Faculté des bioingénieurs

Dynamic Muddy Flood Susceptibility Mapping in Wallonia

Author: Giang Châu Nguyễn

Supervisors: Prof. Charles Bielders

Prof. Pierre Defourny

Readers: Dr. Alice Alonso

Dr. Julien Radoux

Academic Year: 2025–2026

Programme:
MSc Erasmus+ GEM – Geo-information Science and Earth Observation for Environmental Modelling and Management

Dynamic Muddy Flood Susceptibility Mapping in Wallonia

Integrating RUSLE/WaTEM-SEDEM Erosion Modelling, Sentinel-2 Vegetation Dynamics, Daily Rainfall Data, and Positive-Unlabelled Learning

Giang Nguyen

Master's Thesis submitted in partial fulfilment of the requirements for the degree of

Erasmus Mundus Master of Science in Geo-information Science and Earth Observation for Environmental Modelling and Management

UCLouvain & Lund University

Faculty Bioscience Engineering, UCLouvain

Supervisors: Professor Charles Bielders & Professor Pierre Defourny

Jury: Alice Alonso & Julien Radoux

Academic Year 2025–2026

Defence: 26 August 2026

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Prof. Charles Bielders and Prof. Pierre Defourny. I am especially grateful to Prof. Bielders for introducing me to this topic, and for your immense patience in guiding me through a subject that was entirely new to me. I would also like to thank Prof. Defourny for his invaluable insights into remote sensing data and machine learning, and for his dedication in looking out for the GEM students. Together, you have set my standards and expectations incredibly high for academic and scientific environments, shaping the path for my future and inspiring me to always meet those same high standards.

This academic journey would not have been possible without the generous support of the Erasmus Mundus GEM Scholarship. Thank you for providing this life-changing opportunity and making this experience a reality.

To my fellow GEM student friends, thank you for sharing this unforgettable experience with me. Life gets so much more fun when we do things together, and seeing everyone navigate the same challenges made the hardships seem not so hard after all. I will always cherish the memories of us being in our twenties, navigating life between two countries together. You have become living time capsules of this chapter, and I will always remember these days.

I owe an immense debt of gratitude to my parents. Thank you for your unwavering belief in me, especially while I was studying so far away from home. Thank you for holding front-row seats to all my ups and downs, and for your strength in raising me to face these challenges.

Finally, I want to thank my partner. Thank you for your constant motivation and sound advice when I felt blindsided by the hurdles in front of me.

Overall, I feel incredibly lucky and thankful to have been surrounded by such wonderful people who have pushed me to grow, learn, and make lifelong memories.

Statement of Contribution and Use of Generative AI

Stage	Author	Co-Supervisors	Artificial Intelligence
Conceptualisation	Developed research objectives and study design	Scientific guidance and feedback	Literature exploration
Methodology	Developed and implemented methodology	Methodological advice and feedback	Python coding assistance
Investigation	Conducted data processing and analysis	Scientific guidance and feedback	Python coding assistance
Formal analysis	Performed statistical and machine learning analyses	Discussion and interpretation support	Python coding assistance
Validation	Conducted model validation and result assessment	Review of results and methodological feedback	-
Writing, original draft	Wrote and assembled the thesis draft	Scientific review and feedback	Drafting and editing assistance
Writing, review & editing	Revised and edited the thesis	Review and detailed feedback	Editing assistance
Visualisation	Prepared maps, figures, tables, and visual outputs	-	-
Supervision	-	Prof. Charles Bielders; Prof. Pierre Defourny	-

1. Conceptualisation: conception of the main research idea, formulation of objectives and hypotheses;
2. Methodology: design of the research protocol, choice of analysis methods;
3. Investigation: conduct of the research process and data collection;
4. Formal analysis: application of statistical, mathematical, or other formal techniques to analyse or synthesise research data;
5. Validation: verification and validation of results, replication of analyses;
6. Writing – original draft: writing of the first version of the manuscript;
7. Writing – review & editing: revision and correction of intellectual content as well as form;
8. Visualisation: preparation, creation and/or presentation of the published work, specifically the visualisation of data;
9. Supervision: responsibility for leading and organising the research activities, including planning and execution.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Meaning
C-factor	Cover-management factor (RUSLE)
CLR	Conditional Logistic Regression
DEM	Digital Elevation Model
DTM	Digital Terrain Model
EI30	Erosion index (30-minute rainfall intensity × kinetic energy)
Fcover	Fraction of vegetation cover
GISER	Gestion Intégrée Sol Érosion Ruissellement, Cellule GISER, SPW, the Walloon technical advisory unit for runoff- and erosion-related flood risk, and the source of the muddy-flood occurrence database used in this study
GPR	Gaussian Process Regression
K-factor	Soil erodibility factor (RUSLE)
LIDAXES	Related Walloon LiDAR-based hydrological/erosion product, used here as a methodological reference point
LODO-CV	Leave-One-Date-Out Cross-Validation
LOOCV	Leave-One-Out Cross-Validation
LPIS	Land Parcel Identification System
LS-factor	Slope length–steepness factor (RUSLE)
MAJA	Multi-temporal atmospheric correction and cloud-detection processor (CESBIO/CNES/DLR)
NDVI	Normalized Difference Vegetation Index
OPW	Organisme Payeur de Wallonie, the Walloon paying agency that manages the LPIS reference-parcel layer
P-factor	Support-practice factor (RUSLE)
PICC	Projet Informatique de Cartographie Continue
PR-AUC	Area under the Precision–Recall curve
PU-RF	Positive-Unlabelled Random Forest
R-factor	Rainfall-runoff erosivity factor (RUSLE)

Acronym	Meaning
RMI	Royal Meteorological Institute of Belgium
ROC-AUC	Area under the Receiver Operating Characteristic curve
ROI	Region of Interest
RUSLE	Revised Universal Soil Loss Equation
SIGeC	Système Intégré de Gestion et de Contrôle, the Walloon system, managed by the OPW, used to track agricultural land and process CAP subsidy applications; produces the LPIS reference-parcel layer used in this study
SPI	Stream Power Index
SPW	Service Public de Wallonie
TWI	Topographic Wetness Index
WALOUS	WALlonie Occupation et Utilisation du Sol, the Walloon land-cover/land-use dataset
WaTEM/SEDEM	Water and Tillage Erosion Model / Sediment Delivery Model

1. Introduction

Muddy floods are sediment-laden runoff events generated on cultivated land, where intense rainfall over bare or sparsely vegetated fields detaches soil particles and transports them downslope before depositing mud in roads, buildings, drainage systems, or other downstream infrastructure (Boardman, Verstraeten, & Bielders, 2006; Boardman & Vandaele, 2020). They are hillslope- and small-catchment-scale processes controlled by the interaction of rainfall, soil surface condition, crop cover, topography, and sediment connectivity to exposed locations downslope; this interaction, and the recurrence of muddy flooding specifically in Wallonia, are examined in detail in Chapter 2. Regional muddy-flood prediction nonetheless remains difficult, for three main reasons, examined below from the most fundamental to the most immediately observable.

Muddy floods are event-dependent.

A catchment may be structurally susceptible because of its slope, soil erodibility, and connectivity to infrastructure, yet a muddy flood will only occur when triggering rainfall coincides with vulnerable land-cover and soil-surface conditions. Static susceptibility indicators are therefore insufficient on their own; conversely, rainfall or vegetation indicators alone do not explain why some catchments produce damaging off-site impacts while others do not.

Reported occurrence data cannot be treated as presence/absence.

Muddy-flood inventories are based on reported damage rather than systematic monitoring: recorded events represent confirmed presences, but unrecorded locations cannot be treated with certainty as true absences, since a flood may have occurred but gone unreported, affected only agricultural land, or fallen below the threshold for inclusion in the database. This resembles a positive-unlabelled learning problem, in which positive cases are known but the remaining observations contain an unknown mixture of true negatives and unobserved positives (Elkan & Noto, 2008), and it means that apparent model accuracy may partly reflect reporting patterns rather than transferable physical understanding.

Static erosion maps do not indicate off-site impact.

Existing erosion maps identify where soil loss is likely under average or scenario-based conditions, but not whether that erosion will become an off-site muddy flood at a particular exposed location and time. High soil loss on an isolated hillslope may never reach a road or settlement, while moderate erosion in a well-connected catchment may cause repeated flooding. Muddy-flood susceptibility is therefore not simply a function of gross erosion, but of erosion, runoff concentration, sediment connectivity, land-cover dynamics, rainfall timing, and exposure at the outlet together.

The combined challenge is to link relatively persistent catchment characteristics with temporally variable Earth-observation and rainfall information, in a way that remains interpretable for regional risk prioritisation, while working with an occurrence record that is inherently incomplete. This thesis addresses these three gaps together by examining whether multitemporal Earth-observation data, catchment connectivity, and reported muddy-flood occurrences can improve regional-scale muddy-flood susceptibility modelling in Wallonia, integrating dynamic Sentinel-2 vegetation information, rainfall erosivity, soil-erosion-model outputs, land-use characteristics, and watershed-based reported-event records in place of static erosion indicators alone. Its contribution is therefore twofold: it tests whether combining dynamic environmental indicators with catchment connectivity can better distinguish erosion-prone areas from locations where erosion becomes an off-site muddy flood, and it explicitly treats the reported-event inventory as positive-unlabelled rather than presence/absence data, evaluating muddy-flood prioritisation under this more defensible but more difficult labelling structure.

The aim of this thesis is to develop and evaluate a regional muddy-flood susceptibility model for exposed infrastructure in Wallonia, in which each upstream watershed is used to constrain and summarise the rainfall and land-cover conditions relevant to a given exposed point, while the model's output is assigned at the scale of that exposed point rather than as a watershed-wide susceptibility surface. The central research question is: to what extent can multitemporal Earth-observation indicators, rainfall erosivity, catchment connectivity metrics, and reported muddy-flood occurrences improve the regional-scale prioritisation of off-site muddy-flood risk in Wallonia, compared with static erosion-based approaches, evaluated using spatial, temporal, and spatio-temporal validation frameworks appropriate for a rare and spatially clustered geomorphological hazard.

2. Literature Review

2.1 The Muddy-Flood Phenomenon

2.1.1 Definition and Distinction from Fluvial Flooding

Muddy flooding is formally defined within the geomorphological and hydrological literature as flooding generated by runoff from agricultural land that carries substantial quantities of eroded soil, subsequently depositing this sediment beyond the area where it was originally detached (Boardman et al., 1994; Boardman and Vandaele, 2020). From a strict hydro-geomorphological perspective, this phenomenon is distinct from conventional fluvial or riverine flooding. Conventional flooding typically occurs when prolonged, basin-wide rainfall causes a permanent river system to exceed its channel capacity, resulting in overbank flow and the inundation of adjacent floodplains. In contrast, muddy flooding is fundamentally a localized hillslope and zero-order catchment process. The damaging process does not require the presence of a permanent or even semi-permanent watercourse. Rather, intense or persistent rainfall generates overland flow directly on agricultural land, mobilises exposed topsoil, concentrates the resulting fluid-sediment mixture into ephemeral topographic flow paths (such as thalwegs, agricultural ditches, or dry valleys), and transports it rapidly towards downstream receptors (Evrard et al., 2007a).

The hazard consequently emerges from a complex, multi-scalar interaction between meteorological forcing, intrinsic soil pedology, dynamic agricultural surface conditions, local topographic gradients, and the structural connectivity of the anthropogenic landscape. The fluid dynamics of muddy flooding are also distinct from clear-water flooding. The high concentration of suspended sediment and bedload alters the rheology of the runoff, shifting it from a Newtonian fluid toward a non-Newtonian, hyper-concentrated flow. The fluid becomes more viscous and dense, granting it elevated shear stress and transport capacity relative to clear water. This elevated density allows muddy runoff to transport debris, scour infrastructure, and cause structural damage to buildings and roads that shallow, clear-water flooding of comparable depth would not produce.

2.1.2 Terminology and Conceptual History

The phenomenon predates the scientific terminology now used to describe it. Historically, different European regions used various vernacular expressions for sediment-laden agricultural runoff, reflecting the fact that rural communities recognised and suffered from the phenomenon long before a unified scientific nomenclature emerged. Boardman and Vandaele (2020) note earlier terms including *mud deluges* in British literature, *inondations boueuses* in French, and *modderoverlast* in Dutch. For much of the twentieth century, these events were treated as localized, unavoidable consequences of extreme weather. They were under-reported in local administrative records, often dismissed as isolated agricultural nuisances rather than studied as a systemic, predictable regional environmental issue.

Boardman, Ligneau, De Roo and Vandaele (1994) played a pivotal role in shifting this paradigm, publishing the first systematic review of muddy flooding in western Europe and establishing it as a coherent, transnational research topic. Their paper documented the geographical distribution, causes, and consequences of agricultural runoff across Northwest Europe, consolidating previously isolated local incidents into a recognised environmental problem and demonstrating that these events shared common physical drivers across national borders. The specific term “muddy flooding” subsequently became entrenched in the literature, defined explicitly by Boardman, Verstraeten and Bielders (2006) as an off-site impact of soil erosion, deliberately distinguished from the on-site agronomic loss occurring within the field itself.

2.1.3 Muddy Flooding versus On-Site Erosion

This terminological distinction carries paramount importance for spatial hazard modelling: soil erosion estimates cannot be automatically interpreted as muddy-flood hazard. A given agricultural parcel can experience severe on-site soil loss without ever producing a damaging flood if the mobilized sediment is deposited safely in a vegetative buffer zone, woodland, or topographic depression before reaching a settlement. Conversely, a field with only moderate erosion potential can contribute disproportionately to muddy flooding if it is highly functionally connected to a road, sunken lane, or municipal drainage network (Verstraeten et al., 2003). Muddy-flood susceptibility therefore depends heavily on the hydrological pathways and the spatial separation between the source of sediment production and the downstream receptor, a distinction that motivates the connectivity-based approach developed later in this thesis (Section 2.5).

2.1.4 Spatial Distribution, Landscape Evolution, and the Loess Belt

The significance of this phenomenon has been thoroughly documented across northwestern Europe, with early syntheses identifying central Belgium, northern France, southern England, and the South Limburg region of the Netherlands as particularly susceptible geographic zones (Boardman et al., 1994; Boardman, 2010). The Walloon region of Belgium serves as a widely studied example of this hazard, combining highly erodible loess-derived soils, intensive mechanized agriculture, rolling topography, and rural settlement networks situated predominantly in valley bottoms. In a region-wide assessment of the Belgian Walloon silt-loam and sandy-loam belt, Bielders, Ramelot and Persoons (2003) surveyed all municipalities in the region together with 1,500 farmers, and found that 68% of municipalities in the silt-loam and sandy-loam regions had experienced at least one muddy-flood event attributable to agricultural runoff during the preceding ten years. A separate, more detailed survey of central Belgium by Evrard, Bielders, Vandaele and van Wesemael (2007a) found that 79% of municipalities (n = 201) had been affected at least once in a ten-year period, and 22% of the affected municipalities had experienced more than ten separate flood events over that decade. Together, these figures establish muddy flooding not as a localized anomaly but as an endemic regional issue intrinsic to the local agricultural geography.

The frequency, magnitude, and spatial distribution of muddy flooding are strongly controlled by long-term anthropogenic changes to the agricultural landscape, largely driven by post-WWII agricultural policies and the European Common Agricultural Policy (CAP). In Wallonia and similar European regions, the push for agricultural modernization and mechanization drove decades of intensive land consolidation (often referred to as *remembrement* in Francophone regions). The progressive enlargement of agricultural parcels to accommodate increasingly large machinery necessitated the systematic removal of hedgerows, vegetative buffer strips, and traditional earthen banks. These landscape modifications increased the effective slope length over which water and sediment can travel without interruption, maximizing the structural and hydrological connectivity of the terrain.

Simultaneously, demographic shifts and counter-urbanisation, the expansion of residential settlements, commuter towns, and paved infrastructure into rural valleys, increased societal exposure to this hazard. The number of potential human and infrastructural receptors located immediately downstream of sediment-producing land has grown substantially over the last fifty years.

2.1.5 Economic, Environmental, and Psychosocial Impacts

The socio-economic and environmental impacts of muddy flooding are severe, persistent, and multifaceted. Off-site, sediment-laden water infiltrates residential properties, damaging interiors, electrical systems, and structural foundations. It buries transport infrastructure in thick, slippery mud, causing traffic disruptions and road accidents, and it obstructs municipal drainage and sewer systems, creating secondary localized flooding that can persist after the rain has stopped.

Environmentally, the transported sediment acts as an efficient vector for agrochemicals. Fine silt and clay particles possess high specific surface areas and cation exchange capacities, allowing them to bind with nitrates, orthophosphates, and pesticides. When muddy runoff enters receiving watercourses, it introduces spikes in these agrochemicals, contributing to eutrophication, algal blooms, and degradation of local aquatic habitat.

Crucially, the economic cost of off-site damage far outweighs the purely agronomic value of the soil lost from the source fields. Boardman et al. (1994) and Verstraeten et al. (2003) documented that while the loss of topsoil might cost an individual farmer a marginal amount in lost yield or fertilizer replacement, the transport of that same soil volume into a village can cause substantial property damage, emergency response, and municipal cleanup costs. Evrard et al. (2007a) estimated the annual cost of erosion-related damage for the Belgian loess belt at approximately €16–172 million. The literature also increasingly recognizes the psychological toll on residents: the repetitive nature of muddy flooding can create lasting anxiety among homeowners in susceptible valley bottoms, as heavy summer thunderstorms bring a renewed threat of property damage. This economic and psychosocial disparity underlies why muddy flooding is now recognized as a distinct societal hazard rather than a mere consequence of agricultural soil degradation.

2.2 Physical and Hydrological Drivers of Muddy Flooding

2.2.1 Rainfall Triggers, Kinetic Energy, and Overland Flow Generation

Muddy flooding is a sequential chain of linked physical processes, beginning with meteorological forcing. Rainfall acts as the immediate trigger, supplying both the kinetic energy required to detach consolidated soil particles and the transport medium required to generate overland flow. Cumulative rainfall amount (e.g., total millimetres per day) is, however, a poor descriptor of a storm's actual erosive potential. The intensity, duration, drop-size distribution, and temporal structure of rainfall dictate both the generation of overland flow and the kinetic energy available for detachment.

In classical hydrology, surface runoff is generally categorized into two mechanisms: Dunnean (saturation-excess) overland flow and Hortonian (infiltration-excess) overland flow. Dunnean flow occurs when the entire soil profile becomes saturated from a rising water table, typically during prolonged, low-intensity winter rainfall. Muddy floods in agricultural landscapes are, by contrast, overwhelmingly driven by Hortonian flow, which occurs when instantaneous rainfall intensity exceeds the infiltration capacity of the surface soil, regardless of whether the deeper soil profile is saturated. High-intensity convective storms, characterised by large, fast-falling raindrops, deliver substantial kinetic energy to the soil surface, triggering Hortonian flow.

Verstraeten, Poesen, Demarée and Salles (2006) demonstrated substantial temporal variability in rainfall erosivity using a 105-year (1898–2002) high-resolution rainfall record from Ukkel, Belgium. Using the EI_{30} index, which multiplies the total kinetic energy of a storm (E) by its maximum 30-minute intensity (I_{30}), and their analysis derived a long-term rainfall erosivity (R-factor) of 871 $MJ \cdot mm \cdot ha^{-1} \cdot h^{-1} \cdot yr^{-1}$ for central Belgium, and found that erosivity is strongly seasonal, with intense summer convective storms contributing disproportionately to the annual erosive forcing despite representing a smaller share of total annual precipitation volume.

The relationship between rainfall and muddy flooding is therefore highly conditional. A moderate, low-intensity frontal rainfall event in winter that produces little damage when winter wheat provides surface protection may generate considerably more destructive runoff if the same rainfall volume falls as a short, high-intensity convective burst on an exposed summer field. Evrard et al. (2007a) confirmed this seasonal dependence, reporting that the daily rainfall threshold associated with flood occurrence in central Belgium was markedly lower in late spring than later in the summer.

2.2.2 The Pedology and Mechanics of Soil Crusting in Loess Landscapes

The conversion of rainfall into Hortonian surface runoff is fundamentally governed by the physical, chemical, and structural condition of the topsoil. While heavy rainfall initiates the process, the dynamic infiltration capacity dictates whether that rainfall enters the root zone or accumulates as overland flow. Infiltration can be rapidly, and sometimes irreversibly, reduced by surface sealing, compaction from heavy machinery, and structural crusting.

This pedological dynamic is a central physical mechanism driving hazard in the Walloon loess belt. Loess-derived soils are aeolian deposits dominated by silt-sized particles. Unlike clay-rich soils, which possess strong electrochemical cohesion binding aggregates tightly together, or sandy soils, which maintain large, stable macropores for rapid drainage, silty soils possess notoriously weak aggregate stability. Under the repeated impact of falling raindrops, the structural aggregates of bare topsoil physically shatter, a process known in soil physics as mechanical slaking and physicochemical dispersion.

The detached fine silt and clay particles are subsequently washed into surrounding soil pores, clogging them. As the surface dries between showers, it forms a dense, largely impermeable surface (or structural) crust. Once a crust forms, the hydraulic conductivity of the topsoil drops sharply (for example, from roughly 30 mm/hr down to 2 mm/hr), substantially increasing the proportion of subsequent rainfall converted into surface runoff.

Rainfall simulation experiments conducted by Evrard, Vandaele, van Wesemael and Bielders (2008) on 65 cultivated fields across central Belgium quantified this effect directly. Using a portable simulator to apply a heavy rainfall event (60 mm/hr over 30 minutes) representative of muddy-flood-triggering storms in the region, they found that runoff was not generated on bare, uncrusted, freshly ploughed soil, but that mean runoff coefficients on fields with an established crust ranged from 13% (winter wheat in July) up to 58% (sugar beet or maize in May and June), with grassed buffer strips and grassed waterways showing higher coefficients still (62% and 73% respectively), reflecting their role in concentrating rather than infiltrating flow once saturated. Complementary work by Ciampalini, Follain and Le Bissonnais (2012), later applied to a cultivated catchment in central Belgium, identified dynamic infiltration capacity, linked directly to crusting, as one of the most sensitive inputs in the LandSoil erosion-deposition model. Antecedent rainfall therefore influences muddy-flood susceptibility in a dual manner: it elevates antecedent soil moisture, but more importantly it physically degrades and seals the soil surface structure, preconditioning the field for higher runoff during the next storm.

2.2.3 Agricultural Land Use and the May–June Vulnerability Window

While soil physics dictates intrinsic erodibility, agricultural land cover represents the most dynamic and influential human control on muddy-flood susceptibility. Treating all cropland as a single, static land-cover class overlooks profound phenological and structural differences between crop types. Bielders et al. (2003) established that the regional prevalence of wide-row crops was associated with increased runoff and erosion, whereas densely sown winter cereals generally provided greater landscape-level protection. Evrard et al. (2007a) linked historical increases in muddy flooding to the agricultural transition toward spring-sown crops, particularly silage maize, potatoes, and sugar beet, at the expense of traditional winter cereals, a shift encouraged by European agricultural policies supporting livestock feed and bio-energy production.

The underlying mechanics of this relationship matter for hazard prediction. Winter cereals are sown in autumn; by the time erosive summer convective storms arrive in May and June, these fields have a dense, mature canopy that intercepts rainfall and mature root systems that bind the soil matrix. Spring-sown row crops, by contrast, are planted in April or May with wide row spacing (e.g., roughly 75 cm for maize) to accommodate harvesting machinery, leaving substantial tracts of bare soil exposed during their early stages of development.

Walloon research has therefore identified late spring, specifically May and June, as a critical “vulnerability window,” during which spring-sown crops have not yet developed sufficient canopy cover to shield the soil. This creates a dangerous intersection of two independent processes: rainfall erosivity rises with the onset of intense summer convective storms at precisely the moment when agricultural surface protection is weakest. The physical vulnerability of a maize field is therefore not determined by its static crop identity, but dynamically by its stage of phenological development on the day a storm occurs. A model intended to predict muddy flooding on a daily basis accordingly requires a dynamic representation of this changing agricultural surface.

2.3 Representing Agricultural Susceptibility: The Cover-Management Factor

2.3.1 The RUSLE2 Theoretical Framework and Subfactors

The Revised Universal Soil Loss Equation (RUSLE) provides the most widely implemented empirical framework for representing spatial variation in soil erosion. Within the RUSLE mathematical framework, the cover-management factor (the C-factor) is the sole component explicitly designed to capture the influence of agricultural activity, human management, and vegetation development. Topography (the LS-factor) and intrinsic soil erodibility (the K-factor) are comparatively stable at the timescale of a single agricultural season; the C-factor is therefore the primary mechanism through which the rapidly changing condition of agricultural land can be incorporated into a predictive environmental model.

The RUSLE2 science documentation (USDA-ARS, 2013) provides a detailed, deterministic interpretation of the C-factor, departing from the simplified, static annual values commonly used in basic GIS analyses. In RUSLE2, the cover-management factor is expressed as a continuous soil loss ratio (SLR) comparing the soil loss expected from a specifically managed field, on a given day, to the soil loss expected from a continuously bare, tilled fallow plot under identical rainfall. It is calculated on a daily timestep as the product of up to seven subfactors (USDA-ARS, 2013):

$$SLR = PLU \times CC \times SC \times SR \times RH \times SB \times SM$$

where PLU is the prior-land-use subfactor, CC the canopy-cover subfactor, SC the ground- (surface-) cover subfactor, SR the soil-surface-roughness subfactor, RH the ridge-height subfactor, SB the soil-biomass subfactor, and SM the antecedent-soil-moisture subfactor (USDA-ARS, 2013). The variables correspond to:

1. Canopy-cover subfactor (CC): the interception of raindrops by live vegetation above the soil surface.

2. Ground-cover subfactor (SC): the protection offered by dead residue, litter, or mulch in direct contact with the soil.

3. Soil-surface-roughness subfactor (SR): accounts for micro-topography (clods and depressions left by tillage), which pond water and slow flow velocity. Roughness decays as rainfall progressively flattens the clods.

4. Ridge-height subfactor (RH): accounts for the orientation and depth of tillage marks; ridges along the contour trap water, while down-slope ridges accelerate flow.

5. Soil-biomass subfactor (SB): represents the binding effect of live and dead plant roots within the topsoil, increasing soil cohesion.

6. Soil-consolidation subfactor: accounts for the settling and physical stabilization of soil over time following a disruptive tillage event.

7. Antecedent-soil-moisture subfactor (SM): adjusts expected runoff based on the hydrologic condition of the soil profile prior to the storm.

This formulation illustrates that a dynamic C-factor is considerably more complex than simply scaling a satellite-derived vegetation index. RUSLE2 explicitly recognises that protective capacity fluctuates continuously as a function of biological growth, mechanical intervention, and cumulative weather.

2.3.2 Effective Canopy Cover versus Ground Cover Dynamics

While the full RUSLE2 formulation is appropriate for field-scale agronomy, not all subfactors can be reliably estimated at regional or national scale using currently available Earth observation data. For a first-order approximation in regional hazard modelling, the two most influential components governing protection against rainfall detachment are the canopy-cover and ground-cover subfactors.

The canopy subfactor quantifies the effect of live, photosynthetic vegetation suspended above the soil surface. In the RUSLE2 formulation, canopy cover reduces erosivity as an approximately exponential function of the fraction of soil shielded from direct raindrop impact, moderated by the effective fall height of drops from the canopy (USDA-ARS, 2013):

$$CC = 1 - Fc \cdot exp(-0.328 \cdot H)$$

where F_C is the fraction of ground surface covered by canopy when viewed vertically from above, and H is the effective fall height of drops from the canopy, in metres (USDA-ARS, 2013). The inclusion of effective fall height is a critical physical nuance: while a canopy intercepts raindrops, the water eventually drips from the leaves. If the canopy is tall (e.g., mature maize reaching heights of 2.5 m), droplets coalesce on the leaves and fall as larger drops, regaining substantial terminal velocity and impact energy before striking the soil. A tall canopy with 80% coverage therefore provides less protection than a low-lying canopy of the same coverage.

A defining feature of RUSLE2 is its strict physical distinction between canopy cover and ground cover: effective canopy cover is calculated after accounting for the portion of soil already protected by ground cover, since vegetation growing above an already-shielded surface (e.g., green leaves over a layer of post-harvest straw) does not provide the same additional reduction in soil detachment as vegetation over bare soil. Ground cover is defined as material in direct physical contact with the soil surface, including dead crop residues, leaf litter, and applied mulch, and is highly effective because it both absorbs raindrop impact energy and directly obstructs and slows overland flow, reducing the shear stress available for rill erosion.

2.3.3 Simplifying the C-Factor for Regional Spatial Modelling

This strict physical distinction between canopy and ground cover has significant implications for remote sensing methodologies. An optical satellite sensor can estimate the proportion of green vegetation above the soil surface reasonably accurately by measuring chlorophyll absorption, but it faces substantial difficulty in quantifying the volume, orientation, and arrangement of non-photosynthetic crop residue lying on the soil, whose spectral signature often closely resembles bare soil, particularly in the shortwave infrared. This limitation is especially problematic in conventional tillage systems, where residue may be incorporated into the soil by a mouldboard plough shortly after harvest, rendering it invisible from space while still physically altering the topsoil matrix.

If ground cover is set to zero as a necessary, data-driven approximation given these observational constraints, and if the effective fall-height correction is neglected due to the difficulty of continuously estimating dynamic crop height from space, the RUSLE2 canopy relationship simplifies to an approximately linear function of canopy cover alone:

$$C \approx 1 - Fc$$

This simplified formulation is attractive for regional predictive studies because it establishes a direct, computationally efficient transformation from remotely estimated canopy cover to a daily erosion-protection proxy. Almagro et al. (2019) compared several NDVI-based C-factor estimation approaches against field-measured values in a tropical setting and found that the relationship between vegetation indices and the C-factor is not universal but depends on the specific empirical formula and local conditions applied, underscoring that any canopy-based C-factor proxy used in this thesis should be treated as a regionally calibrated approximation rather than a directly transferable constant.

Within the context of empirical muddy-flood modelling, this canopy-based C-factor must therefore be documented explicitly as a first-order proxy for the dynamic cover effect. Setting ground cover to zero inherently assumes no direct erosion reduction from post-harvest residue. This assumption is most defensible during the May–June vulnerability window, when active canopy growth is the primary protective mechanism and fields are typically cleared of old residue, but it becomes a known source of uncertainty in the post-harvest autumn months, where the approximation may overestimate hazard.

2.4 Remote Sensing of Dynamic Vegetation

2.4.1 Vegetation Indices, Sensor Capabilities, and the Soil Background Problem

Continuous, direct field measurement of canopy cover is not feasible at regional scale, so satellite-derived vegetation indices serve as the most viable operational alternative for deriving spatially and temporally varying C-factors. The Normalized Difference Vegetation Index (NDVI) has historically been the most widely applied metric for this purpose (Van der Knijff et al., 1999), owing to its computational simplicity and its reliance on the contrast between the red band, which chlorophyll absorbs strongly, and the near-infrared band, strongly reflected by the spongy mesophyll structure of the leaf.

Modern Earth observation constellations, such as the European Space Agency's Sentinel-2 mission and NASA/USGS Landsat 8/9, provide spatial resolutions (10–30 m) suited to capturing intra-field agricultural variability. However, NDVI is not strictly equivalent to physical canopy cover: its relationship with actual canopy coverage is distorted by confounding optical variables, chief among them soil background reflectance. During early crop development, when vegetation is sparse, the spectral signature of exposed bare soil strongly influences the pixel NDVI value independently of actual plant biomass, precisely during the period when muddy-flood hazard is highest. NDVI is also known to saturate at high Leaf Area Index (LAI), losing sensitivity once a crop reaches full canopy closure and treating moderately dense and extremely dense fields similarly.

The remote sensing literature has proposed alternative indices to address these limitations. The Soil-Adjusted Vegetation Index (SAVI) and the Enhanced Vegetation Index (EVI) introduce correction factors intended to reduce soil brightness influence and mitigate saturation at high biomass. Almagro et al. (2019) demonstrated, in a comparison of NDVI-based C-factor estimation methods for tropical regions, that the mathematical relationship between vegetation indices and erosion protection is method- and region-dependent, cautioning against uncritically transferring empirical formulas between different agroecological settings without local validation.

A methodologically distinct alternative to deriving canopy cover indirectly from a vegetation index is to use fCover (the Fraction of green Vegetation Cover), a Sentinel-2 biophysical variable produced directly by ESA's own Level-2B processor. Unlike NDVI, SAVI, and EVI, which are index formulas subsequently converted to a cover estimate through a separate empirical relationship, fCover is generated by inverting a canopy radiative-transfer model (typically PROSAIL) using a neural network trained on simulated reflectance spectra, and is defined directly as the fraction of ground covered by green vegetation when viewed from above (Weiss and Baret, 2016). This makes it conceptually closer to the RUSLE2 Fc term than an index such as NDVI, since it is already expressed as a cover fraction rather than requiring a further conversion step. Validation studies indicate that this processor performs comparatively well over agricultural crops: Kamenova and Dimitrov (2021) reported lower retrieval error for fCover than for LAI in winter wheat in Bulgaria, while Fernandes et al. (2023) found that the same underlying algorithm meets user accuracy requirements over homogeneous canopies such as cropland but underestimates values over structurally heterogeneous canopies such as forests, a distinction favourable for application to the largely homogeneous arable fields of the Walloon loess belt. Because it is a standard product generated automatically within the SNAP toolbox alongside LAI and FAPAR, fCover offers a more direct, less assumption-laden route to a dynamic C-factor proxy than an NDVI-based empirical formula, at the cost of depending on a global, generically trained radiative-transfer inversion rather than a locally calibrated index-to-cover relationship.

2.4.2 The Challenge of Cloud Contamination and Temporal Leakage

While the pursuit of dynamic, frequent optical vegetation data is scientifically important, it introduces a significant methodological obstacle: cloud contamination. High-resolution optical satellites cannot penetrate cloud cover, and muddy floods are by definition triggered by substantial weather systems.

Consequently, direct satellite observations of the Earth's surface are frequently unavailable on the exact day a muddy flood occurs.

Synthetic Aperture Radar (SAR), such as Sentinel-1, has been explored as a cloud-penetrating alternative for measuring crop structure and biomass. Because radar wavelengths penetrate cloud cover, they offer more consistent observation opportunities. However, SAR backscatter is strongly influenced by the dielectric constant of the surface (soil moisture) and by surface roughness, which complicates the isolation of a pure vegetation signal for C-factor derivation and makes it less stable for routine daily modelling than optical indices.

Consequently, researchers commonly rely on temporal interpolation of optical data to fill gaps between clear satellite overpasses. Techniques such as linear interpolation, Whittaker smoothing, or Savitzky–Golay filtering are used to construct continuous daily vegetation index time-series. This introduces the risk of *temporal leakage* in predictive frameworks. Suppose a storm triggers a muddy flood on 15 June, and the nearest clear satellite images are from 1 June and 25 June. Interpolating the vegetation state for 15 June using the 25 June image relies on information from after the event, an image that may itself reflect the flood's impact (e.g., a washed-out or flattened crop, or accelerated growth from the added moisture).

A rigorous methodology must therefore enforce strict temporal boundaries: the dynamic C-factor assigned to any given event should be derived exclusively from lagged observations or backward-looking composites, ensuring that predictions rely only on information that would realistically have been available prior to the storm's onset.

2.5 Spatial Routing and Landscape Connectivity

2.5.1 Beyond RUSLE: A Comparative Overview of Erosion and Sediment-Routing Models

While the RUSLE framework is valuable for representing the potential for on-site soil detachment, it is insufficient on its own for the muddy-flood prediction problem. RUSLE was designed to estimate long-term, average annual soil loss from an individual hillslope; it contains no runoff mechanism, no sediment-routing component, and no representation of landscape connectivity, and it was never intended to calculate the probability that mobilized sediment is routed out of the field, across the wider landscape, and delivered to a downstream receptor during a single storm event (Renard et al., 1997). In heavily modified landscapes such as Wallonia, topography and drainage infrastructure exert substantial control over the post-detachment movement of sediment, so a muddy-flood susceptibility framework must draw on models that go beyond RUSLE's original scope. Table 2.1 compares the principal candidate models against the criteria most relevant to this thesis: whether the model represents runoff generation explicitly, whether it routes sediment across the landscape rather than only estimating detachment, whether it accounts for landscape connectivity, its native temporal scale, and its main practical limitation for regional application.

Table 1. Comparison of erosion and sediment-transport models relevant to muddy-flood susceptibility mapping.

Model	Main purpose	Temporal scale	Runoff	Sediment routing	Connectivity	Main limitation
RUSLE	Soil loss	Annual	No	No	No	Not event-based
Curve Number	Runoff	Event	Yes	No	No	Limited sediment representation
WaTEM/SEDEM	Erosion + sediment delivery	Mainly annual	Limited	Yes	Yes	Temporal dynamics
EROSION 3D	Process/event erosion	Event	Yes	Yes	Yes	Data / complexity
LandSoil	Erosion/deposition	Event/process	Yes	Yes	Yes	Parameterisation
WaterSed	Runoff + erosion, multi-year	Event, multi-year	Yes	Yes	Yes	Parameterisation
OpenLISEM	Runoff/erosion, physically-based	Event	Yes	Yes	Yes	Data / complexity
Machine learning	Susceptibility	Flexible	Depends on inputs	Depends on inputs	Yes	Training data / interpretability

RUSLE / RUSLE2

RUSLE estimates long-term average annual soil loss from five empirical factors and contains no runoff or routing component. Its annual timestep is a poor match for the sub-daily, event-driven dynamics that trigger muddy floods (Renard et al., 1997), so in this thesis it functions strictly as a source of the dynamic C-factor, not as a stand-alone predictive model.

The Curve Number (CN) Method

The SCS Curve Number method estimates event-scale direct runoff from rainfall depth, land use, and hydrologic soil group (USDA-SCS, 1972), addressing the runoff-generation process RUSLE omits. It has no native sediment or routing component, so in this thesis CN-derived runoff potential is used as a hydrologic predictor layer alongside the dynamic C-factor rather than as a description of sediment behaviour.

WaTEM/SEDEM

WaTEM/SEDEM is the model this thesis relies on most directly for sediment routing, and its origin and structure warrant closer attention than the other models surveyed here. It was developed at KU Leuven's Laboratory for Experimental Geomorphology (now the Geography and Tourism Research Division within the Department of Earth and Environmental Sciences) as a merger of two originally separate models: WaTEM (the Water and Tillage Erosion Model, representing the combined effect of water and tillage-driven soil redistribution; Van Oost, Govers and Desmet, 2000) and SEDEM (the Sediment Delivery Model, representing catchment-scale sediment yield; Van Rompaey, Verstraeten, Van Oost, Govers and Poesen, 2001).

Structurally, WaTEM/SEDEM combines a RUSLE-type empirical detachment formulation with a spatially explicit sediment-routing algorithm operating on a two-dimensional grid. Rather than using the standard, one-dimensional RUSLE LS-factor, it applies the modified algorithm of Desmet and Govers (1996), which computes slope length and steepness directly from a DEM and a parcel map using a multiple-flow-direction approach, later adjusted by Takken et al. (2001) to account for tillage direction; this two-dimensional treatment allows the model to represent concentrated flow along field boundaries and thalwegs rather than only diffuse sheet erosion. At each grid cell, the model compares gross erosion (from the RUSLE-type formulation) against a locally defined transport capacity: where gross erosion exceeds transport capacity, sediment is exported downslope; where transport capacity is exceeded by the incoming sediment load, deposition occurs. This distinction between gross erosion and net erosion or deposition is what allows WaTEM/SEDEM, unlike RUSLE alone, to identify not just where soil is detached but where it is actually delivered (Verstraeten et al., 2002).

The model has been applied extensively across the Belgian loess belt since its development (e.g., Van Rompaey et al., 2001; Verstraeten et al., 2006) and remains actively maintained by KU Leuven, with recent releases adding a Curve Number extension for estimating runoff and sediment concentration alongside the core erosion-routing module, an extension of direct relevance to this thesis given its own pairing of RUSLE-derived erosion potential with CN-derived runoff. Its principal limitation here is temporal: WaTEM/SEDEM was designed to estimate mean annual sediment yield rather than to resolve the dynamics of an individual storm, so representing the daily variability central to a muddy-flood susceptibility approach requires adapting its inputs to a finer timestep than the model was originally built to accept.

EROSION 3D

EROSION 3D (Schmidt et al., 1996; von Werner, 1995) is a physically-based, raster-distributed model simulating infiltration, runoff, flow routing, detachment, transport, and deposition explicitly, at a temporal resolution as fine as ten minutes within a single storm. This makes it the most complete process representation among the models considered here, validated across several European loess and loam catchments (e.g., Saxony and Lower Saxony, Germany) against measured runoff and sediment yield. Its cost is substantial: EROSION 3D requires site-specific soil physical parameters typically obtained through dedicated rainfall-simulation campaigns rather than existing regional databases, difficult to justify at the scale of an entire Walloon catchment within a thesis timeline.

LandSoil

LandSoil (Ciampalini et al., 2012) is an event-based model built on the STREAM expert-rules approach (Cerdan et al., 2001, 2002), simulating interrill, rill, and tillage erosion at fine spatial resolution (1–10 m) with lower parameterisation demands than EROSION 3D. Its STREAM lineage is directly relevant here: Evrard, Cerdan, van Wesemael and Bielders (2009) applied STREAM itself to two Belgian catchments with soil textures similar to the Walloon loess belt and found it reliable specifically where Hortonian overland flow dominates, the same runoff regime characteristic of Walloon muddy-flood triggering, and LandSoil's later Belgium-specific validation builds directly on that transfer. It nonetheless remains demanding to run at the catchment extent and rainfall-event frequency this thesis requires.

WaterSed

WaterSed (Landemaine, 2016) is a more recent development in the same STREAM lineage as LandSoil, combining concepts from STREAM (Cerdan et al., 2001) and the LISEM model (De Roo et al., 1996) into a raster-based, event-scale runoff and erosion model. Its principal advance over its predecessors is the ability to represent both saturation-excess (Dunnean) and infiltration-excess (Hortonian) runoff within the same framework, and to simulate large catchments over multiple years without excessive equifinality issues, addressing a limitation noted for LandSoil and STREAM, whose expert rules were calibrated primarily for Hortonian-dominated conditions. This matters for Wallonia specifically, since winter and early-spring runoff in the region is not exclusively Hortonian in origin, and a model restricted to that single mechanism may misrepresent the shoulder seasons around the May–June vulnerability window.

A Multi-Model Comparison on Belgian Loess: Baartman et al. (2020)

Rather than relying on any single model's self-reported validation, Baartman, Nunes, Masselink, Darboux, Bielders, Degré, Cantreul, Cerdan, Grangeon, Fiener, Wilken, Schindewolf and Wainwright (2020) directly compared five of the models discussed above on the Chastre catchment in the Belgian loess belt, the same catchment used by Cantreul et al. (2018) for the connectivity-resolution study. Across 53 scenarios varying field size, orientation, and conservation measures, they found that rainfall amount was the dominant control on connected sediment and runoff export in every model tested, and that *functional* aspects of connectivity (how water and sediment actually moved during a given event) mattered more than *structural* connectivity (the static arrangement of fields and landscape features) in explaining model output. Notably, the five models showed limited agreement with one another on the spatial patterns of connectivity and on the effects of field size, crop allocation, and conservation practices, a result the authors describe as unexpected, and which they attribute to differences in how each model parameterises runoff generation and routing rather than to the underlying physics being genuinely ambiguous.

This finding carries a direct methodological implication for the present thesis. It cautions against treating a static, structural connectivity index as sufficient on its own to represent muddy-flood susceptibility, since Baartman et al.'s comparison, conducted on Belgian loess soils under conditions closely resembling the Walloon study area, found structural connectivity alone to be a comparatively weak predictor relative to the dynamic, rainfall-driven behaviour of the models. It reinforces the rationale for pairing a structural connectivity layer with dynamic, event-specific predictors (rainfall forcing and the dynamic C-factor) rather than relying on connectivity structure in isolation.

Machine learning models differ qualitatively from the process-based models above: rather than encoding a physical description of infiltration or routing, they learn a statistical mapping from a flexible set of input predictors to observed susceptibility. This flexibility is precisely what allows a hybrid approach, in which RUSLE-, CN-, or connectivity-derived variables are supplied as engineered predictors, to substitute for the runoff and routing components that RUSLE itself lacks, without requiring the intensive parameterisation of EROSION 3D or LandSoil. The trade-off is reduced physical interpretability and a dependence on the quality and completeness of the (positive-unlabelled) training data.

Implications for This Thesis

None of the fully physically-based models (EROSION 3D, LandSoil, WaterSed) are feasible to parameterise across the full extent of this thesis's study area within the available timeframe, given their reliance on site-specific rainfall-simulation data or multi-year calibration. WaTEM/SEDEM offers a Belgium-validated, connectivity-aware routing framework but at an annual rather than event timestep. The Baartman et al. (2020) comparison further suggests that even where such models are deployed, structural connectivity indices alone should not be treated as a reliable stand-in for event-driven, rainfall-dependent behaviour. Together, these findings motivate the hybrid physical-empirical strategy: using RUSLE-derived dynamic erosion potential and CN-derived runoff potential as engineered predictors, combined with a structural connectivity index and a machine learning classifier, as a pragmatic middle path between the data demands of full process-based simulation and the physical blindness of a purely statistical model.

Research Traditions in Other Affected Regions

Central Belgium is not the only setting in which these models have been applied, and briefly situating the Belgian, regional-mapping tradition against a few contrasting approaches clarifies why this thesis follows the path it does. In southern England, Boardman and collaborators have favoured long-term, site-level monitoring over regional statistical mapping (Boardman, Evans and Ford, 2003; Evans and Boardman, 2003), and identified a seasonal driver, autumn runoff from winter cereal stubble, that differs from the Belgian May–June window . In northern France, the STREAM model has been applied less to rank existing sites by susceptibility than to simulate future land-use scenarios for participatory planning with farmers and local authorities (Souchère et al., 2005). In South Limburg, the Netherlands, LISEM was developed and commissioned directly by regional water authorities and municipalities for scenario testing (De Roo, Wesseling and Ritsema, 1996a, 1996b; De Roo, 2000), reflecting the most institution-intensive of these traditions.

Two further contrasts are closer to home. Within Belgium itself, Flanders adopted a binding “Soil Erosion Decree” in 2001 that ties compliance to CAP income support and assigns each parcel a mandatory erosion rating (Vandaele, 2020), whereas Wallonia's GISER programme remains diagnostic and voluntary; a susceptibility model of the kind developed here would feed into that discretionary process rather than a regulatory one. In Germany and Slovakia, EROSION 3D applications illustrate the limits of transferring a physically-based model across regions: Hänsel et al. (2019) used it for short-lead-time flood forecasting in Saxony, while Slovak applications first required reconciling German and USDA soil-classification systems and re-deriving crusting parameters for local soils, a concrete illustration of the reparameterisation burden discussed above. Taken together, these traditions confirm that no single methodology has become an international standard, and that this thesis's hybrid, connectivity-aware approach draws on more than the Belgian literature alone.

2.5.2 Structural versus Functional Connectivity and the Index of Connectivity (IC)

In modern geomorphology, connectivity is generally divided into structural and functional connectivity. Structural connectivity refers to the static, physical linkages within the landscape: the topography, field arrangement, and presence of roads. Functional connectivity refers to the dynamic, episodic linkage of the landscape during a specific storm event: how water actually flows given momentary rainfall intensity and soil saturation.

Sediment connectivity describes the degree to which sediment sources are physically linked to transport pathways and downstream sinks. Borselli, Cassi and Torri (2008) formalised this through the Index of Connectivity (IC), which estimates the relative probability of sediment transfer from an upslope contributing-area component (the driving force) and a downslope flow-path resistance component (the routing resistance).

The mathematical representation of connectivity is notably sensitive to the spatial resolution of the underlying terrain data. In a study conducted on a 1.24 km² agricultural site in central Belgium, Cantreul, Bielders, Calsamiglia and Degré (2018) compared DEM resolutions ranging from 0.25 to 10 m and found that IC values are systematically biased by pixel size, recommending a 1 m DEM as a suitable compromise between accuracy and computational cost. Fine-scale landscape features, such as agricultural ditches, grass strips, raised earthen banks, and sunken lanes (*chemins creux*), exert substantial control over actual flow pathways. A 10 m or 30 m resolution DEM acts as a low-pass filter that can erase these micro-topographic features and artificially smooth the landscape. Without hydro-conditioning the DEM (mathematically enforcing culverts and drainage lines), routing algorithms can misrepresent flow accumulation, meaning the scale and preparation of the elevation data materially affect whether a spatial model correctly identifies the true source-to-receptor pathway.

2.5.3 The Dual Role of Artificial Surfaces

Within these flow pathways, artificial surfaces occupy a complex, dual role. On one hand, expanding urbanisation increases societal exposure by placing residential buildings directly in the natural topographic paths of agricultural valleys.

On the other hand, artificial surfaces actively modify the hydrology itself. Paved roads act as efficient, low-friction conduits for fast-flowing water, sometimes functioning as artificial channels during intense storms, while concrete curbs and municipal drainage systems can intercept and divert runoff out of its natural catchment. De Walque, Degré, Maugnard and Bielders (2017) demonstrated that the precise spatial configuration of artificial surfaces, not merely their raw extent, is a key predictor of muddy-flood hazard in Wallonia, as it forms the direct methodological precedent for this thesis. The structural anatomy of the downstream landscape ultimately determines whether mobilized sediment dissipates harmlessly into a forest or becomes a damaging flood in a village square. Practical mitigation measures that alter this connectivity, grassed waterways and earthen dams, have shown measurable effect: Evrard, Vandaele, van Wesemael and Bielders (2008b) documented that a 12 ha grassed waterway and earthen dams installed in the thalweg of a 300 ha catchment near Velm, Belgium reduced peak discharge by over 40% and prevented flooding in the downstream village despite subsequent rainfall events with return periods up to 150 years.

2.6 Empirical Modelling and Machine Learning

2.6.1 De Walque et al. (2017): A Direct Methodological Precedent

De Walque, Degré, Maugnard and Bielders (2017) provide the direct methodological precedent for the empirical, presence/background approach to muddy-flood modelling followed in this thesis, and their study warrants closer examination than the other individual studies reviewed in this chapter. Working from a database of 442 muddy-flood-affected sites across Wallonia, they constructed an equal number of homologous non-flooded control sites, matched by contributing-area characteristics, and extracted eleven candidate explanatory variables spanning relief, land use, sediment production, and sediment connectivity for each site pair. Using simple and multiple logistic regression, they identified a final model requiring five explanatory variables: mean slope of the contributing area, a sediment connectivity index, and three metrics describing artificial surfaces specifically, their proportion, their spatial aggregation, and their proximity to the site outlet.

The model achieved a prediction quality of 76% on the calibration dataset and 81% on an independent validation dataset of 48 further site pairs, and the inclusion of artificial-surface characteristics improved model fit substantially (p-values ranging from 10^{-11} to 10^{-5}), consistent with the argument developed that the spatial configuration of urbanised land, not merely its extent, is a key control on hazard. This result matters for the present thesis in two respects. First, it demonstrates that a relatively compact set of static, persistent landscape variables, dominated by connectivity and artificial-surface metrics rather than erosion potential alone, can achieve strong discriminative performance for muddy-flood occurrence in Wallonia specifically, which is the same outcome and broadly the same geographic setting this thesis addresses. Second, because the model is built entirely from static, time-invariant predictors, it necessarily answers where muddy floods are generally more likely to occur, on average, rather than when a flood is likely on a given day. A daily prediction model shifts the question toward the dynamic meteorological and agricultural condition of the landscape on a specific day, identifying *where and when* a muddy flood is actually likely to occur, and this distinction is the central methodological gap this thesis addresses by extending De Walque et al.'s static, connectivity-based framework with dynamic, event-specific predictors.

2.6.2 Machine Learning and Hybrid Physical-Empirical Architectures

Answering this requires conceptualising muddy-flood occurrence as an interaction between static spatial susceptibility and dynamic environmental conditions (daily rainfall forcing, dynamic C-factor). These interactions are non-linear: a given rainfall threshold may cause no damage on a fully vegetated, low-slope field but trigger a flood on a bare, highly connected field, and the threshold itself shifts with the season and with antecedent soil moisture. Linear models such as De Walque et al.'s logistic regression represent this behaviour only through explicitly specified interaction terms, which become unwieldy as the number of candidate dynamic predictors grows; a model with daily rainfall, dynamic C-factor, connectivity, and their pairwise and higher-order interactions would require a prohibitively large set of manually engineered terms to approximate what a non-parametric model can learn directly from the data.

Ensemble decision-tree algorithms are consequently the most widely used class of model in comparable susceptibility-mapping applications (e.g., gully-erosion susceptibility mapping; Arabameri, Chen, Loche, Zhao, Li, Lombardo, Cerda, Pradhan and Bui, 2020), and are well suited to the muddy-flood problem for the same reasons. Random Forest constructs many decision trees on bootstrapped subsets of the training data and randomly restricted subsets of predictors, then averages their predictions, which naturally represents hierarchical threshold logic (e.g., "if slope exceeds X and canopy cover is below Y, hazard is high") without the analyst having to specify such rules in advance, and its use of many largely decorrelated trees makes it comparatively robust to multi-collinearity among predictors, a property that matters given the moderate correlation expected between the dynamic C-factor, rainfall erosivity, and connectivity variables. Gradient Boosting methods, and in particular Extreme Gradient Boosting (XGBoost), build trees sequentially, with each new tree fitted to the residual error of the ensemble so far, which typically improves predictive accuracy over Random Forest at the cost of being more sensitive to hyperparameter choices and more prone to overfitting on a small or heavily imbalanced positive class, a risk directly relevant to the muddy-flood class-imbalance problem.

A promising direction, and the one adopted in this thesis, is the "hybrid physical-empirical" architecture, in which outputs from established physical models are used as explicitly engineered predictors for a machine learning algorithm rather than being fed to it as raw spectral or meteorological data. Daily RUSLE erosivity and the dynamic C-factor, Curve Number runoff potential, and a structural connectivity index are computed first using established physical or empirical relationships, and only these derived, physically interpretable quantities are passed to the classifier. This combines the physical interpretability of process models, since each predictor has a known geomorphological meaning, with the pattern-recognition capacity of statistical learning to capture non-linear interactions between them, and it directly extends De Walque et al.'s static predictor set with the dynamic, event-specific predictors that a daily susceptibility model requires.

2.6.3 Class Imbalance, Spatio-Temporal Validation, and Interpretability

Developing a daily hazard model introduces substantial challenges for statistical validation due to extreme class imbalance. Muddy floods are rare events relative to the number of location-days over which they could occur but do not; the ratio of "no-flood" to "flood" observations in a regional, multi-year dataset can be extremely large. In such settings, overall accuracy is not a meaningful evaluation metric: a model that predicts "no flood" every day can achieve very high accuracy while being entirely uninformative. Models must instead be evaluated using metrics suited to extreme class imbalance, such as the F1-score, Cohen's Kappa, or the Area Under the Precision-Recall Curve (PR-AUC), which focus on the minority hazard class.

In addition, a conventional random train–test split can produce optimistic performance estimates in spatial environmental modelling, owing to spatial and temporal autocorrelation (Tobler's First Law of Geography). If a field is randomly assigned to the training set while an adjacent field, sharing similar slope, soil, and rainfall conditions, is assigned to the test set, the model is effectively evaluated on conditions it has already seen. Robust evaluation of a daily muddy-flood model therefore requires spatio-temporal validation strategies, such as spatial block cross-validation and temporal holdouts, in which entire geographic regions (e.g., watersheds) or distinct time periods (e.g., a full year) are withheld from training. Strong performance on these unseen regions and dates provides stronger evidence that the model has learned generalisable relationships rather than memorised local conditions.

Finally, the use of SHAP (SHapley Additive exPlanations) values allows researchers to inspect the internal logic of otherwise opaque machine learning models, checking that the relationships an algorithm has learned (e.g., higher rainfall associated with higher predicted hazard) are physically consistent with known geomorphological processes rather than artefacts of the training data. Interpretability of this kind is particularly important given the class-imbalance and validation constraints described above: a model that scores well under spatial block cross-validation but whose learned relationships contradict known process understanding warrants caution regardless of its reported performance metrics.

2.7 Hazard Mitigation and Catchment Management

The applied purpose of a predictive muddy-flood model is to inform targeted, efficient mitigation. Because off-site damage costs far exceed the agronomic value of the soil lost on-site, Boardman and Vandaele (2020) argue that muddy flooding cannot be addressed by sewer upgrades or concrete barriers alone, but requires an integrated approach across three stages: the source, the pathway, and the receptor.

Source-level measures reduce the dynamic C-factor during the May–June vulnerability window through conservation tillage, winter cover crops, contour ploughing, and residue retention, which absorb rainfall kinetic energy and reduce Hortonian runoff (Evrard et al., 2008). Adoption is limited by short-term yield penalties and machinery costs, which CAP eco-schemes increasingly offset. Pathway-level measures (vegetative buffer strips, grassed waterways, and fascines) sever structural sediment connectivity by acting as hydraulic brakes that slow flow and encourage deposition before sediment reaches a settlement (Verstraeten et al., 2003); the Velm case study (Section 2.5.3) demonstrated a peak-discharge reduction exceeding 40% from such measures. Receptor-level measures, such as small earthen retention dams and micro-basins at sub-catchment outlets, provide a final line of defence by storing peak runoff before it reaches infrastructure.

Implementing mitigation everywhere is not economically feasible, so predictive susceptibility models of the kind developed in this thesis serve a targeting function: identifying the most severe, recurrent pathways and the periods of peak vulnerability so that limited mitigation resources are directed where they will have the greatest effect, shifting catchment management from reactive cleanup toward proactive resilience.

This shift toward predictive, targeted management is also visible at the institutional level, independently of the present thesis. The Interreg NWE project MUDCAP (“MUDdy flood mitigation through transnational action-oriented CAPacity building,” 2026–2029) brings together researchers, public authorities, and water managers from Belgium, France, and Germany to build a transnational database of pluvial muddy flood events, predictive tools, and a digital decision-support toolkit for local and regional authorities, with one of its four pilot areas located in the Dyle catchment in Wallonia. This thesis is not part of MUDCAP or any of its work packages, but the project's existence illustrates that the transition from reactive cleanup to proactive, predictive catchment management described above is an active priority for public authorities in the Walloon loess belt at the time of writing, not merely a theoretical direction.

3. Research Objectives

Muddy-flood prediction in Wallonia is constrained by the joint difficulty of an event-driven physical process and an imperfectly observed occurrence record: reported locations are confirmed occurrences, but unreported locations cannot be assumed to be true absences, and inventories rarely record eroded soil volume, sediment load, deposit thickness, or damage cost. This study therefore models relative risk prioritisation rather than absolute mud quantity, using reported muddy-flood occurrences, algorithmically generated background pour points, and predictors derived from terrain analysis, soil-erosion modelling, rainfall erosivity, land use, and Sentinel-2 vegetation indicators.

Table 2. Research questions and how each is addressed.

RQ	Research question	Addressed by
RQ1	Which persistent landscape characteristics explain where muddy floods are more likely to occur?	Predictor-importance analysis: static terrain, soil, land-use and connectivity variables.
RQ2	Which daily environmental conditions explain when muddy floods occur?	Predictor-importance analysis: daily rainfall erosivity and antecedent conditions added as time-varying predictors .
RQ3	Does dynamic crop/vegetation information improve prediction compared with static land-cover information?	Predictor-importance analysis: dynamic Sentinel-2 C-factor proxy vs. static land-cover baseline .
RQ4	Does a Positive-Unlabelled Random Forest provide more robust prediction than a conventional classification approach treating background locations as true non-events?	Model comparison: PU-RF vs. conventional Random Forest on held-out data .
RQ5	Does model performance remain strong when tested on unseen locations and unseen periods?	Same model comparison, evaluated under spatial and temporal.

4. Materials and Methods

This chapter combines the description of each input dataset with the processing steps applied to it, so that each data source is presented once, followed immediately by how it was prepared for modelling. The chapter then turns to the modelling framework itself: how these prepared layers are combined into a watershed-level, event-date prediction dataset, and how the resulting classification models were selected and validated.

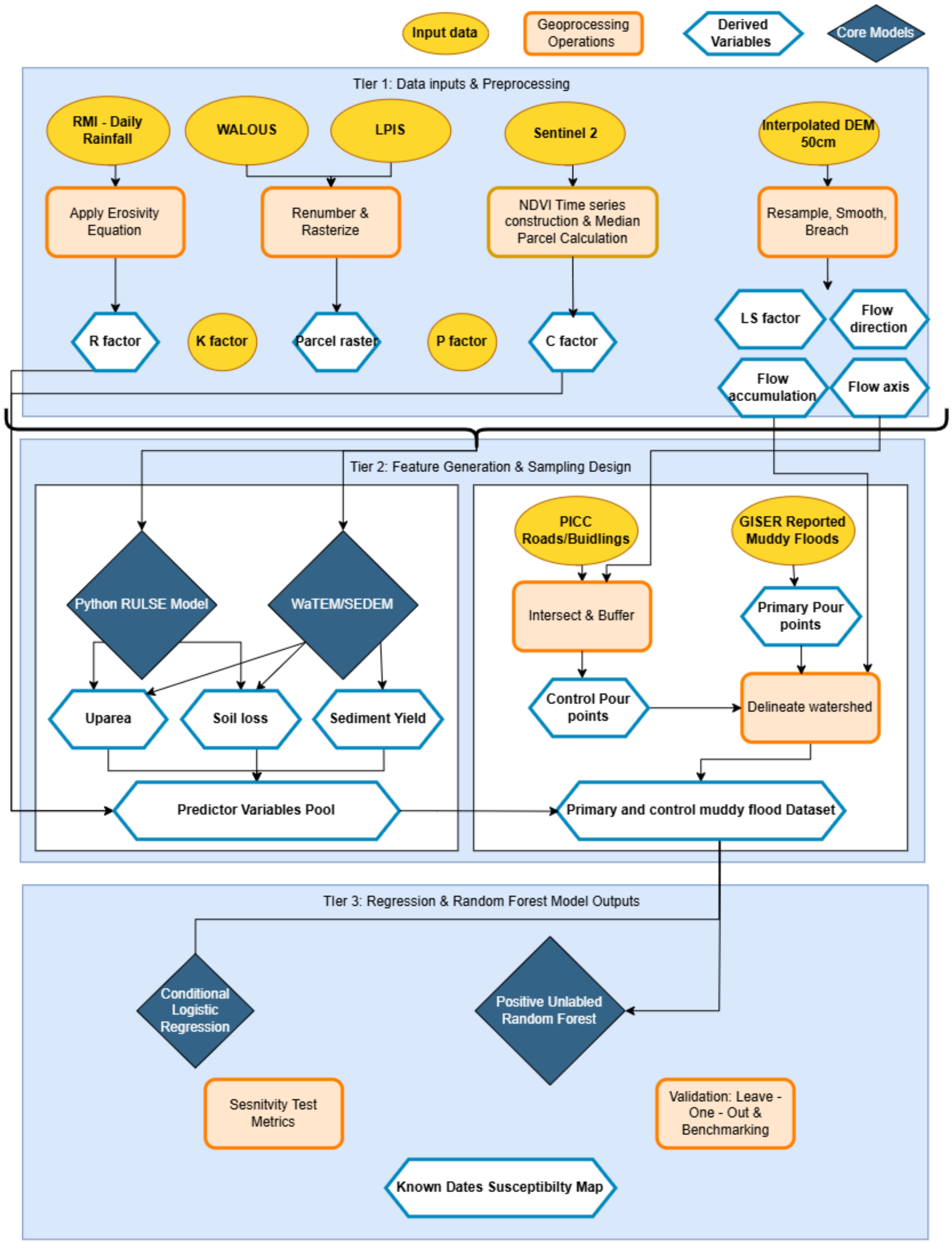

Figure 1. Overall workflow diagram.

4.1 Study Area

The study covers the Walloon Region of Belgium, with particular attention to the loess belt, where highly erodible silt-loam soils, cultivated slopes, and a dense network of small dry valleys combine to produce conditions highly conducive to muddy-flood generation . The region has a temperate oceanic climate, with mean annual precipitation in the loess belt typically in the range of 750–850 mm, distributed relatively evenly through the year but punctuated by short, high-intensity convective events in late spring and summer . Soils in the loess belt are predominantly silt-loam Luvisols, which are highly erodible and prone to surface crusting when left bare or sparsely covered. Land use is dominated by intensive arable agriculture, including winter cereals and summer row crops (maize, potatoes, sugar beet), interspersed with expanding peri-urban settlement, a combination identified in Chapter 2 as a principal driver of both erosion and the connectivity of eroded sediment to exposed infrastructure.

Although all environmental input layers used in this work are available for the whole of Wallonia, a smaller rectangular region of interest (ROI) was selected as a test site for model development and validation. The ROI boundary shapefile used for this study contains one polygon in EPSG:31370 (Belgian Lambert 1972), with an area of 2,331 km², approximately 70 km wide and 33 km high. The ROI lies within Sentinel-2 tile 31UFS and contains a denser muddy-flood hotspot around Liège together with more sparsely affected surrounding agricultural catchments, together totalling approximately 30 documented muddy-flood sites.

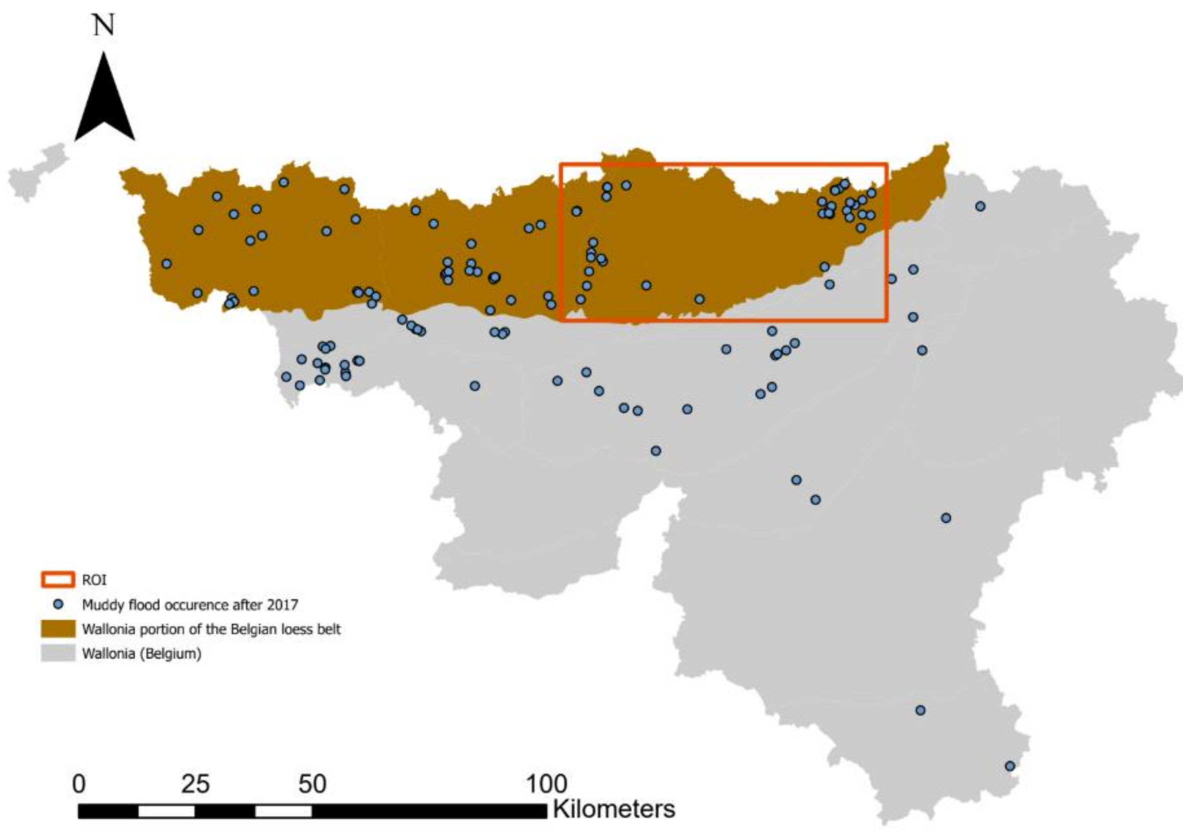

Figure 2.Study area map where the muddy flood occurrence polygons are represented by blue points

4.2 In Situ Muddy-Flood Occurrence Data

The primary source of muddy-flood occurrence data is the muddy-flood-affected-sites database maintained by the Service Public de Wallonie (SPW), Cellule GISER (Gestion Intégrée Sol Érosion Ruissellement), the Walloon technical advisory unit for runoff- and erosion-related flood risk. The database records the location, as a polygon, and the date of each reported muddy-flood event across Wallonia. Reported polygons trace the affected extent along roads and adjacent garden or yard areas rather than marking a single point: the smallest cases cover one property's yard and the frontage of road immediately outside it, while the largest cover an entire affected street, drawn as several connected segments running the length of a cluster of houses whose backyards and adjoining road were all impacted by the same event.

Several data-quality issues had to be addressed before use. First, when a site has been affected more than once, the existing polygon is retained rather than updated, so a single polygon can correspond to more than one reported date, and event dates are not always recorded in chronological order. Second, reports are not independently field-verified; each reported event date was therefore cross-checked against the rainfall record, and only dates with measurable rainfall within a three-day lag were retained. Third, because Sentinel-2 observations only become reliably available from 2018 onward, and because this type of muddy flood is predominantly triggered by spring and summer convective storms over bare or sparsely vegetated fields, the working dataset was restricted to the months of May to October for the years 2018–2021.

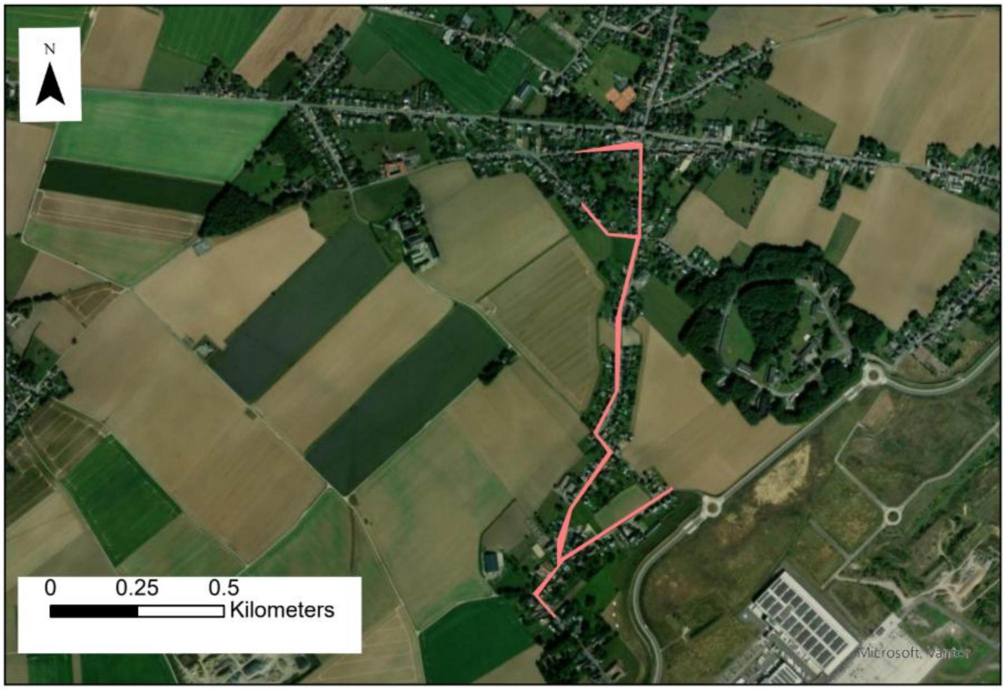

Figure 3. Examples of muddy flood occurrence polygons from GISER dataset illustrated in red

4.3 Digital Elevation Model

The terrain dataset used for hydrological pre-processing is the 50 cm digital terrain model (DTM) derived from the Wallonia-wide LiDAR survey flown between 19 February 2021 and 5 March 2022, at an average point density of 6 points/m². Dense vegetation and open-water surfaces return little or no signal to the LiDAR sensor, so the resulting point cloud contains data gaps in these areas; the DTM used here is interpolated (nearest-neighbour) to remove these gaps, which makes the surface somewhat sensitive to edge effects in densely vegetated zones.

This differs from the DTM used for LIDAXES, the Service Public de Wallonie's regional LiDAR-derived flood-routing product, which maps potential surface-water flow paths across Wallonia from the same LiDAR survey; in that product, water-body edges were treated with custom processing and building footprints were deliberately left as no-data, so that simulated flow could pass through built-up areas.Harmonisation of elevation values with neighbouring regions and countries was also not performed for the dataset used here, so any watershed extending beyond the Walloon border would carry additional elevation uncertainty. None of the watersheds used in this study cross the regional boundary, so this is noted only as a constraint for any future upscaling beyond Wallonia, not as a limitation of the current results.

The 50 cm DTM was resampled to 5 m resolution and processed with WhiteboxTools, a hydrological-analysis package chosen for its fast, parallelised handling of high-resolution rasters. A 3×3 mean filter was applied to the resampled DEM, followed by stream breaching rather than depression filling, following the approach recommended in the LIDAXES processing report. Breaching removes obstructions to flow by cutting through the DEM along the most efficient path while leaving the surrounding terrain largely unchanged, whereas filling raises the elevation of an entire depression to the level of its lowest outlet, which can create artificially flat zones; breaching was preferred here because it disturbs the surrounding terrain less while still resolving blockages. A D8 flow-direction and flow-accumulation raster was then generated from the breached, filtered surface, forming the basis for the pour-point and watershed-delineation procedure that follows.

A small sensitivity analysis compared 3×3, 5×5, and 9×9 smoothing windows by examining the resulting flow-accumulation pattern and the watershed area obtained from the same set of snapped pour points. The 3×3 (minimal) smoothing window was retained, since more aggressive smoothing altered watershed boundaries without a clear benefit. One consequence of the ten-fold resampling from 50 cm to 5 m is the appearance of large, very flat areas, particularly over flat agricultural fields, where the D8 flow-routing algorithm has limited information with which to resolve a single preferred flow direction; this produces visible parallel flow-routing artefacts converging into the main channels. This pattern persists under D8 routing regardless of the smoothing window tested, and should be kept in mind when interpreting flow-accumulation results over the flattest parts of the landscape, since more aggressive smoothing would resolve it only at the cost of distorting the wider terrain.

4.4 Land-Cover Data: LPIS and WALOUS

Two annually updated geodatabases from the Wallonia Geoportal were used to characterise land cover and agricultural parcels. The Land Parcel Identification System (LPIS) provides the agricultural-parcel boundaries used as the spatial unit for the Sentinel-2 NDVI/C-factor time-series reconstruction. LPIS parcels are themselves produced within the Système Intégré de Gestion et de Contrôle (SIGeC, Integrated Management and Control System), the regional system used to track agricultural land and process Common Agricultural Policy subsidy applications; the reference parcel layer is managed by the Organisme Payeur de Wallonie (OPW) and distributed publicly via the Wallonia Geoportal as the anonymised agricultural-parcel layer used in this study. Each muddy-flood event date was matched to the LPIS parcel layer corresponding to its calendar year.

WALOUS (WALlonie Occupation et Utilisation du Sol) provides regional land-cover information and was used in this thesis in the form of WAL_OCS (Occupation du Sol), a 1 m resolution raster classifying the physical cover of the land surface into 11 principal classes: artificial ground surfaces, above-ground artificial constructions, railway networks, bare soils, surface waters, rotating herbaceous cover, continuous herbaceous cover, coniferous trees, broadleaf trees, coniferous shrubs, and broadleaf shrubs. WAL_OCS was used operationally as a land-cover mask restricting the erosion calculations to agricultural land: cells classified as rotating or continuous herbaceous cover were retained as the active, erodible domain for the RUSLE and WaTEM/SEDEM workflows, while artificial-surface, water, and woody-vegetation cells were excluded entirely. This is a static, categorical boundary, distinct in role from both the annually updated LPIS parcel geometry, the spatial unit for NDVI reconstruction, and the C-factor itself, which is derived independently from the Sentinel-2 NDVI time series rather than from WAL_OCS: WAL_OCS restricts *where* erosion is calculated, while the *seasonal intensity* of erosion susceptibility on that retained land is governed separately by the dynamic, remotely sensed C-factor.

4.5 Remote Sensing Data: Sentinel-2 NDVI and the Dynamic C-Factor

All optical Earth-observation inputs used to derive the seasonal cover-management (C) factor are based on Sentinel-2 imagery processed to Level-2A surface reflectance using MAJA (MACCS-ATCOR Joint Algorithm), an atmospheric-correction and cloud-detection processor that converts raw satellite imagery into surface-reflectance values comparable across dates by removing the distorting effects of the atmosphere. Sentinel-2 imagery in this study used Processing Baseline 05.00 (N0500), the ESA-assigned version identifier for the ground-processing software in effect at the time of acquisition.

Fmask (Function of mask) is the cloud, cloud-shadow, and snow/ice detection mask distributed alongside the MAJA L2A product, generated automatically for every acquisition. Several Fmask variants are distributed for the same acquisition, labelled -10pct, 0pct, 10pct and 20pct; only the 0pct variant was used in this study. This label refers to the specific masking variant selected, not to the cloud cover of the image itself, and should not be read as meaning that the corresponding image has 0% cloud cover.

Two vegetation indicators were initially considered: NDVI, calculated from the red and near-infrared Sentinel-2 bands as NDVI = (NIR − Red) / (NIR + Red); and the fraction of vegetation cover (Fcover), produced by the Sen4CAP processing chain (a Copernicus-funded system generating agricultural monitoring products from Sentinel-2 data) from MAJA L2A inputs combined with Sen4CAP L3 product-validity masks. Fcover is conceptually closer to the crop-cover information that the RUSLE C-factor is intended to represent; however, it was not used in the final workflow. Sen4CAP applies stricter biophysical quality thresholds than the general-purpose Fmask used for NDVI, so a smaller proportion of Fcover pixels passed as valid on any given date, leaving fewer clear-sky observations available to reconstruct each parcel's time series. This is reflected directly in reconstruction quality: even the best-performing method and parameter combination for Fcover (Whittaker smoothing, λ = 10,000) achieved an RMSE of 0.454 across 5,126 validation points, roughly seven times higher than the equivalent best-case NDVI RMSE of 0.0642. NDVI was therefore adopted as the operational vegetation-cover indicator, using the processing-baseline-05.00 (N0500) MAJA product line.

4.5.1 NDVI time-series reconstruction: method comparison

Sentinel-2 observations are affected by frequent gaps caused by cloud cover, particularly outside the summer months; this is also the mechanism by which rainfall can indirectly bias a remotely sensed C-factor, since heavy-rainfall periods are also cloudy periods. To obtain continuous, daily NDVI profiles suitable for event-based C-factor calculation, three gap-filling and reconstruction methods were compared at the LPIS agricultural-parcel scale: Savitzky–Golay filtering (window sizes of 15, 31, 45, and 61 days), Whittaker smoothing (λ = 10, 100, 1,000, 10,000), and Gaussian Process Regression (GPR) with an optimised kernel.

In plain terms, Savitzky–Golay filtering reconstructs each point by fitting a simple curve to a moving window of nearby observations, similar to a smoothed trendline. Whittaker smoothing instead fits one continuous curve across the whole season at once, balancing how closely it follows the observed points against how smooth the overall curve is. Both return a single reconstructed value with no indication of how confident that value is. GPR is different in kind, not just in formula: it is a probabilistic method that returns both a predicted value and a measure of how uncertain that prediction is, which grows the further the prediction sits from an actual observation.

The three methods were compared using a parcel-level leave-one-observation-out validation against LPIS-delineated agricultural parcels. For each eligible parcel, one valid NDVI observation was withheld, the remaining observations were used to reconstruct the value on the withheld date, and the predicted value was compared against the observed one. The first and last observations of each time series were excluded from validation, to avoid testing extrapolation rather than interpolation. Two separate thresholds were applied, at two different levels. At the level of a single date, an individual NDVI observation was retained only if at least 30% of the pixels within that parcel were classified as valid (cloud-free) on that date, and if the resulting parcel-mean NDVI fell within the physically plausible range of −0.2 to 1.0. At the level of the parcel's full time series, a parcel was included in the comparison only if it had at least six such valid observation dates remaining after this per-date filtering; parcels with fewer than six valid dates did not provide enough data to support a meaningful leave-one-observation-out comparison and were excluded entirely.

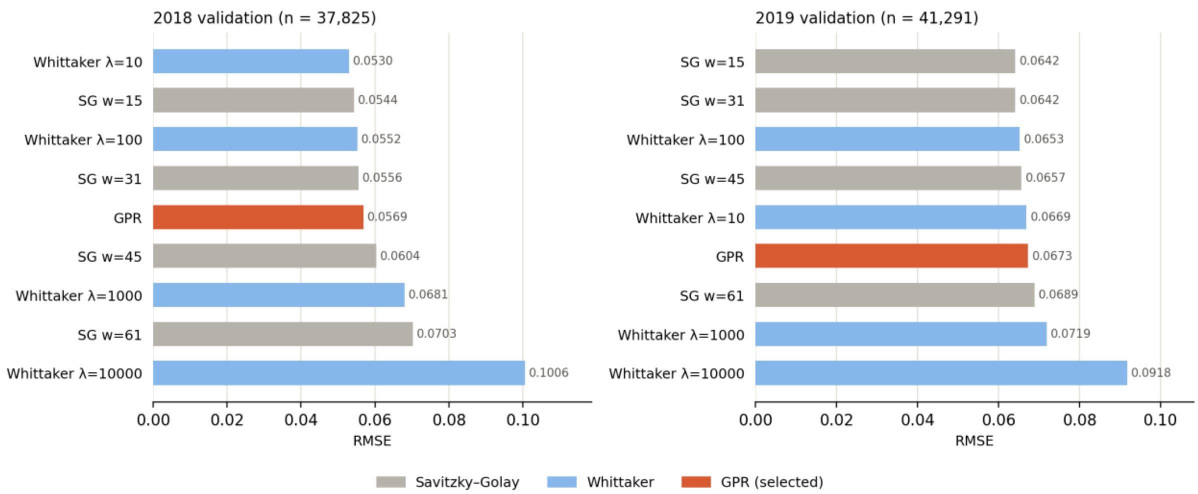

Figure 4. Leave-one-observation-out RMSE for all nine method/parameter combinations tested, ranked lowest to highest, for the 2018 and 2019 validation sets. GPR (highlighted) ranks fifth of nine in 2018 and sixth of nine in 2019.

Figure X shows the full result of this comparison for both validation years. GPR was not the lowest-error method in either year: it ranked fifth of nine configurations in 2018 (RMSE 0.0569, against a best of 0.0530 for Whittaker λ=10, a difference of +7.3%) and sixth of nine in 2019 (RMSE 0.0673, against a best of 0.0642 for Savitzky–Golay with a 15-day window, a difference of +4.9%). Savitzky–Golay with a 15-day window was in fact one of the strongest performers in both years, and on accuracy grounds alone would have been the more defensible choice (Table 3).

Table 3. Summary of best NDVI time-series reconstruction methods against GPR, at the LPIS agricultural-parcel scale.

Validation set	Best method	Best RMSE	GPR RMSE	Difference	GPR bias
2018 file	Whittaker λ = 10	0.0530	0.0569	+7.3%	−0.0014
2019 file	Savitzky–Golay w = 15	0.0642	0.0673	+4.9%	−0.0001

GPR was nonetheless retained as the operational reconstruction method, because, unlike Savitzky–Golay or Whittaker smoothing, it is the only one of the three methods that natively returns a predictive uncertainty for every reconstructed value. This uncertainty was intended to be propagated into the C-factor estimate, so that periods reconstructed with low confidence, typically long gaps between clear-sky observations, would carry a wider C-factor confidence interval than periods reconstructed close to an actual observation. This property, GPR's flexibility for time-series interpolation combined with its capacity to quantify uncertainty on the reconstructed values, is well established in the vegetation-monitoring literature (Belda et al., 2020) and has been applied directly to C-factor mapping in a regional context immediately adjacent to this thesis's study area: Matthews, Kasprzak and Verstraeten (2025) used the same GPR NDVI-reconstruction approach, for the same reason, in a soil-erosion risk assessment of Flanders commissioned by the Departement Omgeving. Their implementation is considerably more developed than the one adopted here: NDVI is converted into a percentage canopy cover using crop-specific empirical relationships (Tenreiro et al., 2021) rather than the single generic exponential formula used in this thesis, and a parallel GPR pipeline is applied to a tillage-sensitive index (NDTI) to estimate residue cover via a Random Forest classifier, feeding directly into their C-factor's uncertainty framework. The uncertainty propagation adopted in this thesis is consequently a considerably simpler, single-step version of an approach already validated at operational scale for this exact purpose in a directly comparable regional setting.

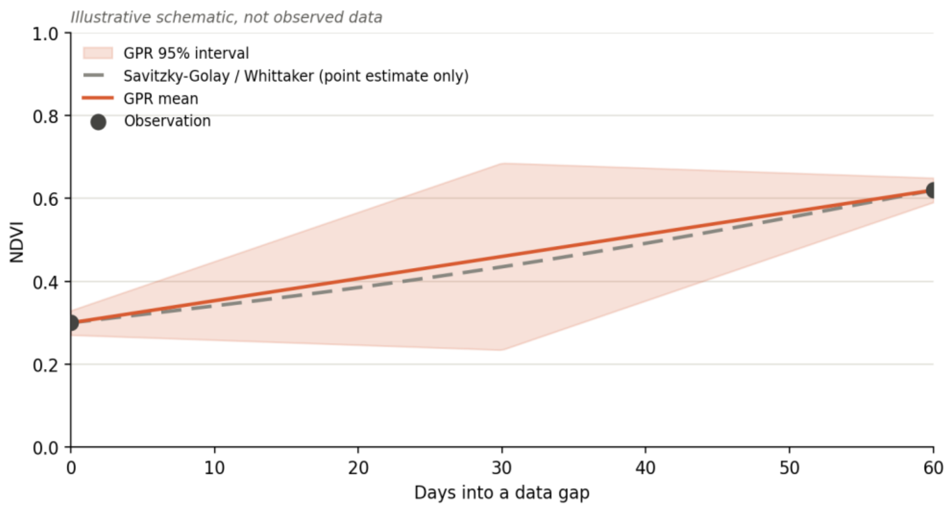

Figure 5. Illustrative schematic (not observed data) of why GPR was retained despite not achieving the lowest RMSE. Point-estimate methods (Savitzky–Golay, Whittaker) return a single reconstructed value with no signal of confidence, even during a long data gap. GPR returns a mean value plus an interval that widens the further the prediction sits from an actual observation, and narrows to near zero at each observation.

Figure 5 illustrates this mechanism schematically. Where an observation exists, all three methods converge to a similar value; in the middle of a data gap, Savitzky–Golay and Whittaker smoothing still return a single, undifferentiated value, giving no signal that the reconstruction is less reliable there than elsewhere. GPR's confidence interval visibly widens in the same gap, which is the property this thesis's C-factor propagation step depends on.

It should be stated plainly that this uncertainty propagation was implemented at the level of the NDVI-to-C-factor conversion, but the resulting C-factor confidence intervals were not carried through into the final susceptibility model or reported in the results of this thesis, owing to time constraints. GPR was therefore accepted at a modest, consistent accuracy cost, in the range of 5 to 7% higher RMSE than the strongest point-estimate alternative in both validation years, for an uncertainty-quantification capability that this thesis was not ultimately able to exploit downstream. This is revisited as a direction for future work.

4.5.2 Operational NDVI reconstruction and C-factor derivation

For the final event-based extraction, a GPR model was fitted independently for each LPIS parcel, that is, at the individual agricultural-parcel scale rather than per pixel, using time (days since the parcel's first valid observation) as the only predictor. Duplicate observations on the same date were aggregated by their median value, and a model was fitted only when at least four valid observation dates were available. The model used a constant kernel multiplied by a radial-basis-function kernel, plus a white-noise kernel:

$$y(t) = f(t) + \varepsilon,\ \ f(t) \sim GP(m(t), k(t,t')) \tag{1}$$

$$k(t,t') = \sigma f^2 \exp[-(t - t')^2 / (2l^2)] + \sigma n^2 \delta tt' \tag{2}$$

where y(t) is the observed NDVI at time t, f(t) is the latent smooth vegetation signal, ε is observation noise, l is the temporal length scale, σf^2 is the signal variance, σn^2 is the noise variance, and δtt′ is the white-noise (Kronecker delta) term. Kernel hyperparameters were fixed rather than optimised per parcel, for computational efficiency across the full parcel dataset: length scale l = 45 days, white-noise level = 0.02, and constant-kernel value = 1.0, based on values found to generalise well across the leave-one-observation-out comparison. Target values were normalised before fitting.

For every muddy-flood event date, the fitted parcel-level GPR model was used to predict (i) the event-day NDVI and (ii) the mean NDVI over a 22-day window centred on the event (11 days before to 10 days after), together with the corresponding GPR standard deviation. The window length reflects the approximate timescale over which NDVI can meaningfully shift during the growing season, short enough to capture the vegetation state relevant to a specific event, but long enough to average over short-term noise in individual reconstructions. This standard deviation reflects the model's distance from the nearest valid observation: it is expected to be higher where the nearest clear-sky observation is further away in time, and also higher during periods of fast crop growth or senescence, when NDVI changes quickly between observations. Each reconstruction was assigned a confidence class based on this standard deviation and on whether at least one valid observation fell within the 22-day window.

Predicted NDVI values were clipped to the valid range and converted to the RUSLE C-factor using the Van der Knijff exponential relationship:

$$C = \exp[-\alpha \cdot NDVI / (\beta - NDVI)],\ \ \text{with } \alpha = 2, \beta = 1 \tag{3}$$

NDVI was clipped between 0.0 and 0.95 prior to applying this formula. This upper clip avoids allowing NDVI to approach the singularity at NDVI = β = 1 in the denominator, which would otherwise be numerically unstable and does not correspond to a physically realistic vegetation-cover condition. The 95% GPR confidence interval on NDVI (±1.96σ) was propagated through Equation 3 to obtain an associated uncertainty range for the C-factor; because C decreases as NDVI increases, the lower NDVI bound yields the upper C-factor estimate, and vice versa.

An additional sensitivity check was performed to test the assumption that individual LPIS parcels are internally homogeneous in crop cover for a given declared crop type, since within-parcel heterogeneity not captured by the parcel-level LPIS declaration could otherwise bias the reconstructed NDVI time series; this check did not indicate a systematic bias large enough to change the operational choice of GPR reconstruction at the parcel scale.

4.6 Rainfall Data and the Dynamic Rainfall-Erosivity (R) Factor

Daily rainfall totals were obtained from the Royal Meteorological Institute of Belgium (RMI) gridded observational dataset. This product interpolates irregularly distributed station observations onto a regular 5 km grid (1,360 pixels covering Belgium), is available from 1961 onward for most variables, and is updated daily. Each grid value represents the rainfall accumulated over the 24-hour hydrological day, from 08:00 to 08:00 the following day.

An alternative source, the PAMESEB/Agromet automatic weather-station network operated by the Walloon Agricultural Research Centre (CRA-W; approximately 30 stations, the exact number varying over time as stations are commissioned or decommissioned), was also considered. It was not adopted because several reported muddy-flood events occurred more than 10 km from the nearest station, which would have left large parts of the study area poorly constrained. The RMI gridded product was therefore judged to offer the best available compromise between spatial coverage and temporal resolution for this study, even though it provides daily rather than sub-daily rainfall, a limitation returned to later in this thesis.

Since RUSLE requires a rainfall-erosivity (R) term rather than a raw rainfall depth, daily rainfall totals (P, mm) were converted to a daily erosivity proxy using the empirical event-depth relationship of Verstraeten, Poesen, Demarée, and Salles (2006), calibrated for Ukkel, Belgium:

$$EI_{30} \approx 0.1089 \times P\hat{}1.8067 \quad (4)$$

This relationship was used because sub-daily (1–60 minute) rainfall-intensity records, which would be needed to calculate a true EI30 value directly, were not available at the spatial coverage required for this study. The resulting daily values should be interpreted as a daily erosivity proxy rather than a directly measured EI30. Following Renard et al. (1997), the long-term mean annual R-factor (MJ·mm·ha^{-1}·h^{-1}·yr^{-1}) is normally obtained by averaging EI30 over all rainfall events across multiple years; in this study, the same event-level relationship was instead applied on a daily basis, to obtain a dynamic, date-specific erosivity layer for the event-based modelling framework introduced later in this chapter.

The resulting daily R-factor rasters were resampled from their native 5 km grid to the 5 m modelling resolution using nearest-neighbour interpolation, in order to preserve the original station-derived values exactly. This is computationally simple and value-preserving, but it produces visible block-like discontinuities in the final raster output, since each 5 km cell is reproduced unchanged across many thousands of 5 m cells; this is discussed further as a limitation later in this thesis.

4.7 PICC Infrastructure Data

Building and road locations were taken from the Projet Informatique de Cartographie Continue (PICC), the official, continuously updated 3D digital cartographic reference for the whole of Wallonia, managed by the SPW and distributed via the Wallonia Geoportal. The PICC captures identifiable landscape elements (buildings, the hydrographic network, roadways, and engineering structures) with a stated positional precision better than 25 cm (SPW Géoportail de la Wallonie, PICC metadata catalogue). In this study, PICC layers were used specifically to locate buildings and roads, which served to define potentially exposed infrastructure locations used later in this chapter.

4.8 RUSLE and WaTEM/SEDEM Erosion Models

4.9 Modelling Concept: From Erosion Maps to Watershed-Date Risk Scores

The overall objective of the modelling stage was to develop a dynamic muddy-flood risk-prioritisation framework for exposed infrastructure locations, rather than a general soil-erosion map. Infrastructure exposure is defined using the PICC (Projet Informatique de Cartographie Continue), the 3D digital mapping reference for the entire Walloon region, which records the x, y, z coordinates of all identifiable landscape elements, including roads, railways, buildings, and settlement edges, at a positional accuracy of better than 25 cm. The approach focuses specifically on point-scale locations where concentrated surface runoff can intersect or reach mapped PICC infrastructure, not on watersheds in general. The basic spatial modelling unit is defined as:

flow axis × PICC exposure feature → representative pour point → upstream watershed

For every representative pour point , an exposed outlet where muddy runoff could plausibly affect infrastructure , an upstream watershed was delineated and used as the spatial unit over which rainfall-erosivity and vegetation-cover-vulnerability metrics were extracted. This links the dynamic risk score to the entire contributing area upstream of each exposed location, rather than to the outlet point alone. Because each pour point is itself derived from a D8 flow-routing raster, its exact position is approximate rather than exact; this should be borne in mind when interpreting any individual pour point's precise location, as opposed to its general upstream contributing area, which is more robust to this uncertainty.

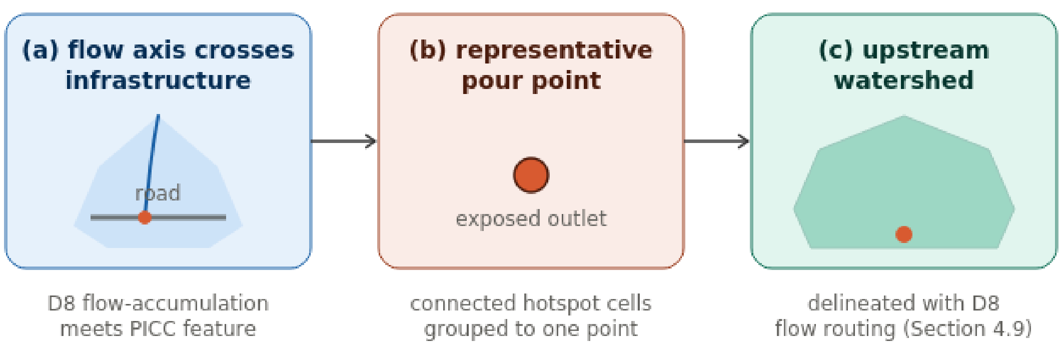

Figure 6. Modelling-concept diagram: (a) a flow axis crosses a PICC infrastructure feature, (b) connected high-flow/PICC contact cells are grouped into a single representative pour point, and (c) the upstream watershed is delineated from that point

4.10 Creation of Representative PICC-Flowaxis Pour Points

Candidate exposed pour points were identified by intersecting concentrated-flow cells, derived from the D8 flow-accumulation raster using a minimum drained-area threshold, with rasterised PICC infrastructure features (roads, railways, buildings, building-proximity buffers, and settlement edges).

An initial exhaustive approach, in which every PICC–flow contact was treated as a separate pour point, was tested and rejected: because a D8 watershed raster assigns each cell to a single downstream outlet, placing pour points too close together fragmented the terrain into an artificially large number of small, competing watersheds.

The final approach instead retained only representative pour points, meaning one pour point per connected cluster of high-flow/PICC contact cells, rather than one pour point per individual contact. Connected high-flow/PICC hotspot cells were grouped, and a single representative pour point was assigned to the cell with the highest flow accumulation within each connected hotspot. Near-duplicate candidates were then thinned using a minimum spacing threshold. Pour points corresponding to reported muddy-flood sites were kept separately throughout this process and combined with the representative background pour points only after the candidate-cleaning step, so that no reported site was inadvertently merged with, or removed by, the thinning procedure.

4.11 Watershed Delineation

Watersheds were delineated from the final representative pour-point set using the hydrologically prepared DEM and D8 flow routing in WhiteboxTools. Pour points were rasterised to the DEM grid, after which a D8 flow-direction (pointer) raster and a watershed raster were generated. Two outputs were retained: a watershed-ID raster, used for the zonal extraction of dynamic predictors, and a binary watershed mask, used to define the active modelling area. Because each watershed is delineated explicitly from its pour point, rather than derived from a generic binary mask, every dynamic raster layer can be summarised per upstream watershed and joined back to its corresponding pour point.

4.12 Dynamic Predictor Extraction

Dynamic predictors were derived from the daily R-factor raster, the daily C-factor raster, and the RUSLE/WaTEM-SEDEM erosion and sediment-output rasters. For each event date and each representative watershed, zonal statistics were extracted from all rasters. In addition to these process-based predictors, static topographic and connectivity variables were extracted per watershed, including the Stream Power Index (SPI), the Topographic Wetness Index (TWI), and upstream contributing area, since these are established structural correlates of runoff concentration and are unaffected by date. The most relevant resulting dynamic variables were:

- the watershed-mean rainfall erosivity on the event date
- the 95th-percentile RUSLE-based erosion rate within the watershed
- the 95th-percentile cover-management factor within the watershed

These three raw variables together make up the best-performing feature set identified by the rainfall-constrained search and represent the three complementary elements of the muddy-flood process introduced above: the erosive rainfall trigger, high-end erosion intensity within the contributing watershed, and high-end cover-management vulnerability. A separate rank-based interaction model, using within-date ranks of rainfall erosivity and the cover-management factor together with a rainfall-gated interaction score, was also developed and tested, but did not outperform this simpler three-raw-variable specification on the top-10% case-capture metric used to select the final model, and was therefore not carried forward. They were selected from a substantially larger candidate pool: Table 4 below lists every variable named in the model-family, grouped by data source, together with how each was derived and whether it was retained in the final model. This is provided so the selection process is fully traceable, rather than only reporting the variables that happened to survive into the best model.

Table 4. Full candidate predictor pool tested during model selection.

Variable	Source	How it was derived	Retained in final model?
Mean rainfall erosivity	Rainfall	Watershed-mean daily R-factor (rainfall erosivity) on the event date	Yes , final PU-RF model
95th-percentile rainfall erosivity	Rainfall	95th-percentile daily R-factor within the watershed on the event date	Tested; not in final model
Within-date rank of the 95th-percentile rainfall erosivity	Rainfall	Within-date rank of the 95th-percentile rainfall erosivity across all candidate watersheds active that day	Tested in rank-based and CLR models
95th-percentile cover-management factor	Sentinel-2 NDVI	95th-percentile daily C-factor (cover-management vulnerability) within the watershed	Yes , final PU-RF model
Mean cover-management factor, cultivated land only	Sentinel-2 NDVI	Watershed-mean C-factor, restricted to cultivated land-cover classes	Tested in static-only family; not in final model
95th-percentile cover-management input value	Sentinel-2 NDVI	95th-percentile C-factor input value within the watershed (pre-interaction)	Tested; not in final model
Within-date rank of the 95th-percentile cover-management factor	Sentinel-2 NDVI	Within-date rank of the 95th-percentile cover-management input value across all candidate watersheds active that day	Tested in rank-based and CLR models
Rainfall-cover rank interaction	Rainfall × NDVI interaction	Product of the R-factor rank and the C-factor rank for the same watershed-date	Tested in rank-based interaction model
Rainfall-gated interaction score	Rainfall × NDVI interaction	Rainfall-gated R×C interaction score; zero/inactive below the rainfall-activation threshold	Tested in rank-based interaction model; not retained
95th-percentile RUSLE-based erosion rate	Python RUSLE	95th-percentile RUSLE soil-loss estimate (t/ha) within the watershed on the event date	Yes , final PU-RF model

Total erosion proxy	Python RUSLE	Total estimated soil loss (tonnes) summed across the watershed	Tested in process-only family; not in final model
Modelled sediment output at the pour point	WaTEM/SEDEM	Modelled sediment output at the pour point itself	Tested; decreased validation performance when added
Maximum Topographic Wetness Index	DEM-derived connectivity	Maximum Topographic Wetness Index within the watershed	Tested in rainfall-constrained plus static family; not in final model
Mean and 95th-percentile upstream contributing area (pixel count)	DEM-derived connectivity	Mean / 95th-percentile upstream contributing area (pixel count) within the watershed	Tested in rainfall-constrained plus static and rainfall-interaction families; not in final model
Rainfall-connectivity interaction	Rainfall × connectivity interaction	Product of rainfall erosivity and upstream contributing area, tested as an alternative interaction term	Tested in R-interactions family; not in final model

The 95th percentile, rather than the watershed mean, was used for the R- and C-factor summaries because muddy-flood generation is disproportionately driven by the most erosive sub-areas of a watershed rather than by its average condition; a watershed with a small but highly vulnerable zone of bare, steep, high-erosivity land can generate a damaging muddy flood even if its area-weighted mean R- or C-factor is unremarkable. The area-weighted mean of each factor was also retained as a separate predictor, such as mean rainfall erosivity, used in the final selected model, precisely because total sediment production scales more closely with the surface-weighted average than with a single high percentile; the two summaries were therefore treated as complementary rather than substitutable, and both were made available to the model-selection search .

Before the three raw variables described above were identified as the strongest candidate, an alternative rank-based formulation was also developed and tested, in which rainfall erosivity and the cover-management factor were each re-expressed as within-date ranks rather than used as absolute values. Using within-date ranks allowed each watershed to be compared against all other exposed watersheds under the same rainfall and vegetation-cover conditions on a given day, directly reflecting the intended use of the model: prioritising locations on a given day, rather than estimating an absolute, long-term muddy-flood probability. This within-date, relative framing was a deliberate design choice in that variant rather than a limitation to be corrected: the practical decision the map is intended to support is which watersheds to prioritise today, given today's rainfall and cover conditions, not what the region-wide long-term probability of a muddy flood is. Within-date ranking is not, however, the only relevant comparison. On days when rainfall is well above an activation threshold, it also remains useful to distinguish more extreme rainfall days from marginally erosive ones, which is why this rank-based variant paired its ranking step with the rainfall-activation gate described next. This rank-based formulation was tested through the same leave-one-date-out evaluation as every other candidate but did not outperform the simpler three-raw-variable model on top-10% case capture, and was therefore not carried forward as the final model.

In this rank-based variant, a rainfall-activation gate was added to prevent the model from assigning elevated relative risk on dry or near-rainfall-free days. If the daily R-factor was at or below a selected activation threshold, the corresponding risk score was set to inactive/very low, regardless of how the watershed ranked among other watersheds that day; the relative R×C interaction score was only used to assign a risk class once the rainfall gate was passed. This two-step structure was designed to avoid the artefact whereby a purely relative ranking model would otherwise still produce a 'highest-ranked' watershed even on days too dry for muddy flooding to be physically plausible; the derivation of the specific activation threshold used, based on the long-term daily R-factor distribution. The final three-variable model described above does not use this rank-and-gate structure: because it operates on absolute predictor values rather than within-date ranks, a day with negligible rainfall erosivity already receives a correspondingly low raw score without requiring a separate gating step.

4.13 Training and Validation Dataset

The model was trained and validated using a manually curated event-date case/background dataset. Each row represents one candidate PICC-crossing watershed on one of 20 reported muddy-flood dates. The dataset contains 2,720 watershed-date observations across 136 candidate watersheds, with the response variable coded as 1 for the 42 watershed-date observations corresponding to a reported muddy-flood impact, and 0 for the remaining 2,678 non-reported background observations (a base rate of approximately 1.5%).

Because the background (non-reported) observations cannot be confirmed as true negatives , the absence of a report does not guarantee the absence of a muddy flood, the dataset is more accurately described as a positive-unlabelled (case/background) structure than as a conventional presence/absence dataset. This terminology, positive-unlabelled or case/background, is used consistently throughout the remainder of this thesis to refer to this specific labelling structure. The modelling task was accordingly framed as a within-date prioritisation problem: for a given event date, which exposed watersheds most closely resemble the locations where a muddy flood was actually reported?

4.14 Model Selection and Validation Strategy

Predictors were drawn from two parallel workflows: the Python RUSLE workflow (R-factor, C-factor, LS-factor, and hydrological-connectivity metrics such as SPI, TWI, and upstream area) and the WaTEM/SEDEM workflow (sediment-output summaries). A range of model families and predictor combinations was tested and compared; each is described briefly here:

- a full Random Forest using the complete predictor set, treating background rows as true negatives;
- a reduced, weighted Random Forest using only three physically motivated predictors, also treating background rows as true negatives;
- a Positive-Unlabelled Random Forest (PU-RF), which repeatedly resamples the unlabelled/background pool to build each bagged tree, without assuming that background rows are true negatives;

- date-matched Conditional Logistic Regression (CLR), stratified by event date to compare watersheds within the same event-date context;
- sediment-output-only models, based on WaTEM/SEDEM summaries alone;
- a simple rank-based model, using only the within-date ranks of R-factor and C-factor, without an explicit interaction term;
- a rank-based rainfall–cover interaction model, combining the R-factor rank, the C-factor rank, and their interaction;
- MaxEnt/maxnet, evaluated as a presence-background alternative.

All models were evaluated using leave-one-date-out cross-validation (LODO-CV), in which all watershed-date observations from one event date are withheld for testing while the model is fit on the remaining dates. This directly tests whether the model generalises to unseen rainfall/vegetation-cover conditions, rather than simply learning date-specific patterns; LODO-CV was preferred over a single train/test split because the number of distinct event dates is small, and a one-off split would have wasted scarce positive cases on a held-out set rather than using every date for both training and testing.

Performance was assessed using ROC-AUC, PR-AUC, and the proportion of reported cases captured within the top 10% and top 20% highest-ranked watersheds per date (higher rank meaning higher assigned risk) , the latter being the most directly relevant to the model's intended prioritisation use (see Section 6.3 for the rationale behind emphasising PR-AUC and top-percent capture over ROC-AUC alone).

4.14.1 Evaluation metrics

The four metrics used to compare candidate models are defined here before being applied. The area under the receiver operating characteristic curve, or ROC-AUC, summarises how well a model separates cases from background observations across all possible score thresholds, ranging from 0.5 for a model with no discriminative skill to 1.0 for perfect separation. The area under the precision-recall curve, or PR-AUC, instead summarises the trade-off between precision (the proportion of predicted-positive watershed-dates that are true cases) and recall (the proportion of true cases correctly identified), and is more informative than ROC-AUC under severe class imbalance, since it is sensitive to the low base rate of reported cases in this dataset. PR-AUC lift expresses a model's PR-AUC relative to the dataset's base rate, so that a lift above 1.0× indicates the model concentrates true cases into higher-ranked predictions more than chance would. Top-n% case capture is defined directly in terms of the model's intended operational use: for a given event date with a fixed number of candidate watersheds, the top-n% threshold selects the n% highest-ranked watersheds on that date, and top-n% case capture is the proportion of that date's reported cases falling among those selected watersheds, summed across all retained event dates. Throughout this thesis, top-n% capture is interpreted against an explicit random baseline: because a small number of reported cases are distributed across a much larger set of watershed-date observations, a random within-date ranking would be expected to place approximately n% of cases within any top-n% selection purely by chance; the enrichment factor, the ratio of a model's observed top-n% capture to this random expectation, expresses how much better than chance the model performs at a given selectivity .

4.14.2 Positive-Unlabelled Random Forest implementation

The PU-RF implementation was designed to match the case/background nature of the muddy-flood inventory . Reported muddy-flood watershed-date observations were retained as confirmed positives, while non-reported watershed-date observations were treated as unlabelled background rather than confirmed negatives. This avoids the stronger assumption made by a conventional Random Forest, in which every non-reported location would be interpreted as a true absence: because the model is never told that a background row is definitely negative, it cannot learn to treat under-reported but genuinely affected locations as reliable counter-examples, which is the central risk of applying a standard classifier to this type of data.

Within each training fold, all available positive cases were retained in every balanced training subset. The background observations were divided into batches of approximately the same size as the number of positives in that fold , for example, with 25 positive cases and 1,700 background observations in a fold, the background pool is divided into roughly 1,700/25 ≈ 68 batches, each treated in turn as a one-to-one pseudo-negative set alongside the full set of 25 positives for one Random Forest , and this batching procedure was repeated five times, drawing a fresh random partition of the background pool on each repeat, to avoid the result depending on a single arbitrary split. The final batch in each partition was filled where necessary to maintain a one-to-one positive/background ratio.

This one-to-one batching was repeated five times per fold, as noted above. For the best feature set, this produced between 250 and 350 Random Forest models per withheld date, and 3,690 Random Forest models across the full leave-one-date-out evaluation. The withheld observations were scored by all models trained in the corresponding fold, and their final PU-RF score was obtained by averaging the ensemble predictions. Random Forest models were implemented in R using the ranger package, chosen for its efficient handling of the repeated bagging structure described above across many leave-one-date-out folds.

The rainfall-constrained search required every candidate feature set to include at least one direct rainfall-erosivity variable, because muddy floods are event-triggered processes and should not be modelled using static susceptibility alone. The direct R variables were combined with process and static predictors such as RUSLE erosion summaries, cover-management factor summaries, SPI, TWI, and upstream-area indicators. This search design was used to identify physically interpretable feature sets that retained a rainfall trigger while allowing non-linear interactions to be learned by the Random Forest ensemble.

The final rainfall-constrained PU-RF evaluation used the cleaned event-date dataset after removing low-rainfall dates. Six low-R dates were removed from the original 20-date dataset: 2018-06-30, 2018-07-03, 2019-05-15, 2019-05-22, 2019-06-29, and 2021-06-15. Each removed date contained one reported case and 135 background observations. The retained dataset therefore contained 14 event dates, each with 136 candidate exposed watersheds, giving 1,904 watershed-date observations. Of these, 36 were reported muddy-flood cases and 1,868 were background observations, corresponding to a positive rate of 1.89%. This same rainfall-thresholded, 14-date dataset was used consistently for every model family compared in Section 5.2, so that the model comparison is made on a like-for-like basis.

The validation strategy described above is temporal, not spatial: leave-one-date-out cross-validation withholds an entire event date, testing whether the model generalises to unseen rainfall and vegetation-cover conditions, but it does not withhold entire geographic regions. Within any given training fold, watersheds that are geographically adjacent, and therefore likely to share similar terrain, soil, and connectivity characteristics, can appear in both the training set and the test set simultaneously, differing only in the event date under which each observation is evaluated. This is a recognised limitation for spatial environmental models generally, and it applies to the validation performed here: the top-percent capture and enrichment figures reported in Section 5.3.1 demonstrate that the model generalises across unseen rainfall and cover conditions, but they do not, on their own, demonstrate that it generalises across unseen terrain. A spatial block cross-validation, withholding geographically clustered groups of watersheds in addition to individual dates, was carried out as a supplementary check and is reported in Section 5.3.5.

4.15 Application to the Full Watershed Inventory

Following model selection, the validated rainfall–cover risk relationship was applied to the full representative PICC-flowaxis watershed inventory covering the wider study area introduced in Section 4.1: 11,122 representative watersheds across the full 2,331 km² ROI (compared with the 136 candidate watersheds used for model training and validation), each evaluated for the 20 selected event dates, giving 222,440 watershed-date rows in the full prediction table. Of the watersheds in this full inventory, 15 correspond to reported event IDs, represented by 29 case-date rows , fewer than the 42 cases in the original training dataset, because not every originally reported pour point was retained in the final, thinned representative spatial base .

The validated risk score was applied to every exposed representative pour point in this table, including background points, and the results were exported as a point shapefile with one record per date and pour point. In ArcGIS Pro, the date field was used to enable time-aware display, with point symbology assigned according to the resulting risk class.

5. Results

5.1 Descriptive Statistics of the Case/Background Dataset

Before presenting model performance, this section summarises the reported muddy-flood cases and their candidate watersheds used in training and validation. Across the 20 reported event dates, watershed size among the 136 candidate exposed watersheds ranges from small, single-hillslope contributing areas to larger, multi-hectare catchments; reported (case) watersheds are drawn from across this size range rather than being systematically the largest or smallest. Land use within candidate watersheds is dominated by arable cropland, consistent with the loess-belt setting described in Section 4.1; the remaining non-cultivated share includes built-up land concentrated near the PICC exposure points that define each watershed's outlet by construction, alongside other non-agricultural land cover. Reported case watersheds do not show an obvious minimum-area or degree-of-urbanisation cutoff that would justify excluding background points on that basis alone; watershed size and land-use composition were therefore retained as candidate predictors in the model-selection search rather than used as a pre-filtering rule.

Table 5 summarises watershed area, cultivated-area proportion, rainfall erosivity, and mean slope for the cleaned 36-case, 1,904-observation evaluation set used for the final model, split by reported case and background watersheds. Reported case watersheds are on average larger than background watersheds, with a substantially higher mean area (55.7 ha versus 37.3 ha), although their median area is smaller (4.5 ha versus 7.3 ha), reflecting a right-skewed distribution in both groups in which a small number of large contributing catchments raise the mean well above the median. Case watersheds also show a higher proportion of cultivated land on average (69.9% versus 58.4%) and a substantially higher mean rainfall erosivity on their reported event dates (31.4 versus 15.8), consistent with the combination of erosive rainfall and cultivated cover assumed by the muddy-flood process model in Section 4.9. Mean slope differs only marginally between the two groups (2.4 degrees versus 2.3 degrees), suggesting that slope alone does not strongly separate reported cases from background watersheds in this dataset, which is consistent with slope-related terms not being selected in the final rainfall-constrained PU-RF model. Given the skewed distributions evident in the mean-median gap for watershed area, these summary statistics are intended to characterise the case/background dataset descriptively, rather than to imply that any single variable in isolation separates cases from background by a fixed threshold.

Table 5. Descriptive statistics for reported case and background watersheds, cleaned 36-case evaluation set.

Variable (group)	Mean	Median	Min	Max	n
Watershed area, ha (case)	55.7	4.5	0.18	310.2	36
Watershed area, ha (background)	37.3	7.3	0.18	324.2	1,868
Cultivated area, % (case)	69.9	76.4	23.5	97.0	36
Cultivated area, % (background)	58.4	62.3	0.0	100.0	1,868
Rainfall erosivity, R-factor mean (case)	31.4	6.1	0.23	123.2	36
Rainfall erosivity, R-factor mean (background)	15.8	2.5	0.0	217.7	1,868
Mean slope, degrees (case)	2.4	2.0	0.77	12.5	36
Mean slope, degrees (background)	2.3	1.9	0.64	12.5	1,868

5.2 Representative Watershed Base

The representative-pour-point approach produced a spatially continuous and hydrologically more meaningful watershed inventory than the rejected exhaustive approach, which had produced a highly fragmented mosaic of small, competing D8 catchments. The final representative workflow reduced this fragmentation by grouping connected PICC-flowaxis hotspots and thinning near-duplicate points prior to watershed delineation.

The resulting watershed base contains 11,122 representative exposed watersheds. These watersheds form a fixed spatial framework: their boundaries do not change through time, while the rainfall- and vegetation-cover-derived predictors associated with each one change by date.

5.3 Model Comparison and Validation Outcome

Table 6 summarises the main modelling outcomes rather than every exploratory run. The date-matched CLR and PU-RF branches answer related but not identical questions, and some metrics are not directly comparable because slightly different subsets and validation framings were used, as noted in the table below. The table is therefore used to compare the role of different model families, while the final rainfall-constrained PU-RF result is reported separately .

Table 6. Summary of different model families.

Model group	Main result	Interpretation
Full Random Forest (all predictors)	ROC-AUC ≈ 0.54	Weak leave-one-date-out generalisation. The model overfit the training data and treated all background rows as true negatives, which is not appropriate for reported-event data.
Conventional RF variants	ROC-AUC ≈ 0.67–0.75; PR-AUC ≈ 0.04–0.05	Class-weighted and balanced RF variants improved over the full RF, but both still rely on a presence/absence framing in which non-reported background rows are treated as negatives. Retained only as baseline comparisons.
Positive-Unlabelled Random Forest	Best model combines mean rainfall erosivity, the 95th-percentile RUSLE-based erosion rate, and the 95th-percentile cover-management factor. ROC-AUC is approximately 0.620, PR-AUC is approximately 0.024, and top-10% capture is 15 of 36 cases.	Main machine-learning result. Although conventional classification metrics are modest, the model captured 41.7% of reported cases within the top 10% of ranked watersheds per date, making it useful as a prioritisation model.
Sediment-output-only models	ROC-AUC ≈ 0.52–0.59; PR-AUC ≈ 0.02–0.03	Sediment-output variables contain physically relevant information but were weaker on their own and did not improve the final combined model.
Rainfall-cover CLR / rank-based models	Rank R×C CLR: ROC-AUC ≈ 0.719; PR-AUC ≈ 0.047; top-10% capture = 15/42	Interpretable statistical benchmark showing that the interaction between rainfall erosivity and cover vulnerability has meaningful ranking skill. Not directly comparable to the final 36-case

		PU-RF result, evaluated on the 42-case set (Section 5.3, note below).
MaxEnt / maxnet	ROC-AUC ≈ 0.68–0.69; PR-AUC ≈ 0.03–0.04	Presence-background sensitivity check. Results were broadly comparable but did not outperform the main prioritisation models.

The conventional RF values refer to related but distinct baseline variants. The weighted RF used class weights, whereas the balanced RF used a different balancing strategy and reached higher ROC-AUC and PR-AUC. Because both variants still treat sampled or weighted background observations as negatives, they are grouped here as conventional RF baselines rather than as final models.

The rank-based CLR evaluation used the full 42-case, 20-date event dataset. The final rainfall-constrained PU-RF evaluation removed six low-rainfall dates and was evaluated on 36 reported cases across 14 event dates. For this reason, the CLR and PU-RF numbers should be interpreted as complementary evidence from related but distinct validation subsets, rather than as directly interchangeable estimates from the same evaluation.

5.3.1 Best rainfall-constrained PU-RF prioritisation result

The best-performing PU-RF model in the rainfall-constrained search came from the rainfall-constrained plus process/static predictor family. It used three predictors: mean rainfall erosivity, the 95th-percentile RUSLE-based erosion rate, and the 95th-percentile cover-management factor. These variables represent three complementary elements of the muddy-flood process: the erosive rainfall trigger, high-end erosion intensity within the contributing watershed, and high-end cover-management vulnerability.

In conventional classification terms, model performance was modest, with ROC-AUC = 0.620 and PR-AUC = 0.024. This low PR-AUC is expected in a dataset with only 36 reported cases among 1,904 observations, especially because the background class is unlabelled rather than a set of confirmed absences. For the purpose of this thesis, the more important question is whether the model places reported cases within the highest-risk fraction of watersheds on each event date.

Table 7. Best rainfall-constrained PU-RF result for within-date watershed-date prioritisation.

Metric	Value
Model family	Rainfall-constrained plus process/static PU-RF
Predictors	Mean rainfall erosivity, 95th-percentile RUSLE-based erosion rate, 95th-percentile cover-management factor
Validation strategy	Leave-one-date-out cross-validation
Event dates used	14
Candidate watersheds per date	136
Total test observations	1,904
Reported cases	36
Background/unlabelled observations	1,868

Base rate	1.89%
PU background sampling	Repeated one-to-one positive/background batching
PU repeats	5
Random Forest models per fold	250–350
Total Random Forest models across folds	3,690
ROC-AUC	0.620
PR-AUC	0.024
PR-AUC lift over base rate	1.28×
Top 5% capture	8/36 = 22.2%
Top 10% capture	15/36 = 41.7%
Top 20% capture	17/36 = 47.2%
Median date-level case rank (0 = highest risk)	0.211

The top-percent capture metrics show that the model has useful prioritisation skill. Higher risk-score rank corresponds to higher assigned risk throughout this thesis: the top 10% of a date's watersheds are its 10% highest-risk watersheds, not its lowest. Since each date contains 136 candidate watersheds, the top 5%, top 10%, and top 20% thresholds correspond to approximately 7, 14, and 28 watersheds per date. Across the 14 retained event dates, these thresholds select 98, 196, and 392 watershed-date observations, respectively. The best model captured 8 of 36 reported cases in the top 5%, 15 of 36 cases in the top 10%, and 17 of 36 cases in the top 20%.

The top 10% result is the most operationally relevant. By selecting only the 14 highest-ranked watersheds per date, the model captured 15 of 36 reported cases, or 41.7% of the cleaned test-set cases. A random ranking would be expected to place approximately 10% of cases, or about 3.6 of the 36 cases, in the top 10%. The observed top-10% capture is therefore more than four times the random expectation, indicating that the model concentrates reported impacts into a relatively compact priority set.

Expanding from the top 10% to the top 20% does not provide a proportional improvement. It doubles the number of selected watersheds per date, from about 14 to about 28, and doubles the selected test set from 196 to 392 watershed-date observations. However, it recovers only two additional reported cases, increasing capture from 15/36 to 17/36. The additional band between 10% and 20% therefore contains only 2 reported cases among 196 additional selected observations. This marginal case concentration is about 1.0%, which is lower than the overall 1.89% case rate in the cleaned PU-RF dataset.

This result supports using the top 10% as the main operational priority class. The top 20% class can still be shown as a broader, lower-confidence screening zone, but it should not be interpreted as providing twice as much useful information. For the final map, the top 10% is the more efficient and defensible threshold because it captures most of the cases recovered by the top 20% threshold while requiring only half as many locations to be inspected or prioritised.

A related rainfall-constrained model using only mean rainfall erosivity and the 95th-percentile RUSLE-based erosion rate achieved a higher top-20% capture of 20/36 and a slightly higher ROC-AUC of 0.643, but captured only 9/36 cases in the top 10%, and is therefore less suitable for the targeted prioritisation objective of this thesis. Conversely, a standard one-to-one PU-RF run without the repeated rainfall-constrained setup selected the same three-variable combination as its best top-10% model, but captured 13/36 cases in the top 10% rather than 15/36. This confirms that the same physically meaningful variables are robust, while the repeated rainfall-constrained setup gives the stronger operational top-10% result.

5.3.2 Random baseline and efficiency of top-percent selection

To interpret the top-percent capture metrics, the best PU-RF result was compared with a random within-date ranking. Because 36 reported cases were present in the cleaned PU-RF test set, a random ranking would be expected to place approximately 1.8, 3.6, and 7.2 cases in the top 5%, 10%, and 20% of selected watersheds, respectively.

The observed result is substantially above this random expectation for the most selective thresholds. The best model captured 8 cases in the top 5% and 15 cases in the top 10%, corresponding to enrichment factors of 4.44× and 4.17× relative to random selection. The top 20% also remains above random, but its enrichment is lower because the selected set becomes much larger while only two additional cases are recovered beyond the top 10%.

Table 8. Random baseline comparison for the best rainfall-constrained PU-RF model.

Selection threshold	Selected rows	Random expected cases	Observed cases	Enrichment	Case concentration
Top 5%	98	1.8/36	8/36	4.44×	8.16%
Top 10%	196	3.6/36	15/36	4.17×	7.65%
Top 20%	392	7.2/36	17/36	2.36×	4.34%
Marginal 10–20%	196	3.6/36	2/36	0.56×	1.02%

The marginal 10–20% band confirms, via the random-baseline framing, the same conclusion reached in Section 5.2.1: it selects the same number of additional watershed-date observations as the first 10% band, but its case concentration (about 1.02%) falls below the overall 1.89% base rate, whereas the top 5% and top 10% bands both exceed it substantially. The enrichment factor makes this concentration explicit , 4.44× and 4.17× above random for the top 5% and top 10% respectively, falling to 0.56× for the marginal 10–20% band, meaning that band performs worse than a random selection of the same size.

[Figure 6 , Cumulative gain curve for the best rainfall-constrained PU-RF model, compared with a random within-date ranking. Muddy-flood cases captured (%) on the vertical axis, per your Comment #116.]

Figure 6. Cumulative gain curve for the best rainfall-constrained PU-RF model, compared with a random within-date ranking.

5.3.3 Date-level behaviour of the best PU-RF model

The leave-one-date-out design also makes it possible to examine whether the overall top-10% result is spread across multiple event dates or dominated by a single date. The date-level summary (Table 6) shows that the model captures at least one reported case in the top 10% on 10 of the 14 retained event dates. The largest event dates, 2018-05-27 and 2018-06-01, contribute 4 and 3 top-10% cases, respectively, but they do not account for the entire result. This supports the interpretation that the model has some temporal prioritisation skill, although performance remains uneven between dates.

Table 9. Date-level capture summary for the best rainfall-constrained PU-RF model.

Event date	Reported cases	Cases in top 10%	Cases in top 20%	Median case rank
2018-05-01	2	1/2	1/2	29.0%
2018-05-09	2	1/2	1/2	33.5%
2018-05-24	1	0/1	0/1	62.5%
2018-05-27	8	4/8	5/8	12.9%
2018-05-28	1	1/1	1/1	2.2%
2018-06-01	11	3/11	4/11	29.4%
2018-07-04	2	1/2	1/2	20.2%
2018-08-09	1	1/1	1/1	5.1%
2018-09-05	2	0/2	0/2	60.7%
2019-05-18	1	0/1	0/1	34.6%
2020-08-14	1	1/1	1/1	0.7%
2021-07-01	2	1/2	1/2	19.5%
2021-07-14	1	0/1	0/1	22.1%
2021-07-24	1	1/1	1/1	5.1%

The same table also shows where the model fails. Some single-case dates, such as 2018-05-24, 2019-05-18, and 2021-07-14, are not captured within the top 20%. This is important for interpretation: the model is useful for prioritising a subset of exposed watersheds, but it is not a deterministic detector of every reported event.

5.3.4 Comparison with simpler predictor families

The rainfall-constrained search also helps clarify the added value of combining rainfall with process and cover variables. Rainfall-only, static-only, and process-only models all captured fewer reported cases in the top 10% than the final rainfall-constrained plus process/static model (Table 7). This does not prove that the selected model captures the complete muddy-flood process, but it does show that the strongest prioritisation result was obtained when the rainfall trigger was combined with erosion intensity and cover vulnerability.

Table 10. Best model within each rainfall-constrained search family.

Family	Best feature set	ROC-AUC	PR-AUC	Top 10%	Top 20%
Rainfall only	Mean rainfall erosivity	0.591	0.021	10/36	12/36
Static only	Mean cover-management factor (cultivated land)	0.505	0.023	9/36	9/36

	and 95th-percentile cover-management factor				
Process only	Total erosion proxy	0.569	0.028	9/36	12/36
Rainfall-constrained plus static	Mean rainfall erosivity, maximum Topographic Wetness Index, and mean upstream contributing area	0.505	0.022	12/36	13/36
Rainfall-constrained plus process/static	95th-percentile cover-management factor, mean rainfall erosivity, and 95th-percentile RUSLE-based erosion rate	0.620	0.024	15/36	17/36
R interactions	Rainfall-connectivity interaction, together with each variable individually	0.383	0.017	9/36	10/36

The process-only model based on total erosion proxy has a slightly higher PR-AUC than the final top-10% model, but it captures only 9 of 36 reported cases in the top 10%. This illustrates why PR-AUC alone is not sufficient for selecting the final map variable: for an operational prioritisation layer, the ability to place reported cases within the highest-ranked watersheds per date is more relevant than small differences in PR-AUC.

5.3.5 Spatial validation of the best model

A conventional random train-test split can give optimistic performance estimates in spatial environmental modelling because of spatial and temporal autocorrelation (Tobler's First Law of Geography), which is why spatial block cross-validation was carried out here as an additional check. The leave-one-date-out design used throughout Section 5.3 withholds an entire event date but does not withhold entire geographic regions: watersheds that are spatially adjacent, and therefore likely to share similar terrain, soil, and connectivity characteristics, can appear in both the training set and the test set within the same fold, differing only in the event date under which each observation is evaluated. To test whether the best rainfall-constrained PU-RF model (Section 5.3.1) depends on this spatial overlap, the same three-predictor model was re-evaluated under two additional validation schemes that withhold geographic clusters of watersheds rather than, or in addition to, individual dates. The 136 candidate watersheds were grouped into eight spatial clusters using k-means clustering on their pour-point coordinates. Under pure spatial block cross-validation, one cluster was withheld at a time, with the model trained on all remaining clusters and evaluated across every date. Under nested spatial-within-date cross-validation, each date's leave-one-date-out fold additionally excluded, from training, any watershed sharing a spatial cluster with that date's reported case watershed, so that both the date and the surrounding geographic neighbourhood of each case were held out simultaneously.

Table 7a compares the two spatial validation schemes against the temporal-only leave-one-date-out result reported in Table 4. ROC-AUC and PR-AUC remained broadly stable across all three schemes, ranging from 0.614 to 0.633 and from 0.023 to 0.025 respectively, indicating that the model's overall ability to separate reported cases from background observations does not depend primarily on spatial overlap between training and test watersheds. Top-percent capture, by contrast, fell substantially under spatial holdout. Top-10% capture dropped from 15 of 36 cases (41.7%) under temporal-only validation to 9 of 36 (25.0%) under nested spatial-within-date validation and to 6 of 36 (16.7%) under pure spatial block validation. Both spatial schemes remained above the approximately 10% capture expected under a random within-date ranking, but the operational prioritisation skill reported in Section 5.3.1 is therefore only partly attributable to the dynamic rainfall-erosion-cover mechanism the model is intended to capture; part of it reflects persistent local terrain or soil characteristics near the reported case watersheds that a purely temporal holdout cannot detect.

Table 11. Comparison of temporal and spatial validation schemes for the best rainfall-constrained PU-RF model.

Validation scheme	ROC-AUC	PR-AUC	Top 10% capture	Top 20% capture
Temporal (leave-one-date-out), Table 4	0.620	0.024	15/36 (41.7%)	19/36 (52.8%)
Nested spatial-within-date CV	0.633	0.025	9/36 (25.0%)	13/36 (36.1%)
Pure spatial block CV	0.614	0.023	6/36 (16.7%)	12/36 (33.3%)

With only 36 reported cases distributed across eight spatial clusters, some clusters necessarily contain very few case watersheds, so the spatial-block estimates in Table 7a are based on smaller effective sample sizes per fold than the temporal-only result and should be read as indicative rather than as precise replacements for Table 4. The comparison nonetheless represents a genuine limitation of the validation strategy used elsewhere in this thesis, and is discussed further as a priority direction for future work in Chapter 6.

WaTEM/SEDEM sediment-output variables were tested as additional predictors but did not improve on the validated rainfall–cover model. The single strongest sediment-derived indicator was sediout_at_point (sediment output at the pour point itself); adding it to the rainfall–cover model produced a small decrease in validation performance rather than an improvement (Table 3, rows 3 and 5). Sediment-output variables therefore contain physically relevant information, but were not the strongest predictors of reported, event-date muddy-flood impact in this dataset. They are accordingly retained as a supporting visual variable (e.g. point size) in the final risk map, rather than as a component of the main risk score.

5.5 Dynamic Risk Map Output

The final output of the modelling framework is a dynamic risk-prioritisation layer for exposed PICC-flowaxis pour points. Each pour point receives a model score for each evaluated date based on the rainfall and land-surface conditions in its upstream contributing watershed. The score is produced by the final rainfall-constrained PU-RF model using mean rainfall erosivity, the 95th-percentile RUSLE-based erosion rate, and the 95th-percentile cover-management factor. The score is interpreted as a relative prioritisation measure rather than a calibrated probability of muddy-flood occurrence.

To make the output suitable for practical interpretation, the continuous model score was grouped into five qualitative risk classes. The classes are ordered from lower to higher priority and describe how strongly a location is prioritised relative to the other exposed locations under the same date-specific conditions:

- Very low / inactive: rainfall conditions are inactive or the model score is near zero.
- Low: rainfall is active, but the model assigns a relatively low score.
- Moderate: the location falls within the middle range of scores among active locations.
- High: the location falls within the upper range of scores, approximately corresponding to the top 20% priority level used in the model evaluation.
- Very high: the location falls within the highest-ranked group, approximately corresponding to the top 10% priority level used as the principal operational priority class.

The classification is dynamic because the model inputs change from date to date. A pour point that is classified as high on one date may therefore be moderate or low on another date if rainfall erosivity or vegetation-cover vulnerability changes. Conversely, a location with persistent structural susceptibility can become a high-priority location when the rainfall and surface conditions become more favourable for muddy-flood generation. The output should therefore be understood as a date-specific ranking of exposed locations rather than as a static susceptibility label.

The full prediction output contains one record for each representative pour point and evaluated date, allowing the results to be inspected spatially and temporally in a GIS environment. In operational use, the same framework can be applied to new dates for which the required rainfall, vegetation, and erosion inputs are available. On dates with little or no erosive rainfall, the output naturally concentrates locations in the inactive or very-low classes rather than forcing a relative ranking that would imply meaningful muddy-flood risk under physically unsuitable conditions. On dates with active erosive rainfall, the output provides a within-date prioritisation of the exposed locations that warrant the greatest attention.

The resulting product is therefore intended to support screening and prioritisation of exposed infrastructure, field inspection, local mitigation planning, and further hydrological investigation. It should not be interpreted as an early-warning probability map or as evidence that a muddy flood will occur at every location assigned to the highest class. Its purpose is to identify where, given the environmental conditions of a particular date, the available evidence suggests that limited monitoring or mitigation resources can be prioritised most efficiently.

6. Discussion

6.1 Why the Weaker Models Underperformed

Several of the tested model families underperformed for reasons that are informative for interpreting the final model choice.

The full Random Forest, using the complete predictor set, suffered from a combination of highly correlated erosion-related variables, very few positive cases, and the conceptual problem of treating every unreported watershed-date row as a confirmed true absence. Its leave-one-date-out performance (ROC-AUC ≈ 0.54) was only marginally better than random, indicating substantial overfitting to the training dates rather than transferable physical signal. This overfitting behaviour is a known property of Random Forests trained on highly imbalanced, correlated predictors rather than an artefact specific to this dataset; it does not, on its own, indicate a problem with hidden feature interactions.

Reducing the Random Forest to three physically motivated predictors improved its behaviour, but it remained conceptually weak for the same reason: the absence of a report does not mean the absence of a muddy flood, only that the location is unlabelled rather than confirmed negative.

Sediment-output-only models were too narrow on their own: muddy floods require both an erosive trigger and a connected, vulnerable landscape, and sediment output alone does not capture the interaction between rainfall erosivity and surface-cover vulnerability.

The simple rank model, ranking R and C separately without an explicit interaction term, was useful but incomplete: ranking each factor separately is physically meaningful, but without an explicit interaction term and an explicit routing/connectivity signal it cannot represent the underlying process, in which high rainfall only becomes dangerous where the surface is also vulnerable and sediment can actually be delivered to the outlet.

The strict pairwise-matched CLR is useful mainly as a robustness check. Because it relies only on a set of matched controls, it does not exploit the full background dataset, and it answers a narrower question, what differs within matched sets, rather than the broader prioritisation question the final map is intended to answer.

MaxEnt is a maximum-entropy method originally developed for species-distribution modelling, where only confirmed presences and a background sample of the study area are available, with no confirmed absences; maxnet is the R implementation used here. Muddy-flood cases share this same case/background structure, so MaxEnt was included as an established, off-the-shelf presence-background method, distinct from the PU-RF and CLR approaches developed specifically for this thesis. Evaluated as this presence-background alternative, MaxEnt/maxnet produced lower ROC-AUC and PR-AUC than the CLR models, and was therefore retained only as a sensitivity check rather than a candidate for the final model.

6.2 Strengths of the Date-Matched CLR and PU-RF Models

Two model families stood out as the strongest candidates, for complementary reasons.

The date-matched conditional logistic regression, using the rainfall-erosivity rank, the cover-vulnerability rank, and their interaction (and, in one tested variant, the sediment-output summary), directly tests a physically motivated question: on the same event date, are reported muddy-flood watersheds associated with the combination of high rainfall erosivity, vulnerable cover, and sediment delivery. Matching by event date is the key strength of this approach, because it compares watersheds within the same rainfall context, rather than simply learning that some dates were worse than others overall , in effect, it asks which watersheds were relatively more susceptible on a given day. The explicit interaction term is physically meaningful: a high R-factor over well-covered land, or a high C-factor without erosive rainfall, should not by itself indicate elevated muddy-flood risk, whereas their interaction directly represents the combination of an erosive trigger and an exposed, vulnerable surface. This model's main strength is interpretability , its coefficients can be related directly to a physical mechanism, and its main limitation is that it assumes a comparatively simple functional relationship and may miss non-linear thresholds in the underlying process.

Ensemble decision-tree methods are well suited to rare-event susceptibility mapping of this kind: their hierarchical threshold structure and robustness to the moderate collinearity expected among rainfall, cover, and erosion predictors made them a natural fit for the muddy-flood problem. Building the final model as a hybrid physical-empirical architecture, in which RUSLE- and rainfall-derived quantities are computed first and only then passed to the classifier as engineered predictors, rather than raw spectral or meteorological rasters, is what keeps the final predictor set interpretable in the way discussed below.

The Positive-Unlabelled Random Forest (PU-RF) is the strongest machine-learning candidate, mainly because it matches the actual structure of the labels: positive rows are confirmed reported events, while all remaining rows are unlabelled rather than confirmed negatives. Rather than fitting one highly imbalanced Random Forest to all background rows as if they were true absences, the PU-RF repeatedly compares the known positives with balanced one-to-one samples of the unlabelled background pool. In the final rainfall-constrained PU-RF search, the best prioritisation model combined mean rainfall erosivity, high-end RUSLE erosion intensity, and high-end cover vulnerability. Its ROC-AUC was modest (0.620) and its PR-AUC was low (0.024), but it captured 15 of 36 reported cases within the top 10% of ranked watersheds per date. The main strength of this result is therefore not probability calibration, but its ability to concentrate a disproportionate share of reported muddy-flood impacts into a small operational priority set.

Taken together, these results point to a two-track modelling interpretation: the date-matched CLR provides the main interpretable rainfall-cover benchmark, while the PU-RF provides the main machine-learning prioritisation benchmark under positive-unlabelled labels. The central conclusion is not that any one model perfectly detects muddy floods, but that the better-performing approaches share three features: they compare watersheds within the same event-date context, they avoid treating unreported locations as confirmed absences, and they represent the process chain linking rainfall erosivity, cover vulnerability, and erosion generation. The PU-RF results further suggest that useful information is concentrated in the highest-ranked watersheds, which favours threshold- or rank-based prioritisation over a calibrated probability interpretation.

6.3 PR-AUC and Top-Percent Capture Matter More Than ROC-AUC

Overall accuracy is not a meaningful evaluation metric under extreme class imbalance; metrics focused on the minority class, such as PR-AUC, are needed instead. Because reported muddy floods are rare , 42 of 2,720 rows (1.5%) in the full training dataset, or 42 of 2,020 rows (2.1%) in the cleaned Random Forest/PU-RF comparison subset , the choice of evaluation metric matters greatly.

ROC-AUC compares the true-positive rate (TP / all positives) against the false-positive rate (FP / all negatives). With a rare positive class, the false-positive rate can look modest even when the absolute number of false positives is large. For example, a model that correctly flags 22 true muddy floods alongside 205 false positives has a false-positive rate of only 205/1,978 ≈ 10.4%, which does not look alarming in ROC space, yet its precision (the share of flagged rows that are genuinely reported events) is only 22/(22+205) ≈ 9.7%, meaning roughly one in ten flagged 'high-risk' rows actually corresponds to a reported muddy flood.

PR-AUC, based on precision (TP/(TP+FP)) and recall (TP/(TP+FN)), exposes this issue directly, because it asks the more operationally relevant question: among the watersheds ranked as risky, how many are genuinely reported muddy floods, and how many of the reported muddy floods are recovered? This is much closer to the practical goal of this study, which is to rank or flag a manageable subset of high-priority watersheds, rather than to classify every background row correctly.

In practice, this is why a model can show what looks like a respectable ROC-AUC (e.g. 0.72) alongside a modest PR-AUC (e.g. 0.05): the model has genuine ranking skill , it is able to place reported cases above background, on average , but a large share of the rows it would flag as 'high risk', if a hard threshold were applied, would still be unreported/background locations.

For this reason, PR-AUC is treated as an important metric in this study: muddy floods are rare; the background class is very large; the practical use of the map is to prioritise the highest-risk watersheds rather than to correctly classify the much larger set of low-risk ones; false positives reduce the practical usefulness of the risk map; and because no report is not a confirmed absence, precision is informative but inherently difficult to push very high. PR-AUC alone is nonetheless not sufficient on its own, and the final model evaluation also reports ROC-AUC, precision, recall, F1, the proportion of reported cases captured in the top 5%, 10%, and 20% of ranked watersheds, and the median rank of reported cases per date. The top-percent capture rates are the most directly relevant to the map's intended use, since they answer the operational question: on a muddy-flood day, was the reported watershed placed among the small set of highest-priority locations?

The final PU-RF result illustrates why this distinction matters. Expanding the selected set from the top 10% to the top 20% doubles the number of selected watersheds per date, from approximately 14 to 28, and doubles the total selected test observations from 196 to 392. However, the best model gained only two additional reported cases, increasing capture from 15/36 to 17/36. The marginal 10–20% band therefore contained only 2 cases in 196 additional selected observations, a concentration of about 1.0%, lower than the overall 1.89% case rate in the cleaned PU-RF dataset. This shows that the useful ranking signal is concentrated mainly in the top 10% class, while the broader top 20% class adds many background locations with limited additional case recovery.

6.4 Interpretation of the Final Model

The validation results indicate that reported muddy-flood impacts in the study area are best explained by the interaction between rainfall erosivity and crop-cover vulnerability, rather than by erosion potential or sediment-delivery output alone. This is physically plausible: muddy floods are event-driven, requiring both sufficient rainfall energy and a vulnerable upstream surface at the same time. High rainfall over well-covered land, or a high C-factor in the absence of erosive rainfall, should not by itself generate the same risk as the combination of the two. The rainfall-cover interaction, and its shifts with season and antecedent soil moisture, is plausibly non-linear in a way a fixed logistic interaction term can only partly capture. A Random Forest can learn this threshold-like behaviour directly from the data rather than requiring it to be specified in advance, which may explain why the PU-RF ultimately outperformed the CLR interaction term on the operational top-10% result.

The rank-based formulation is also well suited to the available data, since the reported muddy-flood database is sparse and impact-biased. Rather than estimating an absolute muddy-flood probability, the model ranks exposed infrastructure locations by their similarity to reported impact locations under the same daily rainfall and vegetation-cover conditions, a result better described as dynamic relative-risk prioritisation than as an occurrence-probability map.

The dominance of the rainfall and cover-vulnerability terms in the final model is, in part, a direct consequence of the modelling design: topography, watershed geometry, and most static soil properties do not change between dates, whereas rainfall and vegetation cover do. The model is therefore primarily sensitive to temporal variation in storm erosivity and cover vulnerability, appropriate for a tool whose explicit purpose is to show how risk changes through time, but it also means the model should be read as an event-date prioritisation tool rather than as a deterministic, independent predictor of sediment volume.

RUSLE and WaTEM/SEDEM nonetheless remain important to the overall framework, because they provide physically meaningful intermediate variables (erosion potential, sediment transport) against which the reported-impact patterns can be interpreted, even though full erosion or sediment-output metrics were not, in this dataset, the strongest final predictors of reported impact. Erosion intensity is expected to carry information beyond the R- and C-factors individually, since it is not a simple product of the two: it also integrates the LS-factor (slope length and steepness) and the K-factor (soil erodibility), both of which vary spatially across a watershed in ways that R and C alone do not capture, and it is the variable through which watershed-level connectivity effects. The fact that erosion intensity was nonetheless not the single strongest predictor in this dataset is plausibly because reported muddy-flood occurrence reflects not only sediment production but also reporting bias, local drainage connectivity, infrastructure exposure, and the precise position of the pour point relative to the flow path. The final framework therefore uses the rainfall × cover interaction as the main dynamic risk score, while retaining sediment-output variables as a supporting indicator of potential sediment-delivery intensity, combining physically based erosion modelling with the predictor combination that validated best against the reported data.

6.5 Spatial and Data Limitations

The final map shows visible block-like boundaries in rainfall-driven risk. This is an expected consequence of resampling the comparatively coarse 5 km RMI rainfall grid to the 5 m modelling resolution by nearest-neighbour interpolation: each rainfall value is reproduced unchanged across many thousands of output cells, producing tile-like discontinuities. This should be read as a limitation of the rainfall data's native spatial resolution, rather than as an error in the broader modelling workflow. It is also compounded by the resolution finding of Cantreul et al. (2018): the 5 m modelling grid adopted here is already coarser than the 1 m optimum identified for connectivity mapping in this landscape type, so fine linear features that disrupt or redirect flow are only partially represented.

Watershed delineation is similarly influenced by the choice of D8 flow routing, which creates a single downstream flow path for every cell and thereby enables straightforward watershed delineation from a pour point, but which can fragment the landscape artificially when pour points are placed too close together. This was mitigated, though not entirely eliminated, by using representative rather than exhaustive PICC-flow pour points; the resulting watershed base is therefore a deliberate compromise between spatial coverage and hydrological interpretability.

More generally, the modelling framework's future transferability is constrained by: the 5 m working resolution (itself a ten-fold upsampling of the original 50 cm LiDAR DTM); the absence of harmonisation between the Walloon DTM and neighbouring regions' elevation data at the regional border (not a concern for the watersheds used here, all of which lie within Wallonia, but relevant for any future extension beyond it); the pairwise/matched-comparison structure used in some candidate models, which limits how much of the available background data they exploit; and the partly manual delineation and curation of the representative watershed base, which would need to be revisited if the workflow were extended to new infrastructure datasets or time periods.

The principal data limitation is the scarcity and bias of the reported muddy-flood observations. Reported events generally reflect impacts noticed near infrastructure or settlements, not all muddy runoff occurring across the landscape; an absence of report therefore does not establish an absence of muddy flooding, and background points should be interpreted as non-reported exposed locations rather than as confirmed non-events. A related, smaller-scale limitation is that only a subset of the original reported pour points survived into the final representative watershed base used for full-scale application: of the original reported sites, only 15 reported event IDs correspond to 29 case-date rows in the full 11,122-watershed inventory (Section 4.15). This does not invalidate the final risk map, since the underlying model was trained and validated on the earlier, more complete case/background dataset (Section 4.13); it does mean that the full representative inventory should be treated primarily as the prediction space for the validated model, rather than as a replacement training dataset.

6.6 Derivation of the Rainfall-Activation Threshold

The rainfall-activation gate introduced in Section 4.12 requires a specific threshold value, derived here independently from the long-term daily R-factor distribution, using RMI 5 km daily precipitation totals for May–September over 2000–2025, converted to daily erosivity via $R = 0.1089P^{1.8067}$ (Section 4.6). To avoid dry days dominating the distribution, the threshold was calculated from wet pixel-days only ($P \geq 2$ mm); the 95th percentile of this wet-day distribution ($R95 \approx 29.84$, equivalent to roughly 22.4 mm/day) was adopted as the high-R activation threshold, with the 99th percentile ($R99 \approx 72.20$, equivalent to roughly 36.5 mm/day) retained as an optional very-high threshold.

This derived threshold is not directly comparable to the fixed erosivity value used by De Walque et al. (2017), who assumed a single, spatially uniform R-factor of 23,300 $MJ \cdot mm \cdot km^{-2} \cdot yr^{-1} \cdot h^{-1}$ for their entire regression, corresponding to a 30-minute design storm of 17.4 mm with a 5-year return period; because their study did not have the date of each reported event, this fixed value serves only to rank sites relative to one another and was not intended to represent a real, dated rainfall condition. The threshold derived here instead reflects a historical percentile of the region's actual daily rainfall-erosivity distribution, and is used specifically to gate the dynamic, date-specific risk score described in Section 4.12. Six reported events in the present database could not be matched to sufficient rainfall erosivity within a 0–2-day window, and were treated as uncertain rainfall-mismatch cases rather than excluded outright.

6.7 Practical Value and Recommended Cartography

The final product is a dynamic point map of exposed muddy-flood impact risk, intended to identify which PICC-flowaxis crossings become relatively more critical under specific rainfall and vegetation-cover conditions. This supports the prioritisation of field inspection, local mitigation planning, and further hydrological investigation, rather than providing an absolute probability of occurrence.

For operational use, the same extraction and scoring workflow could in principle be extended beyond the 20 historical event dates, since daily R- and C-factor rasters can be generated for a much longer period. Because a daily point layer covering several years across the full watershed inventory would generate tens of millions of records, it would be more practical to retain the full daily output as a tabular (CSV) dataset and export point shapefiles only for active-rainfall days or other dates of specific interest. Applying the model to a date with no reported muddy flood is precisely the intended operational use, not an edge case to be avoided: on a genuinely dry or low-rainfall day, the rainfall-activation gate forces every watershed to inactive/very-low regardless of its relative R×C rank, so the model does not manufacture a spurious 'highest-risk' watershed out of noise on days too dry for muddy flooding to be physically plausible. On a day with erosive rainfall but no reported event, in contrast, the model is expected to behave exactly as it does on the 14 retained training/validation dates: it produces a within-date ranking of relative risk, without asserting that a muddy flood will occur at the top-ranked locations. This is a direct consequence of the positive-unlabelled framing , the model was never trained to output a calibrated occurrence probability, only a relative prioritisation conditional on rainfall being active , and it has not been separately validated against independent, unreported rainfall dates, since no independent ground truth exists for such dates in the current dataset .

Based on the validation results, the rainfall-constrained PU-RF model, combining mean rainfall erosivity, the 95th-percentile RUSLE-based erosion rate, and the 95th-percentile cover-management factor, is recommended as the primary analytical model and the main cartographic variable, since it produced the stronger operational top-10% result of the two leading candidates identified . The rank-based rainfall–cover interaction CLR remains a useful, more directly interpretable secondary benchmark , and its coefficients can still be inspected to sanity-check that the PU-RF's learned relationships are physically consistent, but it is not the recommended cartographic variable. Because absolute probabilities are not well supported by this modelling structure, relative risk percentiles are recommended for stakeholder communication instead of probability values, for example: Low (bottom 50%), Moderate (50–80%), High (80–95%), Very High (top 5%), consistent with the five risk classes defined for the PU-RF-based map. Although sediment output did not improve the statistical model, it retains value as a physically meaningful process indicator, and a bivariate cartographic approach is recommended, using point colour for the PU-RF risk score and point size for sediment-delivery intensity.

6.8 Toward a True Muddy-Flood Risk Map

The results are best interpreted as a relative, event-based muddy-flood risk ranking, rather than a fully predictive probability model. Although the framework combines dynamic rainfall erosivity, daily-updated seasonal cover information, and sediment-output metrics, the validation metrics show that these predictors only partly explain reported muddy-flood events. This is expected, given that muddy floods are governed by several short-lived, local processes , rainfall intensity, runoff generation, soil crusting, crop growth stage, surface roughness, flow concentration along roads and ditches, and the presence of exposed infrastructure at the outlet , that are only partly captured by daily rainfall totals, Sentinel-derived vegetation cover, and RUSLE/WaTEM-SEDEM sediment proxies. This does not invalidate the workflow, but it does change how its output should be read: as an indication of which watersheds are relatively more likely to be problematic under given rainfall and cover conditions, which remains useful for prioritising attention during or after an erosive rainfall event.

Several directions could move the framework closer to a true muddy-flood risk map.

Sub-daily rainfall intensity

Muddy floods are usually triggered by short, intense convective storms; two days with identical daily rainfall totals can have very different muddy-flood potential depending on whether the rain fell gradually or within a 20–30-minute burst. A more complete event-based hazard model would ideally use hourly rainfall, radar rainfall, or maximum 30-minute intensity rather than daily totals. This is consistent with De Walque et al.'s (2017) observation that rainfall intensity is an important muddy-flood-hazard variable, conditional on knowing the precise event date and having rainfall data close to the affected site.

Antecedent soil moisture and runoff generation

RUSLE-type approaches estimate erosion potential but do not explicitly represent runoff generation; the same rainfall falling on dry, rough soil versus already-wet or crusted soil can produce very different runoff volumes. A related and mechanistically distinct gap is soil compaction from agricultural machinery, which reduces infiltration capacity independently of surface crusting and is not represented by the C-factor at all: Boardman and Vandaele (2025) report infiltration falling from around 50 mm/hr on uncrusted soil to as low as 2 mm/hr on the same soil once crusted, and note that compaction from heavy vehicles operating on wet soil, particularly post-harvest, is a further, separate driver of runoff generation. The present workflow has no compaction proxy, so any compaction-driven runoff is implicitly absorbed into the rainfall-erosivity term rather than attributed to its actual cause. Adding antecedent rainfall, a soil-moisture proxy, or a simple curve-number-style runoff index would strengthen the physical basis of the event trigger and make the model less dependent on RUSLE alone.

Drainage and flow-concentration features

Roads, ditches, sunken lanes, culverts, and sewer inlets strongly influence whether runoff and sediment actually reach exposed infrastructure. In the current workflow, PICC infrastructure is mainly used to define pour points and exposure locations; a more complete model would also represent how this infrastructure modifies flow connectivity. De Walque et al. (2017) similarly found that artificial-surface characteristics and sediment connectivity were important muddy-flood-hazard predictors, including the proportion, aggregation, and proximity of artificial surfaces and a sediment-connectivity index, suggesting that a future version of this model should ask not only whether a road or building is exposed, but whether roads, ditches, or drainage networks actively connect sediment sources to the outlet. The municipality of Voeren, in the Flemish part of Belgium, illustrates what acting on this connectivity looks like in practice: following a documented rise in muddy flooding linked to conversion of protective grassland to arable land on steep slopes, ten retention structures were installed over the past decade at points identified as critical in the local drainage network, and these have proved effective in the zones where they were installed, though flooding continued in areas the plan had not yet reached (Boardman & Vandaele, 2025). This is broadly the logic the present framework is built toward, applied at a single-catchment scale with local knowledge of the drainage network rather than a Wallonia-wide model, and it suggests that a future connectivity-aware version of the present workflow could be used similarly to prioritise where such structures would be most effective.

A clearer separation of susceptibility, hazard, and risk

The current model combines elements of all three. Susceptibility describes where the landscape is prone to muddy flooding given a triggering rainfall; hazard describes the event-specific likelihood under a given rainfall and cover condition; risk additionally requires exposure and potential consequences (buildings, roads, gardens, economic assets, critical infrastructure). A future version should therefore include an explicit exposure/vulnerability layer, so that, for example, a watershed with high sediment output but no exposed infrastructure registers high hazard but lower risk, while a watershed with moderate sediment output directly connected to buildings or roads registers higher risk.

A two-step classification

The first step should remain a rainfall-activation trigger, so that watersheds are never classified as high risk on days with negligible rainfall; the second step should apply the CLR or PU-RF score to rank watersheds only once that trigger is passed. This is more defensible than using the relative model score alone, since a purely relative ranking will always produce a 'highest-ranked' watershed, even on days when rainfall is too weak for muddy flooding to occur.

Moving beyond RUSLE/WaTEM-SEDEM as the central event predictor

RUSLE and WaTEM/SEDEM are useful for deriving sediment-production and -delivery proxies, but they are not complete muddy-flood simulators: they do not fully represent rainfall intensity, infiltration-excess runoff, temporary ponding, road drainage, sewer interception, or sediment deposition at infrastructure. In an operational setting, their outputs should be treated as physically informed predictors rather than as final predictions; a more complete event-based model would couple rainfall intensity, runoff generation, sediment detachment, routing, deposition, and exposure explicitly.

Improved validation data

Because reported muddy-flood data are inherently biased toward damaging events, the unreported class cannot be treated as a true absence class, a limitation that cannot be fully removed without independent, systematic field observation. It can, however, be made more transparent: for the present data, PR-AUC and top-5/10/20% capture remain more relevant than absolute probability accuracy, since the model is intended for hotspot prioritisation. A future validation dataset incorporating independent event reports, field observations of deposits, road-cleaning records, emergency calls, photographs, or municipal intervention records could improve confidence that high-ranked watersheds correspond to genuine impact locations, even though some reporting bias would likely remain.

Overall, the workflow developed here is best presented as a process-informed, relative risk-ranking framework, rather than a deterministic muddy-flood prediction model. It improves on static, long-term susceptibility mapping by incorporating event-specific rainfall, dynamic vegetation cover, and sediment-delivery information; to become a fully operational muddy-flood risk map, it would additionally require sub-daily rainfall intensity, runoff and soil-moisture information, crop-specific surface-condition data, explicit drainage-connectivity representation, and a clearer exposure/vulnerability component.

6.9 Synthesis: Answering the Research Objectives

The results can be summarised in relation to the three research objectives defined in Chapter 3.

The first objective was to compare a Positive-Unlabelled Random Forest approach with conventional approaches that treat background locations as true non-events. The PU-RF is the more appropriate machine-learning formulation for this dataset because the non-reported watersheds are not confirmed absences. Its performance is not strong enough to support calibrated occurrence probabilities, but it provides a defensible prioritisation benchmark under uncertain absence conditions. Conventional Random Forest variants are less conceptually appropriate because they require treating the full background set as negative, which is not supported by the reported-event inventory.

The second objective was to identify which watershed-level predictors best distinguish reported muddy-flood locations from background locations. The best rainfall-constrained PU-RF model selected mean rainfall erosivity, the 95th-percentile RUSLE-based erosion rate, and the 95th-percentile cover-management factor. These variables correspond to the main process chain expected for muddy-flood generation: an erosive rainfall trigger, high-end erosion intensity in the upstream contributing area, and high-end cover-management vulnerability. The repeated appearance of these variables in both the rainfall-constrained and standard PU-RF searches suggests that the result is physically interpretable rather than arbitrary, and is consistent with the process understanding established in Chapter 2.

The third objective was to test whether the resulting ranking is useful for regional prioritisation compared with a random baseline. The best PU-RF model captured 15 of 36 reported cases in the top 10% of ranked watersheds per date, compared with about 3.6 cases expected under random ranking. This corresponds to a more than fourfold enrichment over random selection. Expanding the selected class to the top 20% captured only two additional cases, while doubling the number of selected watershed-date observations.

Therefore, the top 10% threshold is the most efficient operational priority class, whereas the top 20% should be treated only as a broader, lower-confidence screening zone.

Overall, the thesis questions are answered positively but cautiously. Dynamic rainfall erosivity, cover vulnerability, and erosion intensity can improve the prioritisation of exposed watersheds relative to random selection and simpler predictor families. However, the low PR-AUC and uneven date-level performance show that the method should be interpreted as a relative risk-prioritisation framework, not as a deterministic or probability-calibrated muddy-flood prediction model.

7. Conclusion

Muddy floods are complex hazards driven by the interaction of rainfall, land-surface conditions, and hydrological connectivity. Recognizing that deterministic prediction is hindered by data limitations, this thesis developed a dynamic, process-informed framework to prioritise 11,122 agricultural watersheds based on relative risk. By integrating RUSLE-based erosion modelling, WaTEM/SEDEM sediment routing, daily rainfall erosivity, and Sentinel-2 vegetation dynamics, the framework successfully moves beyond static susceptibility mapping to provide date-specific hazard prioritisation.

Because historical muddy-flood inventories suffer from severe reporting bias—where the absence of a report does not guarantee the absence of a flood—a Positive-Unlabelled Random Forest (PU-RF) algorithm was utilized. This approach proved theoretically and practically superior to conventional classifiers by treating the data as confirmed positives against an unlabelled background. The PU-RF model consistently identified three paramount predictors: mean rainfall erosivity, the 95th-percentile RUSLE erosion rate, and the 95th-percentile cover-management (C) factor. The selection of these variables confirms that risk is best assessed through a physically meaningful process chain linking the meteorological trigger, hazard intensity, and dynamic surface vulnerability. Crucially, the inclusion of temporal rainfall and vegetation data allows the exact same watershed to receive dynamically updated risk scores based on changing daily environmental conditions.

While standard statistical classification metrics were modest (ROC-AUC = 0.620; PR-AUC = 0.024), the model demonstrated strong operational utility. Under temporal validation, prioritizing just the top 10% of ranked watersheds successfully captured 41.7% of all reported cases, representing an enrichment factor of 4.17 compared to random selection. Expanding this threshold to the top 20% yielded negligible gains, confirming the top 10% as the most efficient operational boundary for targeting limited inspection or mitigation resources. However, strict spatial block validation reduced this capture rate to 16.7%, indicating that while the model excels at temporal prediction within known susceptible areas, its geographic transferability to entirely unseen regions remains constrained by persistent local characteristics.

Several limitations must be considered, primarily the coarse 5 km resolution of the available rainfall data, the physical simplifications of D8 flow routing, and the inherent reporting biases of the inventory. Additionally, the current model evaluates physical hazard without an explicit socioeconomic exposure component. Future research should integrate high-resolution radar precipitation, explicit municipal drainage routing, and independent validation datasets—such as municipal cleanup records—to fully separate physical hazard from societal consequence. Nevertheless, this thesis successfully demonstrates that combining dynamic environmental forcing with Positive-Unlabelled machine learning provides a robust method for prioritizing muddy-flood risk, establishing a practical, date-specific foundation for future early-warning systems and targeted catchment management.

References

Almagro, A., Thomé, T. C., Colman, C. B., Pereira, R. B., & Marques, J. F. (2019). Improving cover and management factor (C-factor) estimation using remote sensing approaches for tropical regions. International Soil and Water Conservation Research, 7(4), 325–334. https://doi.org/10.1016/j.iswcr.2019.08.005

Almagro, A., Thômé, T.C., Colman, C.B., Pereira, R.B., Marcato Junior, J., Rodrigues, D.B.B., & Oliveira, P.T.S. (2019). Improving cover and management factor (C-factor) estimation using remote sensing approaches for tropical regions. International Soil and Water Conservation Research, 7(4), 325–334.

Arabameri, A., Chen, W., Loche, M., Zhao, X., Li, Y., Lombardo, L., Cerda, A., Pradhan, B., & Bui, D.T. (2020). Comparison of machine learning models for gully erosion susceptibility mapping. Geoscience Frontiers, 11(5), 1609–1620.

Ayalew, D. A., Deumlich, D., Šarapatka, B., & Doktor, D. (2020). Quantifying the sensitivity of NDVI-based C factor estimation and potential soil erosion prediction using spaceborne earth observation data. Remote Sensing, 12(7), 1136. https://doi.org/10.3390/rs12071136

Baartman, J.E.M., Nunes, J.P., Masselink, R., Darboux, F., Bielders, C., Degré, A., Cantreul, V., Cerdan, O., Grangeon, T., Fiener, P., Wilken, F., Schindewolf, M., & Wainwright, J. (2020). What do models tell us about water and sediment connectivity? Geomorphology, 367, 107300.

Bai, Y., & Cui, H. (2021). An improved vegetation cover and management factor for RUSLE model in prediction of soil erosion. Environmental Science and Pollution Research, 28, 20421–20435.

Beaumont, B., et al. (2021). Creating Wallonia's new very high resolution land cover maps combining GRASS GIS, OBIA and OTB pixel-based results.

Belda, S., Pipia, L., Morcillo-Palléres, P., Rivera-Caicedo, J.P., Amin, E., De Grave, C., & Verrelst, J. (2020). DATimeS: A machine learning time series GUI toolbox for gap-filling and vegetation phenology trends detection. Environmental Modelling & Software, 127, 104666.

Benito, G., Grodek, T., & Enzel, Y. (1998). The geomorphic and hydrologic impacts of the catastrophic failure of flood-control-dams during the 1996-Biescas flood (Central Pyrenees, Spain). Zeitschrift für Geomorphologie, 42(4), 417–437.

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: A review of existing applications and future research directions. Hydrology and Earth System Sciences, 26(16), 4345–4378.

Bielders, C. L., Ramelot, C., & Persoons, E. (2003). Farmer perception of runoff and erosion and extent of flooding in the silt-loam belt of the Belgian Walloon Region. Environmental Science & Policy, 6(1), 85–93. https://doi.org/10.1016/S1462-9011(02)00121-7

Bielders, C.L., Ramelot, C., & Persoons, E. (2003). Farmer perception of runoff and erosion and extent of flooding in the silt-loam belt of the Belgian Walloon Region. Environmental Science & Policy, 6, 85–93.

Boardman, J. (2010). A short history of muddy floods. Land Degradation & Development, 21, 303–309.

Boardman, J., & Vandaele, K. (2020). Managing muddy floods: Balancing engineered and alternative approaches. Journal of Flood Risk Management, 13(1), e12578. https://doi.org/10.1111/jfr3.12578

Boardman, J., & Vandaele, K. (2025). Muddy flooding: A neglected environmental hazard in western Europe. Geography, 110(2), 81–91. https://doi.org/10.1080/00167487.2025.2506301

Boardman, J., Evans, R., & Ford, J. (2003). Muddy floods on the South Downs, southern England: Problem and responses. Environmental Science & Policy, 6(1), 69–83.

Boardman, J., Ligneau, L., De Roo, A., & Vandaele, K. (1994). Flooding of property by runoff from agricultural land in northwest Europe. Geomorphology, 10, 183–196.

Boardman, J., Ligneau, L., De Roo, A., & Vandaele, K. (1994). Flooding of property by runoff from agricultural land in northwestern Europe. Geomorphology, 10(1–4), 183–196.

Boardman, J., Vandaele, K., Evans, R., & Foster, I. D. L. (2019). Off-site impacts of soil erosion and runoff: Why connectivity is more important than erosion rates. Soil Use and Management, 35(2), 245–256. https://doi.org/10.1111/sum.12496

Boardman, J., Verstraeten, G., & Bielders, C. (2006). Muddy floods. In J. Boardman & J. Poesen (Eds.), Soil Erosion in Europe (pp. 743–755). John Wiley & Sons.

Borselli, L., Cassi, P., & Torri, D. (2008). Prolegomena to sediment and flow connectivity in the landscape: A GIS and field numerical assessment. Catena, 75(3), 268–277.

Cantreul, V., Bielders, C. L., Calsamiglia, A., & Degré, A. (2018). How pixel size affects a sediment connectivity index in central Belgium. Earth Surface Processes and Landforms, 43(4), 884–893. https://doi.org/10.1002/esp.4295

Cantreul, V., Bielders, C., Calsamiglia, A., & Degré, A. (2018). How pixel size affects a sediment connectivity index in central Belgium. Earth Surface Processes and Landforms, 43(4), 884–893.

Cerdan, O., Le Bissonnais, Y., Couturier, A., & Saby, N. (2001/2002). The STREAM expert-based runoff and erosion model.

Cerdan, O., Le Bissonnais, Y., Souchère, V., Martin, P., & Lecomte, V. (2002). Sediment concentration in interrill flow: Interactions between soil surface conditions, vegetation and rainfall. Earth Surface Processes and Landforms, 27(2), 193–205.

Ciampalini, R., Follain, S., & Le Bissonnais, Y. (2012). LandSoil: A model for analysing the impact of erosion on agricultural landscape evolution. Catena / related outputs.

De Roo, A.P.J. (2000). Applying the LISEM model for investigating flood prevention and soil conservation scenarios in South-Limburg, the Netherlands. In Soil Erosion Research Methods (2nd ed.) or associated proceedings.

De Roo, A.P.J., Offermans, R.J.E., & Cremers, N.H.D.T. (1996b). LISEM: A single-event physically based hydrological and soil erosion model for drainage basins. II: Sensitivity analysis, validation and application. Hydrological Processes, 10(8), 1119–1126.

De Roo, A.P.J., Wesseling, C.G., & Ritsema, C.J. (1996a). LISEM: A single-event physically based hydrological and soil erosion model for drainage basins. I: Theory, input and output. Hydrological Processes, 10(8), 1107–1117.

De Walque, B., Degré, A., Maugnard, A., & Bielders, C. L. (2017). Artificial surfaces characteristics and sediment connectivity explain muddy flood hazard in Wallonia. Catena, 158, 89–101. https://doi.org/10.1016/j.catena.2017.06.016

De Walque, B., Degré, A., Maugnard, A., & Bielders, C.L. (2017). Artificial surfaces characteristics and sediment connectivity explain muddy flood hazard in Wallonia. Catena, 158, 89–101.

Desmet, P. J. J., & Govers, G. (1996). A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. Journal of Soil and Water Conservation, 51(5), 427–433.

Desmet, P.J.J., & Govers, G. (1996). A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. Journal of Soil and Water Conservation, 51(5), 427–433.

Douvinet, J., Delahaye, D., & Cavalié, M. (2015). A flash flood hazard assessment in dry valleys (northern France) by cellular automata modelling. Natural Hazards, 75(2), 1601–1631.

Elkan, C., & Noto, K. (2008). Learning classifiers from only positive and unlabeled data. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 213–220). Association for Computing Machinery. https://doi.org/10.1145/1401890.1401920

Evans, R., & Boardman, J. (2003). Curtailment of muddy floods in the Sompting catchment, South Downs, West Sussex, southern England. Soil Use and Management, 19(3), 223–231.

Evrard, O., Bielders, C. L., Vandaele, K., & van Wesemael, B. (2007). Spatial and temporal variation of muddy floods in central Belgium, off-site impacts and potential control measures. Catena, 70(3), 443–454. https://doi.org/10.1016/j.catena.2006.11.011

Evrard, O., Bielders, C.L., Vandaele, K., & van Wesemael, B. (2007a). Spatial and temporal variation of muddy floods in central Belgium, off-site impacts and potential control measures. Catena, 70(3), 443–454.

Evrard, O., Cerdan, O., van Wesemael, B., Chauvet, M., Le Bissonnais, Y., Raclot, D., Vandaele, K., Andrieux, P., & Bielders, C. (2009). Reliability of an expert-based runoff and erosion model: Application of STREAM to different environments. Catena, 78(2), 129–141.

Evrard, O., Persoons, E., Vandaele, K., & van Wesemael, B. (2007). Effectiveness of erosion mitigation measures to prevent muddy floods: A case study in the Belgian loam belt. Agriculture, Ecosystems & Environment, 118(1–4), 149–158.

Evrard, O., Vandaele, K., Bielders, C. L., & van Wesemael, B. (2008). Seasonal evolution of runoff generation on agricultural land in the Belgian loess belt and implications for muddy flood triggering. Earth Surface Processes and Landforms, 33(9), 1285–1301. https://doi.org/10.1002/esp.1613

Evrard, O., Vandaele, K., van Wesemael, B., & Bielders, C.L. (2008). Seasonal evolution of runoff generation on agricultural land in the Belgian loess belt and implications for muddy flood triggering. Earth Surface Processes and Landforms, 33(8), 1285–1301.

Evrard, O., Vandaele, K., van Wesemael, B., & Bielders, C.L. (2008b). A grassed waterway and earthen dams to control muddy floods from a cultivated catchment of the Belgian loess belt.

Fabre, G. (1989). Les inondations du 3 octobre 1988 à Nîmes. Hydrogéologie, causes et conséquences. Revue de Géographie de Lyon, 64(3), 197–202.

Fernandes, R., Brown, L., Canisius, F., Dash, J., He, L., Hong, G., Huang, L., Le, N.Q., MacDougall, C., Meier, C., Darko, P.O., Shah, H., Spafford, L., & Sun, L. (2023). Validation of Simplified Level 2 Prototype Processor Sentinel-2 fraction of canopy cover, fraction of absorbed photosynthetically active radiation and leaf area index products over North American forests. Remote Sensing of Environment, 293, 113600.

Hänsel, P., Douinot, A., & Bronstert, A. (2019). Prediction of muddy floods using high-resolution radar precipitation forecasts and physically-based erosion modeling in agricultural landscapes. Geosciences, 9(7), 302.

Hänsel, P., Kaiser, A., Buchholz, A., Böttcher, F., Langel, S., Schmidt, J., & Schindewolf, M. (2018). Mud flow reconstruction by means of physical erosion modeling, high-resolution radar-based precipitation data, and UAV monitoring. Geosciences, 8(11), 427.

Hänsel, P., Langel, S., Schindewolf, M., Kaiser, A., Buchholz, A., Böttcher, F., & Schmidt, J. (2019). Prediction of muddy floods using high-resolution radar precipitation forecasts and physically-based erosion modeling in agricultural landscapes. Geosciences, 9(9), 401.

IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (V. Masson-Delmotte et al., Eds.). Cambridge University Press.

Kamenova, I., & Dimitrov, P. (2021). Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCover of winter wheat in Bulgaria. European Journal of Remote Sensing, 54(sup1), 89–108.

Landemaine, V. (2016). Modélisation du ruissellement et de l'érosion des sols à l'échelle périannuelle et pluriannuelle: le modèle WaterSed. PhD thesis / associated publications.

Mahdizadeh Gharakhanlou, N., & Perez, L. (2023). Flood susceptible prediction through the use of geospatial variables and machine learning methods. Journal of Hydrology, 617, 129121.

Matthews, F., Kasprzak, A., & Verstraeten, G. (2025). Mapping the soil erosion risk in Flanders: A remote sensing based assessment of the crop management factor (C-factor). Study commissioned by the Departement Omgeving, 81 pp.

Maugnard, A., Bielders, C., & Degré, A. (2022). Cartographie du risque de dégradation des sols et de transfert des pollutions agricoles diffuses aux cours d'eau, Lot 1: Mise à jour des données. Service Public de Wallonie.

Maugnard, A., Bielders, C., & Degré, A. (2022). SPW soil-erosion mapping methodology (Lot 1).

MUDCAP (2026–2029). MUDdy flood mitigation through transnational action-oriented CAPacity building. Interreg North-West Europe programme. https://mudcap.nweurope.eu.

Nearing, M. A. (1997). A single, continuous function for slope steepness influence on soil loss. Soil Science Society of America Journal, 61(3), 917–919. https://doi.org/10.2136/sssaj1997.03615995006100030029x

Nearing, M.A. (1997). A single continuous function for slope steepness influence on soil loss. Soil Science Society of America Journal, 61, 917–919.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190(3-4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. Ecography, 40(7), 887–893. https://doi.org/10.1111/ecog.03049

Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., & Yoder, D. C. (1997). Predicting soil erosion by water: A guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE) (Agriculture Handbook No. 703). U.S. Department of Agriculture.

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

Schmidt, J., von Werner, M., & Michael, A. (1996/1999). EROSION 3D: A process-based model for simulating runoff and soil erosion at the catchment scale.

Souchère, V., Cerdan, O., Dubreuil, N., Le Bissonnais, Y., & King, C. (2005). Modelling the impact of agri-environmental scenarios on runoff in a cultivated catchment (Normandy, France). Catena, 61(2–3), 229–240.

SPW Géoportail de la Wallonie. (n.d.). Projet Informatique de Cartographie Continue (PICC), metadata catalogue entry. Service Public de Wallonie. https://geoportail.wallonie.be/catalogue/b795de68-726c-4bdf-a62a-a42686aa5b6f.html

Takken, I., Govers, G., Jetten, V., Nachtergaele, J., Steegen, A., & Poesen, J. (2001). Effects of tillage on runoff and erosion patterns. Soil and Tillage Research, 61(1–2), 55–60.

Tanyaş, H., Kolat, Ç., & Süzen, M. L. (2015). A new approach to estimate cover-management factor of RUSLE using Google Earth Engine. Soil and Tillage Research, 148, 62–72.

USDA-ARS. (2013). Science Documentation: Revised Universal Soil Loss Equation, Version 2 (RUSLE2). USDA Agricultural Research Service.

USDA-SCS. (1972). National Engineering Handbook, Section 4: Hydrology. USDA Soil Conservation Service.

Van Campenhout, J., Verstraeten, G., & Poesen, J. (2015). Flash floods and muddy floods in Wallonia: Recent temporal trends, spatial distribution and reconstruction of the hydrosedimentological fluxes using flood-based indicators. Belgeo, 2015(4).

Van der Knijff, J. M., Jones, R. J. A., & Montanarella, L. (2000). Soil erosion risk assessment in Europe (EUR 19044 EN). European Soil Bureau, Joint Research Centre.

Van der Knijff, J.M., Jones, R.J.A., & Montanarella, L. (1999). Soil erosion risk assessment in Italy. European Soil Bureau.

Van Oost, K. (2003). Spatial modelling of soil redistribution processes in agricultural landscapes .

Van Oost, K., Govers, G., & Desmet, P. (2000). Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. Landscape Ecology, 15(6), 577–589.

Van Oost, K., Govers, G., & Desmet, P.J.J. (2000). Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. Landscape Ecology, 15(6), 579–591.

Van Oost, K., Govers, G., & Desmet, P.J.J. (2003). Evaluating the effects of grid cell size and slope-length correction on the calculation of the LS-factor for use in WaTEM/SEDEM.

Van Rompaey, A. J. J., Verstraeten, G., Van Oost, K., Govers, G., & Poesen, J. (2001). Modelling mean annual sediment yield using a distributed approach. Earth Surface Processes and Landforms, 26(11), 1221–1236. https://doi.org/10.1002/esp.275

Van Rompaey, A.J.J., Verstraeten, G., Van Oost, K., Govers, G., & Poesen, J. (2001). Modelling mean annual sediment yield using a distributed approach. Earth Surface Processes and Landforms, 26, 1221–1236.

Vandaele, K. (2020). Development of soil erosion control regulations in Flanders (Belgium) based on the effectiveness and practicability of soil erosion control scenarios.

Verstraeten, G., & Poesen, J. (1999). The nature of small-scale flooding, muddy floods and retention pond sedimentation in central Belgium. Geomorphology, 29(3–4), 275–292. https://doi.org/10.1016/S0169-555X(99)00020-3

Verstraeten, G., Poesen, J., Demarée, G., & Salles, C. (2006). Long-term (105 years) variability in rain erosivity as derived from 10-min rainfall depth data for Ukkel (Belgium): Implications for assessing soil erosion rates. Journal of Geophysical Research, 111(D22109). https://doi.org/10.1029/2006JD007169

Verstraeten, G., Poesen, J., Demarée, G., & Salles, C. (2006). Long-term (105 years) variability in rain erosivity as derived from 10-min rainfall depth data for Ukkel (Brussels, Belgium): Implications for assessing soil erosion rates. Journal of Geophysical Research: Atmospheres, 111, D22109.

Verstraeten, G., Poesen, J., Govers, G., Gillijns, K., Van Rompaey, A., & Van Oost, K. (2003). Integrating science, policy and farmers to reduce soil loss and sediment delivery in Flanders, Belgium. Environmental Science & Policy, 6, 95–103.

Verstraeten, G., Van Oost, K., Van Rompaey, A., Poesen, J., & Govers, G. (2002). Evaluating an integrated approach to catchment management to reduce soil loss and sediment pollution through modelling. Soil Use and Management, 18(4), 386–394.

Weiss, M., & Baret, F. (2016). S2ToolBox Level 2 products: LAI, FAPAR, FCOVER (Version 1.1). Algorithm Theoretical Basis Document, European Space Agency.

Wischmeier, W. H., & Smith, D. D. (1978). Predicting rainfall erosion losses: A guide to conservation planning (Agriculture Handbook No. 537). U.S. Department of Agriculture.

Yan, H., Zhan, J., Zhang, T., Li, Z., Fan, S., Sun, W., & Yu, X. (2020). A synthesized approach for estimating the C-factor of RUSLE for a mixed-landscape watershed: A case study in the Gongshui watershed, southern China. Agriculture, Ecosystems & Environment, 301, 107066.

SUMMARY

Muddy floods are sediment-laden runoff events that occur when rainfall interacts with vulnerable agricultural surfaces and transports eroded soil towards downstream infrastructure. They are an important hazard in Wallonia, where cultivated loess soils, agricultural land use, rainfall events, and connections between hillslopes and roads or settlements create conditions favourable to muddy flooding. Predicting these events at regional scale is difficult because they depend on both relatively persistent landscape characteristics and rapidly changing rainfall and vegetation conditions. A further challenge is that available muddy-flood inventories contain reported events but cannot reliably identify true non-events. This thesis therefore investigates whether dynamic environmental information and a positive-unlabelled modelling approach can improve the regional prioritisation of exposed locations compared with conventional static erosion-based approaches.

The study developed a watershed-based framework for 11,122 representative exposed watersheds in Wallonia. PICC infrastructure data were used to identify representative flow-axis pour points, from which upstream contributing watersheds were delineated. Candidate predictors combined rainfall erosivity, Sentinel-2-derived vegetation information, RUSLE-based erosion indicators, WaTEM/SEDEM sediment-output information, terrain variables, and hydrological connectivity. The modelling dataset was constructed around reported muddy-flood event dates, with reported locations treated as confirmed positive cases and non-reported observations treated as unlabelled background rather than confirmed absences. Several model families were compared, including conventional Random Forest, Positive-Unlabelled Random Forest (PU-RF), conditional logistic regression, sediment-output models, rank-based models, and MaxEnt. Validation relied primarily on leave-one-date-out cross-validation, with additional spatial validation used to test geographic transferability.

The final rainfall-constrained PU-RF model selected three predictors: mean rainfall erosivity, the 95th-percentile RUSLE-based erosion rate, and the 95th-percentile cover-management factor. The model was evaluated on 14 event dates containing 1,904 watershed-date observations, of which 36 were reported cases. Conventional discrimination metrics were modest, with a ROC-AUC of 0.620 and PR-AUC of 0.024. However, the model was more useful when evaluated as a prioritisation tool: it captured 15 of the 36 reported cases within the top 10% of ranked watersheds for each date, corresponding to 41.7% of reported cases and an enrichment of more than four times relative to random selection. Expanding the priority area to the top 20% captured only two additional cases, indicating that the top 10% represents the more efficient operational priority class.

Spatial validation showed an important limitation. While temporal validation produced 41.7% top-10% capture, this decreased to 25.0% under nested spatial-within-date validation and 16.7% under pure spatial block validation. This indicates that part of the observed prioritisation skill is associated with persistent local landscape characteristics and that transferability to geographically unseen areas remains limited. Sediment-output variables provided physically relevant information but did not improve the final combined model.

Overall, the thesis demonstrates that a dynamic, watershed-based approach combining rainfall erosivity, vegetation-cover vulnerability, erosion information, connectivity, and Positive-Unlabelled learning can improve the concentration of reported muddy-flood cases into a relatively small set of priority locations. The resulting framework should not be interpreted as a calibrated probability model or deterministic prediction of muddy-flood occurrence. Its main value is as a relative, event-specific risk-prioritisation tool that can support the screening of exposed infrastructure, field inspection, and future mitigation planning. Further development would require higher-temporal-resolution rainfall, improved information on soil moisture and runoff generation, explicit drainage connectivity, and more systematic independent muddy-flood observations.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des bioingénieurs
Croix du Sud, 2 bte L7.05.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/agro**