University of Pavia

Department of Civil engineering

Master of Science in Environmental Engineering-Resilience to Climate Change Effects (REACH)

Cross-Biome Stability of Feature Importance Rankings in Random Forest-Based SAR Flood Mapping

Author

Parisa Havakhor

Student ID: 531433

Supervisor

Prof. Paolo Ettore Gamba

Co-Supervisor

Prof. Dr. Antara Dasgupta

RWTH Aachen University

Academic Year 2025/2026

Abstract

Flood mapping with machine-learning classifiers such as Random Forest (RF) depends on informed feature selection, yet the stability of feature-importance rankings across contrasting environments has not been systematically evaluated for RF-based flood delineation. This study investigates the consistency of RF feature-importance rankings in a binary flood-classification task based primarily on Synthetic Aperture Radar (SAR) imagery. A stack of thirteen features, nine SAR-derived variables (Sentinel-1 VV and VH polarizations, their polarimetric ratios, and temporal-change metrics) and four contextual variables (Slope, Height Above the Nearest Drainage [HAND], Euclidean Distance to Water [EDTW], and land cover) was assembled for fifteen flood events distributed across five global biomes: Deserts and Xeric Shrublands; Tropical and Subtropical Moist Broadleaf Forests; Temperate Broadleaf and Mixed Forests; Temperate Coniferous Forests; and Mediterranean Forests, Woodlands and Scrub. Three feature-attribution methods were compared (Mean Decrease in Impurity [MDI], permutation importance [MDA], and SHapley Additive exPlanations [SHAP]), as well as feature cardinality and correlation, were accounted for during interpretation, with all experiments repeated over ten independent iterations (repeated model runs, each with an independent random seed and training-sample draw). The three attribution methods can rank correlated and low-importance features differently; in this study they agreed on the leading predictor within each biome but diverged for mid-ranked and correlated features, underscoring the value of reporting more than one method. Feature-importance rankings were analyzed first within each biome and then across biomes. Post-event SAR backscatter, particularly VV Post, emerged as the most consistently important and transferable flood predictor, whereas VH Post (though equally important on average) varied strongly between biomes, dominating tropical forests but weakening in coniferous and Mediterranean settings. Crucially, feature importance proved biome-dependent: terrain descriptors outranked SAR backscatter in the Mediterranean (Slope) and Temperate Broadleaf (HAND) biomes, while VH Post and Slope showed the greatest cross-biome variability. These results show that feature-importance patterns do not transfer uniformly across environments, a finding that is central to building reliable, globally applicable RF-based (and, more broadly, machine-learning-based) flood-mapping pipelines and to fostering user trust in the resulting products.

Sommario

La mappatura delle inondazioni mediante classificatori di apprendimento automatico come il Random Forest (RF) dipende da una selezione informata delle feature; tuttavia, la stabilità delle classifiche di importanza delle feature in ambienti contrastanti non è stata valutata in modo sistematico per la delineazione delle inondazioni basata su RF. Questo studio esamina la coerenza delle classifiche di importanza delle feature di RF in un compito di classificazione binaria delle inondazioni basato principalmente su immagini radar ad apertura sintetica (SAR). È stato costruito uno stack di tredici feature nove variabili derivate dal SAR (polarizzazioni VV e VH di Sentinel-1, i relativi rapporti polarimetrici e metriche di variazione temporale) e quattro variabili contestuali (pendenza, altezza sul drenaggio più vicino [HAND], distanza euclidea dall'acqua [EDTW] e copertura del suolo) per quindici eventi alluvionali distribuiti su cinque biomi globali: Deserti e arbusteti xerici; Foreste umide tropicali e subtropicali di latifoglie; Foreste temperate di latifoglie e miste; Foreste temperate di conifere; e Foreste, boschi e macchia mediterranea.

Sono stati confrontati tre metodi di attribuzione dell'importanza (Mean Decrease in Impurity [MDI], importanza per permutazione [MDA] e SHapley Additive exPlanations [SHAP]) e, nell'interpretazione, si è tenuto conto anche della cardinalità e della correlazione delle feature; tutti gli esperimenti sono stati ripetuti su dieci iterazioni indipendenti (esecuzioni ripetute del modello, ciascuna con un seed casuale e un'estrazione del campione di addestramento indipendenti). I tre metodi di attribuzione possono ordinare in modo differente le feature correlate e di bassa importanza; in questo studio essi concordano sul predittore principale all'interno di ciascun bioma, ma divergono per le feature di importanza intermedia e correlate, sottolineando l'utilità di riportare più di un metodo.

Le classifiche di importanza delle feature sono state analizzate dapprima all'interno di ciascun bioma e successivamente tra i diversi biomi. La retrodiffusione (backscatter) SAR post-evento, in particolare VV Post, è risultata il predittore delle inondazioni più costantemente importante e trasferibile, mentre VH Post (pur essendo in media altrettanto importante) ha mostrato forti variazioni tra i biomi, risultando dominante nelle foreste tropicali ma indebolendosi negli ambienti di conifere e mediterranei. Elemento cruciale, l'importanza delle feature si è rivelata dipendente dal bioma: i descrittori del terreno hanno superato la retrodiffusione SAR nei biomi mediterraneo (pendenza) e delle latifoglie temperate (HAND), mentre VH Post e la pendenza hanno mostrato la maggiore variabilità tra i biomi. Questi risultati mostrano che i pattern di importanza delle feature non si trasferiscono in modo uniforme tra i diversi ambienti: un risultato centrale per la costruzione di pipeline di mappatura delle inondazioni affidabili e applicabili su scala globale, basate su RF (e, più in generale, sull'apprendimento automatico), e per rafforzare la fiducia degli utenti nei prodotti che ne derivano.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Paolo Gamba of the University of Pavia, for his continuous guidance, scientific rigor, and constructive feedback throughout the course of this research. His supervision has been fundamental to the development of this thesis, and I am deeply grateful for the knowledge and academic perspective I have gained under his mentorship.

I would also like to extend my heartfelt appreciation to Dr. Antara Dasgupta of RWTH Aachen University, who supervised my work during my Erasmus exchange, in the framework of which this thesis was carried out. Her expert guidance, methodological insight, and consistent support were instrumental to the completion of this study, and I am truly thankful for her dedication and for the opportunity to learn from her expertise.

I gratefully acknowledge the Erasmus+ programme, which enabled me to undertake this research at RWTH Aachen University, and I thank both the University of Pavia and RWTH Aachen University for facilitating this academic collaboration. I further acknowledge the Copernicus Emergency Management Service and the providers of the Sentinel-1 and ancillary datasets employed in this study.

Parisa Havakhor
Pavia, July 2026

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AOI	Area of Interest
CEMSR	Copernicus Emergency Management Service – Rapid Mapping
DEM	Digital Elevation Model
DTM	Digital Terrain Model
EDTW	Euclidean Distance to Water bodies
F1	F1 score (harmonic mean of precision and recall)
GEE	Google Earth Engine
GRD	Ground Range Detected
HAND	Height Above Nearest Drainage
IW	Interferometric Wide (swath mode)
MDA	Mean Decrease in Accuracy (permutation importance)
MDI	Mean Decrease in Impurity
OA	Overall Accuracy
OSM	OpenStreetMap
RF	Random Forest
SAR	Synthetic Aperture Radar
SHAP	SHapley Additive exPlanations

VH	Vertical transmit / Horizontal receive (cross-polarization)
VV	Vertical transmit / Vertical receive (co-polarization)

Chapter 1: Introduction and Motivation

1.1 Background and Motivation

Floods represent the most frequent and economically destructive natural hazards globally, affecting millions of people annually and causing substantial loss of life, infrastructure damage, and agricultural disruption (Tellman et al., 2021). Accurate and timely flood mapping is therefore a critical component of disaster management, supporting emergency response operations, damage assessment, and the development of flood early warning systems. Remote sensing offers a scalable and cost-effective means of mapping flood extent over large areas, particularly in regions where ground-based observations are unavailable or inaccessible during active flood events.

Synthetic Aperture Radar (SAR) has emerged as the primary remote sensing modality for operational flood mapping due to its ability to acquire imagery independently of cloud cover and solar illumination (Moreira et al., 2013). Unlike optical sensors, which are frequently obstructed by the cloud cover that accompanies flood generating precipitation events, SAR systems operate in the microwave spectrum and can penetrate cloud cover to image flood extent in near-real time. The launch of the Copernicus Sentinel-1 constellation has substantially advanced operational SAR-based flood mapping by providing freely available, systematic, dual-polarization C-band imagery with a six-day revisit cycle at global scale (ESA, 2024).

The physical basis of SAR flood detection is the specular reflection of radar signals from calm open water surfaces, which produces strong backscatter reduction in the VV polarization channel compared to surrounding non-flooded terrain (Twele et al., 2016; Shen et al., 2019; Landuyt et al., 2019). In vegetated environments, the cross-polarized VH channel additionally captures double-bounced scattering between floodwater and vegetation stems, providing complementary flood information that improves detection accuracy in forested and agricultural landscapes (Tsyganskaya et al., 2018). Multi-temporal change detection (comparing post-event imagery against a pre-event reference acquisition) further enhances flood discrimination by isolating transient backscatter changes attributable to inundation from land surface characteristics (Chini et al., 2017).

1.2 Machine Learning for SAR-Based Flood Mapping

Machine learning classifiers have become increasingly prevalent in SAR-based flood mapping, offering the ability to integrate heterogeneous multi-source feature sets (combining SAR-derived backscatter features with contextual environmental variables such as topography, hydrology, and land cover) without requiring explicit physical models of the relationships between these variables and flood occurrence (Maxwell et al., 2018). Among the available machine learning approaches, Random Forest (RF) has emerged as one of the most widely applied classifiers in remote sensing applications due to its robustness to overfitting, ability to handle high dimensional feature spaces, and computational efficiency (Breiman, 2001).

RF-based flood mapping systems typically incorporate SAR backscatter features alongside contextual predictors derived from digital terrain models and hydrological datasets. Height Above Nearest Drainage (HAND) (a topographic index quantifying the vertical distance from each pixel to the nearest drainage channel) has been identified as a particularly informative flood susceptibility predictor, as low-lying areas adjacent to drainage networks are inherently more prone to inundation (Rennó et al., 2008). Euclidean Distance to Water Bodies (EDTW), terrain slope, and land cover classification have similarly been incorporated as contextual features that capture physical and landscape controls on flood susceptibility.

Despite the widespread adoption of RF in flood mapping, several methodological challenges remain unresolved. Class imbalance (arising from typically small proportion of flooded pixels relative to non-flooded background) can bias RF classifiers towards the majority class, reducing flood detection sensitivity (He & Garcia, 2009). Training sample design, including the choice of spatial distribution strategy, class distribution strategy, and sample size, has been shown to substantially influence classification performance (Maxwell et al., 2018). Hyperparameter optimization further affects model behavior, with Bayesian optimization offering a principled approach to navigating the RF hyperparameter space more efficiently than grid or random search (Snoek et al., 2012).

1.3 Feature Importance and Model Interpretability

Beyond classification accuracy, the interpretability of machine learning flood mapping models has become an increasingly important consideration for operational adoption and scientific understanding. Feature importance analysis (the quantification of each input variable's contribution to model predictions) provides insights into the physical mechanisms underlying flood classification, supports informed feature selection, and builds user trust in automated mapping systems (Lundberg & Lee, 2017).

The choice of attribution method can substantially influence feature importance rankings, particularly for correlated and low-importance features (Strobl et al., 2007; Altmann et al., 2010). Comparative evaluation of multiple attribution methods is therefore recommended for robust feature interpretation in remote sensing applications (Lundberg et al., 2019). However, systematic comparison of MDI, MDA, and SHAP in the context of SAR-based flood mapping remains limited in technical literature.

1.4 Research Gap and Objectives

A fundamental assumption underlying operational RF-based flood mapping is that a single, standardized feature stack contains the predictors that are informative in every region of the world even if the particular features that rank highest differ from place to place. If this assumption holds, the same RF pipeline can be applied globally, drawing on whichever subset of features is most discriminative in each environment. The degree to which this assumption is valid (that is, whether one feature set transfers across contrasting biomes) is precisely what this study sets out to test.

To start with, different biomes present fundamentally different SAR backscatter environments: arid shrublands exhibit sparse vegetation and exposed bare soil that produce strong backscatter contrast under flood conditions (Garg et al., 2024), while dense tropical forests attenuate C-band SAR signals and produce double-bounce scattering that complicates straightforward flood detection (Tsyganskaya et al., 2018; Grimaldi et al., 2020). Temperate broadleaf forests exhibit seasonal phenological dynamics that alter pre-event reference backscatter, while Mediterranean environments combine heterogeneous land cover with complex terrain (Médail et al., 2019). These biome-specific characteristics may substantially alter the relative importance of SAR-derived and contextual features for flood classification, with implications for the transferability of RF models and the generalizability of feature selection decisions across geographic settings.

It should be noted that feature importance is likely to depend not only on biome but also on flood typology (for example, flash floods in steep catchments versus slow riverine inundation), which influences which physical processes dominate the backscatter signal. This study controls for environmental setting through biome stratification but does not explicitly separate flood types; disentangling the additional effect of flood typology is left for future work.

This research addresses this research gap by systematically evaluating the stability of feature rankings in RF-based SAR flood mapping across five global biomes and fifteen flood events (3 per biome). Specifically, this study aims to answer the following research questions:

RQ1: To what extent do different feature attribution methods (MDI, MDA, and SHAP) agree in their rankings, and how does the choice of attribution method influence conclusions about feature importance stability?

RQ2: Which input features show stable importance rankings within individual biomes (intra-biome stability), and which ones exhibit high variability across flood events within the same biome?

RQ3: How consistent are feature importance rankings across contrasting global biomes (inter-biome stability), and which features constitute universally transferable flood predictors?

1.5 Thesis Structure

Although there are 14 biomes across the world (Olson et al., 2001), this analysis was conducted across fifteen flood events distributed over only five distinct global biomes: Temperate

Broadleaf & Mixed Forests, Deserts & Xeric Shrublands, Mediterranean Forests, Woodlands and Scrub, Temperate Coniferous Forests, Tropical & Subtropical Moist Broadleaf Forests. The reason for this limited number of biomes is twofold. First, these five biomes were deliberately selected to span the widest possible range of SAR backscatter environments (from arid, sparsely vegetated deserts to densely forested tropical and temperate systems) so that any feature whose importance persists across them can be regarded as genuinely transferable rather than site-specific. Second, feature-importance analysis is computationally intensive when repeated across many events, and adding further biomes environmentally similar to those already sampled would substantially increase computation while adding little new information about transferability. For each biome, three flood events were selected to capture variability in hydrological and environmental conditions. Three is the minimum number that allows intra-biome variability to be assessed (it supports a leave-one-site-out cross-site evaluation and a measure of consistency across events) while keeping the total of fifteen events within the limits of available data and the considerable per-event processing cost. The decisive constraint on which events could be included was the completeness of the data required to obtain reliable results: each selected event had to have, all simultaneously available, (i) a reliable ground truth flood delineation from the Copernicus EMS Rapid Mapping catalogue; (ii) a precisely dated flood occurrence, so that the inundation peak could be matched to the imagery; and (iii) a suitable Sentinel-1 acquisition recorded contemporaneously with the event (within a maximum temporal offset of seven hours of the reference mapping) and etc. Events for which any of these elements (ground truth, accurate event timing, or a contemporaneous acquisition by the appropriate sensor and etc) was missing or of insufficient quality were excluded. Among the events meeting these requirements, the three retained per biome were further chosen to span a range of geographic locations, catchment sizes, and event durations, so as to capture genuine intra-biome variability rather than near-duplicate cases. The classification task was formulated as a binary flood mapping problem in which each pixel was labeled as either flooded or non-flooded. A feature stack composed of SAR-derived and contextual variables was assembled for each flood event. Random Forest models were then trained using multiple sampling strategies and a fixed sample size. Feature importance rankings were extracted using three different attribution methods and evaluated for consistency within and across biomes. Moreover, to account for stochastic variability in model training and sampling, all experiments were repeated over ten independent iterations.

The remainder of this thesis is organized as follows. Chapter 2 introduces the study domain, describing the fifteen flood events distributed across five global biomes and the datasets used for SAR imagery, contextual environmental variables, and ground truth flood delineation. Chapter 3 presents the methodological framework, including the RF classification approach, training sample design, feature attribution methods, and experimental design. Chapter 4 presents and discusses the results of the classification performance assessment and the feature importance analysis across biomes. Chapter 5 summarizes the main conclusions of the study and provides recommendations for future research lines.

Chapter 2: Study Area and Datasets

2.1 Study Domain and Flood Events

This study encompasses 15 flood events distributed across five global biomes, selected to represent a broad range of environmental and hydrological conditions. The five biomes were defined following the terrestrial ecoregion classification of Olson et al. (2001) [Figure 2.1].

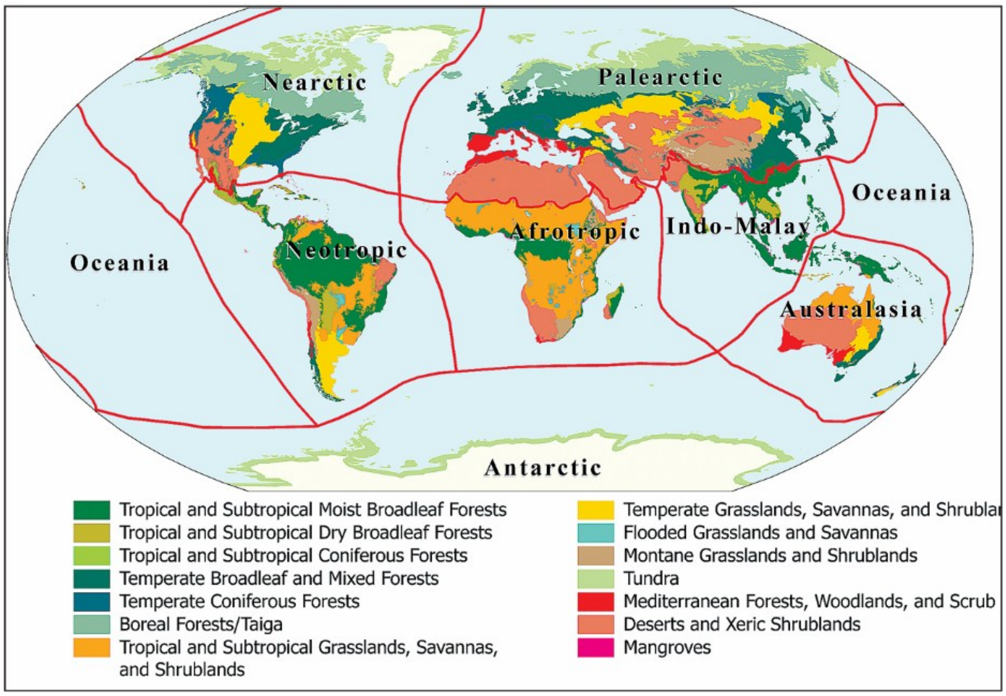

Figure 2.1. Global distribution of terrestrial ecoregions (Olson et al., 2001).

For each biome, three flood events were selected to capture intra-biome variability in hydrological and environmental conditions. All events were sourced from the Copernicus Emergency Management Service Rapid Mapping catalogue (Copernicus EMS, 2025). For events where the primary sensor used by CEMSR was not Sentinel-1, the Copernicus Browser was consulted to verify the availability of a contemporaneous Sentinel-1 acquisition within a maximum temporal offset of seven hours. The reason of this limited offset is that floods are dynamic and rapid phenomena whose boundaries change in a short period of time (Tellman et al., 2021). Seven hours offset have been chosen so that the radar acquisition captures essentially the same inundation state as the Copernicus EMS reference mapping and minimizes differences in flood extent between the two observations. Table 2.1 provides an overview of all fifteen flood events.

As shown in Figure 2.2, the five biomes cover a broad range of ecological and hydrological conditions, each presenting distinct challenges for SAR-based flood mapping:

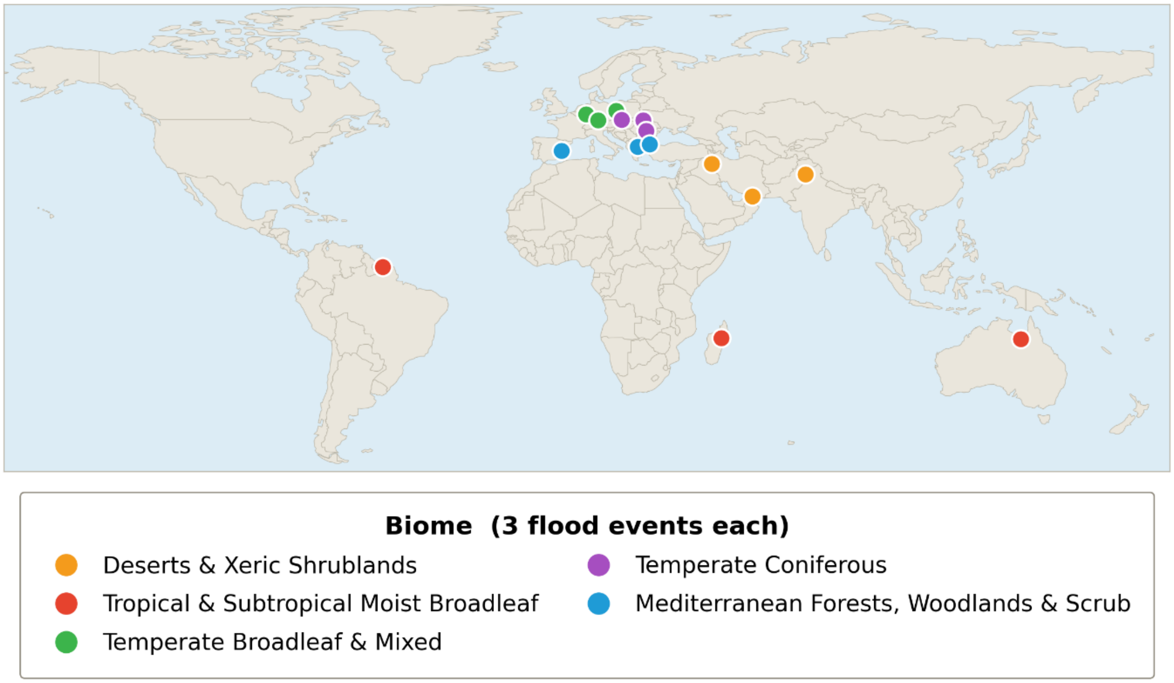

Figure 2.2. Location of all fifteen flood events, colour-coded by biome (three events per biome).

- Temperate Broadleaf and Mixed Forests biome: in Central Europe (represented by the study sites) the mean canopy height ranges from 14.5 m in broad-leaved stands to 21.1 m in middle-leaved forests, with mixed forests averaging 18.5 m (Urbazaev et al., 2022). The dense multi-layer canopy structure strongly influences backscatter through volume scattering and double bounce interaction between floodwater and tree trunks, making flood detection more complex than in open terrain (Tsyganskaya et al., 2018). Three flood events in Central Europe from Germany (Moselle, Stuttgart) and Poland (Wroclaw) represent this biome.

- Deserts and Xeric Shrubland: in arid regions, Synthetic Aperture Radar (SAR) flood detection faces unique challenges because dry sandy surfaces often exhibit low backscatter values similar to water, creating significant ‘sand-water confusion’ (Garg et al., 2024). These difficulties are exacerbated by the nature of episodic flash floods, which are typically rapid and short-lived, posing challenges for both timely satellite acquisition and effective backscatter contrast in complex arid terrains (Garg et al., 2024). Events from Iran (Jask), Pakistan (Rasoo Nagar), and Iraq (Chaq Chaq) represent this biome.

- Mediterranean Forests, Woodlands and Scrubs: The vegetation structure of the Mediterranean-European region is highly diverse and heterogeneous, comprising various types of matorrals (shrublands) and forests (Médail et al., 2019). According to Salem & Hashemi-Beni (2022), detecting water underneath vegetation is more challenging for C-band SAR compared to L-band, because shorter wavelengths (C-band) have a more limited capacity to penetrate the vegetation canopy. In their RF classification, this led to lower accuracy in C-band compared to the higher penetration capabilities of L-band signals. Three events from Greece (Magnesia, Orestiada) and Spain (Valencia) represent this biome.

- Temperate Coniferous Forests: In temperate landscapes of Europe this biome primarily dominates subalpine and mountainous regions, typically occurring at elevations above 1,200 m a.s.l. (Senf & Seidl, 2018). According to Salem & Hashemi-Beni (2022), the presence of water underneath vegetation enhances the SAR backscattered signal due to double-bounced and multi-bounce effects between the flat water surface and vertical vegetation structures, such as trunks and stems. Events from Slovakia (Trencin), Ukraine (Ivano), and Romania (Ghimbav) represent this biome.

- Tropical Subtropical Moist Broadleaf Forests: tis biome has a complex, multi-layered structure, typically comprising five vertical strata: the emergent layer of scattered tall trees, the closed canopy, the understory, the shrub layer, and the forest floor (Richards et al., 1996; Whitmore, 1998). This tall, dense vegetation strongly attenuates the C-band SAR signal, limiting canopy penetration and reducing the detectability of floodwater beneath the canopy (Tsyganskaya et al., 2018). Three events from Madagascar (Mahajanga), Suriname (Diitabiki), and Australia (Burketown) represent this biome.

Table 2.1. Study areas with the captured pre- and post-event SAR imagery dates.

Biome	Country	Flood Site	Activation	Pre-event	Post-event	Flood type
Temperate Broadleaf & Mixed Forests	Germany	Moselle	EMSR517	2021-07-10	2021-07-16	Riverine
	Poland	Wroclaw	EMSR756	2024-08-25	2024-09-18	Riverine
	Germany	Stuttgart	EMSR728	2024-05-09	2024-06-02	Riverine
Deserts & Xeric Shrublands	Pakistan	Rasool Nagar	EMSR838	2025-08-16	2025-08-28	Riverine
	Iran	Jask	EMSR419	2020-01-02	2020-01-11	Flash
	Iraq	Chaq Chaq	EMSR551	2021-11-15	2021-11-21	Flash
Mediterranean Forests	Greece	Magnesia	EMSR692	2023-09-01	2023-09-07	Flash & riverine
	Spain	Valencia	EMSR773	2024-10-01	2024-10-31	Flash
	Greece	Orestiada	EMSR277	2018-03-27	2018-04-01	Riverine
Temperate Coniferous Forests	Slovakia	Trencin	EMSR471	2020-10-13	2020-10-15	Riverine
	Ukraine	Ivano	EMSR444	2020-06-20	2020-06-26	Riverine
	Romania	Ghimbav	EMSR802	2025-05-24	2025-05-29	Riverine
Tropical & Subtropical Moist Broadleaf	Madagascar	Mahajanga	EMSR424	2020-01-26	2020-02-01	Riverine

Biome	Country	Flood Site	Activation	Pre-event	Post-event	Flood type
Tropical & Subtropical Moist Broadleaf	Suriname	Diitabiki	EMSR577	2022-06-09	2022-06-16	Riverine
	Australia	Burketown	EMSR562	2022-01-22	2022-02-03	Riverine

2.2 SAR Imagery

All flood mapping analyses in this study are based on dual-polarization Sentinel-1 (S1) C-band SAR imagery. Sentinel-1 operates in Interferometric Wide (IW) swath mode over land, acquiring data at 5.405 GHz with a 250 km swath width and a nominal spatial resolution of 5×20 m (ESA, 2024). The two-satellite constellation provides a 6-day revisit cycle, ensuring systematic temporal coverage suitable for monitoring dynamic flood events.

Imagery was accessed through Google Earth Engine (GEE) cloud platform Python API as Level-1 Ground Range Detected (GRD) products, preprocessed to backscatter coefficient (σ°) in decibels (Google Developers, 2025). The following preprocessing steps are applied internally by GEE:

- Apply Orbit File
- GRD Border Noise Removal
- Thermal Noise Removal
- Radiometric Calibration
- Terrain Correction (Orthorectification)

For each flood event, two S1 acquisitions were selected: a post-event image acquired during or immediately after the flood peak, and a pre-event reference image acquired before the event under dry hydrological conditions, from the same relative orbit and flight direction (ascending or descending) to ensure geometric consistency.

From each acquisition, three polarimetric features were extracted: VV backscatter, VH backscatter, and the VV/VH polarimetric ratio. The VV channel is primarily sensitive to surface scattering and exhibits strong backscatter reduction over calm open water due to specular reflection (Garg et al., 2024). The VH cross-polarized channel is more sensitive to volume scattering from vegetation and double-bounced scattering, providing complementary flood information particularly in vegetated areas (Salem & Hashemi-Beni, 2022). The Polarimetric ratio VV/VH provides a normalized metric highlighting polarization difference between surface types. Multi-temporal change features were derived by computing the ratio between post- and pre-event backscatter for each polarimetric channel, yielding three additional change layers. The final SAR-derived feature set therefore consists of nine layers: Pre-event (VV, VH, and VV/VH); post-event (VV, VH, VV/VH); and temporal change ratio for VV, VH, and VV/VH.

2.3 Contextual Environmental Variables

Three hydro-topographic variables and one land cover dataset were incorporated as contextual environmental features alongside the SAR-derived inputs. Terrain elevation data were obtained from FathomDEM (Uhe et al., 2025), a global Digital Terrain Model (DTM) available at 1 arc-second (~30 m) resolution between 60°S and 80°N. Unlike raw Digital Elevation Models, FathomDEM represents the bare earth surface, excluding vegetation and built structures, making it particularly suitable for hydrological applications. It is derived from the Copernicus DEM (ESA, 2021) and refined using a hybrid visual transformer model to remove non-terrain elements.

Water body data were obtained by rasterizing OpenStreetMap (OSM) water surfaces (OpenStreetMap, 2025), including features tagged as "natural = water", "waterway = *", and "landuse = reservoir". These data served as the basis for deriving two proximity-based flood susceptibility descriptors: Height Above Nearest Drainage (HAND) and Euclidean Distance to Water Body (EDTW). As demonstrated by Rennó et al. (2008), the HAND descriptor quantifies the relative vertical distance from each grid cell to its nearest drainage outlet along the flow path. This normalized topographic measure serves as a reliable indicator of hydrological conditions, where low HAND values identify low-lying, flat areas adjacent to drainage networks that are inherently prone to being waterlogged or inundated. Terrain slope was additionally derived from FathomDEM, reflecting the gravitational influence on surface water flow and retention. The derivation workflow for these three variables is described in Chapter 3.

Land cover information was obtained from ESRI's 10-meter Annual Land Cover data set, accessed through the Google Earth Engine Python API. The data set classifies Earth's surface into eight thematic classes: Water, Trees, Flooded vegetation, Crops, Built Area, Bare Ground, Snow/Ice, and Rangeland, using a UNet convolutional neural network applied to annual Sentinel-2 imagery (Karra et al., 2021). Land cover type influences both flood susceptibility and SAR backscatter behavior, as different surface types exhibit distinct microwave scattering responses under wet and dry conditions (Dasgupta et al., 2018).

Chapter 3: Methodology

3.1 Data and Feature Construction

3.1.1 Input Feature Stack

The feature stack consists of thirteen variables including nine SAR-derived features and four contextual environmental features.

The SAR-based variables are:

- VV polarization (pre-event)
- VH polarization (pre-event)
- VV/VH ratio (pre-event)
- VV polarization (post-event)
- VH polarization (post-event)
- VV/VH ratio (post-event)
- VV temporal change
- VH temporal change
- VV/VH ratio temporal change

These variables capture both the absolute backscatter properties and their temporal variation between pre- and post-flood acquisitions. An example for the Valencia test case is provided in Figure 3.1.

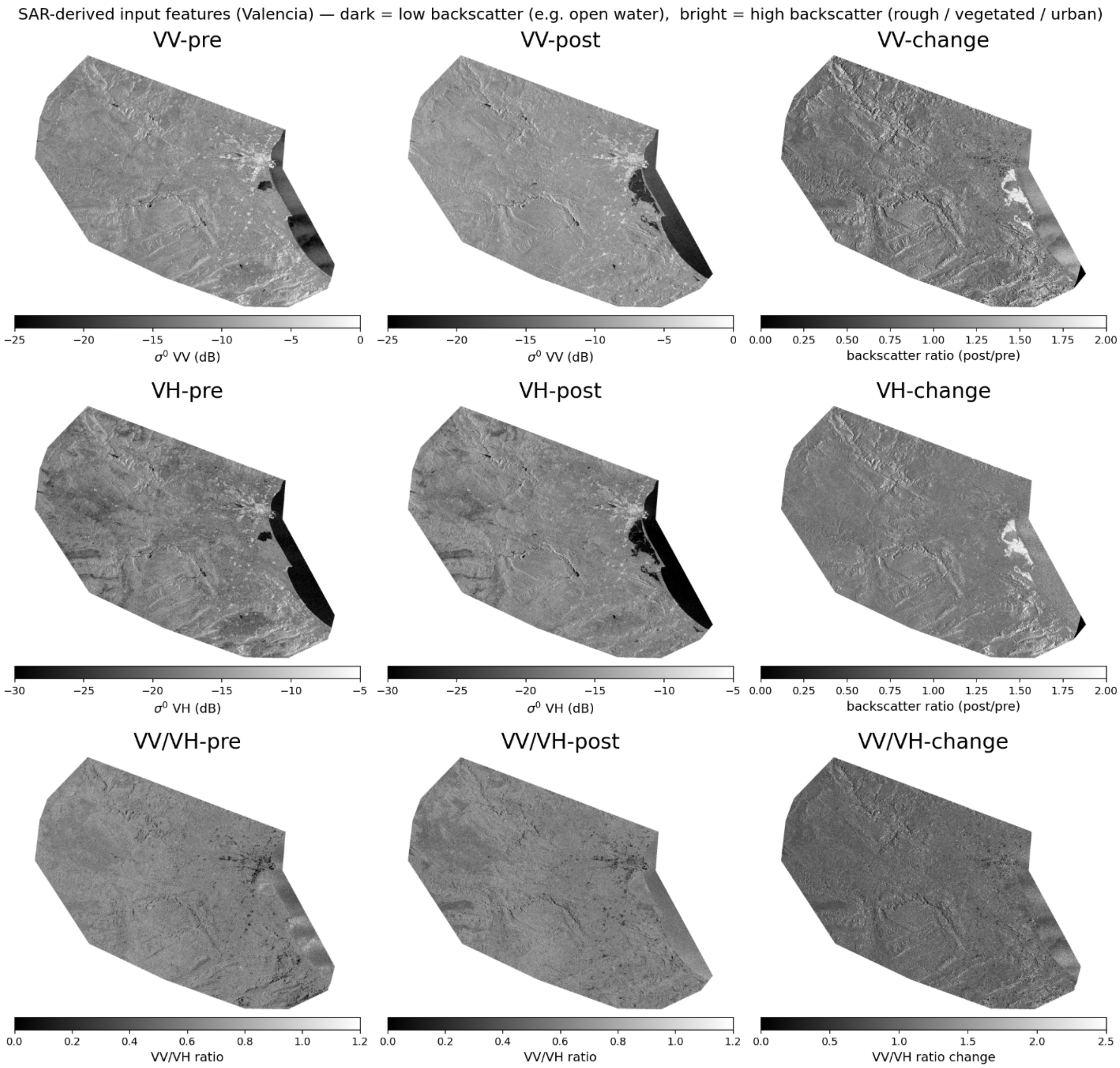

Figure 3.1. SAR-derived input features for the Valencia event, shown in the same order used throughout the study. Backscatter panels (VV/VH, pre/post) are displayed as σ⁰ in decibels; darker tones correspond to lower backscatter (smooth surfaces such as open water, which reflects the radar signal away from the sensor) and brighter tones to higher backscatter (rough, vegetated or built-up surfaces). Temporal-change panels show the post-/pre-event backscatter ratio (1 = no change, <1 = decrease, >1 = increase), and the VV/VH panels the (dimensionless) co-/cross-polarisation ratio. Each colour bar gives the physical value range of its panel.

Four contextual variables are additionally included:

- Terrain slope derived from a digital elevation model
- Height Above the Nearest Drainage (HAND)
- Euclidean Distance to Water (EDTW)
- Land cover derived from the ESRI global land cover dataset

Together, these contextual descriptors provide information about terrain morphology, drainage proximity, and land surface characteristics that may influence flood occurrence and detectability. Figure 3.2 provides examples of contextual variables for the Valencia test case.

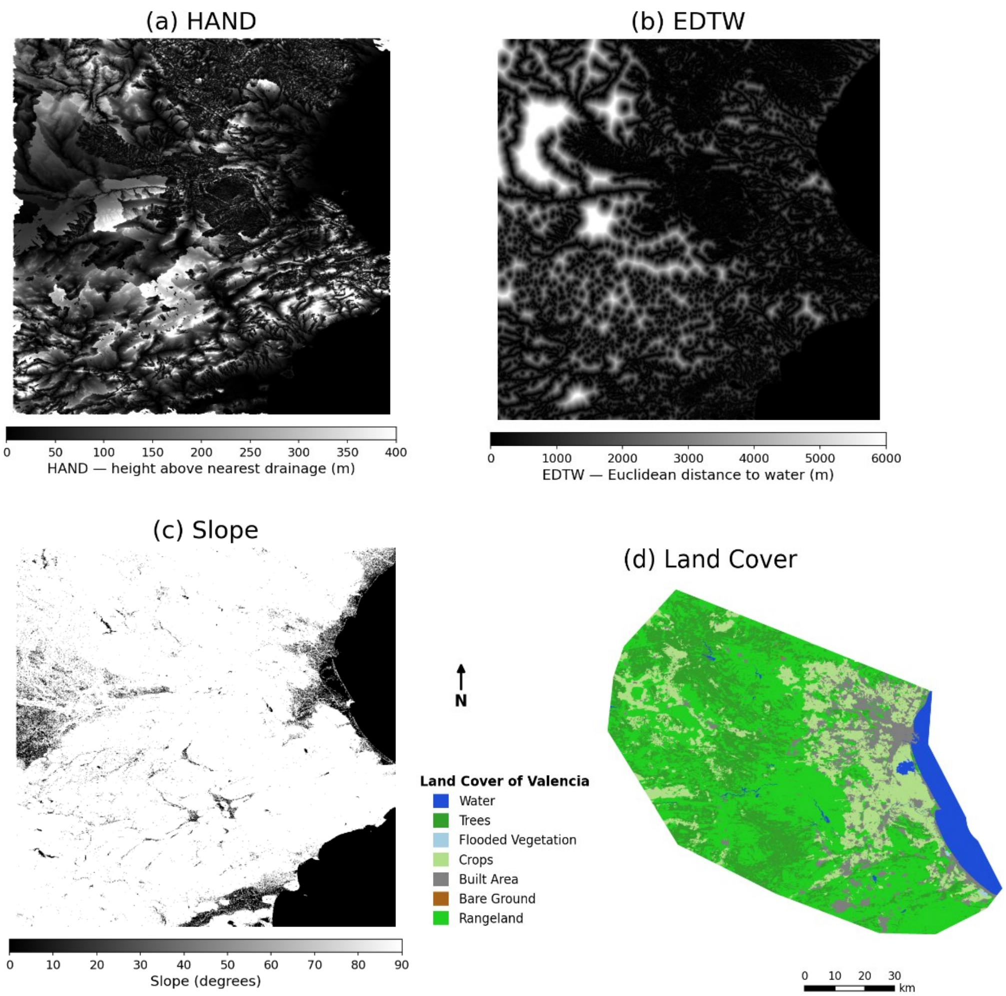

Figure 3.2. Contextual (non-SAR) input variables for the Valencia test case. (a) HAND (Height Above Nearest Drainage), in meters; dark tones mark cells close to the drainage network (low-lying, flood-prone) and bright tones mark areas elevated well above the channels. (b) EDTW (Euclidean Distance To Water), in meters; dark tones are close to permanent water bodies, bright tones are farther away. (c) Slope (terrain slope) in degrees (0–90°). (d) Land Cover (categorical ESRI land-cover classes), shown with the discrete color legend (Water, Trees, Flooded Vegetation, Crops, Built Area, Bare Ground, Rangeland). For the continuous panels (a-c) each grey-scale color bar gives the physical value range; for (d) the legend lists the class of each color. Areas outside the study footprint are left blank.

3.1.2 Ground Truth Data

Reference flood extent maps were obtained from Copernicus Emergency Management Service Rapid Mapping (CEMSR) catalogue (Copernicus EMS, 2025). For each of the fifteen flood events, the corresponding CEMSR activation was identified, and the flood delineation product was downloaded manually from the corresponding activation page of the Copernicus EMS Rapid Mapping portal and used as the ground-truth reference. Across the fifteen events, two types of CEMSR delineation products were used, specifically, the "Delineation products" and "Delineation Monitoring" layers (versions 1 and 2). Delineation Products represent the initial flood extent mapping, while Delineation Monitoring products are updated versions that incorporate additional satellite acquisitions as they become available, providing more accurate and complete flood extent delineation. The specific product type used for each event is reported in Table 2.1 and in Table 3.1.

Table 3.1 All assessed flood events with date and time

Country	**Latitude**	**Longitude**	**Biome**	**Area/ Flood site**	**Activation number**	**Monitoring**	**sensor**	**Acquisition time for each area**	**S1-Acquisition time**	**Difference between acquisition time of original sensor and S-1 sensor**
Germany	50.5444551	7.1186671	Temperate Broadleaf & Mixed Forests	Moselle	EMSR517	Delination product	Sentinel-1A/B	15/07/2021 at 05:50 UTC	–	–
Poland	51.6127858	16.337455		Wroclaw	ESMR756	Delination monitoring 2	Sentinel-2/HR+	18/09/2024 at 9:56 UTC	18/09/2024 at 16:36:05 UTC	6:40′
Germany	43.9357181	11.0941473		Stuttgart	ESMR728	Delination product	sentinel-1/HR+	02/06/2024 at 05:26 UTC	–	–
Pakistan	36.3042502	74.26652003	Deserts & Xeric Shrublands	Rasoo Nagar	EMSR838	Delination product	sentinel-1/HR+	28/08/2025 at 01:07 UTC	–	–
Iran	25.6436156	57.7745644		Jask	EMSR419	Delination product	Sentinel-1A	11/01/2020 at 14:16 UTC	–	–
Iraq	35.5940157	45.36168059		Chaq chaq	ESMR 551	Delineation Monitoring 1	Sentinel-1A/B	21/11/2021 at 03:01 UTC	–	–
Greece	40.6879949	22.8444773	Mediterranean Forests, Woodlands & Scrub	Magnesia	EMSR692	Delineation Monitoring 1	Sentinel-1A/B	06/09/2023 at 04:39 UTC	–	–
Spain	39.4697065	-0.3763353		Valencia	ESMR773	Delination product	Sentinel-1A/B	31/10/2024 at 18:02 UTC	–	–
Greece	41.5030836	26.5329437		Orestiada	EMSR277	Delination monitoring 1	Sentinel-1A/B	01/04/2018 at 04:21 UTC	–	–
Slovakia	48.8922719	18.0387465	Temperate Coniferous Forests	Trencin	EMSR471	Delination product	Sentinel-1B	15/10/2020 at 16:34 UTC	–	–

Country	Latitude	Longitude	Biome	Area/ Flood site	Activation number	Monitoring	sensor	Acquisition time for each area	S1-Acquisition time	Difference between acquisition time of original sensor and S-1 sensor
Ukraine	48.7481718	4.5207477		Ivano	EMSR444	Delination product	WorldView-2	26/06/2020 at 09:55 UTC	26/06/2020 at 16:09:40 UTC	6:14′
Romania	45.6634951	25.5066416		Ghimbav	EMSR802	Delination product	Sentinel-1A/B	29/05/2025 at 04:29 UTC	–	–
Madagascar	-17.474716	48.3507289	Tropical & Subtropical Moist Broadleaf Forests	Mahajanga	EMSR424	Delination product	Sentinel-1A	31/01/2020 at 02:26 UTC	–	–
Suriname	4.1147603	-54.6758547		Diitabiki	EMSR577	Delination product	Sentinel-1A/B	16/06/2022 at 09:21 UTC	–	–
Australia	-17.7400597	139.5465396		Burketown	EMSR562	Delination product	Sentinel-1A/B	03/02/2022 at 20:17 UTC	–	–
Finland	67.9593397	23.66774037	Boreal Forest/Taiga	Muonio	EMSR441	Delination product	Sentinel-1A/B	08/06/2020 at 04:56 UTC	–	–
Sweden	63.4309355	18.4247915		Gottne	EMSR841	Delination Monitoring 2	Sentinel-1	15/09/2025 at 05:06 UTC	–	–
Norway	61.26117795	11.6052056		Amot	EMSR668	Delination Monitoring 1	Sentinel-1A/B	25/05/2023 at 05:39 UTC	–	–

For two events (Wroclaw and Ivano) the primary sensor used by CEMSR was not Sentinel-1 but Sentinel-2/HR+ and WorldView-2, respectively. For these events, a contemporaneous Sentinel-1 acquisition was identified via the Copernicus Browser, with a temporal offset of approximately 6:40′ hours for Wroclaw and 6:14′ hours for Ivano, both well within the seven-hour threshold established as the selection criterion.

3.2 Training Sample Design

The selection of training samples from the data set introduced in the previous section plays a critical role in machine learning techniques. To evaluate the sensitivity of model performance and feature importance rankings to sampling design, multiple sampling strategies were implemented. As class distribution strategy for the training data, **balanced sampling** was considered, with an equal number of flooded and non-flooded samples selected to mitigate class imbalance. As for spatial sampling, the strategy used to distribute samples across the study area is **systematic sampling**. Systematic spatial sampling distributes training samples across the study area at regular spatial intervals, ensuring uniform coverage of the AOI. Unlike random sampling, which may produce spatially clustered samples, systematic sampling guarantees that all parts of the study area contribute equally to model training, reducing the risk of spatial bias in feature importance estimates (Stevens & Olsen, 2004). A fixed training sample size of 1000 samples was used for all experiments. This sample size was selected based on evidence from the remote sensing literature suggesting that performance saturates beyond 500–1000 balanced training samples for binary classification tasks of moderate feature dimensionality (Maxwell et al., 2018).

3.3 Machine Learning Model

The selected Machine Learning model used in this work is Random Forest (RF, Breiman, 2001), an ensemble learning method that constructs multiple decision trees, each trained on a bootstrap sample of the training data. At each node split, only a random subset of predictor variables is considered, which reduces correlation between trees and improves generalization. Final predictions are obtained by majority voting across all trees.

The reason why RF was selected is that it suited to multi-source remote sensing applications as it handles heterogeneous input features such as continuous SAR backscatter, categorical land cover, and topographic indices without requiring feature normalization (Maxwell et al., 2018). It is also data-parsimonious, performing well with limited training data and across a wide variety of environmental contexts (Belgiu & Drăguţ, 2016). In this study, the Random Forest models were implemented using the scikit-learn library in Python, a widely used and well-documented implementation for classification tasks. Model training was conducted for each case study event. The resulting models were then used to generate flood predictions and compute feature importance metrics.

3.4 Feature Importance Methods

To assess the relative contribution of individual predictor variables, three feature attribution methods were applied. To support the interpretation of feature importance rankings, Pearson correlation coefficients were computed between all input feature pairs. This correlation index is critical to a more reasoned interpretation of MDI, MDA, and SHAP values, as correlated features may lead to misleading importance estimates. For instance, when two features share the same predictive information, importance scores may be distributed unevenly between them depending on the attribution method used. Correlation matrices were computed separately for each flood event and subsequently averaged across events within each biome, with variance indicating consistency of correlation patterns across environmental contexts.

3.4.1 Mean Decrease in Impurity (MDI)

The Mean Decrease in Impurity (MDI) quantifies the contribution of each feature by summing the total reduction in node impurity across all splits involving that feature, averaged over all trees in the ensemble (Breiman, 2001). Because MDI is calculated during training, it is computationally efficient. However, Strobl et al. (2007) demonstrated that MDI is biased toward features with many unique values. In this study, the nine continuous SAR-derived features may therefore receive inflated importance scores compared to the binary one-hot encoded land cover variables.

3.4.2 Permutation Importance (MDA)

Permutation Importance (MDA), introduced by Breiman (2001) and formalized by Altmann et al. (2010), evaluates feature importance by measuring the decrease in model accuracy when the values of a single feature are randomly permuted while all other features remain unchanged. This permutation breaks the relationship between the feature and the target variable, such that any resulting drop in accuracy reflects the feature's true predictive contribution. Unlike MDI, MDA is not computed during training and is therefore not biased toward high cardinality or continuous features. However, MDA can underestimate the importance of correlated features: when two features share predictive information, permuting one may not substantially degrade performance if the other retains the same signal (Altmann et al., 2010). This is particularly relevant in this study, as VV and VH backscatter are moderately correlated (Pearson $r \approx 0.67$), which may cause MDA to underestimate VH importance with respect to what is estimated by the other methods.

3.4.3 SHapley Additive exPlanations (SHAP)

SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee (2017), provides a unified framework for feature attribution grounded in cooperative game theory. Each feature is treated as a player in a cooperative game, where the model prediction is the payout. The SHAP value for a feature represents its average marginal contribution across all possible orderings of features. Lundberg and Lee (2017) showed that SHAP is the unique additive feature attribution method satisfying three desirable properties: local accuracy, which ensures that the sum of SHAP values equals the model output; missingness, which guarantees zero attribution for absent features; and consistency, which ensures that if a feature's contribution increases, its importance score never decreases. This consistency property makes SHAP particularly suitable for comparing feature importance across different biomes, as rankings remain interpretable even when the underlying model structure varies between study areas. In this study, TreeSHAP (Lundberg et al., 2019) was implemented, which exploits the tree structure of RF to compute exact SHAP values efficiently without evaluating all possible feature subsets.

3.4.5 Evaluation Metrics

Model classification performance was evaluated using two complementary metrics: Overall Accuracy (OA) and F1 score. OA measures the proportion of correctly classified pixels across all classes:

$$OA = (TP + TN) / (TP + TN + FP + FN)$$

However, OA alone can be misleading in flood mapping applications due to the inherent class imbalance between flooded and non-flooded pixels. In a typical flood scene, where flooded areas represent only a small fraction of the total image extent, a classifier that predicts all pixels as non-flooded would still achieve high OA while failing to detect any flood extent. The F1 score addresses this limitation by combining Precision and Recall into a single metric:

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1 = 2 × (Precision × Recall) / (Precision + Recall)

F1 reaches its maximum value of 1 when both Precision and Recall are perfect and approaches zero when either metric fails. By explicitly penalizing both false positives and false negatives, F1 provides a more sensitive and reliable measure of flood detection performance than OA alone, particularly in the presence of class imbalance (He & Garcia, 2009). In this study, both metrics are reported to provide a comprehensive assessment of classification performance.

3.5 Experiment Design

To evaluate the robustness of feature importance rankings, experiments were conducted across multiple flood events and environmental conditions. For each flood event, the full modeling pipeline including sampling, model training, and feature importance estimation was executed independently. Classification performance was assessed at two levels. In the same-site setting, the Random Forest was trained and tested on spatially disjoint samples drawn from the same flood event (site), quantifying how well the model fits an individual AOI. In the cross-site setting, a leave-one-site-out protocol was applied within each biome: the model was trained on the remaining sites of the biome and tested on the held-out site, quantifying spatial transferability to an unseen location. Here, "site" denotes an individual flood event (AOI), not a biome. Both settings were evaluated using Overall Accuracy (OA) and F1 score and repeated over the ten iterations to obtain mean performance and its variability. All experiments were repeated across ten iterations with different random seeds to capture variability arising from stochastic sampling and model initialization. Feature importance rankings were first analyzed within each biome to assess intra-biome variability. Subsequently, rankings were compared across biomes to evaluate the transferability of feature-importance patterns between different environmental contexts. This experimental design allows systematic assessment of whether SAR-based and contextual features maintain consistent importance rankings across diverse ecological and hydrological conditions.

Chapter 4: Results and Discussion

4.1 Classification Performance

4.1.1 Same-Site Performance

Figure 4.1 summarizes the same-site classification performance, when a model is trained and evaluated on spatially disjoint samples drawn from the same flood event (site), provides a direct measure of how well the classifier fits an individual area of interest (AOI). Overall Accuracy (OA) ranged from 0.85 to 0.99, while F1 scores ranged from 0.79 to 0.99, indicating consistently strong same-site performance with moderate variability across biomes and individual events.

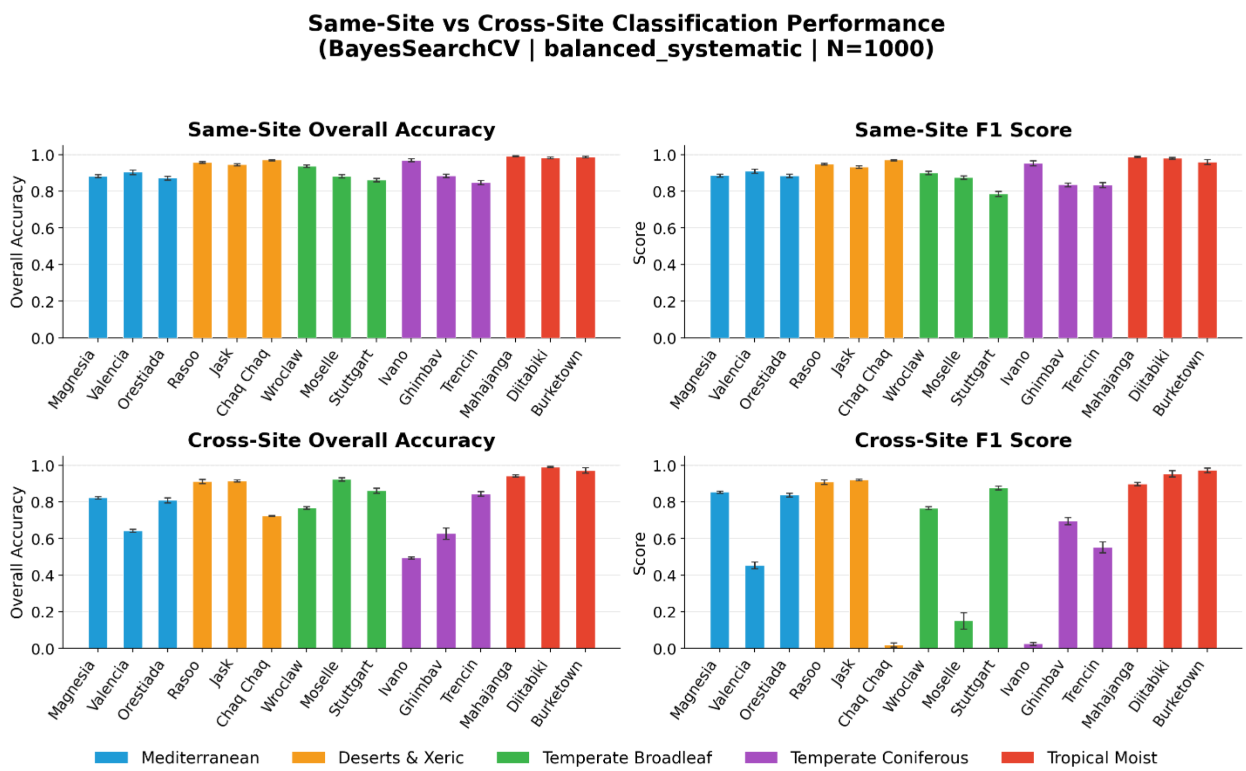

Figure 4.1. Same-site (top) and cross-site (bottom) classification performance

The Tropical and Subtropical Moist Broadleaf Forest biome consistently achieved the highest classification performance, with mean OA values of 0.99, 0.98, and 0.99 for Mahajanga, Diitabiki, and Burketown respectively, and corresponding F1 scores of 0.99, 0.98, and 0.96 [Figure 4.1]. This strong performance is likely attributable to the pronounced backscatter contrast between flooded and non-flooded surfaces in these lowland environments, where flat terrain and open water produce a clear specular reduction in the co-polarized VV channel (Tsyganskaya et al., 2018).

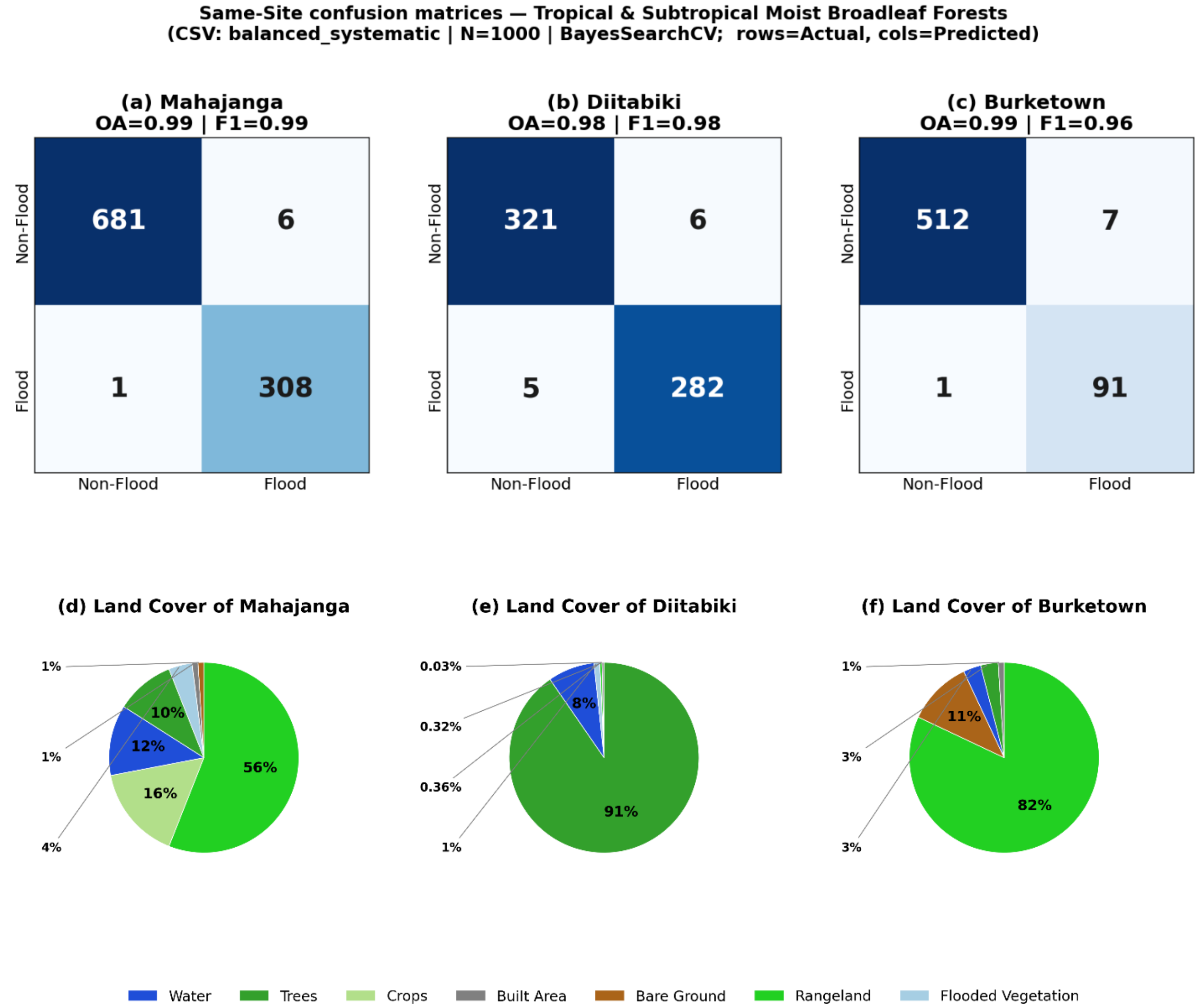

Figure 4.2. Same-site confusion matrices ***(a–c)*** *and land-cover composition* ***(d–f)*** *for the AOIs in the Tropical and Subtropical Moist Broadleaf Forests biome:* ***(a, d)*** *Mahajanga,* ***(b, e)*** *Diitabiki,* ***(c, f)*** *Burketown. In each confusion matrix, rows are the actual class and columns the predicted class.*

The Desert and Xeric Shrublands biome also achieved high same-site performance across all three sites (Chaq Chaq OA=0.97, F1=0.97; Rasoo Nagar OA=0.96, F1=0.95; Jask OA=0.95, F1=0.93). In these arid AOIs the flood signal is carried mainly by the co-polarized VV channel: inundation replaces a rough, moderately scattering dry surface with smooth open water, whose specular reflection sharply reduces VV backscatter and separates it clearly from the surrounding dry terrain (Garg et al., 2024). Notably, even at Rasoo Nagar, where cropland dominates the AOI (82%), the same-site classifier separated flooded from non-flooded pixels reliably (F1=0.95); although emergent crop vegetation can raise backscatter through double-bounce between standing water and plant stems and, in principle, complicate flood detection (Tsyganskaya et al., 2018), this effect did not materially degrade same-site performance in this biome. Figure 4.3 reports the corresponding confusion matrices and land-cover fractions.

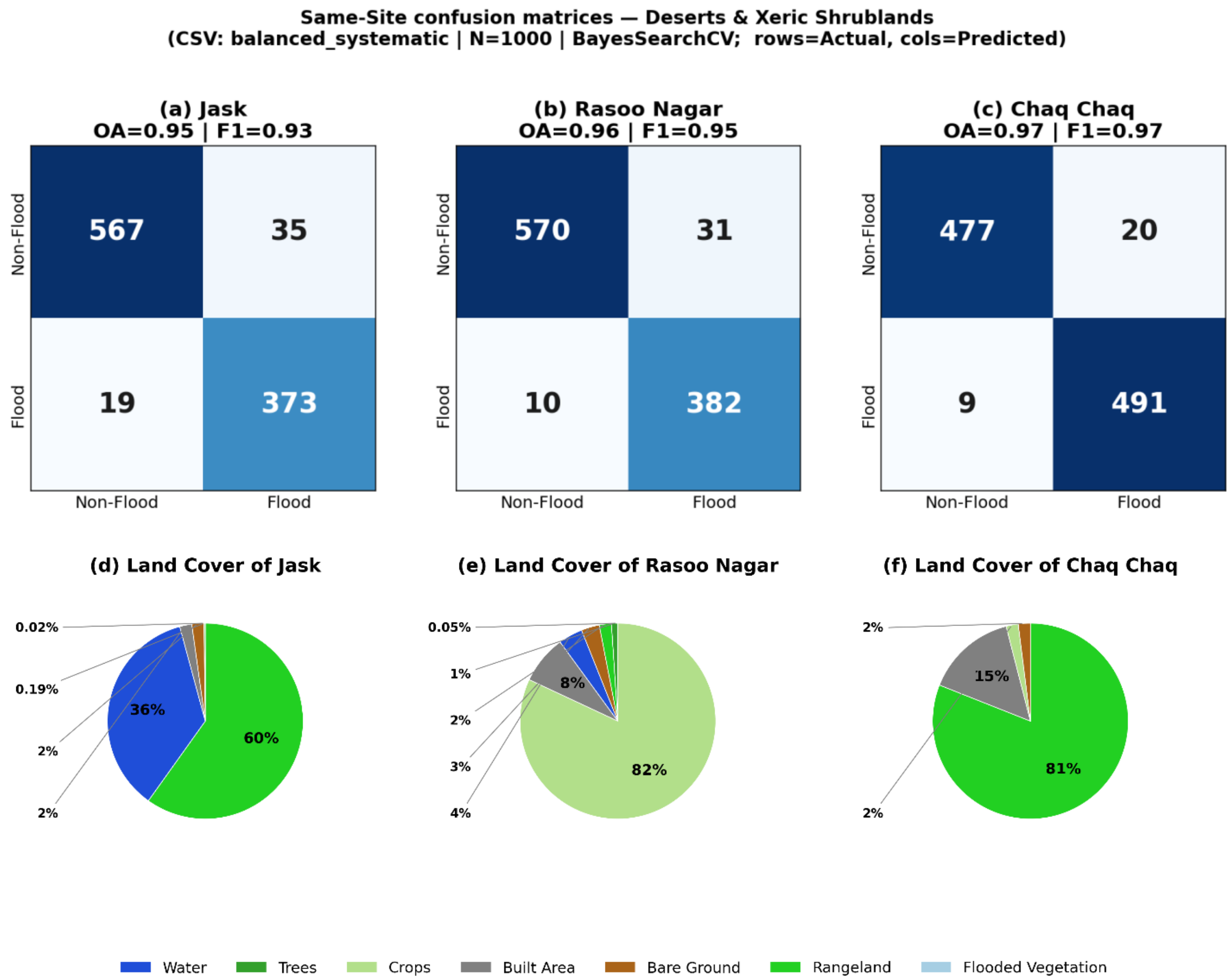

*Figure 4.3. Same-site confusion matrices **(a–c)** and land-cover composition **(d–f)** for the AOIs in the Deserts and Xeric Shrublands biome.: **(a, d)** Jask, **(b, e)** Rasoo Nagar, **(c, f)** Chaq Chaq. In each confusion matrix, rows are the actual class and columns the predicted class.*

The Temperate Broadleaf and Mixed Forests biome showed good same-site performance, with OA between 0.86 and 0.94 across the three sites. F1 scores were more variable, ranging from 0.79 (Stuttgart) through 0.88 (Moselle) to 0.90 (Wroclaw), indicating that while overall pixel-level accuracy is consistently high, the detection of flooded pixels specifically varies more between sites. This variability is consistent with the mixed land-cover composition of these sites, where dense tree canopy (Trees: 49% in Moselle) attenuates the C-band signal and reduces flood detectability beneath the canopy (Tsyganskaya et al., 2018). The per-site confusion matrices and land-cover fractions for this biome are shown in Figure 4.4.

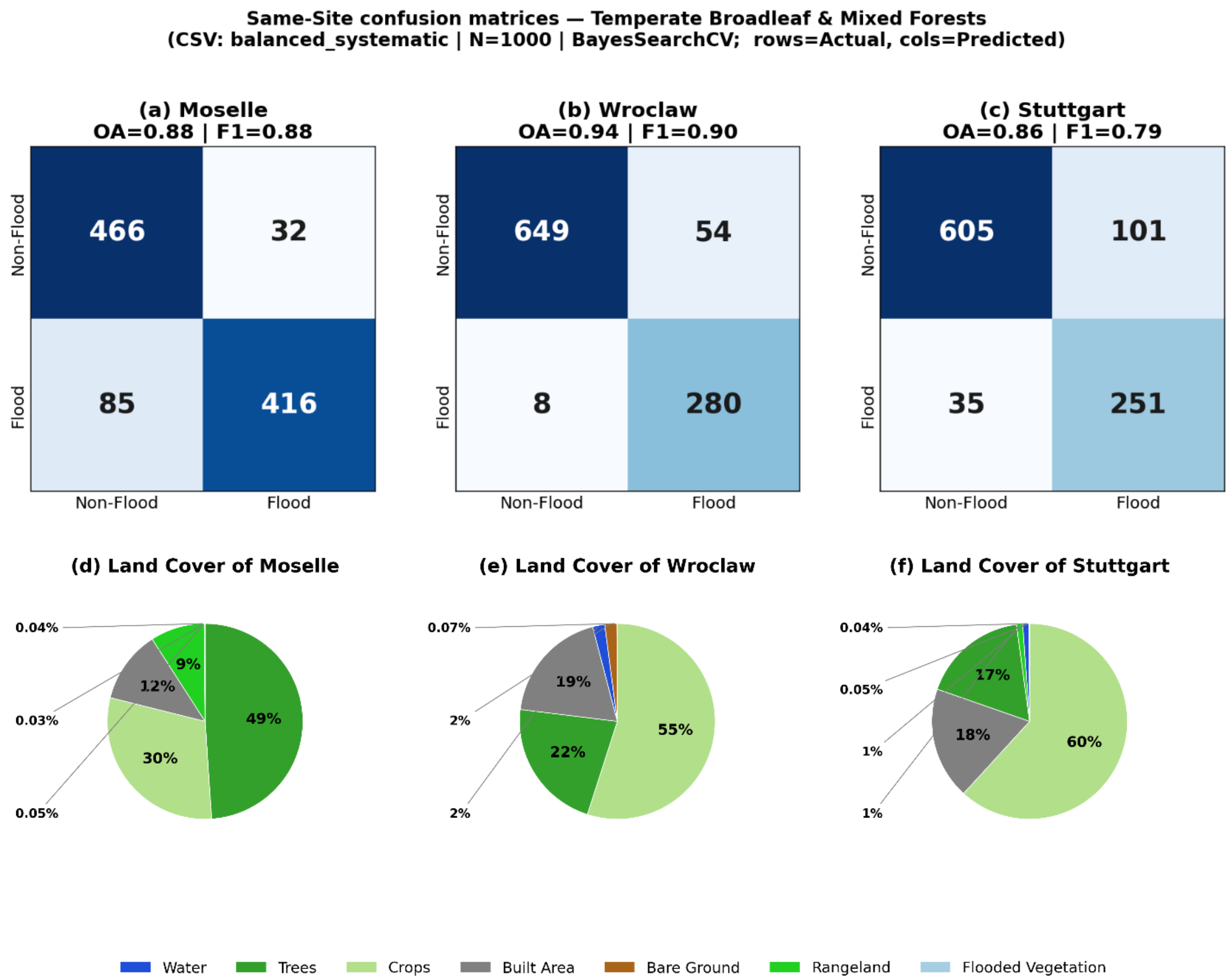

*Figure 4.4. Same-site confusion matrices **(a–c)** and land-cover composition **(d–f)** for the AOIs in the Temperate Broadleaf and Mixed Forests biome: **(a, d)** Moselle, **(b, e)** Wroclaw, **(c, f)** Stuttgart. In each confusion matrix, rows are the actual class and columns the predicted class.*

The Mediterranean Forests, Woodlands and Scrub biome achieved consistently good same-site performance. As shown in figure 4.5, the OA ranges between 0.87 and 0.90 and F1 ranges between 0.88 and 0.91 across the three sites. The comparatively uniform scores indicate that, despite the heterogeneous landscape mosaic of crops, rangeland and trees, the same-site classifier was able to separate flooded from non-flooded pixels reliably when trained and tested on the same event.

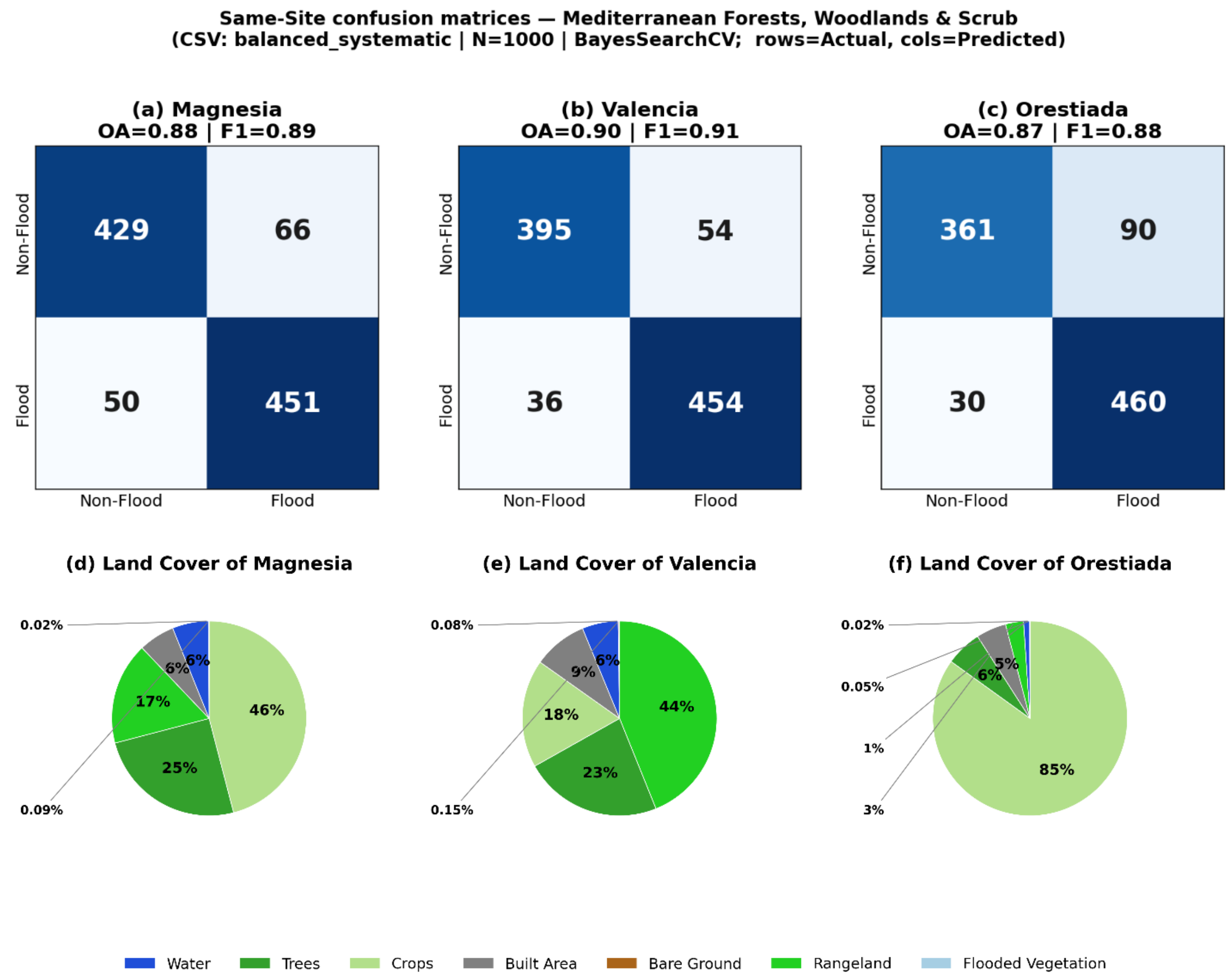

Figure 4.5. Same-site confusion matrices ***(a–c)*** *and land-cover composition* ***(d–f)*** *for the AOIs in the Mediterranean Forests, Woodlands and Scrub biome:* ***(a, d)*** *Magnesia,* ***(b, e)*** *Valencia,* ***(c, f)*** *Orestiada. In each confusion matrix, rows are the actual class and columns the predicted class.*

As shown in Figure 4.6, The Temperate Coniferous Forests biome achieved solid same-site performance (OA 0.85–0.97; F1 0.83–0.95), with Ivano scoring highest (OA=0.97, F1=0.95) and Trencin lowest (OA=0.85, F1=0.83). While same-site classification is therefore reliable, this biome proved the most difficult for cross-site transfer (Section 4.1.2): C-band Sentinel-1 backscatter is dominated by the upper canopy and penetrates dense coniferous forest only weakly, and forest backscatter is temporally variable in response to wind, moisture and structural change, so a model trained on one site generalizes poorly to another. The high cropland fraction in Ghimbav (79%) and Ivano (40%) further introduces backscatter heterogeneity between sites (Tsyganskaya et al., 2018).

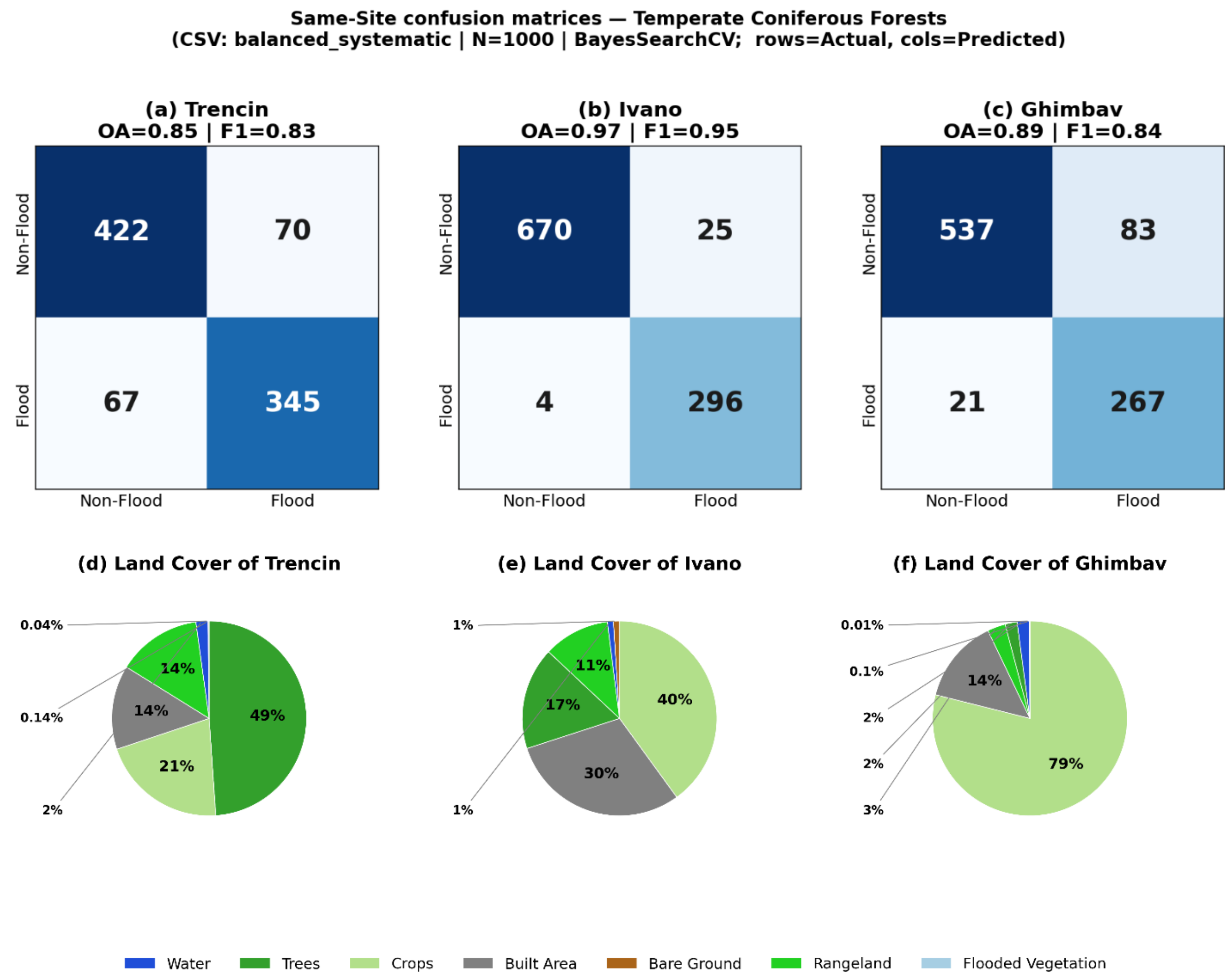

*Figure 4.6. Same-site confusion matrices **(a–c)** and land-cover composition **(d–f)** for the AOIs in the Temperate Coniferous Forests biome: **(a, d)** Trencin, **(b, e)** Ivano, **(c, f)** Ghimbav. In each confusion matrix, rows are the actual class and columns the predicted class.*

4.1.2 Cross-Site Performance

Cross-site performance, when models trained on two sites within a biome are evaluated on the third held-out site, provides a direct measure of intra-biome transferability. Cross-site OA and F1 scores are summarized in Figure 4.7.

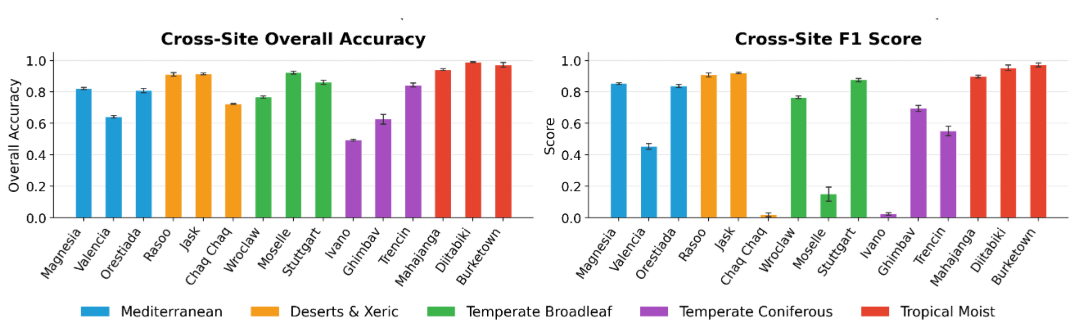

Figure 4.7. Cross-site classification performance.

As shown in Figure 4.8, the Tropical and Subtropical Moist Broadleaf Forest biome demonstrated the most consistent cross-site transferability, with cross-site F1 scores of 0.90, 0.95, and 0.97 for Mahajanga, Diitabiki, and Burketown respectively. This result suggests that flood signatures in tropical environments are sufficiently homogeneous across different geographic locations within the biome to support robust model transfer. The dominance of tree cover across all three sites (Diitabiki 91%, Mahajanga 10% trees with consistent rangeland, and Burketown 82% rangeland) may contribute to consistent SAR response patterns that generalize well across sites.

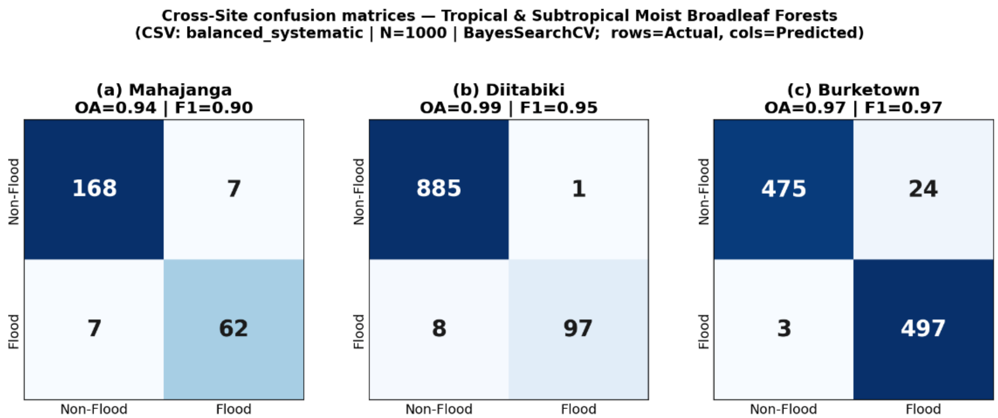

Figure 4.8. Cross-site confusion matrices for the Tropical and Subtropical Moist Broadleaf Forests biome: ***(a)*** *Mahajanga,* ***(b)*** *Diitabiki,* ***(c)*** *Burketown. Rows are the actual class and columns the predicted class.*

As shown in figure 4.9, Stuttgart showed strong cross-site transferability within the Temperate Broadleaf biome (cross-site F1=0.88), while Moselle exhibited a dramatic drop from a same-site F1 of 0.88 to a cross-site F1 of 0.15. This divergence suggests that the flood signature at Moselle is poorly represented by training data from Stuttgart and Wroclaw, potentially reflecting event-specific hydrological characteristics or unique land-cover configurations at the Moselle site.

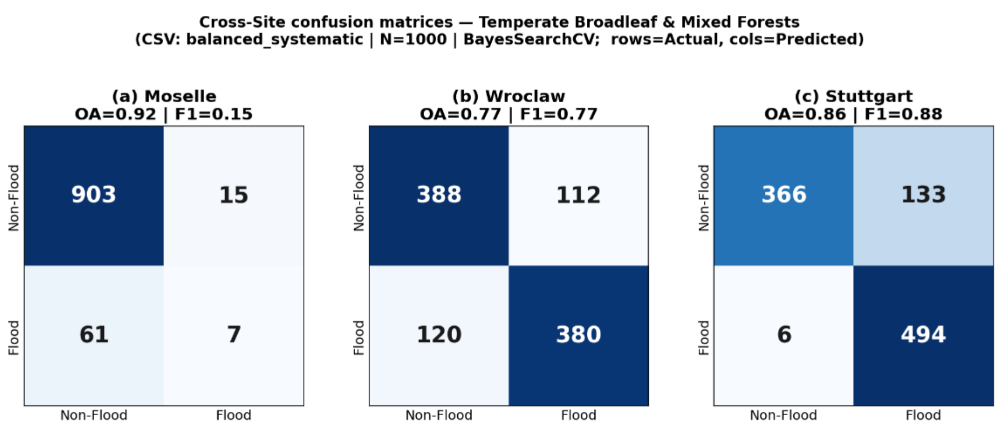

Figure 4.9. Cross-site confusion matrices for the Temperate Broadleaf and Mixed Forests biome: ***(a)*** *Moselle,* ***(b)*** *Wrocław,* ***(c)*** *Stuttgart. Rows are the actual class and columns the predicted class.*

Within the Mediterranean Forests, Woodlands and Scrub biome, cross-site transferability was moderate but uneven. Magnesia and Orestiada transferred well (cross-site F1 = 0.85 and 0.84 respectively), whereas Valencia dropped sharply from a same-site F1 of 0.91 to a cross-site F1 of 0.45. This asymmetry reflects the heterogeneous landscape mosaic of the biome: Magnesia (crops 46%) and Orestiada (crops 85%) share a broadly crop-dominated composition, so a model trained on one generalizes to the other, whereas Valencia (dominated by rangeland (44%) with substantial tree cover (23%) and driven by a flash-flood (DANA) event) is poorly represented by the other two sites, degrading its transfer.

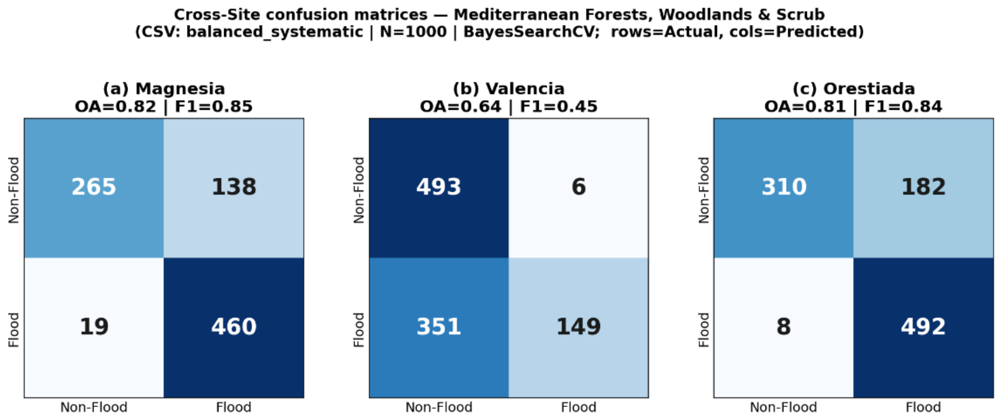

Figure 4.10. Cross-site confusion matrices for the Mediterranean Forests, Woodlands and Scrub biome: ***(a)*** *Magnesia,* ***(b)*** *Valencia,* ***(c)*** *Orestiada. Rows are the actual class and columns the predicted class.*

The most severe cross-site performance degradation was observed in the Temperate Coniferous biome, where Ivano dropped from a same-site F1 of 0.95 to a cross-site F1 of 0.03. This near-complete failure of cross-site generalization reflects the strongly differing land-cover compositions across the three sites, ranging from tree-dominated (Trencin, 49% trees) to cropland-dominated (Ghimbav, 79% crops; Ivano, 40% crops) landscapes. [Figure 4.11]

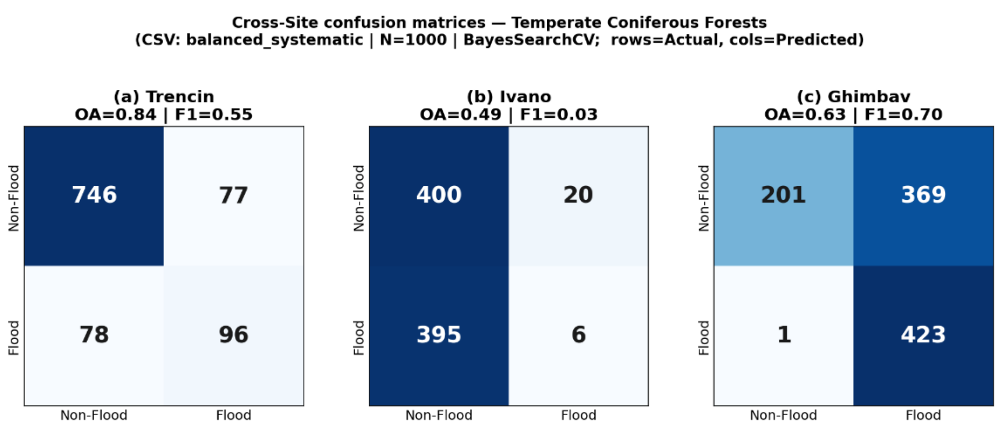

Figure 4.11. Cross-site confusion matrices for the Temperate Coniferous Forests biome: ***(a)*** *Trencin,* ***(b)*** *Ivano,* ***(c)*** *Ghimbav. Rows are the actual class and columns the predicted class.*

Similarly, Chaq Chaq in the Deserts biome exhibited a cross-site F1 of 0.02 despite a same-site F1 of 0.97, corroborating the known data-quality issues identified for this site. The overall pattern across biomes reveals that high same-site performance does not guarantee cross-site transferability, highlighting the importance of geographically diverse training data for operational flood mapping (Maxwell et al., 2018). [Figure 4.12]

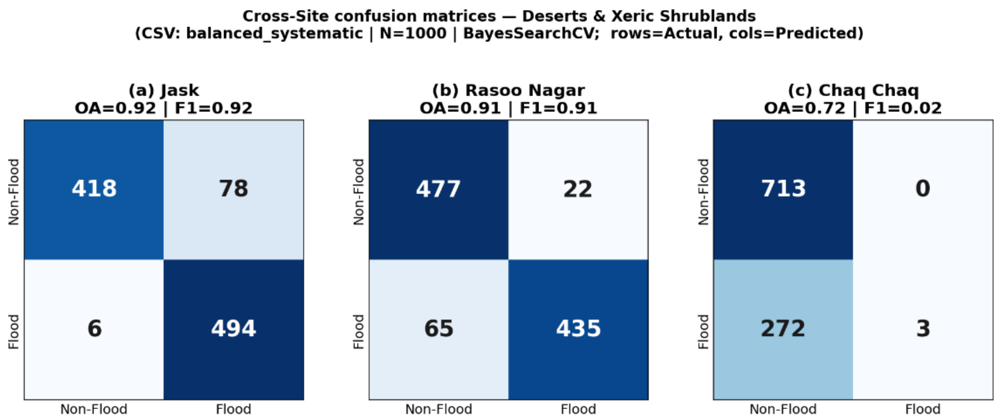

Figure 4.12. Cross-site confusion matrices for the Deserts and Xeric Shrublands biome: ***(a)*** *Jask,* ***(b)*** *Rasoo Nagar,* ***(c)*** *Chaq Chaq. Rows are the actual class and columns the predicted class.*

4.2 Feature Importance Rankings (Intra-Biome Analysis)

Feature importance rankings were computed using three attribution methods (MDI, MDA, and SHAP) for each flood event within each biome. The following section discusses the dominant features and method agreement patterns observed within each biome.

4.2.1 Desert and Xeric Shrublands

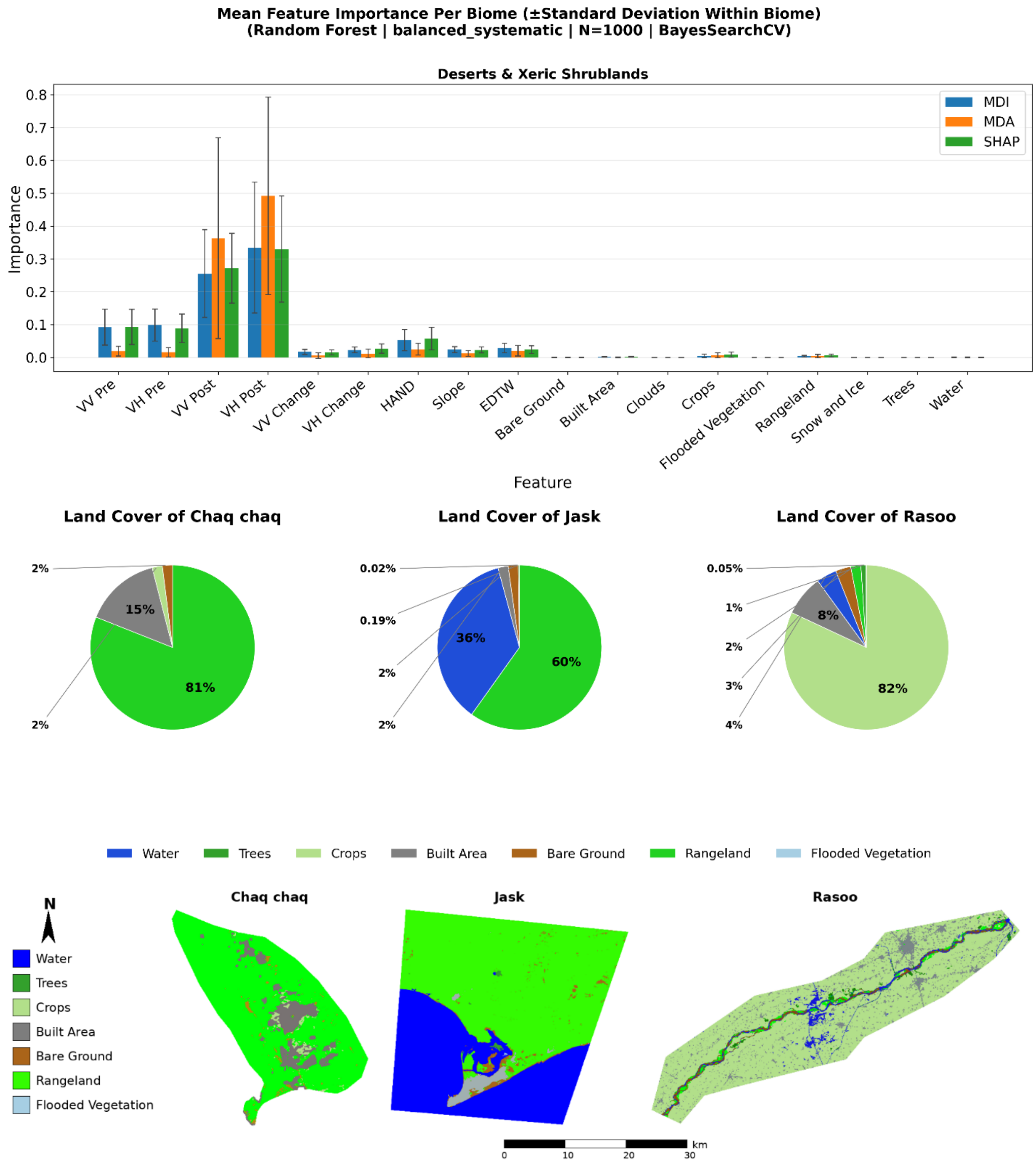

Figure 4.13. Feature importance and land cover in the Deserts and Xeric Shrublands biome.

Within the Desert and Xeric Shrublands biome, post-event SAR backscatter features dominated the feature importance rankings across all three attribution methods. VH Post consistently emerged as the single most important predictor, followed by VV Post, with both features showing substantially higher importance scores than all other variables. This pattern is consistent with the physical characteristics of arid flood environments: the strong, spatially extensive change in the post-event radar return when the dry arid surface is inundated is captured by both channels. Physically, the co-polarized VV channel is the primary open-water discriminator, because the specular reflection of smooth standing water sharply reduces VV backscatter relative to the dry background (Garg et al., 2024); the cross-polarized VH channel adds complementary sensitivity where inundation is partly vegetated, through double-bounce and volume-scattering interactions that VV alone does not capture (Tsyganskaya et al., 2018). Because these AOIs combine open water with sparsely vegetated and agricultural surfaces, both post-event channels carry substantial flood information.

Pre-event backscatter features (VV pre and VH pre) ranked as secondary predictors, reflecting their role as reference baselines for change detection. Among the contextual variables, HAND emerged as the most informative, ranking around fifth overall (and as high as third under MDA), clearly above the temporal change features. Its modest absolute importance is nonetheless consistent with the physical setting of these arid basins, where flooding is dominated by flash floods across flat alluvial plains rather than by topographically controlled riverine inundation (Rennó et al., 2008). The remaining contextual variables (Slope and EDTW), together with temporal changes features (VV Change and VH Change), Showed near zero importance, indicating that in this biome the absolute post-event backscatter level is far more informative than either terrain descriptors or the relative change from pre-event conditions. Land cover features also showed minimal importance, consistent with the homogeneous rangeland-dominated landscape of Chaq Chaq (81% rangeland) and Jask (60% Rangeland) and crop-dominated landscape of Rasoo Nagar (82% crops).

4.2.2 Mediterranean Forests, Woodlands and Scrub

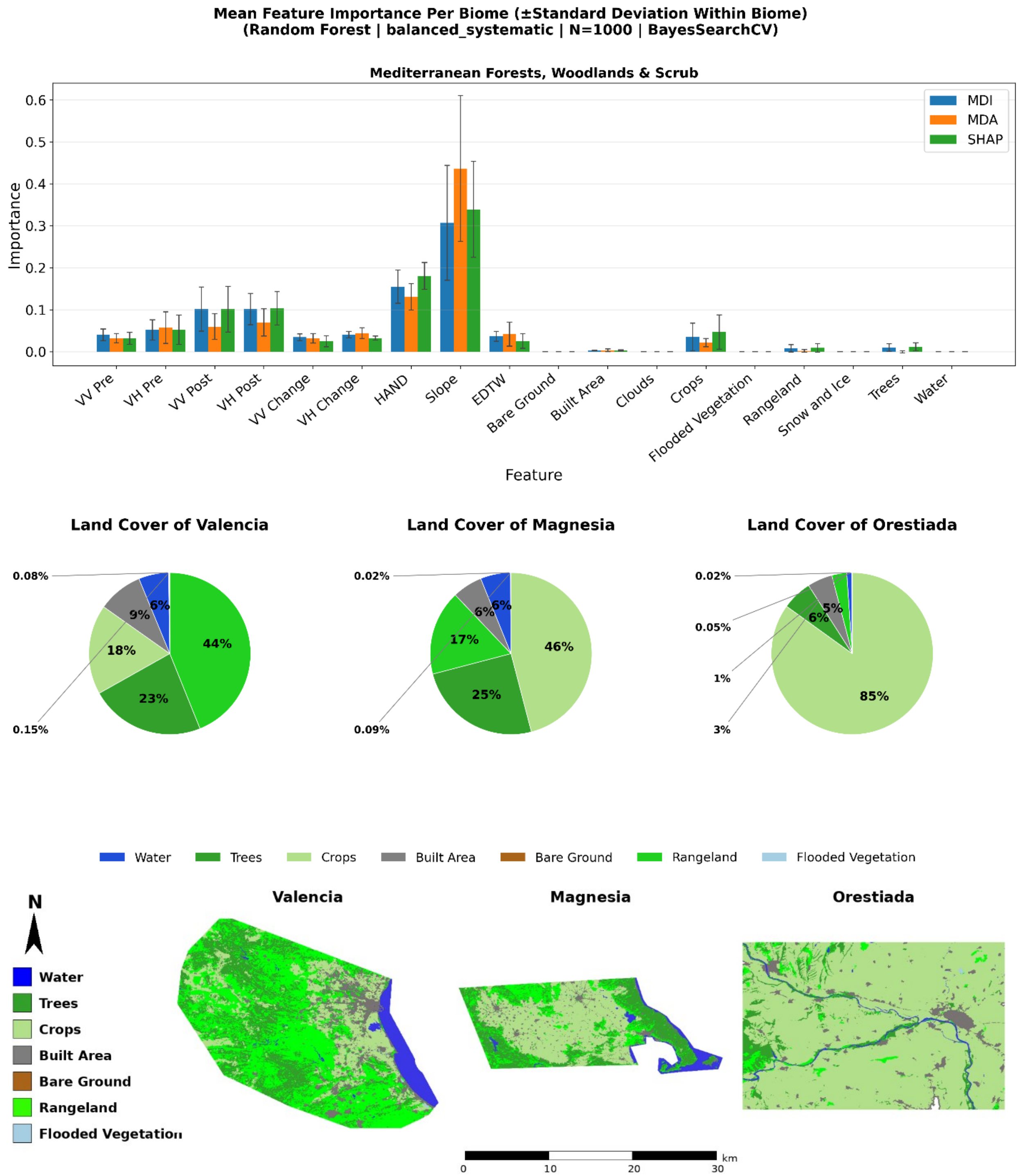

Figure 4.14. Feature importance and land cover in the Mediterranean Forests, Woodlands and Scrub biome.

Among the five biomes, the Mediterranean and the Temperate Broadleaf and Mixed Forests biomes were the two in which contextual terrain features outranked SAR backscatter in the importance rankings. What set the Mediterranean apart was the type of terrain control involved: here Slope was the single most important predictor under all three attribution methods (MDI ≈ 0.31, MDA ≈ 0.44, SHAP ≈ 0.34), followed by HAND in second place (≈ 0.13–0.18) whereas the post-event backscatter features VV Post and VH Post ranked only third and fourth (≈ 0.06–0.10). This contrasts with the Temperate Broadleaf biome, where the drainage-proximity descriptors HAND and EDTW lead the rankings (Section 4.2.3). The three methods agreed closely on Mediterranean ordering, with SHAP assigning HAND its largest value. This dominance of terrain features is consistent with the hydro-geomorphology of Mediterranean flood environments. Extreme floods in the Mediterranean belt are predominantly steep-catchment flash floods, in which orography and hillslope gradient govern the speed of runoff concentration and the spatial pattern of inundation (Marchi et al., 2010; Borga et al., 2014). In such settings, local slope and the elevation of each pixel above the nearest drainage channel are strong physical controls on where floodwater accumulates (Rennó et al., 2008). The study sites reflect this: Valencia and Magnesia combine steep coastal mountain ranges with confined floodplains, so that slope discriminates flood-prone valley floors from non-flooded uplands more sharply than the radar backscatter itself.

Land cover contributed a modest but non-negligible share of importance (SHAP ≈ 0.07), higher than in the Desert and Tropical biomes but well below the cropland-dominated Temperate Coniferous biome (Section 4.2.4), driven chiefly by the Crops class. This is consistent with the heterogeneous landscape mosaic characteristic of the Mediterranean region (Médail et al., 2019): at Valencia (rangeland 44%, trees 23%, crops 18%) and Magnesia (crops 46%, trees 25%, rangeland 17%) the Random Forest was able to exploit land-cover class as a proxy for the low-lying valley-floor agricultural land where floodwater preferentially accumulates, exploiting the spatial correlation between cultivated floodplains and inundation extent.

The comparatively low importance of the SAR backscatter features in this biome is itself informative. The structurally diverse vegetation of Mediterranean matorral and woodland weakens the open-water specular-reflection contrast that makes VV Post and VH Post so discriminative in arid or homogeneous settings, because over vegetated surfaces the C-band return is governed by canopy and double-bounce scattering rather than the underlying water (Salem & Hashemi-Beni, 2022; Tsyganskaya et al., 2018). With the radar signal less decisive, the model relied more heavily on terrain descriptors, producing the slope-dominated ranking observed here. The strong agreement among MDI, MDA, and SHAP (in contrast to the method divergences seen in the more evenly balanced biomes) indicates that this slope dominance is a robust property of the data rather than an artefact of any single attribution method (Strobl et al., 2007; Lundberg & Lee, 2017).

4.2.3 Temperate Broadleaf and Mixed Forests

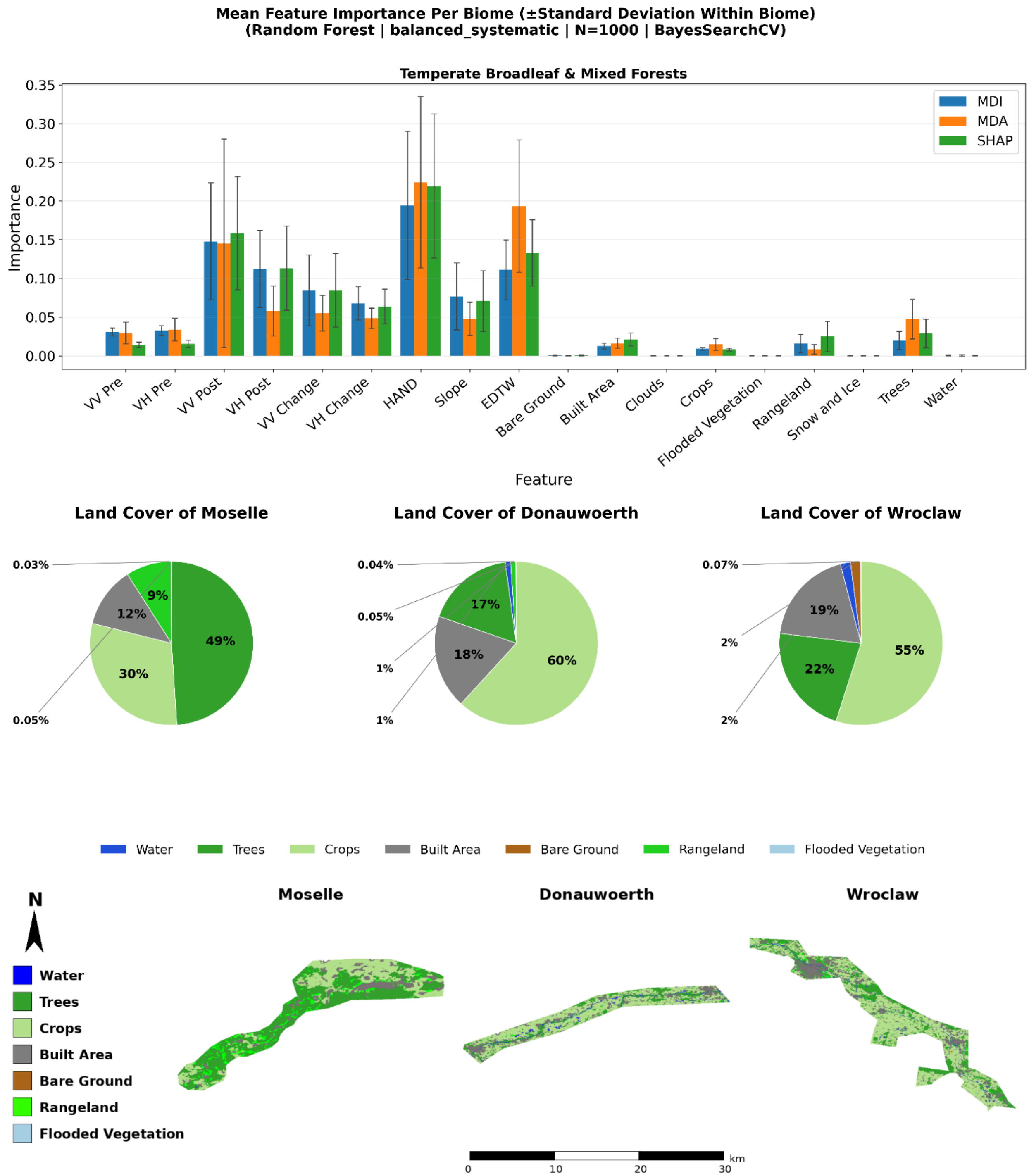

Figure 4.15. Feature importance and land cover in the Temperate Broadleaf and Mixed Forests biome.

The Temperate Broadleaf and Mixed Forests biome showed the most balanced distribution of importance between SAR-derived and contextual variables and (together with the Mediterranean biome) was one of the two biomes in which a terrain descriptor outranked the SAR backscatter features. HAND emerged as the single most important predictor under all three attribution methods (MDI ≈ 0.20, MDA ≈ 0.22, SHAP ≈ 0.22), with EDTW also among the most important features (second only to HAND under MDA, ≈ 0.19). Of the SAR features, VV Post was the strongest (≈ 0.15), followed by VH Post (≈ 0.06–0.11), while the temporal-change feature VV Change retained moderate importance (≈ 0.09).

The leading role of HAND and EDTW reflects the riverine nature of the flood events studied: Moselle, Stuttgart, and Wroclaw all represent major lowland European river floods, in which inundation extent is governed primarily by topographic proximity to the drainage network rather than by local backscatter contrast (Rennó et al., 2008). The mixed forest-agricultural land cover of these sites (Moselle (trees 49%, crops 30%) and Wroclaw (trees 22%, crops 55%)) contributed a moderate land-cover importance (≈ 0.05–0.09), as cropland and forested pixels exhibit distinct SAR backscatter responses under flood conditions (Maxwell et al., 2018).

The temporal-change features (VV Change and VH Change) showed moderate importance (higher than in the desert biome) suggesting that in temperate forest environments, where pre-event backscatter varies seasonally with vegetation phenology, the relative change between pre- and post-event acquisitions provides discriminative information beyond the absolute post-event level (Tsyganskaya et al., 2018).

The three attribution methods diverged in an informative way. All three ranked HAND first, but MDA placed markedly greater weight on the terrain descriptors (ranking EDTW second (≈ 0.19) while simultaneously down-weighting VH Post (MDA ≈ 0.06 versus MDI ≈ 0.11). This is characteristic of permutation importance, which credits a feature only with the unique predictive information lost when it is shuffled: where VH Post is partially redundant with the correlated VV Post, its permutation importance falls, whereas the terrain descriptors (largely uncorrelated with the SAR backscatter) retain their full individual contribution (Altmann et al., 2010). For HAND, MDA and SHAP agreed closely and both slightly exceeded MDI; for EDTW, MDA stood out as substantially higher than either MDI or SHAP, which were close to each other (Strobl et al., 2007). SHAP fell between the extremes, and its ranking (HAND > VV Post > EDTW > VH Post) offers the most balanced and theoretically consistent summary for this biome (Lundberg & Lee, 2017).

4.2.4 Temperate Coniferous Forests

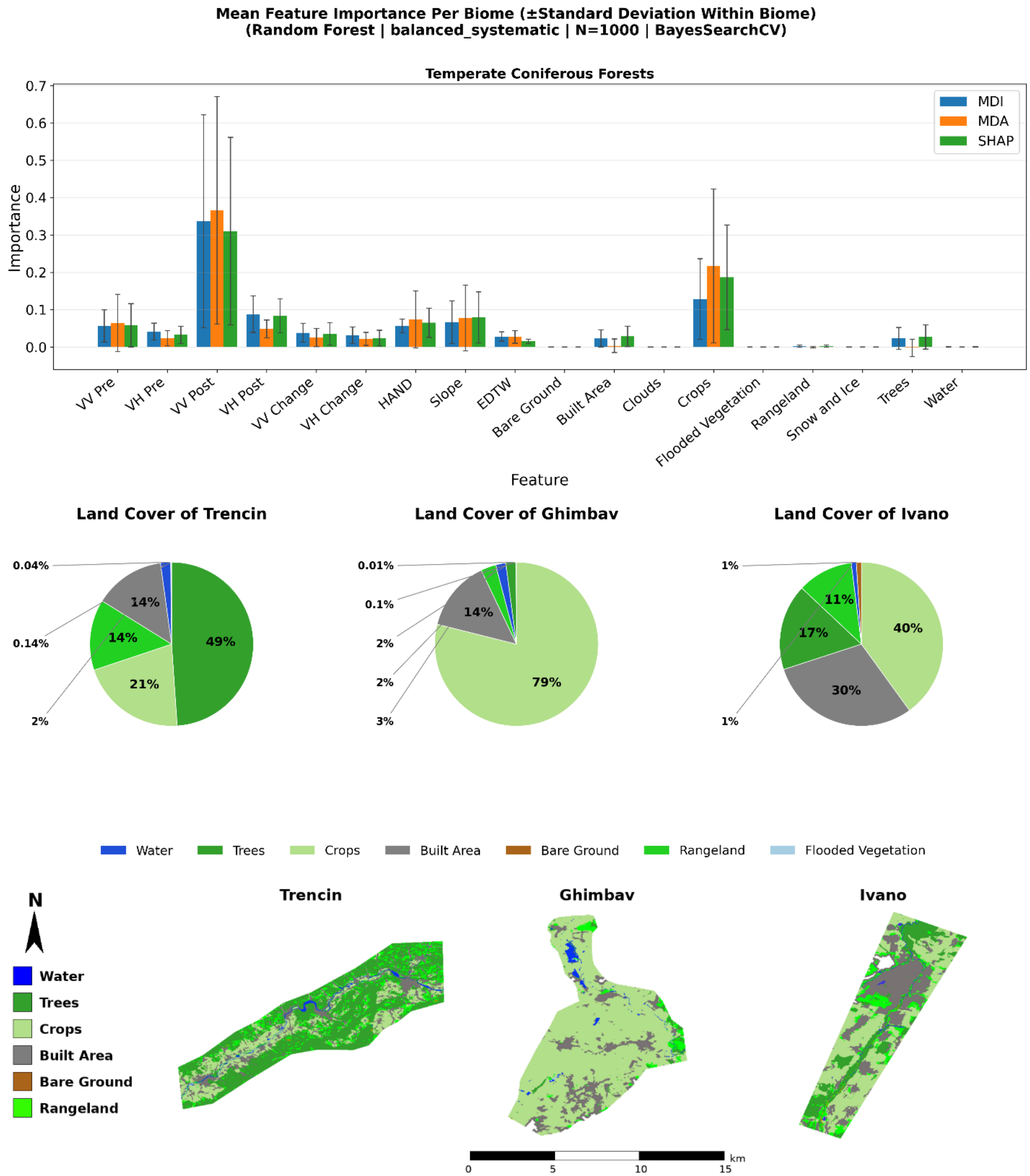

Figure 4.16. Feature importance and land cover in the Temperate Coniferous Forests biome.

The Temperate Coniferous biome showed the most distinctive feature-importance pattern of all five biomes, in two respects. First, VV Post was the single dominant predictor across all three attribution methods (MDI ≈ 0.34, MDA ≈ 0.37, SHAP ≈ 0.31), while its cross-polarized counterpart VH Post was comparatively weak (≈ 0.05–0.09), the largest gap between the co- and cross-polarized post-event channels of any biome. Second, the Crops land cover class emerged as the second most important feature overall (MDA ≈ 0.22, SHAP ≈ 0.19, MDI ≈ 0.13), a prominence not seen in any other biome. The terrain descriptors Slope and HAND followed at modest levels (≈ 0.06–0.08), and the temporal-change features were low.

The elevated importance of the Crops feature is explained by the land-cover composition of the study sites: Ghimbav is dominated by cropland (79%) and Ivano contains 40% cropland. The binary Crops indicator effectively acts as a spatial proxy for the low-lying agricultural floodplains where inundation concentrated, so the Random Forest exploited the strong spatial correlation between cropland and flooding (Maxwell et al., 2018). The same composition helps explain the dominance of VV Post over VH Post: with flooding occurring largely over open agricultural surfaces rather than beneath the coniferous canopy, the co-polarized VV channel (most sensitive to the specular backscatter reduction over smooth standing water) carried most of the flood signal, while the cross-polarized VH channel, dominated by volume scattering from the surrounding vegetation, contributed comparatively little (Tsyganskaya et al., 2018). Consistent with this, HAND was less important here than in the Mediterranean and Temperate Broadleaf biomes, where riverine, topographically controlled flooding prevailed.

The attribution methods diverged most clearly over the Crops feature but not in the direction a simple impurity-bias argument would predict. Rather than MDI inflating the land-cover variable, it was MDA and SHAP that assigned Crops its highest importance (≈ 0.22 and ≈ 0.19), while MDI rated it lower (≈ 0.13). This indicates that the predictive value of cropland is genuine rather than an artefact of impurity-based splitting: permuting the Crops indicator causes a substantial accuracy drop (high MDA), and its game-theoretic contribution is correspondingly large (high SHAP) (Altmann et al., 2010; Lundberg & Lee, 2017). The lower MDI score is consistent with MDI distributing impurity reduction across the correlated, high-cardinality SAR features at the expense of the single binary land-cover indicator (Strobl et al., 2007).

4.2.5 Tropical and Subtropical Moist Broadleaf Forests

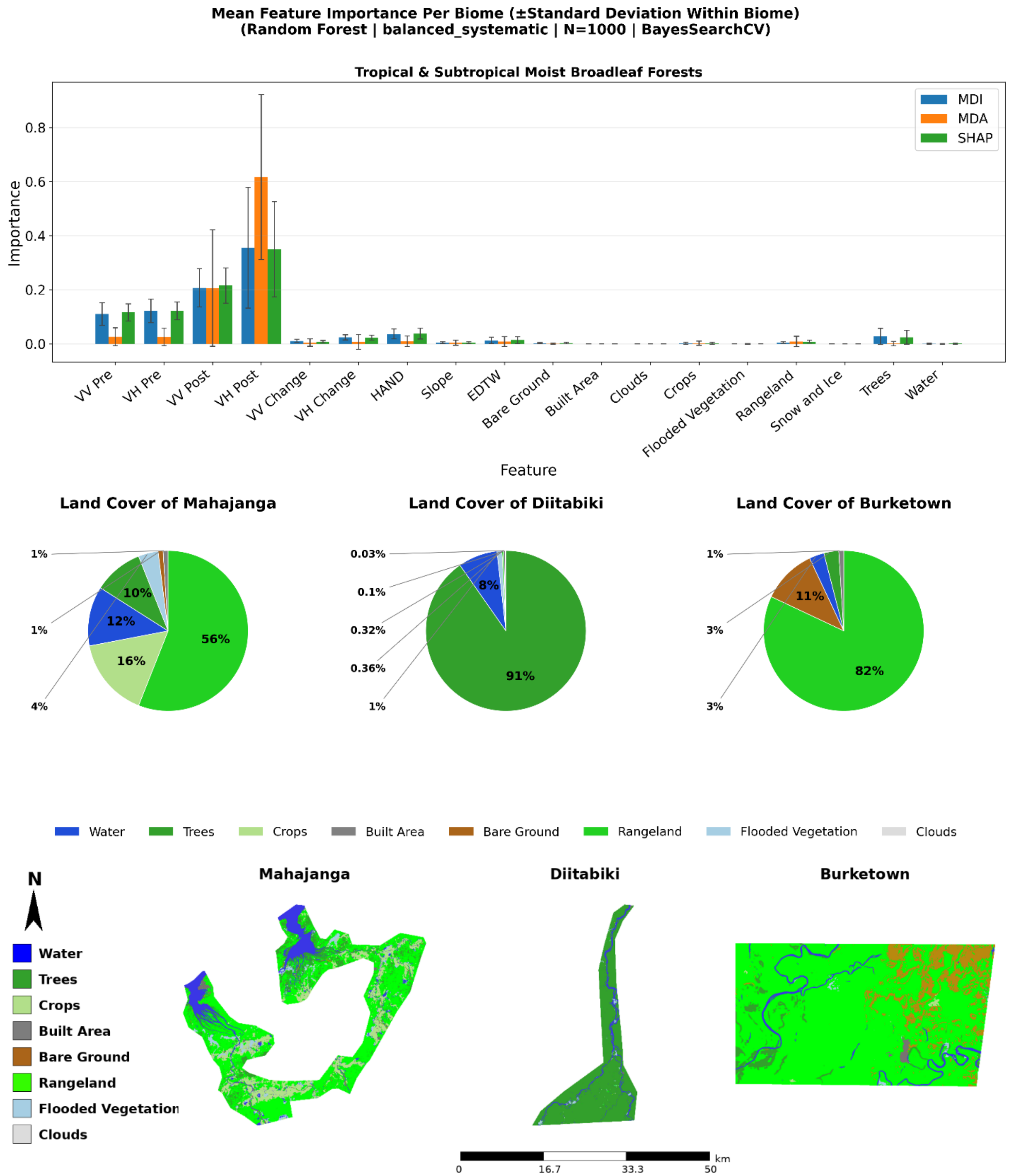

Figure 4.17. Feature importance and land cover in the Tropical and Subtropical Moist Broadleaf Forests biome.

The Tropical and Subtropical Moist Broadleaf Forests biome exhibited the most concentrated importance pattern of all five biomes: VH Post dominated the rankings under all three attribution methods by a wide margin (MDI ≈ 0.36, MDA ≈ 0.62, SHAP ≈ 0.35), with VV Post a distant second (≈ 0.21) and the pre-event channels VH Pre and VV Pre (≈ 0.11–0.12) the only other contributors of note. All contextual variables (HAND, Slope, EDTW, and land cover) showed near-zero importance.

The dominance of VH Post is physically consistent with the role of double-bounce scattering between floodwater and tree trunks in dense tropical forest. The cross-polarized VH channel is particularly sensitive to this mechanism, which produces anomalously high backscatter over flooded forest (rather than the backscatter reduction seen over open water) creating a distinctive and highly discriminative flood signature (Tsyganskaya et al., 2018). The effect was especially pronounced at Diitabiki, where tree cover reaches 91% of the AOI.

The near-zero importance of HAND, Slope, and EDTW reflects the flat topography of tropical lowland floodplains, where weak elevation gradients provide little spatial discrimination between flooded and non-flooded areas. Likewise, the homogeneous land cover of the sites tree-dominated at Diitabiki and rangeland-dominated at Burketown (82% rangeland) contributed to the low land-cover importance.

The three attribution methods agreed on the ranking more closely than in any other biome, all identifying VH Post and VV Post as the two dominant predictors. They differed, however, in magnitude: MDA concentrated importance even more heavily on the single dominant feature, assigning VH Post ≈ 0.62 against ≈ 0.35 for MDI and SHAP. This pattern (strong ranking agreement when one feature clearly dominates, contrasted with the larger divergences in the more evenly balanced Mediterranean and Coniferous biomes) is consistent with the behaviour of permutation importance, which amplifies a single decisive predictor while suppressing partially redundant ones (Altmann et al., 2010).

4.3 Feature Importance Rankings (Inter-Biome Analysis)

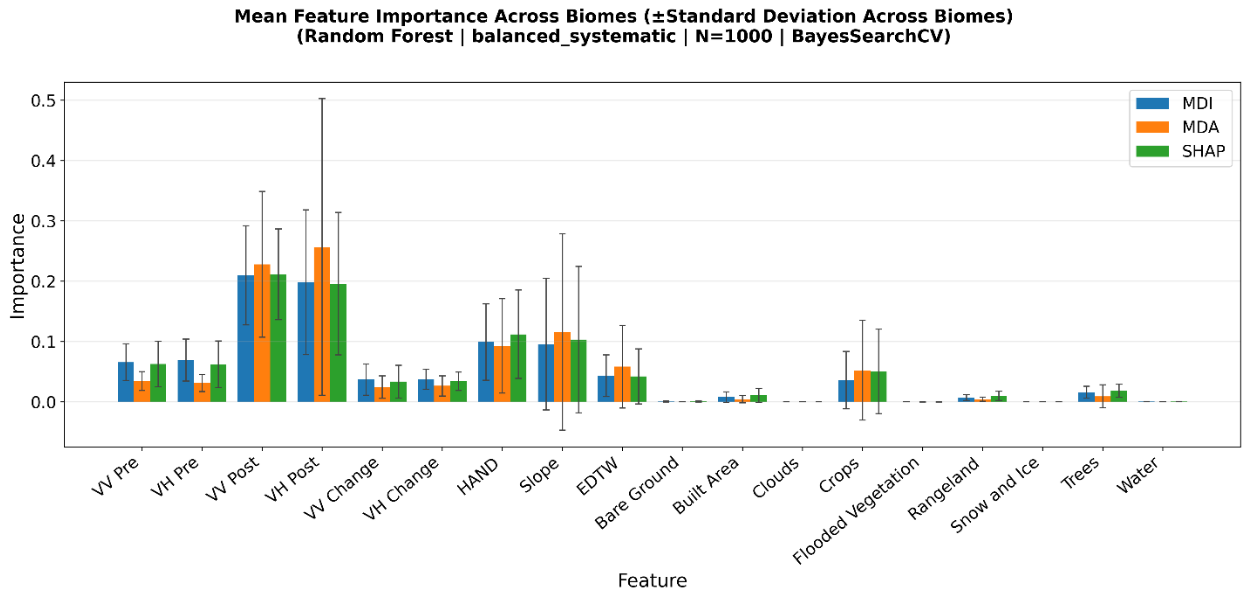

Figure 4.18 Mean feature importance across biomes for the MDI, MDA, and SHAP attribution methods; error bars show the standard deviation across biomes

Figure 4.18 presents the mean feature importance scores averaged across all five biomes, with standard deviation bars reflecting cross-biome variability. This inter-biome comparison enables assessment of which features show stable rankings regardless of environmental context and which are strongly biome-dependent.

4.3.1 Stable Features Across Biomes

Post-event SAR backscatter features (VV Post and VH Post) consistently emerged as the two most important predictors across all five biomes and all three attribution methods (mean ≈ 0.21 and ≈ 0.20, respectively), confirming that post-event C-band backscatter is the primary transferable flood indicator (Twele et al., 2016; Tsyganskaya et al., 2018). Of the two, VV Post was the more consistent across biomes (the shorter error bar), whereas VH Post, though equally important on average, varied far more between biomes (Section 4.3.2) and therefore cannot be regarded as equally transferable.

Pre-event backscatter features (VV Pre and VH Pre) ranked consistently as stable secondary predictors across all biomes, with mean importances of ≈ 0.06 and low variance, providing the dry-condition reference baseline that helps the classifier separate genuine flood-induced change from background backscatter variation (Chini et al., 2017).

Among the contextual variables, HAND was the most consistently important, with a mean of ≈ 0.10 and comparatively low cross-biome variability, confirming the value of drainage-relative elevation as a broadly transferable topographic flood-susceptibility predictor (Rennó et al., 2008) although its importance still rose markedly in the riverine Temperate Broadleaf and Mediterranean biomes.

4.3.2 Variable Features Across Biomes

The features with the largest cross-biome variability were not the contextual descriptors but two predictors that swing strongly with biome type. VH Post showed the highest variability of all features (standard deviation ≈ 0.12 under SHAP and up to ≈ 0.25 under MDA): it dominated the Tropical biome through double-bounce scattering in flooded forest yet fell to low importance in the Coniferous and Mediterranean biomes (Tsyganskaya et al., 2018). Slope showed the next-highest variability, with high importance in the steep-terrain Mediterranean sites but near-zero importance on the flat plains of the Tropical and Desert biomes, confirming that slope should be treated as a biome-specific rather than a transferable predictor.

EDTW, by contrast, was relatively stable but consistently low in importance, contributing little in any biome except for a modest role in the riverine Temperate Broadleaf events; it is therefore a weak predictor rather than a highly variable one. The Crops land-cover class represented the clearest single-biome anomaly: negligible in four biomes, its importance rose sharply in the Temperate Coniferous biome, where the cropland-dominated sites (Ghimbav 79%, Ivano 40%) allowed it to act as a spatial proxy for flood-prone agricultural lowlands (Maxwell et al., 2018) a biome-specific correlation that should not be mistaken for a transferable flood predictor.

Finally, the temporal-change features (VV Change and VH Change) were consistently low and stable across all biomes (mean ≈ 0.04). Despite their theoretical appeal as direct indicators of flood-induced backscatter change, they were systematically outranked by the absolute post-event backscatter, most likely because the post-event values already encode the same flood signal and are preferentially selected at node splits (Strobl et al., 2007).

4.4 Comparison of Attribution Methods

The three attribution methods (MDI, MDA, and SHAP) produced broadly consistent rankings for the top-ranked features across all biomes but exhibited systematic divergences for mid- and low-ranked features that reflect their distinct theoretical foundations and computational mechanisms.

For the dominant features VV Post and VH Post, all three methods produced comparable importance scores across biomes, confirming the robustness of these features as primary flood predictors regardless of the attribution approach used. This convergence for high-importance features is consistent with previous comparative studies demonstrating that attribution methods tend to agree when one or two features strongly dominate the prediction (Lundberg & Lee, 2017).

The most systematic divergence between methods was observed for MDI versus MDA and SHAP for contextual variables. MDI consistently assigned higher relative importance to continuous SAR features and lower importance to binary land cover indicators compared to MDA and SHAP. This pattern is consistent with the well-documented bias of impurity-based importance measures toward high-cardinality continuous features, which offer more potential split points and thus more opportunities to reduce node impurity during tree construction (Strobl et al., 2007). In contrast, MDA and SHAP (being model-agnostic and post-hoc methods respectively are less susceptible to this bias and provided more balanced importance estimates across feature types.

MDA assigned notably higher importance to VH Post relative to MDI in several biomes, particularly in the Desert and Tropical biomes. This divergence reflects the sensitivity of permutation-based importance to the unique predictive contribution of each feature: when VH Post values are permuted, the resulting accuracy drop is large because VH Post captures flood signatures that are not fully redundant with VV Post, despite their moderate correlation (Pearson $r \approx 0.67$). MDI, by contrast, distributes impurity reduction across correlated features, potentially underestimating the individual contribution of VH Post (Altmann et al., 2010).

SHAP values provided the most theoretically grounded and interpretable feature attributions across all biomes. By satisfying the consistency property (which guarantees that if a feature's contribution to model predictions increases, its SHAP value never decreases) SHAP enabled reliable cross-biome comparisons of feature importance that are not confounded by differences in model structure or feature correlations between biomes (Lundberg & Lee, 2017). The close agreement between SHAP and MDA for most features, combined with their shared divergence from MDI for contextual variables, suggests that permutation-based and game-theoretic attribution approaches provide more robust importance estimates than impurity-based measures for heterogeneous multi-source feature sets such as those used in this study.

Chapter 5: Conclusions

5.1 Summary of Main Findings

This study investigated the stability of feature-importance rankings derived from Random Forest-based flood mapping across five global biomes, using thirty Sentinel-1 SAR acquisitions (a pre- and post-event pair for each of fifteen flood events) as input data. Three feature-attribution methods (Mean Decrease in Impurity (MDI), Mean Decrease in Accuracy (MDA), and SHAP values) were applied to assess both intra-biome and inter-biome consistency of feature rankings.

Across the five selected biomes, post-event SAR backscatter features (particularly VV Post and VH Post) were consistently among the most important predictors of flood extent, confirming the central role of C-band backscatter change as a physical indicator of flooded surfaces (Chini et al., 2017; Tsyganskaya et al., 2018). They were the leading predictors in three of the five biomes (Desert, Tropical, and Temperate Coniferous). This dominance was not universal, however: in the Mediterranean and Temperate Broadleaf biomes, terrain descriptors (Slope and HAND, respectively) outranked the SAR features. Between the two polarizations, VV Post was the more consistent across biomes (the single most transferable feature) whereas VH Post, though equally important on average, varied far more between biomes.

Pre-event backscatter features (VV Pre and VH Pre) ranked consistently as secondary predictors, reflecting their role as reference baselines for change detection. The temporal-change features (VV Change and VH Change), although theoretically capturing the most direct flood signal, showed consistently low importance: the absolute post-event backscatter appears to subsume their contribution, since both encode the same flood-induced reduction and the post-event features are preferentially selected at node splits (Strobl et al., 2007).

5.2 Biome-Specific Patterns

Although post-event SAR backscatter was important in every biome, the single most important feature) and the relative weight of the contextual variables (differed markedly between biomes.

In **Deserts and Xeric Shrublands**, VH Post emerged as the single most important feature across all three attribution methods, with VV Post a close second. Although flood detection in arid environments is widely recognized as challenging because of the "water look-alike" behavior of dry sand surfaces (Garg et al., 2024), the Random Forest found VH-derived post-event backscatter to be the most influential predictor, suggesting that the ensemble can exploit subtle VH backscatter variations that simple thresholding approaches may miss. HAND was the most important contextual variable but remained modest in absolute terms, while Slope and EDTW were low consistent with the flat alluvial plains of Jask and Rasoo Nagar, where flash flooding is not strongly controlled by topography.

In **Mediterranean Forests, Woodlands and Scrub**, Slope was the single most important feature across all three attribution methods, followed by HAND (the only biome in which slope dominated). The post-event SAR features VV Post and VH Post ranked below these terrain descriptors. This reflects the steep, complex terrain of the study sites, where hillslope gradient and drainage-relative elevation strongly control where floodwater accumulates (Marchi et al., 2010; Rennó et al., 2008). Land cover showed a modest importance, consistent with the heterogeneous landscape mosaic of Valencia (crops 18%, rangeland 44%, trees 23%) and Magnesia (crops 46%, trees 25%).

In **Temperate Broadleaf and Mixed Forests**, HAND was the most important feature overall, with EDTW also prominent (especially under MDA). This is consistent with the riverine nature of the floods studied, where inundation extent is strongly controlled by topographic proximity to drainage channels (Rennó et al., 2008). The mixed forest–agricultural land cover of Moselle (trees 49%, crops 30%) and Wroclaw (trees 22%, crops 55%) reflects the typical Central European floodplain landscape.

In **Temperate Coniferous Forests**, VV Post was the dominant feature while its cross-polarized counterpart VH Post was comparatively weak (the largest gap between the two post-event channels of any biome). The Crops land-cover class was the second most important feature, most strongly in the MDA and SHAP rankings, explained by the cropland-dominated sites Ghimbav (crops 79%) and Ivano (crops 40%), where the Crops indicator acts as a spatial proxy for flood-prone agricultural lowlands (Maxwell et al., 2018). The weakness of VH Post is consistent with the setting of these events: because inundation occurred largely over open agricultural surfaces rather than beneath the coniferous canopy, the co-polarized VV channel captured most of the flood signal, while the cross-polarized VH return was dominated by volume scattering from the surrounding vegetation and contributed comparatively little (Tsyganskaya et al., 2018).

In **Tropical and Subtropical Moist Broadleaf Forests**, VH Post dominated the rankings with the highest importance scores observed across all biomes and methods. The dense evergreen canopy of this biome (reflected in the land cover of Diitabiki (trees 91%) and Mahajanga (rangeland 56%, trees 10%)) strongly attenuates the C-band signal, yet the cross-polarized VH channel retains sensitivity to double-bounce scattering between floodwater and tree trunks (Tsyganskaya et al., 2018). HAND and EDTW showed near-zero importance, consistent with the flat terrain of tropical lowland forests, where topographic gradients provide little discriminative power.

5.3 Consistency Across Attribution Methods

A key finding of this study concerns the degree of agreement between the three feature-attribution methods. Across all biomes, MDI, MDA, and SHAP produced broadly consistent rankings for the top-ranked feature within each biome, confirming the robustness of the leading predictors. Notable divergences appeared for lower-ranked features, particularly the contextual variables HAND, EDTW, and land cover.

MDA assigned higher importance to VH Post relative to MDI and SHAP in several biomes, especially the Desert and Tropical biomes, reflecting the sensitivity of permutation-based methods to a feature's unique predictive contribution: permuting VH Post produces a large accuracy drop because it captures flood signatures not fully redundant with VV Post, despite their moderate correlation (Pearson $r \approx 0.67$). MDI, by contrast, distributes impurity reduction across the correlated features, while SHAP (satisfying consistency and local accuracy) provided the most theoretically grounded and balanced attribution, generally falling between MDI and the more extreme MDA estimates (Altmann et al., 2010; Lundberg & Lee, 2017).

The standard-deviation bars in the inter-biome comparison reveal that VH Post and Slope exhibited the highest cross-biome variability, indicating that these features are strongly context-dependent and should be interpreted with caution when transferring RF models across geographic settings; EDTW, by contrast, was low in importance but relatively stable.

5.4 Implications and Recommendations

The results of this study have several practical implications for the development and interpretation of RF-based flood-mapping systems. First, the consistently high importance of post-event SAR backscatter (and its role as the leading predictor in most biomes) supports the use of single-date post-event imagery as a minimum viable input for operational flood mapping, even without pre-event reference data. Second, the strongly biome-dependent behavior of the contextual variables (Slope, HAND, and land cover) shows that feature-importance rankings should not be assumed to transfer directly across geographic settings without validation.

From a methodological perspective, this study demonstrates that the choice of attribution method influences feature-importance rankings, particularly for correlated and low-importance features. The use of multiple complementary attribution methods, as implemented here, is therefore recommended for robust feature interpretation in remote-sensing applications (Lundberg & Lee, 2017; Strobl et al., 2007).

Future research should extend this analysis to additional biomes, including Boreal Forests and Temperate Grasslands, and incorporate a larger number of flood events per biome to improve the statistical robustness of intra-biome variability estimates. The inclusion of alternative machine-learning architectures, such as gradient boosting or deep-learning models, would further broaden the generalizability of the findings.

References

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. Bioinformatics, 26(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Borga, M., Stoffel, M., Marchi, L., Marra, F., & Jakob, M. (2014). Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows. Journal of Hydrology, 518, 194–205. https://doi.org/10.1016/j.jhydrol.2014.05.022

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chini, M., Hostache, R., Giustarini, L., & Matgen, P. (2017). A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case. IEEE Transactions on Geoscience and Remote Sensing, 55(12), 6975–6988. https://doi.org/10.1109/TGRS.2017.2737664

Copernicus EMS. (2025). Copernicus Emergency Management Service Rapid Mapping. European Commission. https://mapping.emergency.copernicus.eu/

Dasgupta, A., Grimaldi, S., Ramsankaran, R., Pauwels, V. R. N., & Walker, J. P. (2018). Towards operational SAR-based flood mapping using neuro-fuzzy texture-based approaches. Remote Sensing of Environment, 215, 313–329. https://doi.org/10.1016/j.rse.2018.06.019

European Space Agency. (2021). Copernicus DEM: Copernicus Digital Elevation Model. ESA. https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model

European Space Agency. (2024). Sentinel-1 overview. ESA Copernicus. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1

Garg, S., Dasgupta, A., Motagh, M., Martinis, S., & Selvakumaran, S. (2024). Unlocking the full potential of Sentinel-1 for flood detection in arid regions. Remote Sensing of Environment, 315, 114417. https://doi.org/10.1016/j.rse.2024.114417

Google Developers. (2025). Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected. Google Earth Engine. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD

Grimaldi, S., Xu, J., Li, Y., Pauwels, V. R. N., & Walker, J. P. (2020). Flood mapping under vegetation using single SAR acquisitions. Remote Sensing of Environment, 237, 111582. https://doi.org/10.1016/j.rse.2019.111582

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239

https://doi.org/10.1109/IGARSS47720.2021.9553499

Landuyt, L., Van Wesemael, A., Schumann, G. J.-P., Hostache, R., Verhoest, N. E. C., & Van Coillie, F. M. B. (2019). Flood mapping based on synthetic aperture radar: An assessment of established approaches. IEEE Transactions on Geoscience and Remote Sensing, 57(2), 722–739. https://doi.org/10.1109/TGRS.2018.2860054

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

https://doi.org/10.1038/s42256-019-0138-9

Marchi, L., Borga, M., Preciso, E., & Gaume, E. (2010). Characterisation of selected extreme flash floods in Europe and implications for flood risk management. Journal of Hydrology, 394(1–2), 118–133. https://doi.org/10.1016/j.jhydrol.2010.07.017

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 39(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

https://doi.org/10.1186/s40663-019-0170-6

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. IEEE Geoscience and Remote Sensing Magazine, 1(1), 6–43. https://doi.org/10.1109/MGRS.2013.2248301

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R.

(2001). Terrestrial ecoregions of the world: A new map of life on Earth. BioScience, 51(11), 933–938.
https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

OpenStreetMap contributors. (2025). OpenStreetMap.
https://www.openstreetmap.org

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., & Waterloo, M. J. (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. Remote Sensing of Environment, 112(9), 3469–3481.
https://doi.org/10.1016/j.rse.2008.03.018

Richards, P. W., Walsh, R. P. D., Baillie, I. C., & Greig-Smith, P. (1996). The Tropical Rain Forest: An Ecological Study (2nd ed.). Cambridge University Press.https://books.google.com/books?isbn=9780521421942

Salem, T., & Hashemi-Beni, L. (2022). Flood extent mapping: An integrated method using deep learning and region growing using SAR imagery. ISPRS International Journal of Geo-Information, 11(5), 279.
https://doi.org/10.3390/ijgi11050279

Senf, C., & Seidl, R. (2018). Natural disturbances are spatially diverse but temporally synchronized across temperate forest landscapes in Europe. Global Change Biology, 24(3), 1201–1211.
https://doi.org/10.1111/gcb.13897

Shen, X., Wang, D., Mao, K., Anagnostou, E., & Hong, Y. (2019). Inundation extent mapping by synthetic aperture radar: A review. Remote Sensing, 11(7), 879.
https://doi.org/10.3390/rs11070879

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. Advances in Neural Information Processing Systems, 25, 2951–2959.
https://proceedings.neurips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics, 8(1), 25.
https://doi.org/10.1186/1471-2105-8-25

Stevens, D. L., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. Journal of the American Statistical Association, 99(465), 262–278.
https://doi.org/10.1198/016214504000000250

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., & Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. Nature, 596(7870), 80–86. https://doi.org/10.1038/s41586-021-03695-w

Tsyganskaya, V., Martinis, S., Marzahn, P., & Ludwig, R. (2018). SAR-based detection of flooded vegetation: A review of characteristics and approaches. International Journal of Remote Sensing, 39(8), 2255–2293. https://doi.org/10.1080/01431161.2017.1420938

Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: A fully automated processing chain. International Journal of Remote Sensing, 37(13), 2990–3004. https://doi.org/10.1080/01431161.2016.1192304

Uhe, P., Hawker, L., Neal, J., Bates, P., & Sampson, C. (2025). FathomDEM: A global high-resolution bare-earth digital terrain model. Scientific Data. https://doi.org/10.5281/zenodo.10067117

Urbazaev, M., Cremer, F., Migliavacca, M., Reichstein, M., Schmullius, C., & Thiel, C. (2022). Potential of multi-temporal ALOS-2 PALSAR-2 ScanSAR data for vegetation height estimation. Remote Sensing, 14(13), 3172. https://doi.org/10.3390/rs14133172

Whitmore, T. C. (1998). An Introduction to Tropical Rain Forests (2nd ed.). Oxford University Press. https://books.google.com/books?isbn=9780198501473