{U}_{prob}$ has been observed to decrease with increased resolution, suggesting a possible trend where higher resolutions lead to overconfident predictions. Higher resolution datasets would aid in investigating this trend as well.

Another limitation of our study is that we have only investigated a restricted set of uncertainty quantification approaches. For example, Bayesian neural networks (Blundell et al., 2015) were excluded from this study since they cannot derive uncertainty estimates from the same model as the other investigated approaches. This would have complicated direct comparisons between the methods, as it is less clear if differences in uncertainty quantification performance are due to differences in the used methods or due to the different models. Nevertheless, exploring Bayesian neural networks would be valuable for future research as they have been shown to outperform MC dropout-based approaches by Hertel et al. (2023). Similarly, deep ensembles have been shown to perform better than MC dropout-based approaches (Lakshminarayanan et al., 2017). Investigating how they compare to the evaluated conformal prediction-based approaches could be worthwhile. However, due to their significant training time requirements, we excluded them from this article; using the recommended number of networks in the ensemble would have quintupled the necessary training time.

It should be noted that none of the investigated uncertainty quantification approaches is able to handle out-of-distribution (OOD) data, i.e., data that is distinctively different from the training data. Alarab et al. (2021) have shown this for network probability and MC dropout-based approaches, while this limitation of conformal prediction has been pointed out, for example, by Angelopoulos et al. (2022). This is not a big problem for the studied dataset, since it has been specifically designed to be representative of the Swedish landscape (Busarello et al., 2025). However, in situations where OOD data is present, the obtained uncertainty estimates may not be reliable. One approach to handle OOD data would be to build on ideas from the "Learn then Test" framework (Angelopoulos et al., 2022).

Our investigation was further limited by focusing solely on conformal regression approaches within either feature space ($\mathcal{U}_{fcp}$) or output space ($\mathcal{U}_{cr}$). The focus on probability ranges rather than actual class predictions may have hindered the utility of generated confidence maps, as they tended to suppress minority class predictions. In the future, this limitation could be addressed by exploring whether the conformal classification approach by Wieslander et al. (2021) can be made more computationally efficient or through further investigation into recent methods proposed by Mossina et al. (2024), Brunekreef et al. (2024). By focusing on conformal classification approaches, the guarantees provided by the conformal predictor would apply directly to the classification outcome, and thus might produce more usable confident maps.

6. Conclusions

In this article, we investigated various uncertainty quantification techniques, including network probability, predictive entropy, mutual information, conformal regression, and feature conformal prediction, and applied them to a specific remote sensing task: identifying ditches and natural streams from elevation data sourced from a digital elevation model (DEM). Additionally, the impact of different DEM resolutions on classification and uncertainty quantification performance was explored. Furthermore, confident maps were generated using conformal prediction methods. Our key findings include:

- Feature conformal prediction (Teng et al., 2023) produces uncertainty estimates most aligned with the actual neural network performance at a reasonable cost to the execution time. However, for correcting misclassified pixels, the network probability output is more suitable, at least for the investigated dataset.
- A higher resolution DEM leads to better classification performance and better uncertainty estimates.

- Conformal regression and feature conformal prediction are not suitable to generate confident maps, since they are overly conservative in their estimates and the model performance is too limited.

CRediT authorship contribution statement

Florian Westphal: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **William Lidberg:** Writing – review & editing, Resources, Funding acquisition, Data curation, Conceptualization. **Mariana Dos Santos Toledo Busarello:** Writing – review & editing, Data curation. **Anneli M. Ågren:** Writing – review & editing, Resources, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Ollama with the llama3.1 model in order to improve the language of the written text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

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.

Acknowledgments

This study was funded by the Swedish research council Formas (proj. no. 2021-00115) and Knut and Alice Wallenberg Foundation (2018.0259 Future Silviculture). It was also partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation. The funding sources had no involvement in study design, collection, analysis and interpretation of data, nor in the writing of the article.

Data availability

I have shared the link to my data/code at the Attach File step.

Uncertainty Quantification for LiDAR-based Maps of Ditches and Natural Streams (Original data) (GitHub)
Automatic Detection of Ditches and Natural Streams from Digital Elevation Models Using Deep Learning (Reference data) (Swedish National Data Service).

References

Alarab, I., Prakoonwit, S., Nacer, M.I., 2021. Illustrative discussion of MC-dropout in general dataset: Uncertainty estimation in bitcoin. Neural Process. Lett. 53, 1001–1011. http://dx.doi.org/10.1007/s11063-021-10424-x, URL http://dx.doi.org/10.1007/s11063-021-10424-x.

Angelopoulos, A.N., Bates, S., Candès, E.J., Jordan, M.I., Lei, L., 2022. Learn then test: Calibrating predictive algorithms to achieve risk control. URL https://arxiv.org/abs/2110.01052. arXiv:2110.01052.

Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Ellison, E., Feng, W., Gong, J., Gschwind, M., Hirsh, B., Huang, S., Kalambarkar, K., Kirsch, L., Lazos, M., Lezcano, M., Liang, Y., Liang, J., Lu, Y., Luk, C., Maher, B., Pan, Y., Puhrsch, C., Reso, M., Saroufim, M., Siraichi, M.Y., Suk, H., Suo, M., Tillet, P., Wang, E., Wang, X., Wen, W., Zhang, S., Zhao, X., Zhou, K., Zou, R., Mathews, A., Chanan, G., Wu, P., Chintala, S., 2024. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In: 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2. ASPLOS'24, ACM, http://dx.doi.org/10.1145/3620665.3640366, URL https://pytorch.org/assets/pytorch2-2.pdf.

Axelsson, S., 2000. The base-rate fallacy and the difficulty of intrusion detection. ACM Trans. Inf. Syst. Secur. 3, 186–205. http://dx.doi.org/10.1145/357830.357849, URL http://dx.doi.org/10.1145/357830.357849.

Benavoli, A., Corani, G., Demšar, J., Zaffalon, M., 2017. Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis. J. Mach. Learn. Res. 18 (1), 2653–2688.

Bishop, K., Allan, C., Bringmark, L., Garcia, E., Hellsten, S., Högbom, L., Johansson, K., Lomander, A., Meili, M., Munthe, J., Nilsson, M., Porvari, P., Skyllberg, U., Sørensen, R., Zetterberg, T., Åkerblom, S., 2009. The effects of forestry on hg bioaccumulation in nemoral/boreal waters and recommendations for good silvicultural practice. AMBIO: A J. Hum. Environ. 38, 373–380. http://dx.doi.org/10.1579/0044-7447-38.7.373.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16. http://dx.doi.org/10.1016/j.isprsjprs.2009.06.004, URL http://dx.doi.org/10.1016/j.isprsjprs.2009.06.004.

Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D., 2015. Weight uncertainty in neural network. In: Bach, F., Blei, D. (Eds.), Proceedings of the 32nd International Conference on Machine Learning. In: Proceedings of Machine Learning Research, vol. 37, PMLR, Lille, France, pp. 1613–1622, URL https://proceedings.mlr.press/v37/blundell15.html.

Boström, H., Johansson, U., 2020. Mondrian conformal regressors. In: Gammerman, A., Vovk, V., Luo, Z., Smirnov, E., Cherubin, G. (Eds.), Proceedings of the Ninth Symposium on Conformal and Probabilistic Prediction and Applications. In: Proceedings of Machine Learning Research, vol. 128, PMLR, pp. 114–133, URL https://proceedings.mlr.press/v128/bostrom20a.html.

Brunekreef, J., Marcus, E., Sheombarsing, R., Sonke, J.-J., Teuwen, J., 2024. Kandinsky conformal prediction: Efficient calibration of image segmentation algorithms. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 4135–4143.

Busarello, M.D.S.T., Ågren, A.M., Westphal, F., Lidberg, W., 2025. Automatic detection of ditches and natural streams from digital elevation models using deep learning. Comput. Geosci. 196, 105875. http://dx.doi.org/10.1016/j.cageo.2025.105875.

Chaudhary, P., Leitão, J.P., Donauer, T., D'Aronco, S., Perraudin, N., Obozinski, G., Perez-Cruz, F., Schindler, K., Wegner, J.D., Russo, S., 2022. Flood uncertainty estimation using deep ensembles. Water 14, 2980. http://dx.doi.org/10.3390/w14192980.

Chicco, D., Tötsch, N., Jurman, G., 2021. The matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Min. 14, http://dx.doi.org/10.1186/s13040-021-00244-z, URL http://dx.doi.org/10.1186/s13040-021-00244-z.

Chowdhuri, I., Pal, S.C., Saha, A., Chakrabortty, R., Roy, P., 2021. Evaluation of different DEMs for gully erosion susceptibility mapping using in-situ field measurement and validation. Ecol. Informatics 65, 101425. http://dx.doi.org/10.1016/j.ecoinf.2021.101425, URL https://www.sciencedirect.com/science/article/pii/S1574954121002168.

Corani, G., Benavoli, A., 2015. A Bayesian approach for comparing cross-validated algorithms on multiple data sets. Mach. Learn. 100, 285–304. http://dx.doi.org/10.1007/s10994-015-5486-z, URL http://dx.doi.org/10.1007/s10994-015-5486-z.

Cortés-Ciriano, I., Bender, A., 2019. Reliable prediction errors for deep neural networks using test-time dropout. J. Chem. Inf. Model. 59, 3330–3339. http://dx.doi.org/10.1021/acs.jcim.9b00297, URL http://dx.doi.org/10.1021/acs.jcim.9b00297.

Dreissig, M., Piewak, F., Boedecker, J., 2023. On the calibration of uncertainty estimation in lidar-based semantic segmentation. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems. ITSC, IEEE, http://dx.doi.org/10.1109/itsc57777.2023.10422384, URL http://dx.doi.org/10.1109/itsc57777.2023.10422384,

Eigen, D., Fergus, R., 2015. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of the IEEE International Conference on Computer Vision. ICCV.

Florinsky, I.V., 2016. Chapter 2 - topographic surface and its characterization. In: Florinsky, I.V. (Ed.), Digital Terrain Analysis in Soil Science and Geology (Second Edition), second ed. Academic Press, pp. 7–76. http://dx.doi.org/10.1016/B978-0-12-804632-6.00002-X, URL https://www.sciencedirect.com/science/article/pii/B978012804632600002X.

Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Balcan, M.F., Weinberger, K.Q. (Eds.), Proceedings of the 33rd International Conference on Machine Learning. In: Proceedings of Machine Learning Research, 48, PMLR, New York, New York, USA, pp. 1050–1059, URL https://proceedings.mlr.press/v48/gal16.html.

Gal, Y., Hron, J., Kendall, A., 2017. Concrete dropout. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. vol. 30, Curran Associates, Inc., URL https://proceedings.neurips.cc/paper/2017/file/84ddfb34126fc3a48ee38d7044e87276-Paper.pdf.

Garey, M.R., Johnson, D.S., 1979. Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman and Company.

Gawlikowski, J., Saha, S., Kruspe, A., Zhu, X.X., 2022. An advanced Dirichlet prior network for out-of-distribution detection in remote sensing. IEEE Trans. Geosci. Remote Sens. 60, 1–19. http://dx.doi.org/10.1109/tgrs.2022.3140324, URL http://dx.doi.org/10.1109/tgrs.2022.3140324.

Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R., Zhu, X.X., 2023. A survey of uncertainty in deep neural networks. Artif. Intell. Rev. 56, 1513–1589. http://dx.doi.org/10.1007/s10462-023-10562-9, URL http://dx.doi.org/10.1007/s10462-023-10562-9. arXiv:2107.03342v3.

Ghosh, S., Belkhouja, T., Yan, Y., Doppa, J.R., 2023. Improving uncertainty quantification of deep classifiers via neighborhood conformal prediction: Novel algorithm and theoretical analysis. Proc. the AAAI Conf. Artif. Intell. 37, 7722–7730. http://dx.doi.org/10.1609/aaai.v37i6.25936.

Goan, E., Fookes, C., 2020. Bayesian neural networks: An introduction and survey. Springer International Publishing, pp. 45–87. http://dx.doi.org/10.1007/978-3-030-42553-1_3, URL http://dx.doi.org/10.1007/978-3-030-42553-1_3,

Gorodkin, J., 2004. Comparing two K-category assignments by a K-category correlation coefficient. Comput. Biol. Chem. 28, 367–374. http://dx.doi.org/10.1016/j.compbiolchem.2004.09.006, URL http://dx.doi.org/10.1016/j.compbiolchem.2004.09.006.

Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On calibration of modern neural networks. In: Precup, D., Teh, Y.W. (Eds.), Proceedings of the 34th International Conference on Machine Learning. In: Proceedings of Machine Learning Research, vol. 70, PMLR, pp. 1321–1330, URL https://proceedings.mlr.press/v70/guo17a.html.

Hertel, V., Chow, C., Wani, O., Wieland, M., Martinis, S., 2023. Probabilistic SAR-based water segmentation with adapted Bayesian convolutional neural network. Remote Sens. Environ. 285, 113388. http://dx.doi.org/10.1016/j.rse.2022.113388, URL http://dx.doi.org/10.1016/j.rse.2022.113388.

Iagaru, D., Gottschling, N.M., 2023. Uncertainty quantification with deep ensemble methods for super-resolution of sentinel 2 satellite images. In: International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. MDPI, p. 4. http://dx.doi.org/10.3390/psf2023009004,

Ilg, E., Cicek, O., Galesso, S., Klein, A., Makansi, O., Hutter, F., Brox, T., 2018. Uncertainty estimates and multi-hypotheses networks for optical flow. In: Proceedings of the European Conference on Computer Vision. ECCV.

Kampffmeyer, M., Salberg, A.B., Jenssen, R., 2016. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

Kasraei, B., Heung, B., Saurette, D.D., Schmidt, M.G., Bulmer, C.E., Bethel, W., 2021. Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. Environ. Model. Softw. 144, 105139. http://dx.doi.org/10.1016/j.envsoft.2021.105139.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations. URL http://arxiv.org/abs/1412.6980.

Koski, C., Kettunen, P., Poutanen, J., Zhu, L., Oksanen, J., 2023. Mapping small watercourses from DEMs with deep learning—exploring the causes of false predictions. Remote. Sens. 15, 2776. http://dx.doi.org/10.3390/rs15112776, URL http://dx.doi.org/10.3390/rs15112776.

Kossmann, D., Wilhelm, T., Fink, G.A., 2021. Towards tackling multi-label imbalances in remote sensing imagery. In: 2020 25th International Conference on Pattern Recognition. ICPR, IEEE, pp. 5782–5789. http://dx.doi.org/10.1109/icpr48806.2021.9412588, URL http://dx.doi.org/10.1109/icpr48806.2021.9412588,

Labuzzetta, C.J., 2022. Practical Methods for the Advancement of Precision Conservation Via Land Cover Classification and Conformal Prediction (Ph.D. thesis). Iowa State University.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. vol. 30, Curran Associates, Inc., URL https://proceedings.neurips.cc/paper_files/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce38-Paper.pdf.

Lidberg, W., Paul, S.S., Westphal, F., Richter, K.F., Lavesson, N., Melniks, R., Ivanovs, J., Ciesielski, M., Leinonen, A., Ågren, A.M., 2023. Mapping drainage ditches in forested landscapes using deep learning and aerial laser scanning. J. Irrig. Drain. Eng. 149, http://dx.doi.org/10.1061/jidedh.ireng-9796, URL http://dx.doi.org/10.1061/jidedh.ireng-9796.

Martínez-Ferrer, L., Moreno-Martínez, Á., Campos-Taberner, M., García-Haro, F.J., Muñoz-Marí, J., Running, S.W., Kimball, J., Clinton, N., Camps-Valls, G., 2022. Quantifying uncertainty in high resolution biophysical variable retrieval with machine learning. Remote Sens. Environ. 280, 113199. http://dx.doi.org/10.1016/j.rse.2022.113199, URL http://dx.doi.org/10.1016/j.rse.2022.113199.

Matthews, B., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. et Biophys. Acta (BBA) - Protein Struct. 405, 442–451. http://dx.doi.org/10.1016/0005-2795(75)90109-9, URL http://dx.doi.org/10.1016/0005-2795(75)90109-9.

Mossina, L., Dalmau, J., Andéol, L., 2024. Conformal semantic image segmentation: Post-hoc quantification of predictive uncertainty. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 3574–3584.

Mukhoti, J., Gal, Y., 2018. Evaluating Bayesian deep learning methods for semantic segmentation. URL http://arxiv.org/abs/1811.12709v2. arXiv:1811.12709v2.

O'Neil, G.L., Goodall, J.L., Behl, M., Saby, L., 2020. Deep learning using physically-informed input data for wetland identification. Environ. Model. Softw. 126, 104665. http://dx.doi.org/10.1016/j.envsoft.2020.104665.

Pagella, T.F., Sinclair, F.L., 2014. Development and use of a typology of mapping tools to assess their fitness for supporting management of ecosystem service provision. Landsc. Ecol. 29, 383–399. http://dx.doi.org/10.1007/s10980-013-9983-9, URL http://dx.doi.org/10.1007/s10980-013-9983-9.

Pakdaman Naeini, M., Cooper, G., Hauskrecht, M., 2015. Obtaining well calibrated probabilities using Bayesian binning. Proc. the AAAI Conf. Artif. Intell. 29, http://dx.doi.org/10.1609/aaai.v29i1.9602.

Pogson, M., Smith, P., 2015. Effect of spatial data resolution on uncertainty. Environ. Model. Softw. 63, 87–96. http://dx.doi.org/10.1016/j.envsoft.2014.09.021, URL http://dx.doi.org/10.1016/j.envsoft.2014.09.021.

Robson, B.A., Bolch, T., MacDonell, S., Hölbling, D., Rastner, P., Schaffer, N., 2020. Automated detection of rock glaciers using deep learning and object-based image analysis. Remote Sens. Environ. 250, 112033. http://dx.doi.org/10.1016/j.rse.2020.112033, URL https://www.sciencedirect.com/science/article/pii/S003442572030403X.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. Springer International Publishing, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28, URL http://dx.doi.org/10.1007/978-3-319-24574-4_28,

Savelonas, M.A., Veinidis, C.N., Bartsokas, T.K., 2022. Computer vision and pattern recognition for the analysis of 2D/3D remote sensing data in geoscience: A survey. Remote. Sens. 14, 6017. http://dx.doi.org/10.3390/rs14236017, URL http://dx.doi.org/10.3390/rs14236017.

Sensoy, M., Kaplan, L., Kandemir, M., 2018. Evidential deep learning to quantify classification uncertainty. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. 31, Curran Associates, Inc., URL https://proceedings.neurips.cc/paper_files/paper/2018/file/a981f2b708044d6fb4a71a1463242520-Paper.pdf.

Singh, G., Moncrieff, G., Venter, Z., Cawse-Nicholson, K., Slingsby, J., Robinson, T.B., 2024. Uncertainty quantification for probabilistic machine learning in earth observation using conformal prediction. Sci. Rep. 14, http://dx.doi.org/10.1038/s41598-024-65954-w, URL http://dx.doi.org/10.1038/s41598-024-65954-w.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Swedish PEFC, 2023. Forest Use Standard. Technical Report, (PEFC SWE 002:5), Swedish PEFC.

Teng, J., Wen, C., Zhang, D., Bengio, Y., Gao, Y., Yuan, Y., 2023. Predictive inference with feature conformal prediction. In: The Eleventh International Conference on Learning Representations. URL https://openreview.net/forum?id=0uRm1YmFTu.

Toth, C., Jóźków, G., 2016. Remote sensing platforms and sensors: A survey. ISPRS J. Photogramm. Remote Sens. 115, 22–36. http://dx.doi.org/10.1016/j.isprsjprs.2015.10.004, URL http://dx.doi.org/10.1016/j.isprsjprs.2015.10.004.

Vovk, V., Gammerman, A., Shafer, G., 2005. Algorithmic Learning in a Random World. Springer-Verlag, http://dx.doi.org/10.1007/b106715, URL http://dx.doi.org/10.1007/b106715.

Wieslander, H., Harrison, P.J., Skogberg, G., Jackson, S., Friden, M., Karlsson, J., Spjuth, O., Wahlby, C., 2021. Deep learning with conformal prediction for hierarchical analysis of large-scale whole-slide tissue images. IEEE J. Biomed. Heal. Informatics 25, 371–380. http://dx.doi.org/10.1109/jbhi.2020.2996300, URL http://dx.doi.org/10.1109/jbhi.2020.2996300.

Wu, L., Xu, Y., Li, R., 2024. Effects of input data accuracy, catchment threshold areas and calibration algorithms on model uncertainty reduction. Eur. J. Soil Sci. 75, http://dx.doi.org/10.1111/ejss.13519.

Xu, Y., Bai, T., Yu, W., Chang, S., Atkinson, P.M., Ghamisi, P., 2022. AI security for geoscience and remote sensing: Challenges and future trends. http://dx.doi.org/10.1109/MGRS.2023.3272825, URL http://arxiv.org/abs/2212.09360v2. arXiv:2212.09360v2. IEEE Geoscience and Remote Sensing Magazine, Volume 11, Issue 2, Pages 60-85, 2023.

Xu, K., Shi, Z., Zhang, H., Wang, Y., Chang, K.-W., Huang, M., Kailkhura, B., Lin, X., Hsieh, C.-J., 2020. Automatic perturbation analysis for scalable certified robustness and beyond. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (Eds.), Advances in Neural Information Processing Systems. vol. 33, Curran Associates, Inc., pp. 1129–1141, URL https://proceedings.neurips.cc/paper_files/paper/2020/file/0cbc5671ae26f67871cb914d81ef8fc1-Paper.pdf.

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 Sens. Environ. 241, 111716. http://dx.doi.org/10.1016/j.rse.2020.111716, URL http://dx.doi.org/10.1016/j.rse.2020.111716.

Yule, G.U., 1912. On the methods of measuring association between two attributes. J. R. Stat. Soc. 75, 579. http://dx.doi.org/10.2307/2340126, URL http://dx.doi.org/10.2307/2340126.

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2025:81

Small streams and ditches have major hydrological and ecological roles in boreal landscapes but remain poorly mapped. In this thesis, a national-scale framework was developed using high-resolution LiDAR-derived topographic data and machine learning. Combining convolutional neural networks, XGBoost, and drainage analyses, the method maps and distinguishes natural streams from ditches. The framework provides consistent, scalable maps that support restoration planning, sustainable forestry, and environmental reporting, offering a reproducible approach for mapping drainage systems globally.

Mariana Dos Santos Toledo Busarello received her PhD education at the Department of Forest Ecology and Management, SLU, Umeå. She holds a Master of Science Degree in Earth Science from Umeå University.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8124-065-8
ISBN (electronic version) 978-91-8124-111-2