Delft University of Technology

MSc Graduation Thesis
Environmental Engineering

Land Cover Change and Hydroclimatic Deviation: A Detailed Examination within the Budyko Framework

Author:
Michael O'Hanrahan (5627796)

Thesis Committee	Dr. Markus Hrachowitz	TU Delft
	Dr. Miren Vizcaino	TU Delft
	Eric Sprokkereef	Rijkswaterstaat
	Jasper Stam	Rijkswaterstaat
	Dr. Laurène Bouaziz	Deltares

September 21, 2023

TU Delft

Rijkswaterstaat
Ministerie van Infrastructuur en Waterstaat

Abstract

Predicting streamflow in a changing climate poses significant challenges for traditional hydrological models. Static parameter sets result from model calibrations over historical data that increasingly encounter the non-stationary impacts on the hydrological system. Endeavouring toward non-stationary model parameters by incorporating time-adaptive ecosystem-scale root zone storage parameters shows promise for modelling systems under change. The ongoing CATAPUC project aims to further develop, refine, and implement this adaptive modelling approach. This paper continues to build upon this body of work by investigating the evidence for land cover change impacts on root zone storage capacity, focusing specifically on uncertainties in the evaporative balance. Combining long-term hydrometeorological data from the primary study area, the Meuse basin, with the large-sample CAMELS datasets (GB and US) analyses were performed across 283 catchments. Applying the vector operations from Velde et al. (2014) and Jaramillo et al. (2018) to decadal changes within the Budyko framework, we separate climate-related evaporative changes. We isolated the residual component of evaporative change, which is the unknown component affecting parameter estimations. Our findings indicate that this residual component is twice as prominent in evaporative change as the climate component. The aim is to test if land cover changes have contributed significantly to the residual component of evaporative change. A multi-scale approach to land cover change analysis is adopted to bridge the data gap from 1984 to 2019. Implementing ensemble machine-learning methods on Landsat imagery with Google Earth Engine, we develop annual timeseries of (30m) high-resolution multi-class data for this period with accuracies up to 86% for 117 catchments (Meuse and GB). Resulting land cover change estimates suggest that Meuse Basin urbanisation rates may have been significantly underestimated for this period. The Meuse Basin over the most recent 20-year period is found to deviate anomalously in actual evaporation compared with the large sample. A distinct spatial pattern reveals a concentration of deviations in the east of the basin. Calculated by Tempel (2023), the anomaly in the basin corresponds with a relative median error in root zone storage capacity change of -14%. Low flow analysis is performed to remove the possibility that deviations are affected by anomalous contribution to streamflow from additional subsurface flow. We observed that the increase in low flow variability over the same period exhibits a spatial pattern similar to the Meuse Basin anomaly. We found no meaningful causal relationships linking multi-scale interdecadal changes in forest, agriculture, and urban land classes to the observed deviations in the large sample datasets. The implication of this study is that land cover change is likely not a significant driver of evaporative changes, specifically, errors, throughout the record available.

I Acronyms

Table 1: Acronyms with Definitions (Alphabetical Order)

Acronym	Definition
AI	Aridity Index
B	Blue
CAMELS	Catchment Attributes and Meteorology for Large-sample
CATAPUC	Climate Analogy Mapping for Temporally Adaptive Parameters Under Change
DEM	Digital Elevation Model
E-OBS	ENSEMBLES gridded observational dataset
EI	Evaporative Index
EVI	Enhanced Vegetation Index
EWT	Equivalent Water Thickness
FSN	Forest and Semi-Natural
G	Green
GB	CAMELS GB
GEE	Google Earth Engine
HC	Hydroclimatic
HM	Hydrometeorological
HILDA+	Historic Land Dynamics Assessment plus
LAI	Leaf Area Index
LCC	Landcover Change
LULCC	Land-Use Land Cover Change
MIR	Mid Infrared
NIR	Near Infra-red
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
NRFA	National River Flow Archive
R	Red
RS	Remote Sensing
SI	Seasonality Index
SMD	Soil Moisture Deficit
SM	Soil Moisture
SLA	Service Level Area
SWIR	Shortwave Infrared
TCA	Tasselled-Cap Angle
TCB	Tasselled-Cap Brightness
TCG	Tasselled-Cap Greenness
TCW	Tasselled-Cap Wetness
TM	Thematic Mapper
US	United States

II Variables

Table 2: Variables with Definitions and Units

Variable	Definition	Units
P	Precipitation	mm/year
Q	Discharge (streamflow)	mm/year or m^3/year
S	Storage	mm^3
E_a	Actual Evaporation	mm/year
E_p	Potential Evaporation	mm/year
T	Temperature	°C
AI	Aridity Index	Dimensionless
EI	Evaporative Index	Dimensionless
ΔEI_r	Residual Change Component	Dimensionless
ΔEI_r	Climatic Change Component	Dimensionless
R_n	Net Incoming Radiation	W/m^2
K	Daily Incoming Shortwave Radiation	W/m^2
Δ	Slope of Temperature-Vapor Pressure Curve	kPa/K
γ	Psychrometer Constant	kPa/K
G	Soil Heat Flux	W/m^2
e_s	Saturated Vapour Pressure	Pa
e_a	Observed (actual) Vapour Pressure	Pa
ρ_a	Mean Air Density at Constant Pressure	kg/m^3
c_a	Specific Heat Capacity of Air	J/kg/K
λ_v	Latent Heat of Vaporisation	J/g
ρ_w	Bulk Density of Water	kg/m^3

Contents

1 Introduction

The escalating repercussions of climate change profoundly influence the hydrosphere, primarily through augmenting the atmospheric water demand. A surge in global temperatures directly amplifies this demand, with a non-linear increase of approximately 7% for each degree of warming (Boer, 1993). This thermal uptick, in turn, has been linked to an elevation in the global concentration of precipitable water, with an average increase ranging from $7-9\%$ since the mid-20th century (Held and Soden, 2006). Changes in water availability and soil moisture deficits cause plants to adapt root zones impacting river flow significantly, the parameterisation of this change poses a challenge to streamflow modelling and projection (Bouaziz et al., 2021, 2022; Gao et al., 2014; Milly, 1994; van Noppen, 2022; Tempel, 2023).

Amplifying the gravity of the situation, the Intergovernmental Panel on Climate Change (IPCC) unequivocally asserts that anthropogenic climate change has led to an increased frequency and intensity of extreme precipitation events, expressing this assertion with a high degree of confidence (Lee et al., 2023; Allen et al., 2010). Compounding this, the confluence of intensified precipitation and rapid urbanisation has been linked to heightened risks for low-lying coastal communities and livelihoods, emphasising the anthropogenic impact exacerbating the vulnerability of these areas (Lee et al., 2023; Jian et al., 2021).

Existing within this interconnected network, the soil moisture-vegetation complex is a critical hydrological interface that facilitates connections between land surface, subterranean, and atmospheric processes, mediated by land-atmosphere feedback (Milly and Dunne, 1994). It presides over crucial functions such as evaporation, transpiration, infiltration, runoff, and flow through the vadose zone (Robinson et al., 2008; Babaeian et al., 2019). Soil moisture is a central influence on transpiration, underscoring its importance for vegetation vitality, especially during dry periods.

Non-stationarity of hydroclimatic systems in the context of climate change is indeed an issue and a well-recognized challenge in hydrology changing the frequency, magnitude and timing of extremes. For example, Blöschl et al. (2017) presents compelling evidence for climate-driven changes in precipitation seasonality over Western Europe that earlier soil moisture maxima lead to earlier and exacerbated winter floods. Traditional hydrological model calibration focuses on past behaviours to achieve short-term streamflow forecasting and longer-term prediction of future behaviour. Static parameters in the context of non-stationary climate are contradictory and result in significant uncertainties (Bouaziz et al., 2021).

The pivotal role of vegetation cover is integral to regulating soil moisture extraction via evapotranspiration. Vegetation, by necessity, maintains evapotranspiration during dry periods, thereby adapting to soil moisture deficit conditions to meet canopy water demand and sustain ecosystem functionality (Milly, 1994; Rodriguez-Iturbe, 2000). Rooting depths are observed to extend approximately 2m below the surface, however, variations are largely attributable to local environmental conditions with the shallowest rooting depths (less than 100 mm) occurring in wet climates with short inter-storm durations (Stone and Kalisz, 1991; Nepstad et al., 1994). In contrast, regions characterised by higher aridity and longer dry periods necessitate deeper roots (Gao et al., 2014). Consequently, incorporating the root zone storage capacity as a parameter in hydrological models is essential for reducing predictive uncertainty.

The root zone storage capacity is a critical parameter in hydrological and land surface modelling. However, it cannot be directly observed at the catchment or ecosystem scale. As such model-independent, ecosystem-scale estimations of the root zone storage capacity are derived from long-term (LT, $\geq 10y$) hydro-meteorological datasets (Gao et al., 2014; Wang-Erlandsson et al., 2016). The benefit of using a model-independent approach foregoes the need for soil or land-cover information. Ecosystem-scale evidence suggests that vegetation adaptive responds to hydro-climatic forcing (Wang-Erlandsson et al., 2016; de Boer-Euser et al., 2019). The root zone depth and vegetation cover extent across the landscape intrinsically determine soil water storage (Desborough, 1997). However, cascading drivers, like heightened atmospheric water demand resulting in drier soils and shifting land uses, present clear non-stationarity for predictive hydrologic models, which traditionally employ stationary defining parameters based on historical records.

Two major, entangled uncertainties remain that impact our ability to project root zone storage capacity (S_r) and more specifically, the underlying actual evaporation. These two uncertainties form the basis for this thesis: *climatic changes* and *land cover changes* and their impact on our ability to project actual evaporation (E_a for determination of S_r). These uncertainties are potentially significant but unquantified in this stream of

research, the 'CATAPUC' project, as it pertains to the Meuse. In this research, we attempt to separate and quantify the non-climate-driven changes (residuals) in actual evaporation and test for any significant impact on this component by changing land cover.

This research builds upon the work of Bouaziz et al. (2021), which investigated how changing climate scenarios could influence root zone storage capacity (S_r) and actual evaporation. Specifically, Bouaziz et al. assumed that catchments would follow their respective trajectories within the Budyko framework under varying climate scenarios. This approach uniquely positions catchments within the Budyko framework based on their long-term water balance, allowing a projection of actual evaporation and, subsequently, root zone storage parameters under over 2K (+2 degrees Kelvin) climate scenarios. For scenario analysis, this effort is useful for showing the sensitivities of different catchments under climate projections.

The effects of land use changes on these projections are accounted for by altering the Fu-parameter ω, as per Fuh (1981). This parameter characterises the shape of the parametric curve within the Budyko framework (outlined in section 4.1), which predicts an evaporative index-based on aridity conditions. By considering future climate changes, a catchment can be projected along the curve associated with its observed ω_{obs} value. However, shifts in land surface properties - including vegetation cover, species composition, forest age, and biomass - can result in changes to the Fu parameter (ω_{change}), as highlighted by Jaramillo et al. (2018). This is a residual error in projection, expressed as a deflection from expectations in the actual evaporation, that complicates the validity of scenario projection.

Bouaziz et al. (2021) used these on-curve projections to run a model demonstrating how the adaptive root zone storage capacity would change under different climatic scenarios, without accounting for land use change (ω_{change}). In this model, the root zone storage capacity ($S_{r,20y}$) increased by 34% compared to historical estimates, due to larger storage deficits in the summer period in a warmer environment. However, when the model incorporated a hypothetical afforestation scenario (with broadleaf forests replacing coniferous ones), the root zone storage capacity decreased slightly by -4%, as broadleaf forests are more water-efficient. This result contrasts with a scenario in which broadleaf forests were replaced by coniferous plantations, resulting in a similar increase in root zone storage capacity (+4%), essentially the inverse of the afforestation scenario. In a dynamic modelling scenario simulating climate-vegetation interactions, future streamflow and groundwater levels decreased by -15% and -10% respectively, with minimal impact from land cover change.

Continuing research spurred the genesis of this CATAPUC project has modeled the implications of climate change on root-zone storage capacity and the subsequent effects on streamflow in the Meuse basin. van Noppen (2022) built upon this work by similarly projecting $2K+$ scenarios above pre-industrial warming conditions, but implementing a time-dynamic root zone storage parameter S_r strategy based on a multiple linear regression relationship using climate analogy principles. The modelled two degree kelvin warming scenario with time-dynamic $S_{r,20y}$ yields decreases in streamflow (Q, -8.6%) and groundwater storage (-4.8%), with increases in actual evaporation (E_a, +6.6%) and increased root zone moisture storage capacity ($S_{r,max}$ +23.6%).

The scenario presented implies future conditions of significant elevated moisture stress, reduced streamflow, and amplified seasonality but doesn't cover low flows or implications of land cover change in the historical data. The predictions are contingent on two assumptions: Firstly, the vegetative species in future climate-analogous catchments are assumed to be the same as that in the future Meuse catchments, such that the Meuse vegetation will equilibrate to the analogy. This is to say that vegetation in the Meuse catchment under change will respond climate changes and emulate the parameter estimation of the catchment's closest analogy. The second assumption is that no human interaction will interfere with this equilibration process. While it is not possible to know with certainty how vegetation will adapt into the future with any great certainty, it however is possible to investigate past changes in land cover and try to attribute their effects to errors observed by quantifying residual errors from the actual evaporation.

In the most recent research preceding this report, Tempel analysed past changes within the Budyko framework to understand the historical implication of inter-decadal climate variability on the root-zone storage capacity and the efficacy of Fu parameter-based projections. Projections are made within a subset of the historic data and compared to subsequent historic data yielding error estimates from the methodology. Actual evaporation in the data is predicted to be smaller than expected. Trends in the evaporative ratio were modelled for streamflow effects and the resulting root zone storage parameters resulting in reduced annual summer evaporation and increased autumn streamflow. Errors in the projection of evaporative ratio are difficult to attribute, and are

becoming more uncertain in a changing climate.

As of yet, there is no direct evidence of whether land cover change has contributed to the residual errors in Budyko framework movements. Further, until recently, there have been data limitations for directly observed, reliable high-resolution land cover change over climatic time scales to address this question. The indication is that big contrasts in forestry incur only minor shifts to the water balance Bouaziz et al. (2021). This is untested with direct land cover change observation and the water-balance methods used in previous studies have cited it as an unquantified uncertainty. What is known is that land use changes since 1960 in Europe and the USA have been primarily by agricultural policies driving the conversion of land to pasture or grain to satisfy global trade (Winkler et al., 2021; Winkler et al., 2020). In Europe, this coincides with the formation of the European Union and the creation of the Common Agricultural Policy in 1957 to fortify European food security and trade through agricultural intensification in the most rural areas (Milczarek-Andrzejewska et al., 2018).

Globally, land cover has been shaped by human intervention over centuries and, in turn, has also influenced climate patterns (Schmidt et al., 2014). Winkler et al. (2021) estimates that global landcover change has affected 32% of the land surface. Vegetation removal or species change occurs at differing paces across the globe, where afforestation is linked to increased evapotranspiration (ET) and reduced streamflow and complicating matters further is the varied ET rates associated with different vegetation species within a catchment (Jaramillo et al., 2018; Buechel et al., 2022)). Concurrently changing climate and evaporative balances affect the evaporative ratio (actual evaporation to precipitation) of a catchment, the contribution of land use vs. climate is still poorly understood and unsatisfactorily quantified.

Separating land surface processes from climatic effects using Budyko decomposition becomes an appealing prospect. The methodology attributing portions of evaporative index EI ratio to changes in land cover, is accomplished usually in the context of afforestation or deforestation Young (2014); Jaramillo et al. (2018). Land cover changes, specifically vegetation surface and biomass changes, throughout history, are related to the evaporative ratio and must be explored to refine better uncertainty in the projection of the root zone storage parameter.

Free and open access to satellite remote sensing from the Landsat missions in a new pre-processed collection (collection 2) has spurred a new wave of land cover change analysis (Woodcock et al., 2008; Jian et al., 2021; Zhu et al., 2016). This reprocessing is enticing to allow us to effectively attempt to attribute a portion of the hydroclimatic uncertainty to land use cover over a significant portion of the Landsat mission since 1984. Remote sensing and its applications in drought assessment and soil moisture deficits allow for some functional relationships to be established between remotely sensed moisture-sensitive products and the water accessible by soil. Moisture-sensitive remote sensing variables effectively determine moisture stress in vegetation, the NDII specifically is found to have a significant linear relationship to equivalent water thickness and root zone storage capacity Sriwongsitanon et al. (2016).

In this study, the aim is to adapt the Landsat data to interdecadal indices on a per-catchment basis to test for causal relationships between land cover changes and residual errors in the Budyko framework. In the absence of the discovery of any causal relationships, the analysis contextualises the Meuse basin against large-sample hydrology datasets. Additionally, the multi-scale land cover change data generation creates new insights into these changes over the period of the study.

1.1 Problem Statement

Land surface alterations and climate change have triggered far-reaching impacts on global hydrological systems, with the effects becoming increasingly pronounced over recent decades. However, there has been an inadequate delineation between the proportional contributions of land use and land cover change (LCC) vis-à-vis climatic changes to the hydrological response. Estimations of future actual evaporation using the Budyko framework are subject to evaporative index deviations resulting in errors in root zone storage capacity and streamflow estimation.

The challenge of quantifying the impacts of these simultaneous land and climate processes necessitates a comprehensive exploration of long-term water balance alterations and the frequency, magnitude, and timing of change. By identifying pertinent climate and land-use descriptors, land cover change observations as a component of error contribution may be assessed.

1.2 Research Questions

This second CATAPUC iteration comprises two theses and is guided by following overarching question:

'What are the individual contributions of climate and land cover changes to root zone storage capacity?'

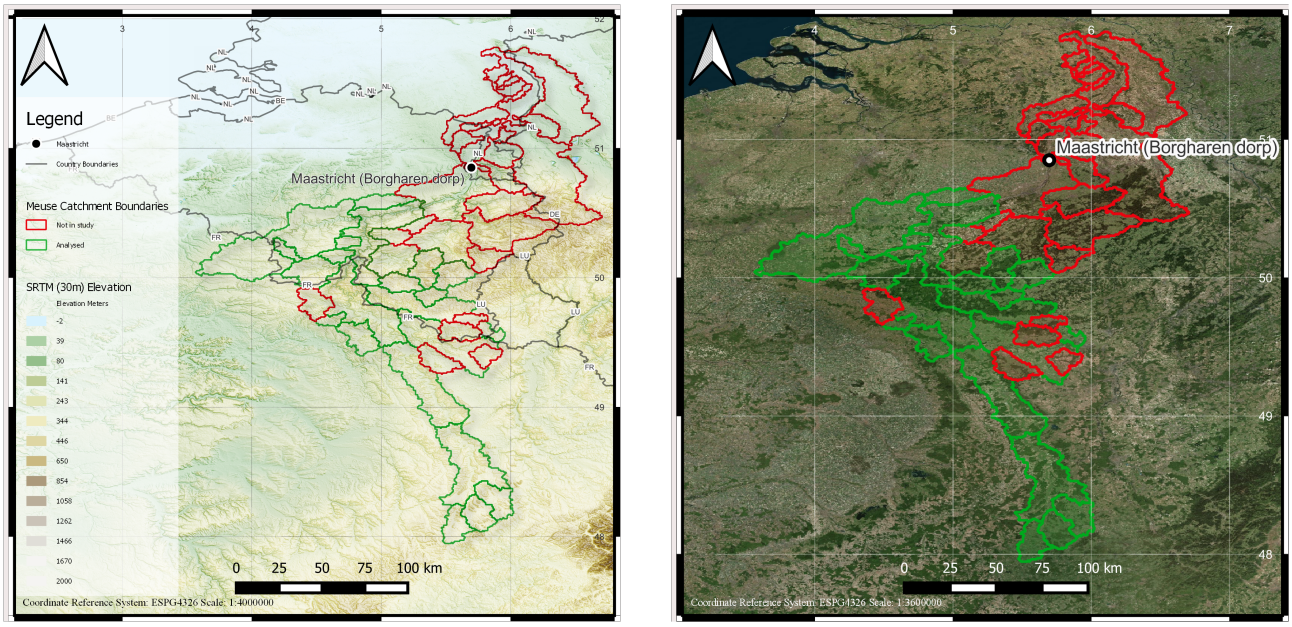
Under this broad question, we should gather data and perform analysis to prove or reject the following hypotheses:

1. *'H1: Budyko framework actual evaporation deviations can be separated effectively as an estimate of climate change and an additional residual effect.'*

2. *'H2: Remote sensing technologies can effectively bridge an existing data gap, to allow land cover observations on a climatic time scale ($>$ 30 years).'*

3. *'H3: The residual component of evaporative index changes, as defined by the Budyko framework, is significantly correlated with observed landcover changes.'*

Tempel (2023) as a recently released parallel thesis topic has focused on the historical climatic changes and the errors in prediction or projection in the Budyko space and models some representative scenarios for implications toward streamflow. The research herein places a greater emphasis on quantifying land cover change portion to compliment the preceding and parallel work but also owes much to, and relies upon data processing efforts by Tempel (2023) and Bouaziz et al. (2022) to enable the results and discussion presented.



(a) (b)

Figure 1: 27 Meuse basin sub-catchments above the Maastricht (Borgharen) gauging station Green catchments (n=22) that have been selected and included in the analysis. Map (a) the Meuse Basin on a digital elevation model (DEM) in meters with country boundaries. The most southerly point at $47.95^{o}N$ and Borgharen at $50.89^{o}N$. Map (b) is the same catchment boundary selection on a (Bing) satellite image.

2 Primary Study Area: The Meuse Basin

The Meuse river basin is situated in Northwest Europe and covers approximately $36,000km^2$. It traverses parts of the Netherlands, Belgium, France, Germany, and Luxembourg. Its origin lies within the Langres Plateau in France, and it travels a course of roughly 925 km before reaching the Rhine-Meuse-Scheldt delta in the Netherlands. The basin is geologically divided into three main zones: the Lorraine Meuse, Ardennes Meuse, and the Dutch and Flemish lowlands. This study focuses on the area of the basin upstream of Maastricht (Borgharen), totalling $21,241km^2$, not including the Dutch lowlands. At a high level, 40.5% of the Meuse basin is dedicated to agriculture, forested areas represent 28% (Copernicus Land Monitoring Service, see appendix for more detailed class area distribution at a low level based on a 2012 example 15, 38). The river is a crucial resource, supplying water to about six million people in Belgium and the Netherlands. With economic growth and population expansion, the demand for water as well as the risk of urban flooding has increased Wit et al. (2007).

Different regions are geologically distinct with the Lorraine Meuse characterized by sedimentary Mesozoic rocks surrounded by hilly landscapes and broad floodplains. The Ardennes Meuse cuts through the Paleozoic rocks of the Ardennes Massif, featuring narrow, steep valleys. On the other hand, the Flemish lowlands and Limburg are underpinned by loosely consolidated sedimentary rocks where the landscape expands to form extensive floodplains and a low river gradient de Wit (2008)Wit et al. (2007) (see maps attached as figure 44 for reference on terrain and location).

Located in a temperate climate zone, the Meuse River experiences frequent weather changes. Its majority rain-fed streamflow shows strong seasonality, with elevated winter flows and low summer flows Wit et al. (2001). Even though precipitation remains relatively constant throughout the year, potential evaporation fluctuates, influencing the seasonal streamflow variation. The summer discharge on average, is one-quarter that of the winter. Snow impact in the basin area is generally low save for some events. Based on historical E-OBS climate and streamflow data, the Meuse basin exhibits approximate annual averages for precipitation, potential evaporation, and streamflow as 968 mm yr^{-1}, 593 mm yr^{-1}, and 397 mm yr^{-1}, respectively (Cornes et al., 2018). Precipitation in the upper reaches of the basin to the low-lying catchments ranges from $1250mmyr^{-1}$ in the Ardennes or Eastern Belgium to $700mmyr^{-1}$ in North Eastern France Wit et al. (2007)Bouaziz et al. (2022).

The catchment selection process is aligned with the parallel research of Tempel (2023) implements and refines

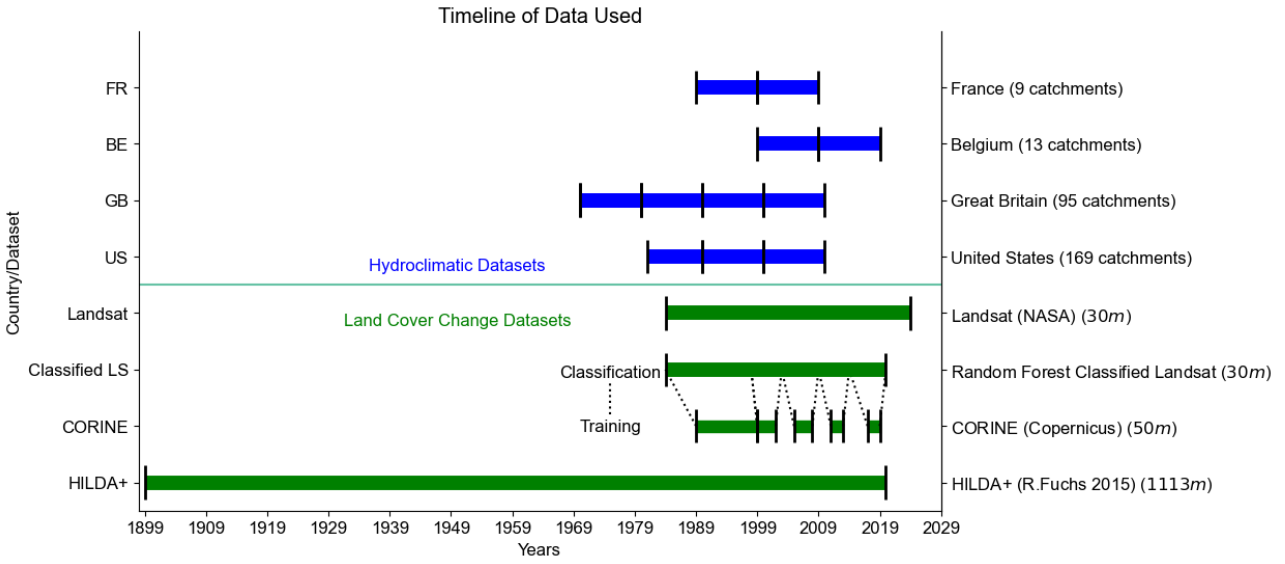


Figure 2: Timeline comparison of datasets employed in the analysis. The blue data (HM timeseries) from four distinct datasets are processed to derive indices for the decadal periods demarcated by black bars. The lower green data pertains to the LCC data showing relative maximum extent of the complimentary timeseries.

the stipulations set by van Noppen (2022). Two primary exclusions for this data is that any streamflow that has been significantly altered (e.g. by canal) has been removed. Additionally, any catchments that experience regular snowfall, which would affect the water balance calculations, have similarly been excluded from the analysis.

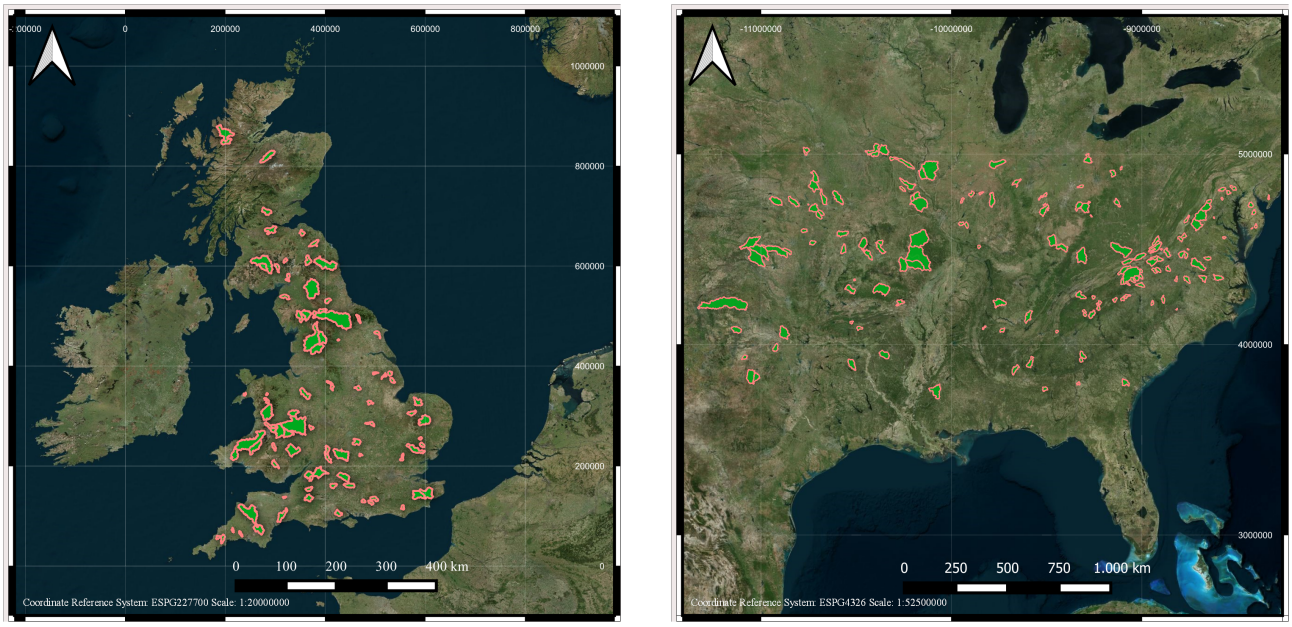
3 Data

The data examined in this research fall into two main categories: hydrometeorological timeseries (encompassing precipitation, evaporation, discharge - 'HM') and remotely-sensed imagery and spatial data, which are processed into complementary timeseries to provide observations on land cover change (LCC). These are transformed into hydroclimatic indices ('HC') over a 10-year timescale. An essential aspect to note is the differential standard potential evaporation calculation methods across datasets, necessitating a consistent calculation approach via the Makkink method for all data, as suggested by van Noppen (2022).

This study utilises three geographically distinct HM datasets corresponding to catchments in: the Meuse Basin, which spans France and Belgium (3.1); the United Kingdom, represented by CAMELS GB (3.1.2); and the USA, represented by CAMELS US (3.1.3). The CAMELS datasets increase the study sample size from 22 catchments (limited to France and Belgium) to 286 catchments that cover a diverse range of climatology, specifically in terms of aridity. Incorporating catchments with varied levels of aridity is beneficial for this study as we employ the Budyko framework for comparative analysis. It's known that vegetation adapts to aridity and seasonality, making such a varied dataset highly relevant (van Noppen (2022)). These historical data are processed to a decadal mean value as outlined in 4.1 to address research hypothesis 2.

3.1 Meuse Data

Historical climatic data has been processed and provided by Bouaziz et al. (2021) from the European daily high-resolution gridded dataset, E-OBS from Cornes et al. (2018). The gridded observation data is providing precipitation, temperature and air pressure at a daily temporal resolution with a grid size of $25km^2$. Temperature data is reprocessed using a DEM and a fixed lapse rate of $0.0065^{oc}m^{-1}$. Bouaziz et al. (2021) notes that $> 20\%$ precipitation underestimation is consistently present in the E-OBS compared to more local observations from the Service Public de Wallonie. This leads to applying the monthly bias correction factor. The Makkink formulation (Appendix Equation: 22) estimates potential evaporation from incoming radiation and temperature. Two distinct datasets, France and Belgium are available for discharge and cover significantly different time windows.



(a) (b)

Figure 3: Mapped location of CAMELS catchment data included in the analysis. (a) Selected Camels GB catchments (red boundaries with a green fill, n=95). (b) Selected Camels US catchments in green (n=169)

The Borgharen timeseries is readily accessible from Rijkswaterstaat and constitutes the longest available timeseries on the Meuse (**?**).

3.1.1 CAMELS Large Sample Datasets

CAMELS (**C**atchment **A**ttributes and **ME**teorology for **L**arge-sample **S**tudies) is a large body of datasets enabling large-sample comparative studies for climate analogies. Though other large-sample datasets exist (CAMELS-CL Alvarez-Garreton et al. (2018), LAMAH Klingler et al. (2021)), we operate a strict catchment selection procedure in an attempt to primarily eliminate the uncertainty from snow-water flux influence and human-flow modifications, this catchment selection process is explained in detail in Tempel (2023).

3.1.2 CAMELS GB

The CAMELS-GB dataset comprises 671 total catchments across Britain (England, Wales, Scotland and Northern Ireland) Coxon et al. (2020). Discharge data is from selected gauging systems within the UK NRFA Service Level Agreement (SLA), meaning that the gauging stations have met a high standard of operation and are considered to contribute most critical information to the gauging network in the U.K. (Dixon et al., 2013). Coxon et al. (2020) compiles meteorological timeseries for each catchment based on $1km^2$ gridded U.K. national, CHESS-met (Robinson et al., 2020) observations, which is either interpolated from daily station data or interpolated from coarser gridded products. Potential evaporation is calculated following the FAO standard, using the Penmann-Monteith method for well-watered grass Allen et al. (1998). The rainfall specifically is derived from the $1km^2$ rainfall dataset CEH-GEAR from 1961 to 2015 for GB and NI.

This compilation work resulted in open-access datasets spanning from October 1970 to September 2015, including flow (Q), rainfall (P), potential evaporation (E_P), temperature (T)

3.1.3 CAMELS USA

The CAMELS USA dataset, as developed by Newman et al. (2014), incorporates 671 catchments spread across the Contiguous United States. The catchment sizes included show significant variance, leaning toward smaller scales (median $336km^2$), with a distinct under-representation of larger eco-regions; fewer than 5 catchments exceed an area of $10,000km^2$.

For the meteorological forcing data, the N15-Daymet source was employed, yielding the necessary variables delineated in sections 3.1.2 and 3.1, covering the period from October 1989 to September 2009. It is important

to note that the potential evapotranspiration, denoted as E_P, is estimated within the N15-Daymet Thornton et al. (2016) using the Priestly-Taylor equation.

3.2 Spatial Data

The examination of landcover change (LCC) and its implications on HC expressions throughout the Meuse basin requires comprehensive LCC data. However, the scale of data selection presents inherent uncertainty. There exists a multitude of pre-processed data sources, each defined by their distinct characteristics. These are primarily distinguished by the diversity of provided information (e.g., land cover classes), their spatial resolution (defined in pixel width $[m]$), and their temporal resolution. These datasets exhibit a significant trade-off: enhanced informational detail often implies a compromise on temporal or spatial resolution.

In this study, the temporal resolution for LCC is desired to be annual. The employed datasets cover a spectrum from high (Landsat Classified), medium (CORINE), to low resolution (HILDA+) data. The Landsat data, a key resource in this study, facilitates the extraction of vegetation or moisture-sensitive bands, offering insights into potential correlations with HC expressions. Additionally, the application of machine learning on Landsat data allows for precise land cover class differentiation, further enabling the study of changes' impacts. Consequently, this bolsters the richness of the available data on vegetation change, presenting a thorough set of annual observational data since 1984.

3.2.1 Landsat Timeseries

Launched in 1982, the Landsat 4 Thematic Mapper (TM) represented a significant step forward in satellite-based Earth observation, capturing images at a 30-meter pixel resolution in a 16-day cycle, with a swath width of 185 km (0.185 degrees) (Chander et al., 2009). This legacy was continued by the nearly identical Landsat 5 in 1985, ensuring the ongoing supply of high-resolution data. In 1999, Landsat 7 introduced the eight-band Enhanced Thematic Mapper Plus (ETM+) sensor, enhancing the program further. Despite a scanning line correction failure in June 2003 that led to inconsistent imaging patterns, Landsat 7 remains one of the most accurate civilian satellite data sources (Chander and Markham, 2003). The program expanded with the launches of Landsat 8 in 2013 and Landsat 9 in 2021, marking a continuous 41-year mission of passive Earth observation on a climatic timescale.

However, one should note that Landsat 8 imagery has narrower red and infrared bands than those of the ETM+, potentially introducing minor errors. According to a study by Xu (2014), these errors are largely negligible for annual landscape change but are relevant for seasonal phenology.

The successive Landsat missions have generated a vast archive of consistently reprocessed, high temporal and spatial resolution imagery, freely available since 2008 (Woodcock et al. (2008)). Nonetheless, no single mission (sensor) can provide a complete timeseries from 1984 to the present (Roy et al., 2014). Therefore, this study compiles Collection Tier-1 atmospherically and geometrically corrected images to create an inclusive timeseries. With the release of Landsat Collection 2 in 2021, all imagery from Landsat 4-9 has been reprocessed, including Quality Assessment (QA) metadata useful for monitoring landscape changes (Survey (2021)). This metadata enables effective masking of clouds and water bodies, ensuring consistency while preventing band saturation. Moreover, the previously time-consuming task of pre-processing is now largely unnecessary.

The open-access cloud computing platform Google Earth Engine Gorelick et al. (2017) offers expansive datasets without requiring local storage or memory, thereby facilitating Land Use and Land Cover (LULC) analyses on an unprecedented scale. Consequently, highly accurate Landsat-based classification products have been developed for global land classes on a yearly basis Zhang et al. (2021), further showcasing the capabilities of the Google Earth Engine platform.

In this study, we access surface reflectance data from the Landsat missions and composite to a yearly timeseries via GEE to analyse land cover changes from 1984-2019. Band arithmetic allows the derivation of vegetation indices, enabling zonal statistics and area-above-threshold computations. We hope to test the hypothesis (3) assuming one or many of the land cover change indices will display some relation to the hydroclimatic variables derived in section 4.1.

Table 3: Six bands are used from the original seven-band TM/ETM+ Landsat sensors that changed band reference in 2013 with the introduction of the OLI sensor on Landsat 8. We remap the bands to the original TM/ETM+ so that band 1 is always blue

	TM/ETM	**OLI**	**Wavelength**	**Wavelength** $[\mu m]$	**Resolution** $[m]$
	Band 1	Band 2	Visible (b)	0.45-0.52	30
	Band 2	Band 3	Visible (g)	0.52-0.60	30
	Band 3	Band 4	Visible (r)	0.63-0.69	30
	Band 4	Band 5	Near-Infrared	0.76-0.90	30
	Band 5	Band 6	Shortwave-Infrared	1.55-0.75	30
	Band 6	Band (10&11)	Thermal	10.4-12.5	120
	Band 7	Band 7	Mid-Infrared	2.08-2.35	30
Sensor Year:	1982	2013			

3.3 Landcover Datasets

3.4 CORINE

The Coordination of Information on the Environment (CORINE) Land Cover (CLC) inventory, launched in 1985 with the reference year of 1990, provides an extensive record of land cover distributed over 44 classes. This catalogue has continuously received updates, specifically in 2000, 2006, 2012, and 2018. To ensure a faithful representation of areal and linear phenomena, the CLC employs a Minimum Mapping Unit of 25 hectares (ha) and a minimum width of 100 m, respectively Buttner et al. (2004).

The CLC database serves as a crucial asset for wide-ranging spatial analyses. It features a homogenous depiction of land cover throughout Europe, allowing for the precise recognition of various land-use categories. One of the key aspects of CLC databases is their frequent updates, which enable the analysis of change dynamics and rates, thereby supporting predictive endeavours. In this study, we have used data from 2006, 2012, and 2018. When juxtaposed with other data sources such as the Urban Atlas and the Global Human Settlement Layer, the CLC database exhibits enhanced adaptability.

However, it is essential to acknowledge certain limitations when utilising CLC data in studies. The in-depth nature of input data and the employed interpretation methods often necessitate a high degree of generalisation. In areas marked by substantial land fragmentation, the results can generally be extrapolated only to dominant land-use types, leading to an inevitable loss of some information. According to scholarly consensus, while the CLC proves extremely useful for small-scale studies, its reliability wanes when applied to larger-scale analyses. Therefore, despite its numerous advantages, the utility of the CLC should be evaluated in the context of these constraints Cieślak et al. (2020). Overall accuracy in the CORINE data is reported as $\approx 85\%$ accurate.

3.5 HILDA+

A novel long-term hybrid product for land cover change analysis is the HILDA+ (**HI**storic **L**and **D**ynamics **A**ssessment+) Winkler et al. (2020). Harmonising multiple open source datasets, the dataset has been impactful on a global scale since 1899 and has been four times greater than previous estimates and more than a third (32%) of land surfaces have been impacted by LCC since 1960. Accuracy is reported to be consistently near 85%, similar to the CORINE data. Globally landcover change is observed to accelerate before 2005 before decelerating into 2019, and the primary driver is due to global trade on agricultural production. The report accompanying the release of this data estimates an overall forest gain in the East of North America since 1960, a lesser proportion of afforestation is apparent in northern Europe.

Significant stability of pasture and rangeland is similarly reported in the period Winkler et al. (2021). The paper discusses the harmonisation of various datastreams reducing uncertainty by change allocation. The data is provided in annual TIFF images since 1899, globally with a scale of 1 degree and constitutes the medium-resolution land cover change product that can be run in parallel to the high-resolution Landsat classification product. HILDA+ reportedly estimates levels of land use change above those of other more common products.

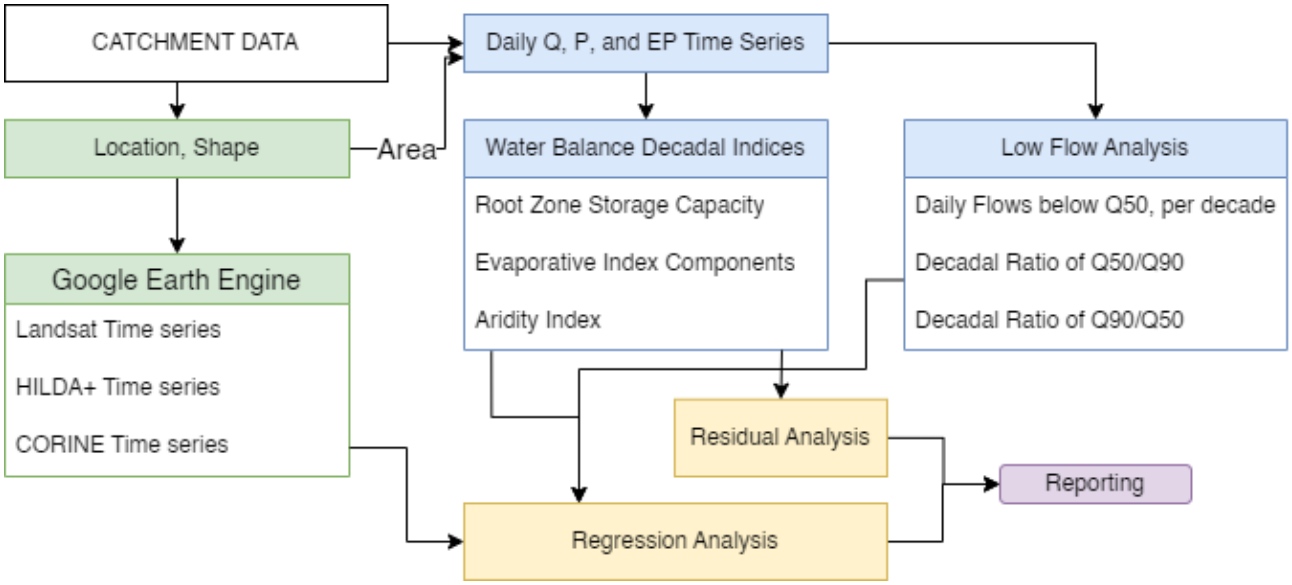


Figure 4: A flowchart depicting a high-level view flow of the processes and flow of data employed to achieve the methodology detailed in this paper. The blue branch indicates hydrometeorological analysis and variables, the green branch of the chart indicates the landcover change-related analysis, and the yellow is the final stages of combining and reporting for this paper.

4 Methodology

This research aims to characterise catchments potentially impacted by land cover change (LCC), focusing on the residual change component of the evaporative index, ΔEI_r. This component has been largely attributed to various near-surface processes, particularly LCC. Our study challenges the assumption that residual deviations in the evaporative index (ΔEI_r) mainly result from land surface changes like deforestation. Combining water balance-based methodologies after Jaramillo et al. (2018), Tempel (2023) and van Noppen (2022) indices are derived on a decadal basis. We develop and employ a bespoke machine-learning and remote-sensing workflow to provide a multi-scale, historical characterisation of annual land cover changes in each catchment. Through regression analysis, hydroclimatic indices are compared to satellite-based observations of LCC to evaluate the significance of LCC influences on ΔEI_r. The overall flow of methods and data is depicted in the flowchart attached as figure 4.

4.1 Decadal Hydro-climatic variables

The initial focus is on separating the land cover from the climate shifts effecting hydrologic extremes and on rootzone storage capacity, which is calculated on a decadal basis,this leads to loss of intra-seasonal variation which is important for understanding phenology. Different indices that describe seasonality over longer terms are employed here instead.

4.1.1 Root Zone Storage Capacity (Mass-Curve Technique)

The root zone storage capacity plays a critical role in understanding land-atmosphere interactions, as it mediates between surface runoff and groundwater recharge, impacts drought occurrence and severity, and influences the transpiration process. This research will calculate root zone storage capacity using a mass balance approach based on annual maximum soil moisture deficits (SMD).

The method, as discussed by Gao et al. (2014), is predicated on three core assumptions: first, the principle of least work suggests that plants do not root deeper than necessary, optimising their growth strategy according to resource availability. Second, all hydrologically active roots are utilised in the event of a water deficit. Lastly,

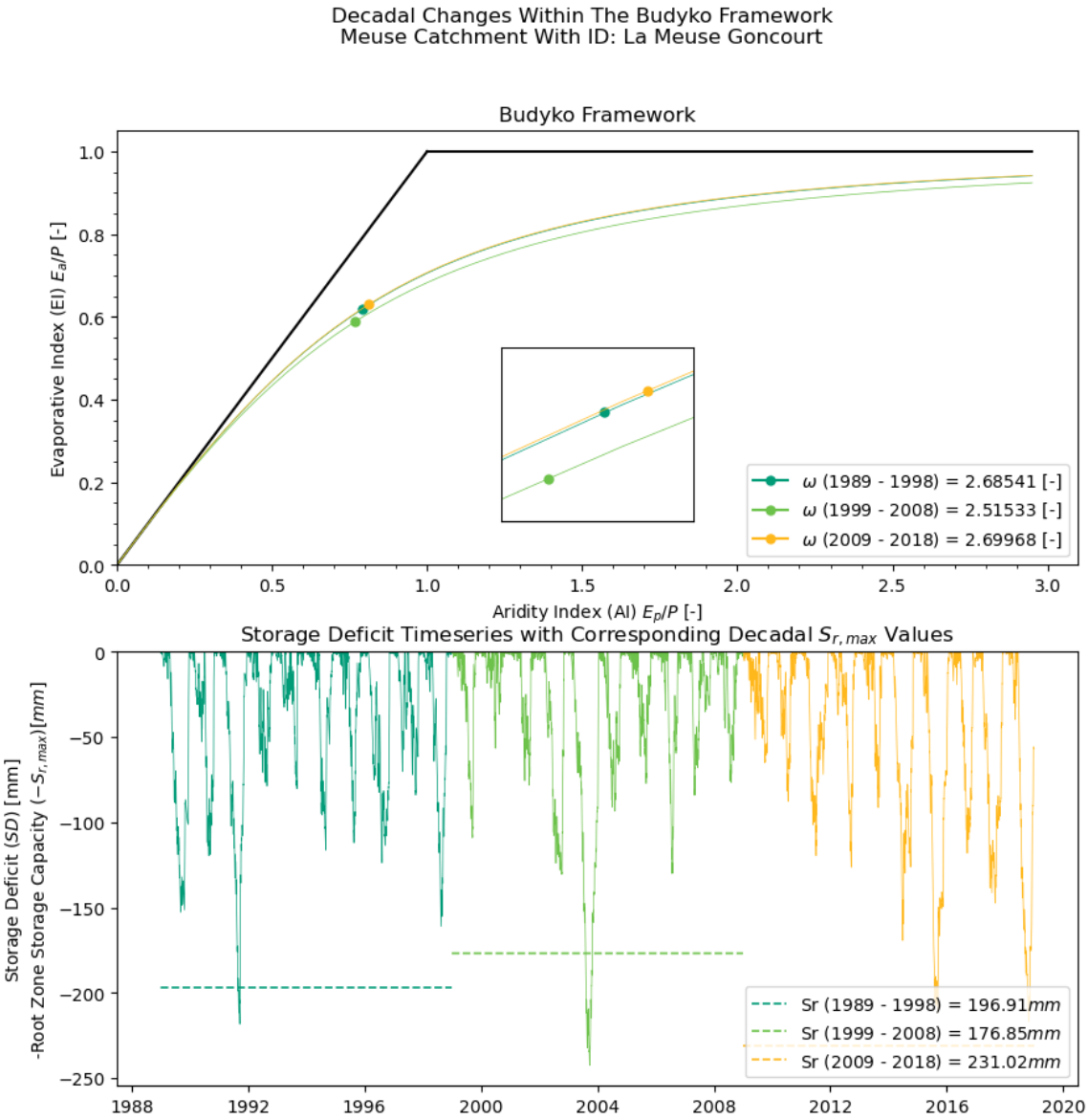


Figure 5: Two complementary subplots are relevant to the entire time series for La Meuse Goncourt. Top: The Budyko framework is presented, where the points represent the positions in the framework corresponding to each decade's respective aridity index (AI) and evaporative index (EI) values. The curves are fitted to the x-y positions and are represented by a derived Fu value (ω) for each decade (Equation 8). Bottom: The derived Meuse dataset is plotted as a time series of soil moisture deficit (Equation 4) in mm and the resulting root zone storage capacity (plotted as negative for overlay purposes, denoted as $-S_{r,20y}$, represented by a horizontal line). Three decades are demarcated in the time series by separate colours: green for the decade ending in 1998, light green for 2008, and yellow for 2018.

ecosystems tend to develop the root zone storage capacity needed to handle droughts with specific return periods, reflecting an adaptation to local climate conditions.

To calculate the root zone storage capacity, a mass balance analysis to yield annual SMD maxima. This approach requires obtaining or estimating the SMD for each year, finding the maximum value, and using it as an estimate of the root zone storage capacity. Long-term mean variables in this case are assumed to be a minimum of 10 years to allow for the water balance (equation 1) to close (i.e. $\frac{dS}{dt} = 0$):

$$\overline{P} - \overline{E_a} - \overline{Q} - \frac{dS}{dt} = 0 \tag{1}$$

Over a sufficient period, this study assumes 10 years, the fluctuations in storage $\frac{dS}{dt}$ are negligible such that the input (precipitation, P) is fully accounted for in actual evaporation (soil, interception, and transpiration) (E_a) and discharge/streamflow (Q). This allows for a representative catchment-scale estimate of E_T:

$$\overline{E_a} = \overline{P} - \overline{Q} \tag{2}$$

With reliable timeseries of daily meteorological data, providing the E_P an estimation of the daily $E_z(t)$ is possible whereby the observed potential daily evaporation ($E_P(t)$) results in a representative approximation that is water balance compliant:

$$E_a(t) = \frac{E_P(t)}{\overline{E_P}} \cdot \overline{E_a} \tag{3}$$

Cumulative storage deficits are calculated where the estimated $E_T(t)$ exceeds the precipitation inputs:

$$S_D(t) = min(0, \int_{t_1}^{t_2} (P(t) - E_a(t))dt) \tag{4}$$

Storage deficits are always negative and are assumed to start from 0 at t=0. Over a select 10-year period the annual maxima of storage deficits are calculated to which a Gumbel distribution is fit yielding a consistent estimate of the storage deficit 20-year return-period. The Gumbel distribution is a common choice of kernel density equations for estimating the probability density function within hydrological data Bakouch et al. (2021). Gao et al. (2014) report that the root zone storage capacity corresponding to the 20 year soil moisture deficit (SMD), (S_{R20Y}) most closely correlated to the unsaturated zone capacity ($R^2 = 0.75$) of a catchment ($S_{U,Max}$) as well as seasonality index (SI, $R^2 = 0.69$), interstorm duration (ID, $R^2 = 0.57$), and weakly to the aridity index (AI, $R^2 = 0.28$). Results from this study show that the plant's available capacity depends on wetness characteristics with deep roots compensating for more arid climates with longer periods between precipitation events.

One of the key benefits of this method is that it is model-independent and indirectly observation-based (Zhao et al., 2016). It uses historical climate and soil moisture data rather than relying on potentially uncertain model outputs. The approach also obviates the need for specific land cover or soil information, which can often be challenging to obtain with sufficient accuracy and spatial resolution. This is especially beneficial in regions where such data are limited and available. By reducing dependency on these factors, the method becomes more universally applicable and cost-effective.

4.1.2 Budyko Framework and Decadal Trajectory Analysis

The Budyko framework, first proposed by Budyko (1974) as a method of organising catchments based on long-term energy and water balance, has become a useful tool in the study of changing hydroclimatic conditions Heidari et al. (2020), this is contains influences from land processes. The x axis describes the ratio of potential evaporation to precipitation, known as the aridity index (AI, equation 5), this is primarily a climate sensitive axis. The y-axis is the evaporative index (EI, equation 7) or evaporative ratio defined as the ratio of actual evaporation (E_a) to precipitation (P).

$$AI = \frac{\overline{E_P}}{\overline{P}} \tag{5}$$

$$\overline{E_a} = 1 - \frac{\overline{Q}}{\overline{P}} \tag{6}$$

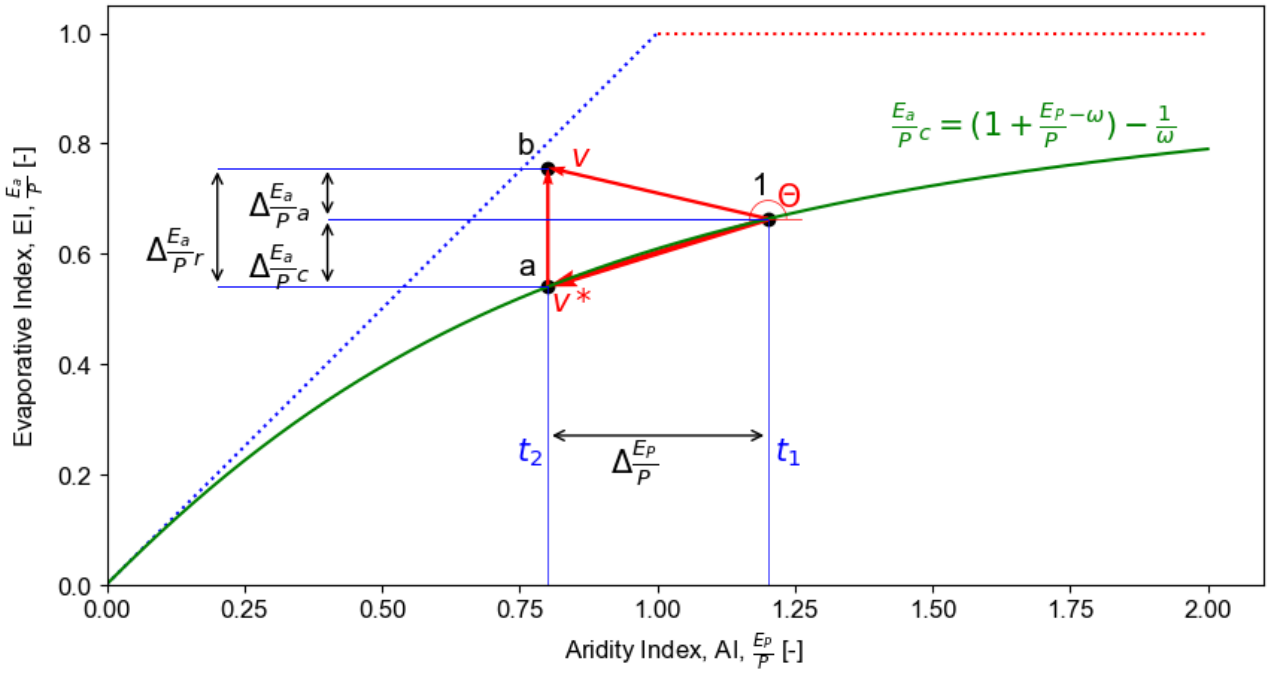


Figure 6: The Budyko framework with a schematic of potential trajectory resulting from inter-decadal changes in the water balance derived variables, adapted from Jaramillo et al. (2018). In this example one catchment moves from its position 1 at t_1, toward its actual position b at t_2 based on the long-term hydroclimatic indices, AI and EI (in this study $dt = 10_y$). The interdecadal movement has an expected, curvilinear, climatic trajectory actual trajectory $1 \rightarrow a$ (vector v^*) defined by equation 8. The positional change defines an actual linear trajectory $1 \rightarrow b$ (vector v) that emerges from the hydroclimatic indices. The EI residual change (ΔEI_r) value for each catchment is then determined by considering the expected climatic change ($\Delta \frac{E_a}{P}_c$), component and the additional ($\Delta \frac{E_a}{P}_a$). Working in all directions, it is best to consider the residual ($\Delta \frac{E_a}{P}_r = \Delta EI_r$) as the length of the vector $a \rightarrow b$.

$$EI = \frac{\overline{E_a}}{\overline{P}} \tag{7}$$

The framework is bound by two upper limits, the first is the water limit, a 1:1 line from the origin to an $AI < 1$ that represents the energy limit, a physical boundary that implies a catchment cannot evaporate more water than there is energy available. Beyond an AI of 1, with more energy available, the upper limit is constant at an EI of 1, this limit is the water limit, meaning that the evaporation (E_a) cannot exceed the water available. Please note that from this point AI is used for the aridity index of 5 and EI is used exclusively for the evaporative index.

The collective influence of climate, topography, soil and vegetation contribute to the location and changes within the framework. The original Budyko method of determining the actual evaporation (equation 6) did not take into account the influences of these factors which are especially relevant in the context of climate projection (Bouaziz et al., 2022). Introduction of a new equation for E_a that implements the exponential "Fu" parameter ω that better represents the expected climatic shift trajectory Fuh (1981); Tang and Wang (2017); Zhang et al. (2015). This equation (8) forms results in the parametric Budyko curve (green in 6), the subscript c is used to denote the climatic portion of EI that it describes.

$$EI_c = \frac{\overline{E_a}}{\overline{P}}_c = 1 + AI - (1 + (AI)^{\omega})^{\frac{1}{\omega}} \tag{8}$$

Equation 8 is solved for the Fu parameter is solved every decade per catchment yielding the same potential trajectory in the absence of additional land surface-driven deviations. After Velde et al. (2014) and Jaramillo et al. (2018) the method for separation of the land surface driven deviations, denoted as $\Delta\frac{E_a}{P}_a$ (additional) and $\Delta\frac{E_a}{P}_c$ (climatic), is first achieved by calculating the magnitude ($r(EI_c, AI)$) and angle (Θ, in degrees) of the movement in the Budyko space:

$$r = \sqrt{\Delta\left(\frac{E_a}{P}_c\right)^2 + \Delta\left(\frac{E_P}{P}\right)^2} \tag{9}$$

$$\Theta = b - \arctan\left(\frac{\Delta(E_a/P)_c}{\Delta(E_P/P)}\right) \tag{10}$$

b is a constant, ($b = 270^o$ if $\Delta AI > 0$, or $b = 270^o$ if $\Delta AI < 0$), this vector is graphically represented in figure 6 as $v*$. The trajectory of v is achieved by replacing $\Delta\frac{E_a}{P}_c$ with $\Delta\frac{E_P}{P}$. Vectors can be plotted in polar histogram (wind rose) plots to allow for first order understanding of the general tendency of the data. Finally the residual, $\Delta\frac{E_a}{P}_r$ is determined by:

$$\Delta EI_r = \Delta\frac{E_a}{P}_r = \Delta\frac{E_a}{P} - \Delta\frac{E_a}{P}_c \tag{11}$$

It is important to note that the omega value is held constant in time for these calculations. The direction and distribution can be plotted in the wind rose diagrams to visualise general tendencies in the space. For a more quantitative view of surface-driven evaporation error, we use the distributions of ΔEI_c vs ΔEI_r. Primarily we focus on the Meuse, secondarily we focus on the rest of the larger data sample to substantiate any emergent results.

The residual effect will be compared with the climatic effect in the results and discussion with additional consideration for the distribution of a population of change values. The datasets will be compared with regards to their overall tendencies and then the tendencies of different aridity groups will be examined. aridity groups are brackets of the aridity index used to slice the data based on the assumption that humid catchments (e.g. group 1) respond differently with respect to arid catchments (e.g. group 5). This can be useful with the Meuse data to compare with catchments of similar characteristics, by analogy of water and energy supply. The brackets used are the same as those employed by Tempel (2023):

Table 4: AI grouping for analysis and the relative number of values available in the dataset

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
Bracket of AI	*0.2-0.4*	*0.4-0.6*	*0.6-0.8*	*0.8-1*	*1.0-1.2*
n Catchments	22	39	110	92	10
n Decadal Datapoints	88	145	358	292	30
n Decadal Change Datapoints	66	104	245	200	20

Further groups were delineated to cluster the data but increasing aridity reduces the frequency of catchments to the point of having no meaningful sample size. After aridity of 1, where catchments are more limited by water than energy availability, group 5 (table 4) drops to a size of 10 and with increasing aridity, using constant bin width, there are no more than 5 catchments per class beyond 1.2.

4.1.3 Flow Duration Curves and Low-Flow Decadal Indices

In light of uncertainties highlighted in the recent paper by Tempel (2023), we compute water balance metrics for the Meuse basin to investigate potential deviations in actual evaporation, denoted as $\overline{E_a}$, and concurrent groundwater losses or gains. To this end, we focus on deriving decadal low-flow indices, vital for understanding hydrological variability. These indices are derived from the Flow Duration Curve (FDC), a well-established graphical representation of river flow characteristics.

The FDC plots the percentage of time that a given flow rate is equalled or exceeded in a stream. The curve, representing exceedance probability against discharge value, provides valuable information for water resource planning and low-flow classification. Indices can be extracted from the FDC in various ways: either by directly referencing a value at a certain probability, by calculating the ratio between two values, or by measuring the area under the curve relative to a specific flow value.

We begin by focusing on the most typical index, the discharge exceeding 50% of the time, denoted as Q50 (Smakhtin, 2001). This index is indicative of the overall groundwater contribution to streamflow. A steep slope in this section of the FDC suggests a significant groundwater influence, as shown in Figure 7.

Two other indices are introduced to assess decadal and interdecadal variability. The first is the ratio of Q50 to Q90 (Q50/Q90), which measures the variability in low-flow discharges, as per (Smakhtin, 2001). The second is the reverse ratio (Q90/Q50), which can be interpreted as an index representing the proportion of streamflow contributed by groundwater while discounting the effects of the catchment area.

In summary, we include three simple decadal changes in common low-flow indices to assess the likelihood that groundwater contribution fluctuations are similarly concurrently acting on the actual evaporation. For example, in the event of a significant anomaly in the derived ΔEI_r, aside from the contributions of landcover change, we would need to assess the possibility of groundwater contribution to the streamflow per decade. If sufficient additional Q ($\Delta Q50 \geq 0$) from subsurface/groundwater then that would contribute to a negative ΔEI_r anomaly according to Equation 2.

4.2 Remote Sensing and Spatial Data Processing

The novel remote-sensing methodology implemented in this study takes advantage of extensive satellite and spatial data available through Google Earth Engine (GEE). GEE provides open access to large geospatial datasets, enabling the monitoring and analysis of land surface changes (Gorelick et al., 2017). Data are accessed, processed, and analysed in a Python Jupyter Notebook environment, making extensive use of API bridging tools from the *Geemap* package Wu (2020). Utilising the Geemap package, we create custom 35-year time series from 1984 to 2019 for each catchment. Timeseries creation is achieved using a customised JavaScript module that is a derivative of the *LandTrendr* module code on the GEE open-access repository and released unde an APACHE license. Initially developed by Kennedy et al. (2018) for the United States Forest Service, *LandTrendr* has succeeded in creating the maximal continuous timeseries from 30m Landsat imagery further reducing the burden of image processing on the user. From this image time series, we extract decadal variables related to land cover change that could influence the residual evaporative index values.

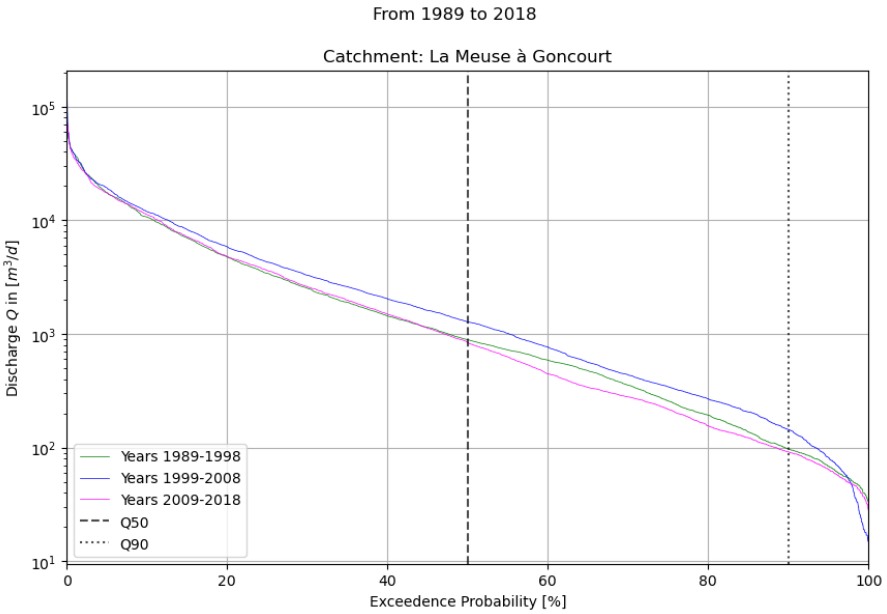


Figure 7: A line chart showing the Frequency Duration Curve from observations at La Meuse Goncourt, a headwater catchment in the Meuse basin. Two additional lines (Q50 and Q90) denoting the positions on the curve where the values for the decadal indices are gathered.

Spatial scale is a critical factor in investigations of land cover change. The three datasets employed in this study—HILDA+, CORINE, and Landsat—operate at vastly different spatial scales. Unlike hydrometeorological data, land cover datasets, especially those with distinct classifications, are often sparser. Users of contemporary datasets commonly face trade-offs among various factors such as data frequency (annual or otherwise), duration (long versus short time series), spatial resolution (measured in meters), geographic coverage (global or local), and accuracy (number of classes and rate of misclassification). An ideal dataset would cover an extended time span and offer high spatial resolution. High-resolution data are particularly desirable as they allow for more accurate characterisation of changes; landscape alterations like landslides and deforestation are more precisely detected at finer scales. Conversely, coarser scales may only detect larger disturbances.

As mentioned previously, the Landsat mission stands out as an attractive option, particularly due to its recent availability for open research. In this section, we explore the land cover data obtained from the annual summer period time series provided by Landsat, spanning the years 1984–2019 and offering a 30m resolution. We will characterise changes in this dataset in three distinct ways.

- Changes to catchment-scale zonal statistics: For example, variations in the mean Normalized Difference Vegetation Index (NDVI) across a catchment.
- Changes in area exceeding a specific threshold: For instance, alterations in the area of a catchment where an index value (e.g., NDVI) is greater than or equal to a specified threshold (≥ 0.2).
- Shifts in land cover classification: For example, modifications in the urban fraction of a catchment area resulting from discrete classifications.

These three characterisation methods offer diverse perspectives on land cover changes, thereby providing a variety of metrics to test our hypthesis against while providing a comprehensive understanding of landscape alterations over time.

4.2.1 Landsat Surface Reflectance Band Arithmetic, Statistics and Areal Calculation

Surface reflectance bands are commonly combined to examine the relationships between bands and to leverage the diverse spectral responses from heterogeneous surfaces. Several indices are added to the six-band time

series, chosen based on their sensitivity to the presence of vegetation or moisture. For each index, mean and standard deviation values per catchment are calculated at a 30m native resolution. Additionally, the area within a catchment exceeding a critical threshold (i.e., area $\geq$ threshold) is calculated and returned on an annual, per-catchment basis. These indices are valuable for the study as they combine bands to capitalise on differences in spectral character between surfaces, such as water and vegetation. The chosen thresholds aim to be physically meaningful; for example, an NDVI value above the threshold will correspond to an area that increases when the vegetation is sufficiently healthy.

Among the indices used, the Normalised Difference Vegetation Index (NDVI) stands out as a dimensionless indicator that is widely utilised in satellite remote sensing. It monitors phenological changes in vegetation and is directly related to vegetation productivity (Kriegler et al., 1969; Reed et al., 1994). The NDVI captures variations in chlorophyll levels in plant leaves, indicative of growth or water stress over a season. Healthy plants with dense cell structures typically show increased photosynthetic activity, absorbing red light while reflecting green and infrared light. We implement a threshold value of $NDVI \geq 0.2$ for areal calculations, providing an estimate of the vegetated proportion of the basin. NDVI is calculated using Landsat bands 3 (R) and 4 (NIR).:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{12}$$

Another noteworthy index is the Normalised Difference Moisture Index (NDMI), which provides information regarding the equivalent water thickness of vegetation (Cahyono et al., 2022; Hardisky, 2008). Using a threshold value of $NDMI \geq 0.2$, we perform areal calculations on the NDMI band, identifying areas with moderate to high canopy cover and low water content. A lower threshold value is favourable in this application, as it allows the index to be more effectively used to investigate its relation to root zone storage. The NDMI employs Landsat bands, specifically TM and ETM band 5 (SWIR).

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \tag{13}$$

The Normalized Burn Ratio (NBR) is used to identify burned areas by harnessing the difference in the near and (mid) short infrared EM spectrum (Eidenshink et al., 2007). A critical NBR threshold of ≥ 0.1 is used, where any value above this threshold indicates an increasing severity of burned ground. NBR calculations utilize Landsat (TM, ETM) band 7 (MIR, 2080 - 2350 nm). NBR is calculated as follows:

$$NBR = \frac{NIR + MIR}{NIR + MIR} \tag{14}$$

The Enhanced Vegetation Index (EVI) was developed to mitigate substantial errors inherent to NDVI, especially under varied (thicker) canopy conditions (Liu and Huete, 1995). The adaptation of NDVI to account for its open-loop structure and the incorporation of coefficients upon conversion to LAI units have reduced the error by up to seven times compared to the original NDVI. Areal calculations are performed using an areal threshold ≥ 0.2. EVI is calculated as follows:

$$EVI = G \cdot \frac{NIR - RED}{NIR + C_1 \cdot Red - C_2 \cdot BLUE + 1} \tag{15}$$

Where G is a gain factor (2.5), $c1$ and $c2$ are coefficients of the aerosol resistance resistance term (6 and 7.5, respectively). The coefficients enable the use of the blue band to compensate for aerosol influence in the red band.

The "Green" Normalized Difference Vegetation Index (GNDVI) is incorporated due to its sensitivity to the "greenness" of a land surface. The GNDVI utilizes the difference between green and infrared reflectance and

Table 5: Tasselled Cap constants for each axis and band

Axis	***B1***	***B2***	***B3***	***B4***	***B5***	***B7***
Brightness	0.2043	0.4158	0.5524	0.5741	0.3124	0.2303
Greenness	-0.1603	-0.2819	-0.4934	0.7940	-0.0002	-0.1446
Wetness	0.0315	0.2021	0.3102	0.1594	-0.6806	-0.6109

Table 6: Areal thresholds per band, pixel values above the threshold are summed to an areal value.

Index	Relation	Threshold
NDVI	$\geq$	0.2
NDMI	$\geq$	0.2
NBR	$\geq$	0.1
EVI	$\geq$	0.1
GNDVI	$\geq$	0.2
TCB	$\geq$	0.15
TCG	$\geq$	0.1
TCW	$\geq$	0.15
TCA	$\geq$	0.2

shows higher sensitivity to chlorophyll variation than the NDVI, by having a higher saturation point **?**.

$$GNDVI = \frac{NIR - GREEN}{NIR + GREEN} \tag{16}$$

The tasselled cap transform was originally derived by Kauth and Thomas (1976) commenting that "The time trajectories of agricultural data points as seen in the Landsat signal space form a pattern suggestive of a tasselled woolly cap.". The tasseled cap transform is a case of principal component analysis originally developed on the Landsat imagery that has been employed in continuous change detection in the forestry space (Cohen et al., 2016). Images are transformed to a new coordinate space with orthogonal axes. Primary axes are Brightness, greenness, wetness and another "angle" index component first derived by (Powell et al., 2010). Brightness is associated with bare soil, rock, and man-made features. Greenness relates to vegetation and wetness is associated with soil moisture. The tasseled cap angle refers to the relative reflectance change magnitude. Huang et al. (2002) derived the coefficients (table 5) for the transform that is applied with code adapted from (Kennedy et al., 2018).

Images are composited to annual summer images, corresponding with peak vegetation and minimal cloud cover. All the photos (approximately six) between June 20th and September 20th per year between 1984 and 2019. The compositing is accomplished via a medoid best pixel approach whereby each pixel band value is compared to the median six-band values of that season, and the values are chosen using Euclidian distance as described in and using modified code from (Kennedy et al., 2018).

4.2.2 Landsat Timeseries Quality Assurance and Control

The resulting Landsat timeseries spanning 1984 to 2019 could allow for direct LCC observations to be attributed to some of these observed changes in EI and specifically EI_r. Over time, the mission instrumentation has been updated or renewed, varying sensitivities and spectral bandwidth with each refresh. To accomplish the timeseries, these must be blended to achieve a continuous timeseries. It is imperative to check for biases in the resulting data for their stability and variance in time across the dataset.

Trends across a dataset should be relatively weak or their frequency of trend occurrence well-balanced with increasing, decreasing and 'no-trend' detected. We test for this stationarity using the mann-kendall non-parametric test for trends in timeseries data (Hu et al., 2020). Any single band should be relatively stable in time or display sufficient variance over the large dataset (all catchments) to avoid a false attribution of the trend from timeseries generation to actual land surface changes. The testing is necessary because the 'new' Landsat Collection 2 reprocessing has managed to remove much of the burden of preprocessing (cloud or cloud

shadow detection and removal) effort from the user. Still, the stability of bands for timeseries generation across all missions has yet to be fully understood.

For each catchment, maximum-length annual timeseries of band values (B1-7 mean and standard deviation) and derived indices (NDVI etc., mean and standard deviation) are generated each component is tested for significant trend, are reported in section 5.2.1. Controls for random errors are implemented in preprocessing the form of a cloud percentage filter (cloudy images over 20% are discarded from the composites) and cloud masking (cloud and cloud shadow pixels are made transparent and filled with the pixel closest in time).

4.2.3 Random Forest Image Classification

Each image in the Landsat consists of multiple bands, which can be envisioned as layers in a three-dimensional data set that is highly heterogeneous laterally, from pixel to pixel, and variable through time. Each pixel in the image series generated per catchment is likely to be dissimilar from its neighbour. This dissimilarity increases as the depth of the image grows, forming a more individual profile for that pixel. By adding bands, we increase the depth and variance of this profile. In addition to the bands described in the above section (Section 4.2.1), to achieve a more individual profile for the image, two more bands are added to the images in the series: elevation and slope, computed from NASA's Shuttle Radar Topography Mission (SRTM) at a 30m resolution.

Building this complexity within an image requires some algorithm to build probability based on the pixel profile, whether it belongs within a certain class or another. A decision tree is a common decision-making tool that assesses one option vs. another, moving down a branch of nodes (decisions) until the most probable outcome is reached at the leaf (outcome). On their own, decision trees are poor models for dealing with complexity and dimensionality. Still, combined in an ensemble via the random forest algorithm, it becomes a powerful classification method.

The process begins by preparing training data. In this case, the CORINE land cover dataset is a 44-class dataset with five levels. Five classes result by grouping all the training data into the highest level available. Those classes are: 1; artificial, 2; agricultural, 3; forest and seminatural, 4; wetlands and 5; waterbodies. The dataset is simplified to what are likely to be sufficiently separable classes, spectrally, meaning that the distinction between broadleaf and conifer forests, for example, is not possible. Additionally, the largest areal proportions in the catchments are training priorities, which are more relevant to the overall rainfall-runoff process. We prioritise the urban, forest and agricultural fraction to the highest degree of accuracy. Serves as a ground truth reference to train the Random Forest classifier. The different land cover types in the CORINE dataset are used as labels for the training data.

Next, features (predictors or independent variables) are extracted from the Landsat imagery. These features are the bands (e.g. B1-7), vegetation indices (e.g. NDVI), and principal components (e.g. TCW). Each pixel in the Landsat image is associated with these features, creating a multidimensional feature space.

The RF classifier is then trained using the prepared dataset. The algorithm generates many decision trees, each built from a random subset of the training data by bootstrap sampling or 'bootstrapping'. The RF algorithm begins by creating several subsets of the original dataset using a method known as bootstrap sampling (or bagging). This process involves randomly selecting observations with replacements, meaning some may be selected multiple times while others might not. Each subset, known as the bootstrap samples, trains a separate decision tree. The number of trees is a parameter set at the beginning (often in the hundreds or thousands). At each node, a random subset of features is selected for each decision tree, and the best split from this subset is chosen based on how much it reduces the impurity (for instance, Gini impurity). This introduces randomness into the feature selection process and helps decorrelate the trees, making the model robust to noise and outliers.

The Random Forest (RF) algorithm has the distinctive capability to estimate its generalisation error (the model's ability to handle unseen data) without needing a distinct validation dataset. This estimation is accomplished through split sampling, in our case, a 70/30 train/test division. The allocated 5000 training points across all classes are divided in such a manner, with 70% used for training. Once classified, the subset is contrasted with the classified product, providing an image representing the model's overall accuracy.

Landsat classification has great potential for landcover change investigation, owing to its speed and effective performance. Zhu et al. (2016) demonstrated that an unbalanced training strategy with weighting proportional

to occurrence is possible with Random Forest training. This means that training data is not evenly distributed per class, but rather a strategy for prioritising one class above another is recommended. In this study, we distribute 70% of 5000 training points between five classes proportional to the area each class in the training data such that the least extensive class (usually wetlands) will receive minimal training prioritisation. The same study further suggests approximately $20,000$ training points to adequately classify a complete Landsat scene, with a minimum of 600 points per class. It becomes quickly apparent that an areally-prioritised imbalanced, split training strategy will lead us to under distribute training to those classes that form a small percentage of the catchments. Those catchments with the highest areal proportion, such as agriculture and forestry will be the most reproducible and reliable.

The CORINE dataset, which includes the Meuse region in mainland Europe, only has data available for 1990, 2000, 2006, 2012, and 2018. These years are considered for model training, and the models generated in these years classify the years not covered by the training data. Unfortunately, the accuracy of the models in these years without training references cannot be directly assessed.

In the case of the Meuse basin, the entire basin is trained and classified during relevant years, and these models are preserved for classification between the training years. For the CAMELS GB dataset, training data is less available, lacking the 1990 set. Thus, the model from 2000 is utilised to classify every year from 1984 to 2003.

4.2.4 Model Calibration

The model calibration process is achieved through an integrated evaluation of overall accuracy and the kappa score. Model hyperparameters are adjusted to locate an optimal parameter space that yields an optimal combined kappa score and accuracy while keeping the model size low. The adopted weighting strategy maximises accuracy on the most significant land cover classes. The 30% of points split from the original 5000 for training are used to create an error matrix called a 'confusion matrix' that compares the model performance by comparing the test split CORINE classes to the model classification prediction.

A confusion matrix is a table layout used in supervised learning for visualising the performance of an algorithm, often a classification model, by showing the actual true instances of each class in the rows and the outcome, or predicted classes, in the columns. The matrix includes four metrics: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), which provide insights into the model's precision, recall, F1-score, and accuracy. False positive (FP) values are the sum along the rows, not including the TP value, FN values by contrast are calculated by summing the columns, less any TP values along the diagonal.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{17}$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{18}$$

The resulting model is then presented in its performance over time, with the use of the f1 score, a measure of precision and recall:

$$F_1 = 2 \cdot \frac{precision * recall}{precision + recall} \tag{19}$$

The F1 score is the harmonic mean of precision and recall, where precision is the number of true positive results divided by the number of all positive results (including those not correctly identified). Recall is the number of true positive results divided by the number of all samples that should have been identified as positive (Chicco and Jurman, 2020). By maximising the F1-score and overall accuracy, we use both objective methods for model calibration.

Hyperparameters are adjusted incrementally beginning with the most impactful on the size, accuracy and significance of the model (generally referred to as 'performance' herein). The hyperparameters calibrated in this model are, in order:

1. The *number of trees* in the forest; increasing the number of trees generally improves the model performance and increases computational cost. Therefore, it is advisable to increase the number of trees until the performance gain becomes negligible.
2. The *number of variables per split*; adjusts the number of features considered when making the best split at each node. Tuning this parameter can help in balancing overfitting and underfitting.
3. The *minimum leaf population*; this is the minimum number of samples required to form a leaf node. This acts as a regularisation method, and higher values can help minimise overfitting.
4. The *maximum number of nodes*; specifies the number of leaf nodes for each tree in the random forest. Limiting the tree to this number of nodes can speed up the training process and control overfitting, but setting it too high may risk overfitting, while setting it too low may result in underfitting.

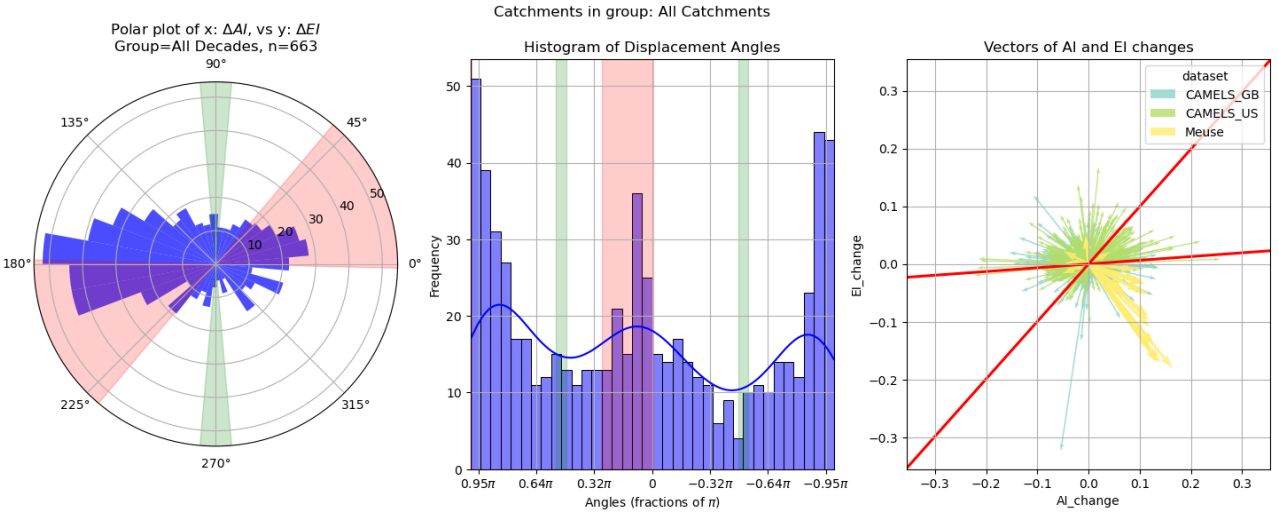
4.3 Linear Regression to Identify Correlation to Residual Effects

In the trajectory analysis carried out as part of this project, the primary derived variables of interest are ΔEI_{res} and ΔEI_{clim}. Our main objective is to discern any meaningful and statistically significant patterns within the extensive datasets gathered for the project. Specifically, we're keen to identify any linear relationships between decadal changes in the components of the evaporative ratio and variations in land cover change indices.

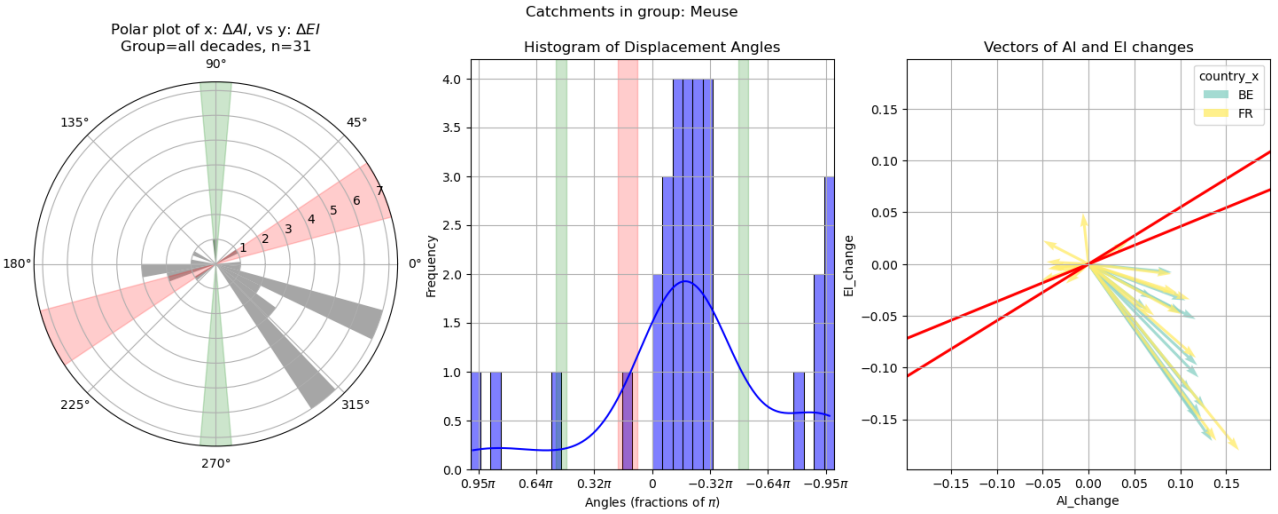
Our working hypothesis at this stage is that the residuals from these relationships can be attributed to one or more changes in land cover occurring during the same periods. To test this, we've crafted a sophisticated correlation strategy, designed to capture a broad range of correlation metrics across our independent data and to mitigate the impact of outliers on our analysis.

We perform linear regression using all data or subsections, implementing a robust resampling methodology to create a representative range, meaning mean and upper and lower bounds, of correlation coefficients (r^2). This bootstrap-based approach helps us build meaningful confidence intervals around our estimates of correlation coefficients and significance values, enhancing the robustness and reliability of our results and reducing the effect of homoscedasticity, where a sequence of variables may have homogeneity of variance that affects regression results (Jeong and Lee, 1999). All populations that are tested for a causal relationship in this manner are normalised by standard deviation and tested for normality using the Shapiro-Wilk test for normality (Shapiro and Wilk, 1965). Any population failing the test for normality ($p_{threshold} > 0.05$ is rejected and not included in the final analysis.

For each set of decadal changes, we apply a Pearson's p-test based on Pearson's product-moment correlation coefficient to assess the significance of our findings. To ensure rigorous statistical standards, we set a significance threshold of $p < 0.05$ for our tests. This approach helps us not only to identify statistically meaningful relationships but also to uphold stringent methodological standards, enhancing the credibility and validity of our study. Finally, for visualisation a regression line is fit using the coefficients resulting from the ordinary least squares method (OLS).



(a) Three different subplots are shown with the same underlying data. All hydroclimatic $\Delta AI, \Delta EI$ pairs are displayed. Plot one, the wind rose displays frequency and direction in polar space. Plot two, the histogram with KDE shows the same in non-polar space. Plot three, the quiver, shows direction and magnitude. A deeper explanation of how to read the graphs in section 5.0.1.



(b) This data is for all decades and the subsection of data relevant to the Meuse basin only.

5 Results

Firstly, in the results section we discuss the data visualisation to aide in interpretability. Secondly, we comment on the qualitative tendencies of the interdecadal changes calculated from the hydroclimatic variables. We report the results of the Budyko trajectory decomposition analysis to answer research hypothesis 1. Subsequently, a comparative assessment of the landcover data generated. Finally, the results of the regression relationship between land cover change observations and ΔEI_r and $\Delta S_{r,20_y}$ and other HC variables.

Firstly, we delve into data visualization techniques to aid in the interpretability of the results, thereby providing an intuitive understanding of the hydroclimatic variables under study. Secondly, we discuss the qualitative and quantitative tendencies associated with interdecadal changes in the Budyko framework for different aridity groups and datasets. This examination serves to identify long-term trends or patterns that have implications for water resource management and climate change adaptation. Within this section we present the findings of the Budyko trajectory analysis, specifically addressing research hypothesis 1. Subsequently, we conduct an assessment of the resulting landcover data looking specifically at changes within the Meuse basin.

Finally, we report the results of the regression analysis, looking for any relationship between observed changes in land cover and variations in the residual change component of the evaporative index (ΔEI_r), root zone storage capacity change ($\Delta S_{r,20_y}$), and other hydroclimatic variables. This regression analysis seeks to establish causal links or strong associations between land cover changes and hydroclimatic variables.

5.0.1 Interpreting Visualisations

Conveying and visualising change within a big-data comparative study is a challenge so it is necessary to explain the unconventional data visualisations carefully. An important foundation is that we are visualising decadal changes, so an interdecadal change for a variable in a catchment is one data point. The first decadal changes are visualised to provide qualitative insight into the overall data tendencies and can be seen in 8a.

Displayed are all catchments from all decades, amounting to 663 data points for interdecadal change. The first subplot, a wind rose diagram, shows a frequency distribution of movement in a specific direction within the Budyko framework (a 2-D plane). The most frequent direction (angle) of movement within the plane is defined by the larger peak in the bimodal distribution (to the left). Since this also corresponds to the Budyko framework, the leftward movement indicates a frequent tendency toward a reduction in aridity ($\Delta AI < 0$), which can result from a reduction in potential evaporation, E_p or an increase in precipitation P. The same orientation principle can be applied to the vertical (90° or 270°, Budyko y-axis), except this is relevant to EI or ΔEI and hence E_a. The width of the red shading corresponds to the width of the ω value distribution in the data displayed.

As detailed in section 4.1.2, the climate-driven path for a catchment is based on the parametric curve described by 8. Figure **??** shows three decades with corresponding curves within an example catchment, interdecadal variation. To visualise whether a catchment is complying with the expected curve direction, we use a colour coding system, red for climate direction (curve) and green (y-axis, EI). Any of the 36 bins (10° tolerance) of the histogram that falls within the red shading imply that those catchments are close to complying with their expected climatic trajectory. Otherwise, some change in the land surface process (LCC, for example) affects actual evaporation E_a to deviate from the expected curve.

A second subplot plot in figure 8a is the more familiar histogram, showing an x-y visualisation of the same directional data as the wind rose. The frequency is on the y-axis showing number of catchments changing in that direction within one of the 36 directional bins. Similar red and green shading is implemented in the histogram. A kernel density estimate (KDE) is plotted on the histogram (blue line), a method of estimating the underlying probability density function of a dataset by smoothing the histogram using a kernel function. It provides a continuous, smooth estimate of the distribution, allowing for a more nuanced analysis of the data's shape and density compared to the discrete representation of the histogram. The x-axis is given in fractions of pi, with 0 (the theta direction) corresponding to a purely rightward direction on the wind rose.

A third subplot, the vector 'quiver' (arrow) plot, is provided that gives conveys the magnitude (length) of movements, as well as the direction. The vectors form arrows radiating from a common origin, with each arrow

representing an interdecadal change. The quiver plot in 8a is colored by the dataset, with yellow representing the Meuse data movements. The green represents the US data and the more blue color represents the GB data. In a similar theme, movement within the confines of the red lines is indicative of movement along the expected curve trajectory.

This section relies on a second type of subplot for a clear representation of the budyko separation results. Boxplots like those in 9a illustrate for all changes measured in all catchments, the three evaporative index components derived in section 4.1.2. The top row represents the distribution of movement and the bottom row represents the distribution in the magnitude of those movements. Three components are presented for consideration, ΔEI (eq:7) the total change in y axis, ΔEI_c (eq:8) the change in y-axis due to a change in E_P (and E_a) and thirdly ΔEI_r (eq:11) the residual, or the error introduced by some change in E_a.

Each plot sub-title provides the median value of the distribution and the standard deviation (σ). Boxplots are constructed in such a way that the box contains the median in the middle (yellow line), the bounds define the interquartile range (IQR, Q1 to Q3), and the whiskers define the IQR $\pm 1.5(IQR)$. The dots are data points that plot outside the whiskers, also known as fliers or outliers.

5.1 Budyko Framework Trajectory and Residuals in the Evaporative Index

5.1.1 Decadal Trajectories within the Budyko Framework

The figure **??** the wind rose shows that the most frequent movement for catchments to make is negative in the aridity AI (left) and neutral to positive concerning the EI. Since most catchments do not fall within the red 'expected', we can see that there is a contribution of ΔEI_r causing deflection. Most catchments display a positive bias, such that we see the majority of movements occur with a positive ΔEI bias. This is seen where most movements are distributed above the $0 - 180$ line and where the KDE is higher in the left 'shoulder' of the middle histogram compared to the right half. The quivers show that the large datasets generally agree, but the Meuse data (yellow) has a notable opposite character albeit a smaller sample of 31. The median catchment will result in a $1.302e-3 \pm 5.581e-2$ displacement (table 7) which is twice the ΔEI_c contribution from climate-related changes.

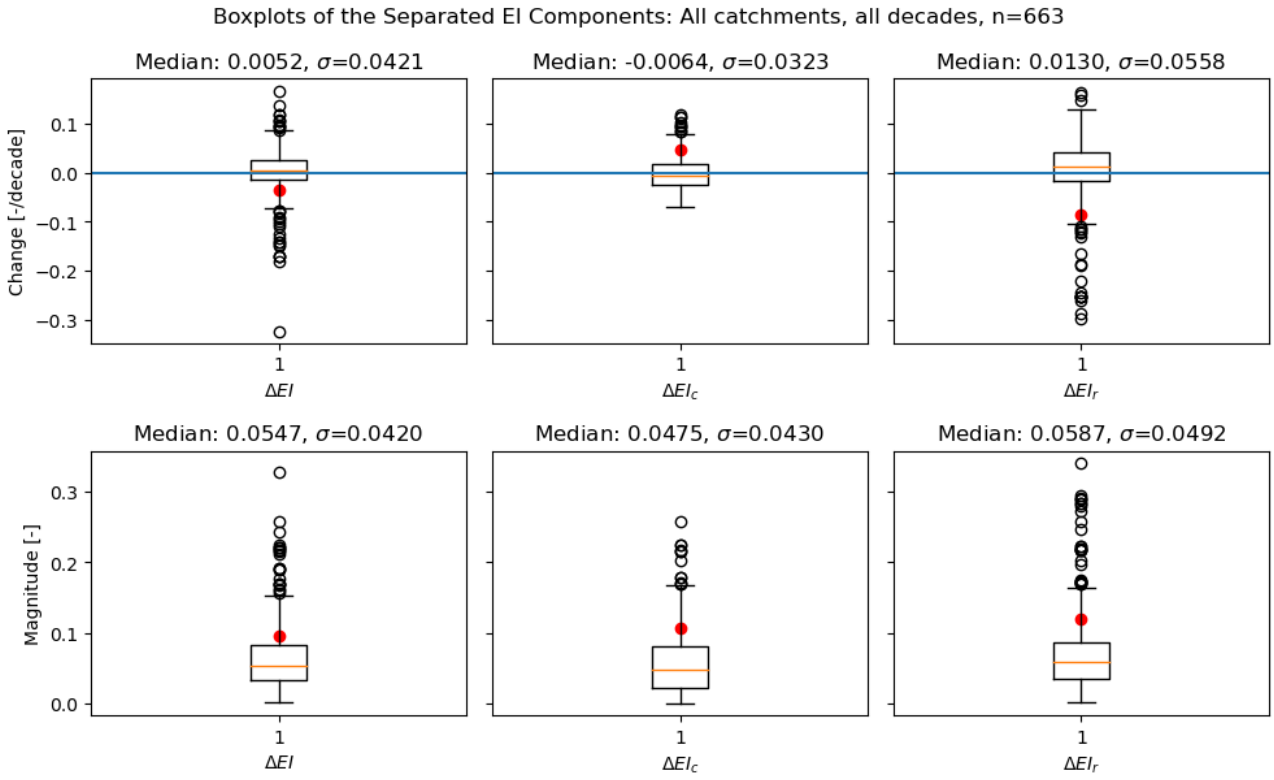
5.1.2 Meuse Evaporative Change Component Results

The Meuse data wind rose for 22 catchments (France and Belgium) for all decadal changes is shown in figure **??**. The wind rose, and quiver (vector) plots show a significant magnitude and directional deviation from the expected, ω distribution defined red (expected) trajectory.

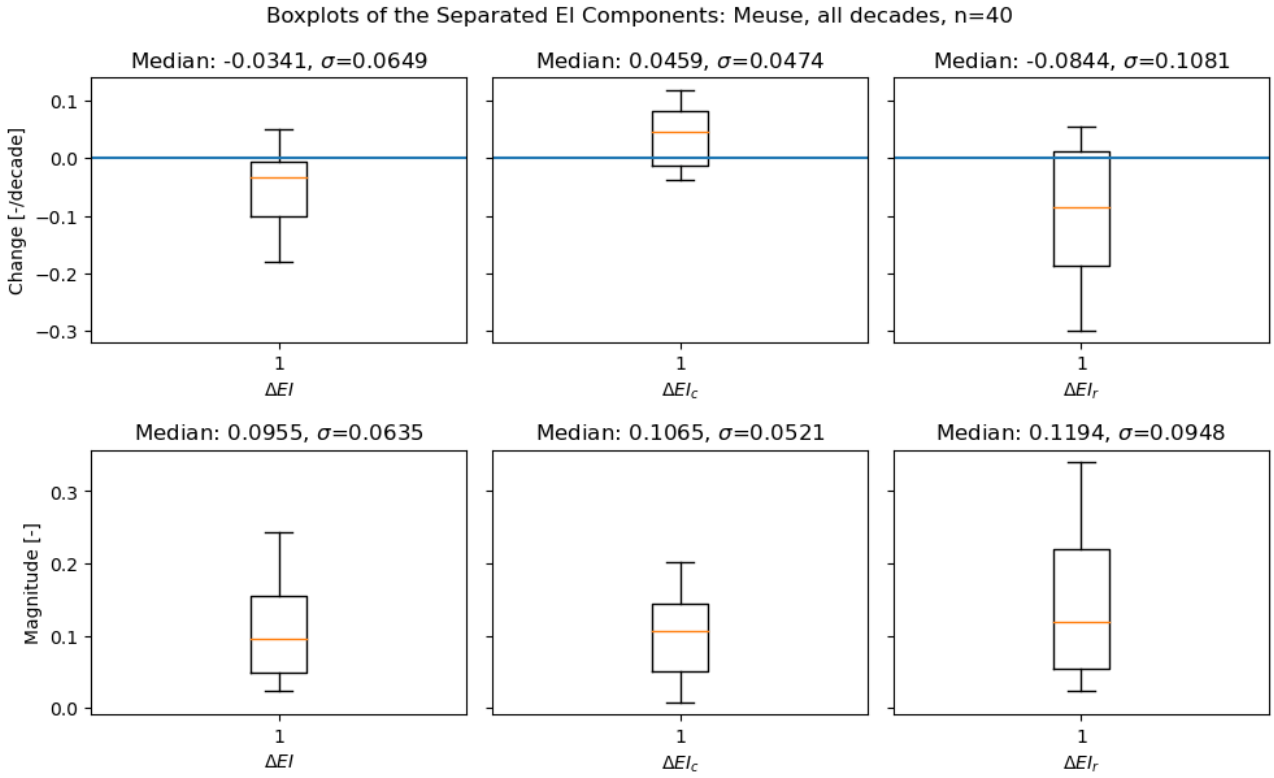
Notably, a large proportion of the data is plotting to the right ($\Delta AI > 0$) with a negative deviation in the ($\Delta EI < 0$). All 13 catchments in the Belgian dataset exhibit this behaviour while only select French catchments show similar magnitude deflections. It is worth noting that there are fewer catchments in the French dataset (9), but there is sufficient data for two decadal changes to be calculated; they are more numerous in the vector (quiver) figure (18 vs 13).

The distribution of the 2008 French data is less negative than the 2018 French (mean: $\Delta EI_{r,2008} = 0.022$ vs $\Delta EI_{r,2018} = -0.125$) data. Belgian results are challenging to directly compare where decadal periods do not perfectly overlap with the French catchments. The timing and agreement are coincident with similarities in the direction and magnitude of the move across the entire basin (mean: $\Delta EI_{r,2008} = -0.164$). That is to say, the deep negative deviation in the EI is an anomaly of the most recent decade (2009-2018). Overall the ΔEI_r is deeply negative with a median -0.085 compared to an all data median $+0.01$ (7). The lower whisker ($Q1 - 1.5(IQR)$) for the 654 total decadal change pairs is $\Delta EI_r = -0.1$. These results align and expand on the significant negative bias in the EI outlined by Tempel (2023).

Spatial distributions help with visualising the intensity of the residual effect revealing a geographic clustering for the most recent anomalous ΔEI_r values as shown in 10. Meuse basin sub-catchments along the Eastern border of the basin showing a concentration of this negative evaporative deviation. *La Chiers Carignan* (map 10, number 10) followed by *Jemelle* (map 10, number 16) show the most negative deviations:$\Delta EI_r = -0.298[-]$ and $\Delta EI_r = -0.262[-]$, respectively. All catchments in this period are shown to deviate negatively for this



(a) Six boxplots with the top row representing the decadal change across three evaporative index components and the bottom row indicating the distribution of the magnitudes of the changes. The data presented is for all datasets for all time, and a blue line is plotted at 0 for ease of reading on the top row. The first component ΔEI, the total change in the evaporative index, the second ΔEI_c the EI component contributed by climate change (ΔEI) and third, ΔEI_r, the residual component supposedly resulting from a change in land surface processes. Per component, the median and standard deviation (σ) is presented. On each distribution (boxplot) is a red dot representing the median Meuse value for the corresponding component.



(b) A repeated figure with a subset of the data displaying only Meuse data (for all decades). The median values here are the same as those plotted as red dots in (a)

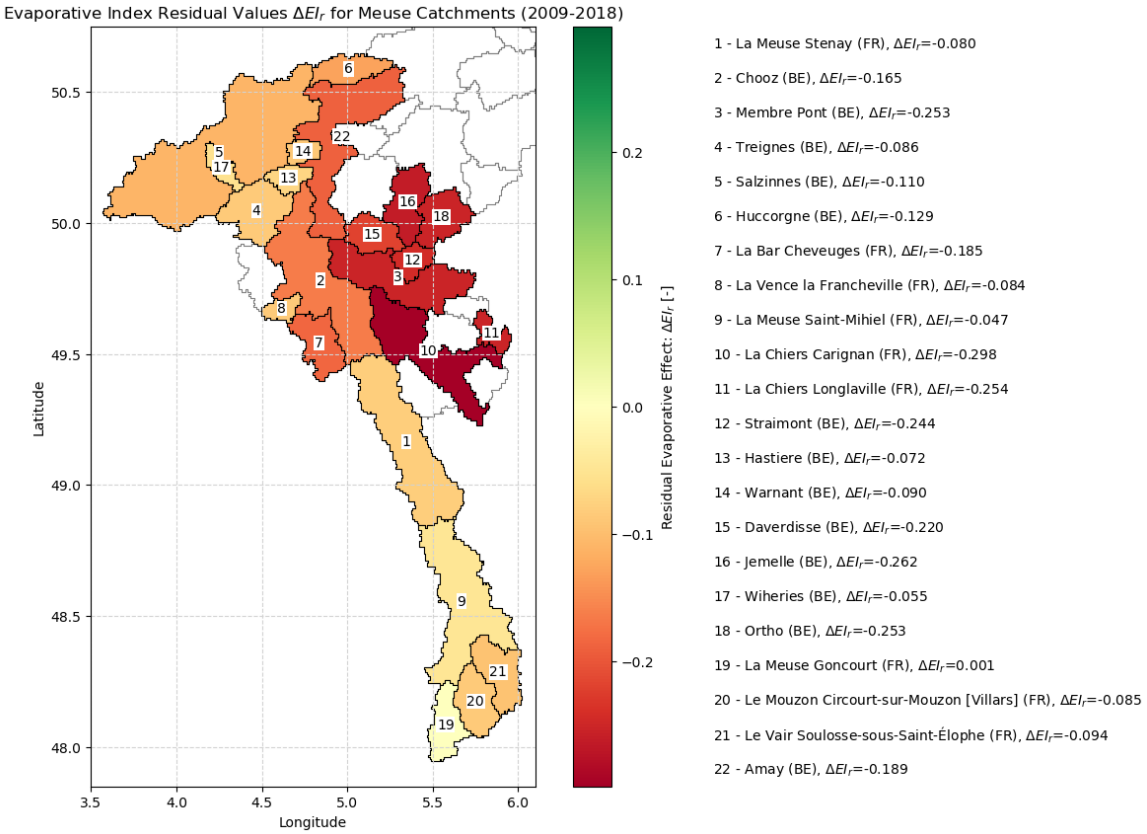


Figure 10: A map displaying the residual EI (ΔEI_r) component across the Meuse basin from the interdecadal ΔEI t_1:1999-2008 to t_2: 2009-2018

period with the exception of La Meuse Goncourt (19, $\Delta EI_r = 0.001$).

Considering appendix Figure 32 we can visually relate the concurrent decadal change in the evaporative index ΔEI to a significant change in $S_{r,20y}$ and the estimations clearly dependent. With such deviations occurring ΔEI_r is the most significant component, it is imperative to know if LCC is a significant contributor to these decade-over-decade changes contributing to a median of $-25mm$ or -13.7% in relative root zone storage capacity change.

5.1.3 CAMELS Residual Component Comparison

The CAMELS datasets US and GB display a similar tendency to each other despite geographic separation giving further evidence for this Meuse observation to be considered an anomaly. Visually the US and GB decrease in aridity index (left) and are positively deviating ($\Delta EI_r > 0$) from the expected climatic trajectory. Less than 10% of the CAMELS GB catchment changes calculated are closely compliant with that expected ω defined direction. Similarly, no notable magnitude of changes are compliant with this path, meaning all major moves (length of the vector) are not along the expected trajectory.

The CAMELS US dataset shows a much higher proportion, compared to GB or Meuse, of catchments complying with the expected climatic shift also with a negative bias in the aridity index and a clear majority positive bias in the evaporative index. The US data shows approximately seven change values complying with an exclusively positive shift in the EI with notable magnitude. The median residual magnitude is higher at 0.07 than the climatic magnitude, and the median residual component of $\Delta EI_r = 0.02$ is moderately more (+0.01) positive than the negative median climatic component (-0.01). The balance is in favour of the residual effects of evaporation being more influential than the climatic effects.

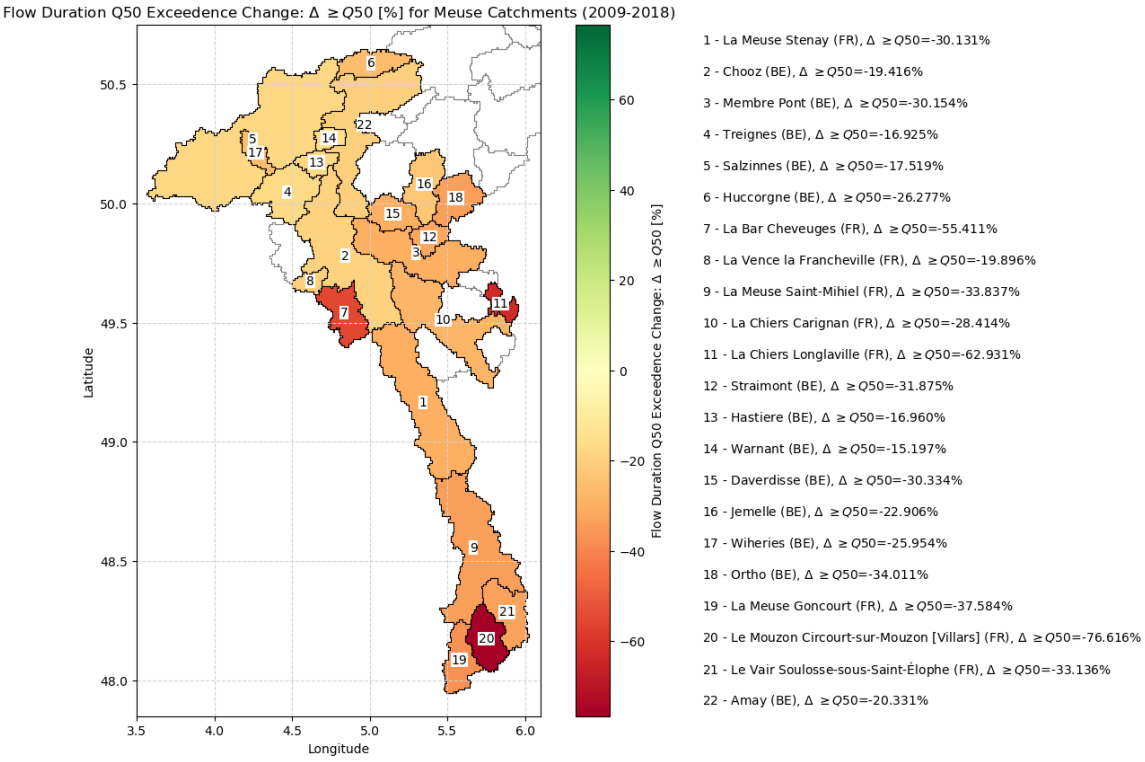


Figure 11: A map displaying the interdecadal change in flow volumes exceeding the 50% threshold t_1:1999-2008 to t_2: 2009-2018

Table 7: Decadal change groupings with the respective group tendencies summarised in terms of the median of both the climatic and residual deviations. The third line is an approximate indication of the level of residual component change vs climatic: $\frac{|\Delta EI_r|}{|\Delta EI_c|}$, the equation used to propagate errors is provided in the appendix as 29.

Dataset	**All Catchments**		**Meuse**	
	median($n = 663$)	σ_{all}	**median**($n = 40$)	σ_{Meuse}
ΔEI_r	1.302e-02	5.581e-02	-8.436e-02	1.081e-01
ΔEI_c	-6.366e-03	3.231e-02	4.587e-02	4.738e-02
$\frac{\lvert\Delta EI_r\rvert}{\lvert\Delta EI_c\rvert}$	2.045e+00	1.907e-03	1.839e+00	2.076e-02

Dataset	**GB**		**US**	
	median($n = 285$)	σ_{GB}	**median** ($n = 338$)	σ_{US}
ΔEI_r	6.716e-03	4.494e-02	1.942e-02	4.114e-02
ΔEI_c	-6.366e-03	2.524e-02	-8.989e-03	3.204e-02
$\frac{\lvert\Delta EI_r\rvert}{\lvert\Delta EI_c\rvert}$	1.055e+00	4.879e-04	2.160e+00	2.347e-03

5.1.4 Aridity Grouping of The Evaporative Index Components

The datasets are combined and subsequently grouped by their aridity brackets, as discussed in the methodology chosen to match those in the preceding study by Tempel (2023). From a qualitative standpoint, in order from most humid to most arid, group 1 (entirely CAMELS GB) is biased toward negative ΔAI movements, i.e. tending to become more humid (appendix figure 21 - 29), with some significant magnitude changes in outside of the expected climate trajectory. The negative tendency in aridity per group is summarised in Appendix table:9 where catchment groups with $AI < 1$ exhibit decreases in between 51.14% and 40.69% of inter decadal changes.

Aridity group (AG) 1 seems more climatic-driven than the moderately more arid AG 2 (0.6-0.8) 65% vs 70% at as residual-driven than the less humid AG 2 further. Climatic effects are consistently negative, and the proportion of residual prominence increases with aridity, except for the most arid (small sample $n = 20$) AG 5. AG5 shows a majority positive climatic shift with a lesser dominance of the residual (table 8, figure 30).

Table 8: Decadal change groupings with the respective aridity group tendencies summarised in terms of the median of both the climatic and residual deviations.

AI Bin	**0.2-0.4**		**0.4-0.6**		**0.6-0.8**	
	median($n = 66$)	$\sigma_{AI,1}$	**median**($n = 104$)	$\sigma_{AI,2}$	**median**($n = 245$)	$\sigma_{AI,3}$
ΔEI_r	3.807e-03	2.703e-02	2.757e-03	7.046e-02	1.548e-02	6.553e-02
ΔEI_c	-5.848e-03	1.387e-02	-3.816e-03	3.363e-02	-9.844e-03	3.658e-02
$\frac{\lvert\Delta EI_r\rvert}{\lvert\Delta EI_c\rvert}$	6.510e-01	1.198e-04	7.226e-01	2.331e-04	1.573e+00	2.161e-03

	0.8-0.1		**1-1.2**	
	median($n = 200$)	$\sigma_{AI,4}$	**median**($n = 20$)	$\sigma_{AI,5}$
ΔEI_r	2.236e-02	4.185e-02	8.547e-03	3.137e-02
ΔEI_c	-9.316e-03	3.221e-02	5.279e-03	3.134e-02
$\frac{\lvert\Delta EI_r\rvert}{\lvert\Delta EI_c\rvert}$	2.400e+00	2.966e-03	1.619e+00	7.019e-04

5.1.5 Meuse Low Flow Analysis

Low flow analysis for each decadal period was accomplished for all Meuse sub-catchments by calculation of the Flow Duration Curve as detailed in section 4.1.3, and a similar mapping visualisation to compare with the residual map directly is attached in Figure 11. In this visualisation, we can interpret negative values as the interdecadal percentage decrease in the sum of flows $\geq Q50$, in other words, the relative change in groundwater/subsurface streamflow contribution.

From this map firstly, it becomes clear that by showing the spatial distribution of $\Delta Q50$, the same spatial pattern of clustering as in ΔEI_r (Figure 10 is not present. Secondly, it is quite clear that inter-decadal changes in groundwater/subsurface-related streamflow changes were commonly on the order of -30% with some showing reductions of almost -76.16%. All sub-catchments within the Meuse basin express a negative change in base flow contribution for this period.

In the same period, the change in the ratio of Q50 to Q90, the measure of low flow variability, had a median change of +15% with a concurrent increase in IQR (appendix table A.4). The increases in the variability of low flows are clustered spatially in a somewhat similar manner to the values associated with the residuals, in the lower catchments and the headwater catchments in the steeper lands to the East. There is no linear relationship between the residuals and this observed increase in variability but the significant spatial coincidence is notable.

5.2 Landcover Change

Generation of the landcover indices is paramount to testing whether the residual evaporative component is land cover change-driven. Satellite observation data consists of two branches, those indices directly derived from individual bands (catchment mean and area above threshold). Second is the discrete classification accomplished via the random forest. First, the generated data is assessed, and the indices that pass the quality assessment

are put forth for regression analysis.

5.2.1 Landsat Image Timeseries Quality assessment

Bands within the Landsat images form timeseries that we test for trend significance using the Mann-Kendall non-parametric test for slope for all statistics and area calculations and surface reflectance zonal statistics bands. Each catchment annual timeseries from 1984 to 2019 is assessed for trend significance (Pearsons p values ¡0.05) and the proportion of those trends qualifying as increasing (green) decreasing (red) or no trend (grey, $p > 0.05$) see methodology section4.2.2 in figure 12. The decadal mean values are a rolling mean window from 1994 to 2019 of the 30m index values in the catchment of interest. What is expected of a sample of 199 catchments over varied geographic zones is that a certain proportion of trends should be increasing or decreasing.

Calculations that accomplish zonal statistics-derived timeseries for each catchment are computationally expensive and slow, and for that reaso,n only 199 out of 286 catchments were successfully assessed for bias. This size sample, however, was sufficiently large to prove the presence of bias and to discard a subsection of landcover change metrics from direct applications. The calculated rend results are illustrated in 12 where the figure shows the calculations only for the core bands and the indices that result from those bands. Trend significances are shown, with red, green or grey indicating a proportion of all timeseries as significant decreasing, increasing or no-trend, respectively.

The results show repeated, significant bias in the core bands 1-7, which form the core of some calculations e.g. NDVI. Bands 1-3 (G, B and R), the visible spectrum bands consistently decrease $83 - 91\%$ or register no trend. The NIR (B4) band is likely the most stable through the series where 78% of the series display no trend. The unfortunate knock-on effect of this artificial trend, resulting from stitching different sensors, is that in the resulting data the proportion of trends in addisional band calculations. The GNDVI, for example, since GNDVI is directly inverse to green reflectance (B2) 84% of the series recorded for 199 individual catchments are increasing. Artificial trends are likely an artifact of stitching different generations of sensors resulting in significant positive trends within the indices to which they are relevant.

The bias in the indices (band ratios) is not as overwhelmingly expressed in the trends of areas above thresholds. Still, their direct dependence on these biased indices results means they are indeed also biased. The least biased indices appear to be NBR, TCW and perhaps NBR, as they show a more balanced proportion of both increasing and decreasing trends, these are retained for regression analysis though any resulting correlation from these will be treated with a strong caveat. Bands and, by extension indices, that show strong bias (artificial trends) are discarded from regression analysis. All bands are used within the random forest classification, the bias in this context is mitigated/minimised by training on several years so that the bias is accounted for in the model.

5.2.2 Landsat Classification Accuracy Assessment

Training and classification are assessed by confusion metrics. The data presented in figure 13 demonstrate class-by-class accuracy metrics resulting from a the 30% split sub-sample of the original 5000 (100%) training points versus what is the resulting classification. Model overall accuracy improved from a baseline $\leq 71\%$ unweighted, unbalanced reference to an impressive $\leq 83.95\%$. The main gain in classification accuracy is due to the addition of the elevation and slope bands as is best illustrated in the ranked feature importance in 45. Feature importance is a useful metric, that can be calculated for Tree-Based Classifiers that indicates the ranked significance of each band's influence in the classification process. The most influential bands in an image leading to successful classification are primarily elevation, followed by slope then EVI and approximate equal importance for all other bands.

An upper accuracy limit is expected when the training data is a more coarse resolution than the resulting classification product (50m to 30m). For example, a training set that classifies a city as a single homogeneous area, then the higher resolution classification could possibly confuse a football field for and agricultural field based on spectral similarity. Indeed that is apparent in an example confusion matrix in figure13. This confusion matrix is from the year 2000 training, and classification instance (also available for 1990, 2006, 2012 and 2018 in appendix section A.5.10).

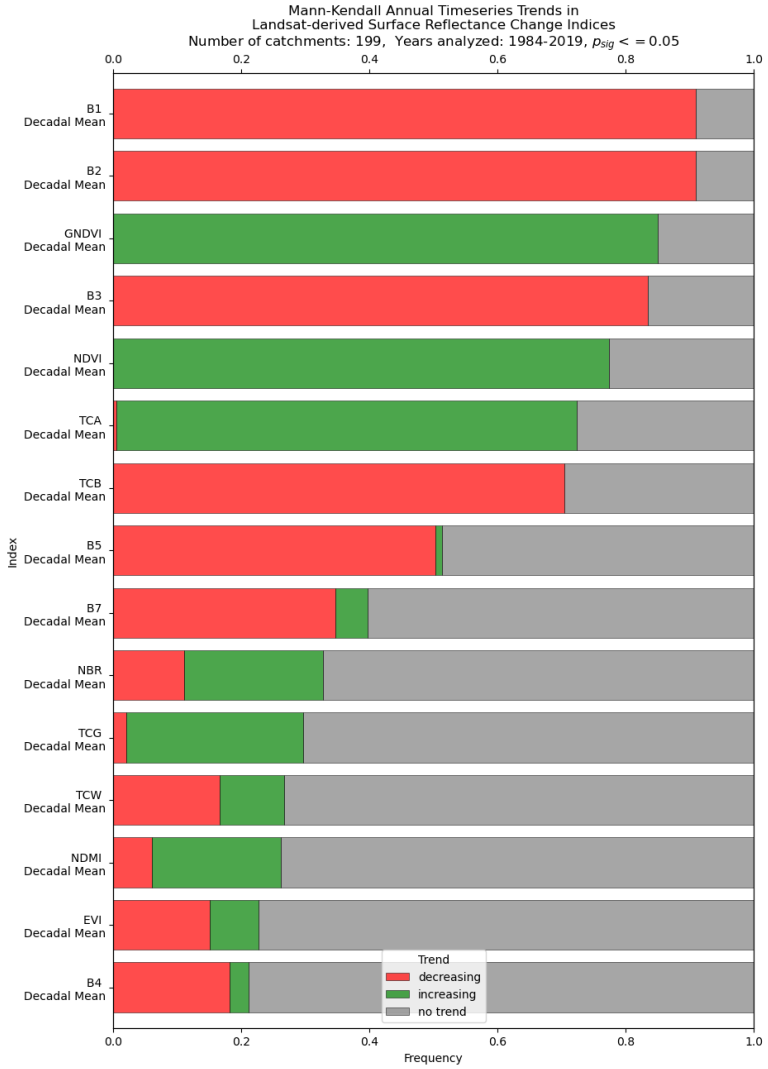


Figure 12: Visualized proportion of significant trend from 1984-2019 in catchment (30m) mean values. A table of each proportion is provided in the appendix as table 11.

An un-balanced approach is used in the sampling and we see that the 30% test split is given to each class, the parentheses on the horizontal axis labels show that agricultural pixels receive the most (proportional to the basin area occupied). True Artificial pixels have been tested over 184 points and commonly (10%) register false negatives for agricultural pixels. This is likely because flat, bare soil and heterogeneous urban surfaces are often spectrally, difficult to distinguish. A second reason is that some vegetation in parks and recreational areas can be detected here due to the resolution in the classification being higher than the training data (30m vs. 50m, respectively).

Classifying agricultural lands mis-classifies much less frequently, but occasional confusion occurs with artificial areas (2.3%) and more frequently with forest and semi-natural (FSN, 6.5%). Misclassifications occur for any number of reasons, most likely vegetated areas are hard to distinguish as semi-natural (e.g. grassland) versus agricultural (pasture).

Forest and Semi-natural pixels are the second most numerous as the second largest class with confusion very occasionally occurring with artificial or agricultural areas. Of the agricultural class pixels tested, 15% are estimated to be confused for agricultural for the same reason the reverse happens, as explained in the previous paragraph. It could also be argued that any forest at low elevation and low slope will be more likely to be confused for agricultural and vice versa for high elevation, steep ground agriculture, in a similar principle to HAND (height above nearest drainage) land classification .

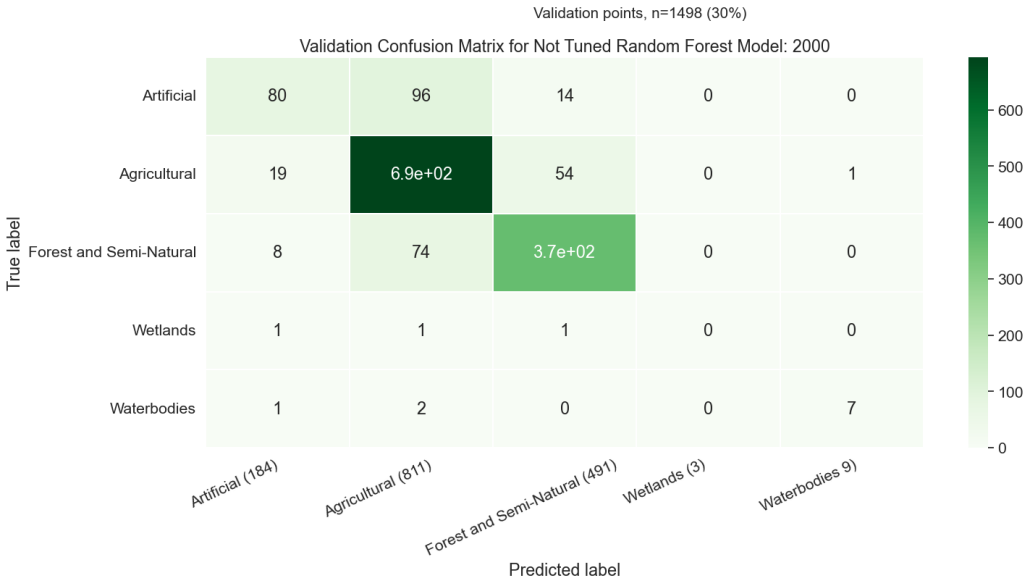


Figure 13: The confusion matrix shows the classification performance on the Meuse basin for the year 2000. The color scale corresponds to number of pixels.

Wetlands and waterbodies are such small proportions of the training area there results very few testing points from the areally balanced training set. Wetlands especially are the least common class, because there is so little training accurate classification is not acheived.

From an individual confusion matrix we can go beyond overall accuracy assessment and understand the precision and recall of a machine-learning model based on its precision and recall of various classes. For this, we present the timeseries of the 'f1' score (figure 14). The maximum overall accuracy of 83.95% is clearly not the most representative metric with an unbalanced multi-class training set as is presented here, some classes are highly accurate and some less so. The true positive rates are higher in the agricultural and FSN classes, the other classes suffer from the lack of training point allocation. It is worth noting that the confusion matrix and the f1 score for artificial surfaces tell us that a significant proportion of the (15/184) points misallocated as a false negative as agricultural, leading to a low f1 score in the artificial.

5.2.3 Resulting Land Classification Products

Three timeseries of classification products are presented for the Meuse basin and CAMELS GB. The areas are the classified proportions of the basin per year above the Maastricht (Borgharen) gauging station. With data from 122 catchments, the total area assessed totals 1.47 million square kilometers. A summary table of changes is attached in the appendix as 12. We present similarly the other two products for HILDA and the Landsat classification product (tables 13 and 14).

The HILDA+ data is the most coarse of the three datasets with a pixel resolution of $1km$. Coarse classification can result in large swings in a short time, which can be especially pronounced in smaller catchments where a change in $1km^2$ from one year to the next is proportionally more than in a whole basin but the relevance of impact on the evaporative balance is as yet unclear. Additionally the classes introduced by the HILDA+ product are the same as the CORINE albeit with slight differentiation. For the FSN class, the corresponding HILDA+ class includes forest and shrubland. For the agriculture class, the corresponding HILDA+ class is a combined cropland and pasture. We now assess the comparative changes in the major land class proportions in the Meuse.

The HILDA+ Forest and semi-natural decline since 1990 in the basin is -1.2% while the CORINE and Landsat products result in -0.2% and -1% declines, respectively. Where HILDA+ estimates a $+25\%$ increase in urban areas for the same period, CORINE and Landsat report 9.4% and 24.9% increases, respectively. The Landsat product agrees best with HILDA+ for Artificial (urban) and forest proportion changes and with CORINE on the agricultural decline. For those primary classes, the Landsat 30m product is in close agreement with at least one other model. The tables on land class area changes are available for CORINE, HILDA+ and Landsat in appendix tables (12, 13 and 14, respectively)

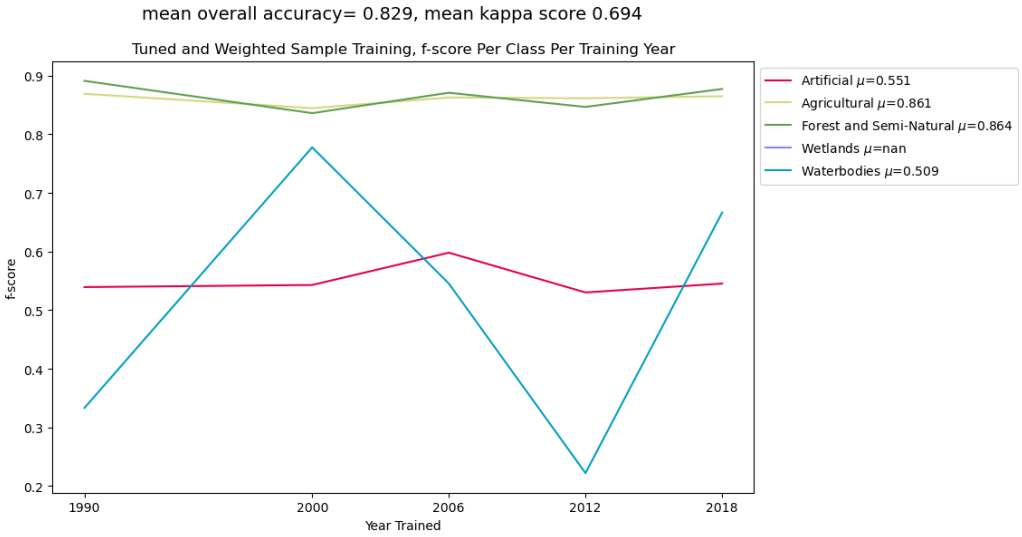


Figure 14: F1 score as a measure of classification model performance with each class plotted as a line over time within the span of the training data.

Agricultural landcover classifications are subject to errors between products. Hilda+ overestimates agricultural landportions compared to Corine on average by 10%. The mean error of the Landsat product shows an underestimation by -1.56% for this class. Removing the -32% error in 2012 which could be the result of a random environmental interference (e.g. cloud), the error is -2.143% (figure 15). For forest and seminatural CORINE is consistently higher in forest fraction than the other datasets, CORINE reports 25% higher forest fraction (FSN) and greater change than HILDA+. HILDA+ and Landsat are in closer agreement regarding mean error and trend, where CORINE shows a significant decrease since 1990 while the others are stably increasing. For FSN Landsat underestimates HILDA+ by approximately 15%.

The mean relative error for Landsat versus CORINE for artificial surfaces for this period is between -36.26% and -73.23%, meaning the higher resolution observations underestimate the urban land cover fraction by approximately half. Both HILDA+ and the Landsat product show to recreate the two vegetated classes reliably for agricultural and forest and seminatural classes. This is important for the regression analysis as we hope to test whether the derived ΔEI_r is related to LCC. We should be able to conclude whether change in these two high confidence classes is related to ΔEI_r and test for other, lower confidence, variables such as artificial surface change, EVI, NBR and NDMI.

5.3 Regression Relationships

Correlation between remote sensing observations and the hydroclimatic variables and this section reports the most prominent results to aid in testing hypothesis 3. Correlation between the decadal changes in catchment area normalised LCC indices and the dependent variable, primarily ΔEI_r, test to establish causal relationships.

Figure 16 presents a joint plot (histograms of the individual populations tested and a scatter visualising their potential linearity). The artificial Landsat decadal changes, normalised by individual catchment areas are presented on the y-axis, this is for all catchments for all decadal changes. The bootstrapped approach to eliminate outliers and homoscedasticity by sub-sampling the distributions for representative ranges returns upper and lower limits of r^2 and p. This is compared to the ΔEI_r component that should contain some as yet undetermined level of influence from changes in land cover. The choice of compared variables presented are those resulting from the regression with the highest $r^2 = 0.142$ value and the lowest $p_{pearson} = 5 \times 10^{-7}$ value for explanatory variables . This pair is the testing the classified artificial change vs the residual changes, it is likely that some large values highly influence the p where the r values are significantly higher due to the center of mass, concentration of data pairs around 0. The two populations are different, as indicated by the histograms on the border. In the data presented all populations compared are sufficiently normally distributed ($p_{shapiro} < 0.05$) for this analysis. Artificial surfaces show very large changes and a large proportion of almost

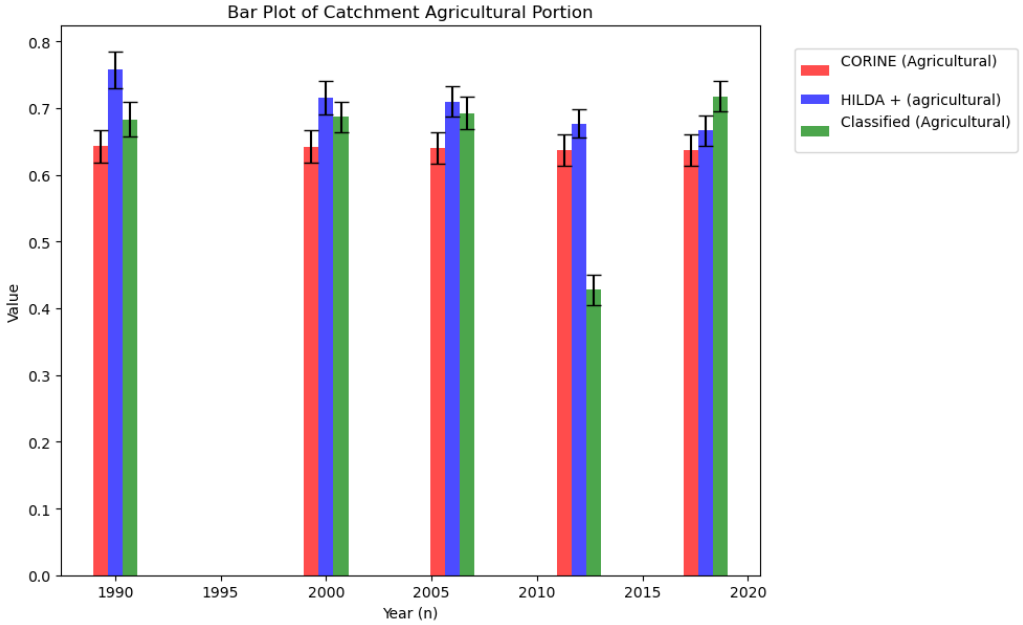


Figure 15: Three models are presented for the Meuse basin and CAMELS GB ($n_{catchments} = 122$). HILDA+ model is in blue, the CORINE data is in red, the classified Landsat product is in green. Each bar is composed of a range of area proportions estimated to be forest and semi-natural land, in the case of the Meuse, each bar and standard error represents 22 catchments upstream of Borgharen.

no change, coinciding with a large bin at zero percentage change per decade.

In both plots the data for classified land cover changes are small and form a large center of mass near around 0. The second example case of Landsat classified forest fraction change vs. the climatic component of evaporation (ΔEI_c) has the second highest p and r^2 values in the analysis, though they are even less . This pair similarly shows that regression results can again be highly influenced by this center of mass.

A dashed $1:1$ line is plotted for reference. The residual component shows a relatively normal distribution, with a majority between $\pm 10\%$ residual deviation. A small population of outliers exist to the left (negative) which will be the eastern Meuse catchments. Most catchments exhibit minimal residual and artificial surface change, with a centre of mass close to zero for both variables. The largest changes in artificial surfaces are not linearly related to the changes in artificial surfaces. The change in artificial catchment proportion decadal change has poor explanatory power concerning ΔEI_r.

A weaker correlation for all other variables is implied, with all correlations tested provided in the appendix and further in the data repository. In reality, there are no linear relations between any hydroclimatic indices included in this analysis. Weaker correlations are similarly implied where the same methods are applied to subgroups, or clusters within the dataset, for example, when testing for relationships across all aridity classes.

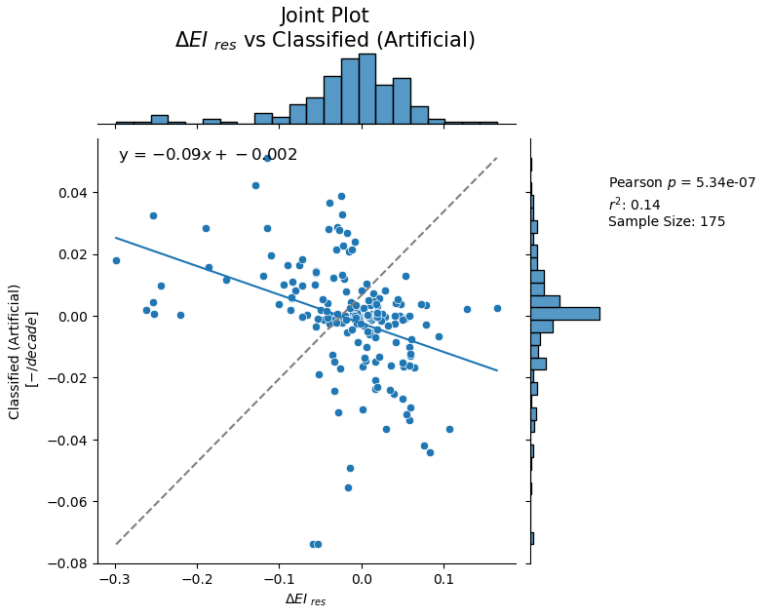


Figure 16: The sample with the highest r^2 value compared to the ΔEI_{res}

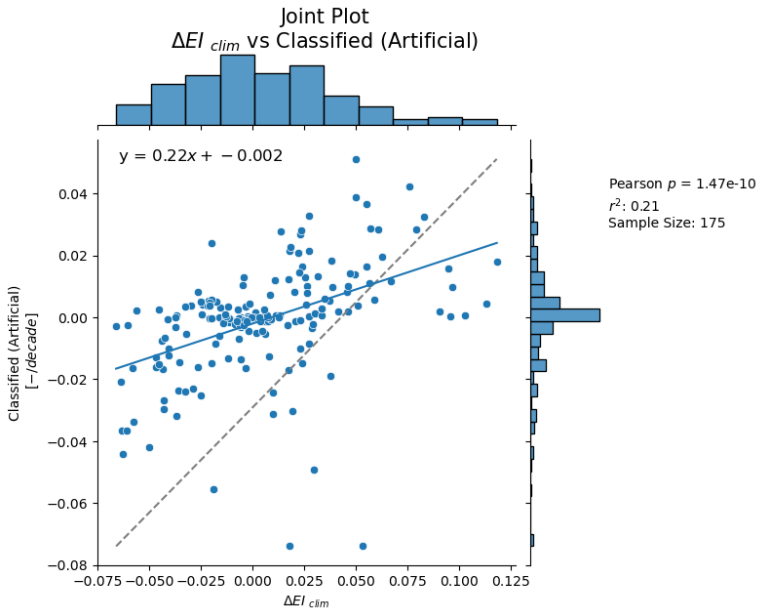


Figure 17: The sample with the highest r^2 value (0.21) compared to the ΔEI_{clim}

6 Discussion

We began this study hoping to unravel the uncertain level of contribution of land cover change to changes in evaporation within the Meuse basin, specifically whether observed LCC changes exhibit evidence of influencing quantified changes in the evaporative index and, by extension our projections of ecosystem-scale root zone storage capacity. As the landscape is modified through natural and man-made processes soil disturbance and vegetation removal occur affecting water balance calculations are affected and following this introduces uncertainties in projections of the future evaporative index. We discuss the possible implications of the results presented for root zone storage capacity and streamflow modelling of the Meuse in general.

We initiated this study to explore the elusive relationship between land cover change (LCC) and unexplained inter-decadal errors in actual evaporation within the Meuse basin. Specifically, we sought to determine whether observed changes in land cover have a causal relationship with the residual component of these errors and, consequently, on our projections of ecosystem-scale root zone storage capacity. The modification of the landscape, whether through natural or anthropogenic processes, leads to soil disturbance and vegetation removal, which in turn affect water balance calculations. These changes introduce uncertainties in future projections of the evaporative index. In the discussion that follows, we explore the ramifications of our findings on root zone storage capacity and streamflow modeling in the Meuse basin.

6.1 Limitations

This study has attempted to include as many steps as possible to ensure the quality, robustness and repeatability of the results presented. Exploratory analyses, like the one presented here, must assume an underlying level of uncertainty that is inherent in all data. The first important consideration is the baseline uncertainty inherent in all hydrologic data studies. **?** found that data uncertainty magnitudes typically range from 10–40% when dealing with hydrological data. Preprocessing steps taken in the preceding bodies of work have similarly taken the utmost care to appropriately handle data and remove unreliable sources (van Noppen, 2022; Bouaziz et al., 2022; Tempel, 2023).

One issue associated with this study and the exploration of limitations of a very individual technology is that complimentary datasets are either non-existent or are themselves full of uncertainties. For the calculation of changes in biomass to be allowed on the same timescale, the area of vegetation would need to be combined with leaf-area-index values. While the Advanced very-high-resolution (AVHRR) radiometer provides $1km$ resolution LAI imagery and has been available as a continuous timeseries since 1981 the quality of this sensor has been noted to drift since 2013. We attempt to harmonise and replace the later years (from 2004) LAI timeseries with newer, resampled resolution imagery from MODIS sensors (example timeseries are attached in Appendix 52). Despite determining coefficients for harmonisation, the series produces unexplained, unpredictable discontinuities consistent with difficulties encountered in the literature by Cao et al. 2008. The conclusion from this aside is that only the area changes and not biomass have been used in this study, which is a limitation.

The method used for land classification in this study has presented various challenges. The workflow for training the land classification models used CORINE data and was applied to classify Landsat imagery. However, due to limitations imposed by Google Earth Engine (GEE) servers, the model size had to be restricted, primarily this resulted in a smaller number of trees in the model than was optimal. Additionally, the maximum number of training points allowed by the server was relatively low, which may have additionally constrained the quality of the results.

Under the current approach, we can effectively train and classify land in areas where adequate training data are available. However, these smaller models do not generalise well to areas that lack training data. For instance, a model trained on data from the Meuse basin would likely perform poorly if applied to a catchment in the United States, even if the underlying conditions were similar. Improving the model size and increasing the number of training points could offer a route to more universally applicable models, as discussed in the work by (Jian et al., 2021).

The effectiveness and accuracy of the Landsat classification strategy are strongly influenced by unbalanced training and by extension the choice of weighting strategy. The choice to weight the model by the areal extent of the classes present was pragmatic and effective for the larger classes without more server computational memory allowed for training. In an ideal case, with more training points a better weighting strategy could be

considered to allow for a more general model that also represents smaller, less extensive classes such as wetlands.

Geologic factors underpin all the catchments in the study that can lead to leakage groundwater leakage and losses or gains from subsurface flows in general Smakhtin (2001); Wit et al. (2001). We address the possibility of subsurface flow influence by introducing some measures to estimate decadal low flow changes, these low-flow results are subject to the same hydrologic errors as mentioned previously by Mc Millan et al. (2018) with the addition that the FDC metric Q50 employed shows extreme values with as much as a 70% reduction in low flows in some catchments. The extremity of these values are, as yet unexplained and should be treated with caution.

In recent decades, there have been notable shifts in commercial forestry management practices, especially within coniferous forests. Although our classification scheme does not distinguish between different tree species, species type significantly influences water balance. Specifically, needle-leafed conifer plantations exhibit lower interception capacities compared to broadleaved trees, potentially affecting runoff ratios during land conversion. Modern forestry practices have not only altered surface drainage patterns but also have led to a trend of thinning conifer plantations, thereby increasing the average age of standing trees (Fenicia et al. 2009). Another potential factor influencing changes near the land surface could be shifts in vegetation phenology. Currently, the most reliable indicator we can derive from the available data is the areal extent of vegetation, which does not offer insights into internal processes specific to the FSN (Forest, Shrubland, and Natural Grassland) class.

Finally, the long-term (decadal) approach of our study was necessary to match the frequency of calculation for the rootzone storage capacity. This has, however, resulted in smoothing over, or averaging over, interseasonal variations in the data, especially those associated with land cover and vegetation dynamics. Rootzone storage has been shown to correlate well with some remote sensing measures (Sriwongsitanon et al., 2016), albeit intraseasonally. Implementation of new indexes that use interseasonal changes in phenology (vegetation), such as differences in plant growth between seasons, could be more sensitive and more informative for continued research.

6.2 Tendencies, Deviations and Residuals

283 catchments processed from three different regions form the backbone of this study. We processed the disparate hydrometeorological data into a series of decadal hydroclimatic indices allowing a comparative approach for catchments within varying geographic regions. The results show that over time the catchments are en-masse becoming more humid and this trend is amplified amongst the more humid catchments than the more arid catchments. Unfortunately, the inclusion of more arid catchments was not possible, therefore the representation of water-limited arid ($AI > 1$) catchments is low ($n_{catchments,AI>1} = 20$). Certainly, it can be seen that the more energy-limited, humid catchments ($AI < 1$) over the 50-year span in observations from 1969 have tended to become less arid. Increasing precipitation in the denominator of both the aridity and evaporative indices is likely the driver (similar to Jaramillo et al. (2018)).

Our study areas area concentrated in the northern hemisphere mid-latitudes where precipitation has been observed to increase echoing the "drier in dry, wetter in wet" (DIDWIW) paradigm, albeit without a lot of dry catchment samples ($AI > 1$). This paradigm according to findings by **?** describes an overall 48% of humid regions becoming more humid. The similar proportion of catchments tending to become more humid (decreasing AI) is generally higher in the less arid (more humid) groupings, the proportions of decreasing aridity below $AI < 1$. Including more arid catchments would allow for a more globally representative sample, limiting the observations in extending the opposite sentiment and seeing conclusively that dry catchments are becoming drier. The Appendix table 9 summarises the per-group proportions.

This study focuses on the separation of a residual evaporative component from the concurrent climatic effects in the evaporative index for all catchments using the Budyko framework. The estimation of future actual evaporation of a catchment or basin using Budyko framework (and Fu-parameter) projections is a common underpinning for modelling streamflow and the hydro-ecological variable root zone storage capacity in potential future climate scenarios. Residuals as a basket term for deviations from the expected curve have been variably attributed to land cover change in the literature. We developed a landcover assessment to test this idea.

First, isolation of the deviating component ΔEI_r was necessary to remove the climatic (expected) compo-

nent of the change in EI to satisfy the initial research hypothesis 1. The relative contribution of this residual component to the change in actual evaporation was surprisingly pronounced, meaning that catchments deviated from the expected more than they adhered concerning the whole dataset ($\frac{|\Delta EI_r|}{|\Delta EI_c|} = 2.05$) the significance is more pronounced for the CAMELS US and Meuse data than with CAMELS GB. The residual effect is relatively significant in deviating from the climatic projection.

The residuals were mostly positive with a concurrent reduction in aridity (AI). This implies that for most catchments in the study, the trend over time has been to become relatively less arid and in response the actual evaporation has increased significantly above expectations. The residual components for all data and subsections of the data such as aridity groups help to contextualise the Meuse basin within this large sample study. The research hypothesis 1 is supported whereby the mathematically separated components are meaningful and consistent across all data and subsets. If this successfully isolated residual effect is influenced heavily by LCC, then such a positive deviation might be associated with increased areal proportions of vegetation or density (biomass) within the landscape.

For all catchments inclusive the relative contribution of the residual component ΔEI_r was twice that of the climatic component ΔEI_c and the relative contribution seems to be higher in the more arid catchments in the sample. The relative contribution (Table 8) is an appealing, simple calculation introduced here that aids in assessing change. It is intuitive based on the examples that a catchment moving along the ω curve within the steep (left) portion would have more vertical displacement due to changing aridity than if it were far to the right of the curve, but the relative size of the move should be insightful. The contrasts show that the US data contains twice as much residual deviation (deviation from the expected) as GB. The Meuse residual data is similar regarding the relative contribution to the evaporative index change. The CAMELS GB data is significantly less arid on average than the US or to a lesser extent, the Meuse basin. Tempel (2023) calculated that the relative median error in estimation for the root zone storage capacity for GB and US data was $+1mm$ and $+14mm$, or 0.8% and 14%, respectively.

The most unexpected and perhaps largest implication for the primary study area is concerning the anomalous residual (ΔEI_r) component evident in the Meuse basin in the most recent period, the change from the decade ending in 2009 to 2018. Evident here is a significant reduction in the evaporative index, a far deviation from the expected. Considering the issue at hand, a slight aridity increase in for the anomalous Eastern cluster of catchments, a median relative root zone storage the capacity error would be $-25mm$ or -14% after Tempel (2023). a severe overestimation of the capacity of the landscape to transpire water. A dramatic increase in the runoff ratio (Q/P, Equation 2) inverse to the evaporative index EI must cause such a deviation, for the interdecadal period ending in 2018. Uncertainty over the additional contribution to streamflow are evident and concerning in the Meuse basin.

To eliminate the potential that an increase in streamflow (Q) has been contributed by subsurface or groundwater flow indicators for decadal changes in the same period are introduced. The contribution of low flows by the Q50 measure seems to indicate that, rather than additional contributions, the low flows have tended to decrease across the Meuse basin, with some extreme reductions in the headwater catchments (e.g. 20 or 11 in Figure 11). No streamflow contributions are evident to explain the spatial pattern in the Meuse residual anomaly. The most similarly spatially distributed metric to the anomalism observed is the measure of low flow variability (Q50/Q90) which seems to increase on the order of 10% in similar catchments that exhibit residual anomalies. The implications of this finding are uncertain, but the results show that groundwater/subsurface contribution to the Meuse negative anomalism is not a factor.

Literature offers some potential plant physiological answers to explain a reduction in EI that include the water use efficiency increase due to increased atmospheric CO_2 essentially fertilizing plants Jaramillo et al. (2018). Changing the imperviousness (infiltration capacity) of the landscape is likely to have a similar effect by increasing the runoff ratio, this could be due to soil compaction from agriculture or increase urbanisation, for example. Without leaf area index (LAI) data that is simultaneously sufficiently long and reliable (see **??**). As it stands, we are limited in assessing changes in imperviousness due to urbanisation (artificial surface change) or vegetation change (agricultural or forest and semi-natural change 'FSN') and whether these changes show evidence for contribution to changes in EI, specifically the residual (deviation) component.

Based on the evidence collected so far, it seems improbable that the residual component is a result of groundwater contributions. The impact of potential land cover changes has not yet been examined. Our preliminary

findings suggest that for land cover changes to have a significant effect on the calculated residuals, the extent of those changes would need to be much smaller in Great Britain (GB) compared to the Meuse region or the United States. While variations in land cover changes could occur from one catchment area to another, it's unlikely that such disparities would be consistent across an entire country with a large population, like GB. To accurately assess any potential contributions from land cover changes, we are conducting independent analyses at various scales for all catchments.

6.3 Landcover Change

A big-data approach to a large sample study can be a complicated path to a simple, definitive answer. A trail of spatial data provides additional insights and creates new value for those who may follow. What is certain is that precise estimates of Land cover change over time continue to prove to be difficult to constrain precisely. To offset this uncertainty a dual-scale, dual-model approach was preferred. Multi-scale studies help to mitigate uncertainties around the scale of change and effects in hydrology. In answer to this, we employ two comparable datasets of different scales (30m vs 1km) with a simplified class structure to form inter-annual timeseries to evaluate change with additional reference to the more intermittent EU standard 50m CORINE landcover dataset. The data processed reveals new estimates for land cover change within the primary study area, the Meuse Basin.

Traditionally remote sensing studies will span much shorter timescales than is required in this 30+ year project. It was necessary to create an unconventional dataset for this purpose that ties together several Landsat sensors to satisfy this timescale and with that some issues became apparent. The results discussed show that some of the bands that form the foundational remote sensing phenological indicators (e.g. NDVI, EVI etc.) are consistently, artificially trending. This eliminates some of the insights that could promise potential FSN class internal phenological insights that might replace LAI biomass calculations. With an artificially trending NDVI, applied to a FSN area to indicate health over time, a false sense of increasing vegetation health would be indicated and as such these insights are not reasonable to pursue.

For high-resolution insights, classification is preferred with regular training intervals over CORINE data to avoid drift resulting from the aforementioned band biases. The outcome can be noisy, with certain years exhibiting significant proportions of masked (unattributed transparent pixels). These unattributed pixels could be allocated to any classes adding to uncertainties. While compositing is a common strategy to compensate for masked pixels - as the images are already composited to form medoid "summertime" images and produced annually - no additional compositing (in other words, averaging or blending) would be responsible.

In the introduction to the study we discussed the overall increase in land conversion for agriculture, usually at the expense of natural grassland. This was a significant change in the European landscape as agriculture became industrialised with conversion incentives beginning in 1960 and accelerating until 2006 primarily increasing the proportion of pasture land for grazing (Milczarek-Andrzejewska et al., 2018; Winkler et al., 2021). The model comparisons show that despite the contrasting scales overall proportions of classes are quite similar. HILDA+ and Landsat show increasing forestry but opposing trends in agriculture. The Landsat data proves to be more noisy and susceptible to spurious error, likely influenced by cloud interferences in the imagery. The overall agreement between models shows that the scale consideration between products can produce differing results. Interestingly, the urban fraction increase between HILDA+ and Landsat is closely in agreement $\approx 25\%$, in contrast to the CORINE data reporting significantly less. This is encouraging, where the concern relating to confusion of the random forest classification over the highly heterogeneous urban surfaces lead to doubts about accuracy in this class. One major caveat is that while trend in urban fraction is highly similar the artificial/urban overall proportion of basins is underestimated by $\approx 40\%$. Offsets in landcover proportion could be attributed to different reasons firstly, the mean error can be a consequence of contrasting scales and second, the datasets share some similarities in their origin with HILDA+ incorporating much more diverse historical data sources.

The two classification products differ slightly in some aspects (e.g. agricultural trend, urban proportion) while agreeing well in crucial aspects (FSN proportion and trend). Using the continuous classification products is likely a reasonable reflection of real land cover change. It is therefore deemed an appropriate approach to consider both LCC products in testing for causal relationships by comparing proportional interdecadal changes in land cover classes with the residual ΔEI_r. In this case, we can accept the hypothesis proposed in 2 where remote sensing has been useful in filling the data gap in land cover change information for this period.

6.4 Regression Relationship

The land cover change data is tested for a direct relation between decadal changes in classes and the decadal change in ΔEI_r. The results show that a significant linear correlation between the decadal changes in residuals and changes in artificial, agricultural or FSN do not exist. This is the primary objective of this thesis and stands as a firm rejection of the hypothesis (3) that LCC is a significant driver of deviations from the Budyko curve, eliminating a significant uncertainty using historical observations. Clustering the data into sub-populations based on aridity

Linear relationships are not evident between the LCC values derived from observations in this study. If the areal proportion of vegetation or change in vegetation type (agricultural vs FSN) does not exhibit sufficient influence upon the evaporative index deviations, then the remaining unexplored options become the next logical steps for research.

7 Conclusions

Hydrological systems under changing climate present streamflow modellers and forecasters with clear problems of non-stationarity Blöschl et al. (2017); Bouaziz et al. (2022); van Noppen (2022). To refine our ability to predict and project one of the key sensitivities in hydrologic models is the root zone storage capacity, a biological parameter adapting to climate variability Tempel (2023). While Tempel was able to show that the relative errors in the root zone storage capacity were not significant enough to dismiss the efficacy of the Fu parameter projection method for significantly affecting the root zone storage capacity, the residual components in the evaporative index remained unexplained.

The residual results from the framework showed that the evaporative index is highly variable over time for all data in this study, and the relative contributions of residuals are twice as impactful as the climate-related errors. Catchments that are already humid are becoming increasingly humid. The Meuse data proves to be an outlier in this analysis concerning the residual error ΔEI_r , the median root zone storage capacity errors were calculated by Tempel (2023) to be -14%. Our analysis showed no evidence for the direct impact of changes in decadal low flows on these residual values in this period. Still, the change in the variability of low flows showed some spatial similarities.

Through the development and analysis of multi-scale classification products, it is clear that land cover change has occurred in the Meuse basin. Both the HILDA+ and Landsat products agree that urban surfaces have increased by an estimated 25% in the Meuse basin since 1990 and forest and semi-natural cover has decreased by an estimated 1%. Despite the development of two multi-scale classification products for the assessment of impacts on the residual values, it was clear from regression analyses that there were no causal relationships would emerge. That is to say that none of the observed decadal land cover changes had any direct impact on the derived evaporative components in the Budyko framework. In response to the overarching research question for this broader project, land cover changes on the ecosystem scale have not been observed to contribute significantly to the changes in root zone storage capacity, while climate and affected seasonality will have a profound effect (Tempel, 2023).

The unequivocal rejection of hypothesis (3) serves as a pivotal moment in our exploration of the drivers affecting the Evaporative Index (EI). This result compels us to confront a difficult truth—that the role of land cover change in influencing EI appears far less significant than previously presumed. The implication is profound: the laborious process of gathering and managing land cover data may no longer be justified as a priority, especially when there are other, less understood variables at play.

In particular, this study unearths a glaring lack of understanding of the residual components impacting EI, especially evident in the Meuse basin. The reliability of future projections hinges on comprehending these elusive elements. This begs the essential question: if land cover change isn't the key factor, then what is?

As we venture forward, three avenues of research emerge as imperatives. First, a concerted effort must be devoted to understanding the intra-annual dynamics of vegetation at the ecosystem level. This nuanced investigation could provide invaluable insights into the enigmatic variables influencing EI. Second, the role of forestry management practices and the standing age of trees, specifically in their alteration of the water balance and root zone storage capacity in the Meuse basin, demands closer scrutiny. And finally, it is necessary to better understand the intricate role of groundwater contributions and losses in the Meuse basin, as they could hold the key to some of the unexplained discrepancies and errors that this study has highlighted.

In closing, while rejecting a hypothesis can often feel like a setback, it acts as a catalyst, compelling us to reconsider established norms and inviting us to explore new questions.

8 Acknowledgements

Completing my Master of Science program in Environmental Engineering at TU Delft has long been an aspiration that at many points seemed almost impossible. Yet, today I stand at the threshold of that dream, profoundly grateful for the journey and the individuals who have guided me along the way.

My academic journey began in Dublin, where my undergraduate years were immersed in studying Earth's complexities and its profound geological history. This period set the stage for a compelling dichotomy in my life: the allure of a successful career in mineral exploration against the backdrop of my fervent love for our planet's delicate ecosystems. Dr. Sean McCleneghan, my professor in economic geology and undergraduate thesis supervisor, once opined that I was too much of a "green-horn" to sustain a career in the mineral exploration sector. His comment, whether intended as a reflection on my inexperience or as a subtle nudge towards my eventual disenchantment with the extractive industries, proved prescient. I have carried this insight with me, finding it accurate in both interpretations as I navigated my professional journey and eventually my exit from the industry.

The years prior to my time at TU Delft were enriched by my discovery of the work of Dr. Suzanne Simard, introduced to me through Peter Wohlleben's book, "The Hidden Life of Trees." Her research invigorated my fascination with subsurface processes and fortified my conviction to seek further education.

Enrolling at TU Delft, I found a sanctuary of intellectual pursuit and the opportunity to delve into hydrology—an intersection of my interests that harmoniously aligned with my core values. Dr. Markus Hrachowitz has been instrumental in this phase of my life, teaching us the importance of root zones in predicting river flows under changing conditions. His patient and composed demeanor gave me the confidence to fully engage with this project. My deepest appreciation extends to the entire faculty of Environmental Engineering, Water Management, and Water Resources at TU Delft, who have imbued me with purpose and conviction.

I owe a special debt of gratitude to Dr. Laurène Bouaziz from Deltares for her meticulous attention to detail in improving the quality of my writing and analysis. The project would not be what it is without her invaluable contributions.

My admiration for Eric Sprokereef and Jasper Stam from Rijkswaterstaat stems from their enduring curiosity and dedication in predicting and protecting the Netherlands' water bodies—a ceaseless source of inspiration for me. It is their curiosity that has enabled this research.

A delightful surprise came in the form of Dr. Miren Vizcaino, whose rigorous scientific approach, although outside her specific field, significantly elevated the quality of this work.

To my roommate Filemon, who deserves a heartfelt thank you for being a constant reminder to strike a work-life balance, while often going above and beyond in maintaining our shared living space.

To my friends, Dutch and International, who have become my second family here at TU Delft: your love and friendship have been my pillars of strength during the challenging moments that are part and parcel of this academic endeavour.

My parents have been the cornerstone of every accomplishment in my life. Supported by a loving circle that includes Nancy, Lucy, Elly, Rob, and Josh, they instilled an insatiable curiosity and a passion for knowledge, gifts that I consider invaluable.

In conclusion, this thesis is a tapestry woven from the myriad influences, moments, and individuals who have significantly shaped my intellectual and personal journey. I earnestly dedicate this work to my parents, Michele and Don; to Geraldine, who has been like a second mother to me; and to my cherished grandparents—Norrie, Christy, Nuala, and Michael. Their unwavering love and encouragement have not only made this academic venture possible but have also cultivated in me qualities that make me genuinely proud of the person I have become.

Thank you for reading.

List of Tables

List of Figures

References

Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. T., Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J.-H., Allard, G., Running, S. W., Semerci, A., and Cobb, N. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, 259:660–684.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. FAO Irrigation and drainage paper 56.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A. (2018). The camels-cl dataset: catchment attributes and meteorology for large sample studies – chile dataset. *Hydrology and Earth System Sciences*, 22(11):5817–5846.

B, P. C. H. and J, T. R. (1972). On the assessment of surface heat flux and ecaporation using large-scale parameters. *Monthly Weather Review*, pages 81–92.

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., and Tuller, M. (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics*, 57(2):530–616.

Bakouch, H. S., Chesneau, C., and Elsamadony, O. A. (2021). The gumbel kernel for estimating the probability density function with application to hydrology data. *Journal of Data, Information and Management*, 3:261 – 269.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., anjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N. L., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M. B., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P. H., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., raj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N. (2017). Changing climate shifts timing of european floods. *Science*, 357:588 – 590.

Boer, G. J. (1993). Dynamical climate change and the regulation of the surface moisture and energy budgets. *Climate Dynamics*, 8:225–239.

Bouaziz, L. J. E., Aalbers, E., Weerts, A. H., Hegnauer, M., Buiteveld, H., Lammersen, R., Stam, J., Sprokkereef, E., Savenije, H. H. G., and Hrachowitz, M. (2021). The importance of ecosystem adaptation on hydrological model predictions in response to climate change. *Hydrology and Earth System Sciences Discussions*, pages 1–39.

Bouaziz, L. J. E., Aalbers, E. E., Weerts, A. H., Hegnauer, M., Buiteveld, H., Lammersen, R., Stam, J., Sprokkereef, E., Savenije, H. H. G., and Hrachowitz, M. (2022). Ecosystem adaptation to climate change: the sensitivity of hydrological predictions to time-dynamic model parameters. *Hydrology and Earth System Sciences*, 26(5):1295–1318.

Budyko, M. I. (1974). Climate and life.

Buechel, M., Slater, L., and Dadson, S. (2022). Hydrological impact of widespread afforestation in great britain using a large ensemble of modelled scenarios. *Communications Earth Environment*, 3:6.

Buttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., and Soukup, T. (2004). The corine land cover 2000 project. *EARSeL eProceedings*, 3(3):331–346.

Cahyono, B. E., Putri, P. O., Subekti, A., Nugroho, A. T., and Nishi, K. (2022). Analysis of soil moisture as an indicator of land quality using vegetation index (savi and ndmi) retrieved from remote sensing data in jember - indonesia. page 020006.

Chander, G. and Markham, B. L. (2003). Revised landsat-5 tm radiometric calibration procedures and post-calibration dynamic ranges. *IEEE Trans. Geosci. Remote. Sens.*, 41:2674–2677.

Chander, G., Markham, B. L., and Helder, D. L. (2009). Summary of current radiometric calibration coefficients for landsat mss, tm, etm+, and eo-1 ali sensors. *Remote Sensing of Environment*.

Chicco, D. and Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21.

Cieślak, I., Biłozor, A., and Szuniewicz, K. (2020). The use of the corine land cover (clc) database for analyzing urban sprawl. *Remote. Sens.*, 12:282.

Cohen, W. B., Yang, Z., Stehman, S. V., Schroeder, T. A., Bell, D. M., Masek, J. G., Huang, C., and Meigs, G. W. (2016). Forest disturbance across the conterminous united states from 1985–2012: The emerging dominance of forest decline. *Forest Ecology and Management*, 360:242–252.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D. (2018). An ensemble version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17):9391–9409.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R. (2020). Camels-gb: hydrometeorological time series and landscape attributes for 671 catchments in great britain. *Earth System Science Data*, 12:2459–2483.

de Boer-Euser, T., Meriö, L., and Marttila, H. (2019). Understanding variability in root zone storage capacity in boreal regions. *Hydrology and Earth System Sciences*.

de Wit, M. (2008). *Van regen tot Maas: grensoverschrijdend waterbeheer in droge en natte tijden.* Veen Magazines.

Desborough, C. E. (1997). The impact of root weighting on the response of transpiration to moisture stress in land surface schemes.

Dixon, H., Hannaford, J., and Fry, M. J. (2013). The effective management of national hydrometric data: experiences from the united kingdom. *Hydrological Sciences Journal*, 58(7):1383–1399.

Eidenshink, J. C., Schwind, B. E., Brewer, K., liang Zhu, Z., Quayle, B., and Howard, S. (2007). A project for monitoring trends in burn severity. *Fire Ecology*, 3:3–21.

Fuh, B.-P. (1981). On the calculation of the evaporation from land surface. *Chinese Journal of Atmospheric Sciences*, 5:23–31.

Gao, H., Hrachowitz, M., Schymanski, S. J., Fenicia, F., Sriwongsitanon, N., and Savenije, H. H. (2014). Climate controls how ecosystems size the root zone storage capacity at catchment scale. *Geophysical Research Letters*, 41:7916–7923. Large sample catchment analysis of SRxY showing SR20y to correlate best with SUmax, NDVI. Also showing the distribution of SR/SUmax in 7 CEC geographic zones. Using Mass curve technique.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27.

Hardisky, M. A. (2008). The influence of soil salinity, growth form, and leaf moisture on-the spectral radiance of spartina alterniflora canopies.

Heidari, H., Arabi, M., Warziniack, T., and Kao, S. (2020). Assessing shifts in regional hydroclimatic conditions of u.s. river basins in response to climate change over the 21st century. *Earth's Future*, 8.

Held, I. M. and Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming.

Hiemstra, P. H. and Sluiter, R. (2011). Interpolation of makkink evaporation in the netherlands.

Hu, Z., Liu, S., Zhong, G., Lin, H., and Zhou, Z. (2020). Modified mann-kendall trend test for hydrological time series under the scaling hypothesis and its application. *Hydrological Sciences Journal*, 65(14):2419–2438.

Huang, C., Wylie, B., Yang, L., Homer, C., and Zylstra, G. (2002). Derivation of a tasselled cap transformation based on landsat 7 at-satellite reflectance. *International Journal of Remote Sensing*, 23(8):1741–1748.

Jaramillo, F., Cory, N., Arheimer, B., Laudon, H., Velde, Y., Hasper, T. B., Teutschbein, C., and Uddling, J. (2018). Dominant effect of increasing forest biomass on evapotranspiration: interpretations of movement in budyko space. *Hydrology and Earth System Sciences*, 22:567–580.

Jeong, J. and Lee, K. (1999). Bootstrapped white's test for heteroskedasticity in regression models. *Economics Letters*, 63(3):261–267.

Jian, W., Li, S., Lai, C., Wang, Z., Cheng, X., Lo, E. Y. M., and Pan, T. C. (2021). Evaluating pluvial flood hazard for highly urbanised cities: a case study of the pearl river delta region in china. *Natural Hazards*, 105:1691–1719.

Kauth, R. J. and Thomas, G. S. P. (1976). The tasselled cap - a graphic description of the spectral-temporal development of agricultural crops as seen by landsat.

Kennedy, R., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W., and Healey, S. (2018). Implementation of the landtrendr algorithm on google earth engine. *Remote Sensing*, 10:691.

Klingler, C., Schulz, K., and Herrnegger, M. (2021). Lamah-ce: Large-sample data for hydrology and environmental sciences for central europe. *Earth System Science Data*, 13(9):4529–4565.

Kriegler, F. J., Malila, W. A., Nalepka, R. F., and Richardson, W. (1969). Preprocessing transformations and their effects on multispectral recognition.

Lee, H., Calvin, K., Dasgupta, D., and Krinner, G. (2023). Synthesis report of the ipcc sixth assessment report (ar6).

Liu, H. Q. and Huete, A. (1995). A feedback based modification of the ndvi to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing*, 33:457–465.

Milczarek-Andrzejewska, D., Zawalińska, K., and Czarnecki, A. (2018). Land-use conflicts and the common agricultural policy: Evidence from poland. *Land Use Policy*, 73:423–433.

Milly, P. C. D. (1994). Climate, soil water storage, and the average annual water balance. *Water Resources Research*, 30:2143–2156.

Milly, P. C. D. and Dunne, K. A. (1994). Sensitivity of the global water cycle to the water-holding capacity of land. *Journal of Climate*, 7:506–526.

Nepstad, D. C., de Carvalho, C. R., and Davidson, E. A. (1994). The role of deep roots in the hydrological and carbon cycles of amazonian forests and pastures.

Newman, A. J., Clark, M. P., Sampson, K. M., Wood, A. W., Hay, L. E., Bock, A. R., Viger, R. J., Blodgett, D., Brekke, L. D., Arnold, J. R., Hopson, T. M., and Duan, Q. (2014). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous usa: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19:209–223.

Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., and Ohmann, J. L. (2010). Quantification of live aboveground forest biomass dynamics with landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, 114:1053–1068.

Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., and Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, 5:703–714.

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., Ogden, F., Selker, J., and Wendroth, O. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, 7(1):358–389.

Robinson, E., Blyth, E., Clark, D., Comyn-Platt, E., and Rudd, A. (2020). Climate hydrology and ecology research support system meteorology dataset for great britain (1961-2017) [chess-met].

Rodriguez-Iturbe, I. (2000). Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamies. *Water Resources Research*, 36:3 – 9.

Roy, D., Wulder, M., Loveland, T., C.E., W., Allen, R., Anderson, M., Helder, D., Irons, J., Johnson, D., Kennedy, R., Scambos, T., Schaaf, C., Schott, J., Sheng, Y., Vermote, E., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Hipple, J., Hostert, P., Huntington, J., Justice, C., Kilic, A., Kovalskyy, V., Lee, Z., Lymburner, L., Masek, J., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R., and Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145:154–172.

Schmidt, M., Klein, D., Conrad, C., Dech, S., and Paeth, H. (2014). On the relationship between vegetation and climate in tropical and northern africa. *Theoretical and Applied Climatology*, 115:341–353.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52:591–611.

Smakhtin, V. (2001). Low flow hydrology: a review. *Journal of Hydrology*, 240(3):147–186.

Sriwongsitanon, N., Gao, H., Savenije, H. H., Maekan, E., Saengsawang, S., and Thianpopirug, S. (2016). Comparing the normalized difference infrared index (ndii) with root zone storage in a lumped conceptual model. *Hydrology and Earth System Sciences*, 20:3361–3377.

Stone, E. and Kalisz, P. (1991). On the maximum extent of tree roots. *Forest Ecology and Management*, 46:59–102.

Survey, U. G. (2021). Landsat collection 2.

Tang, Y. and Wang, D. (2017). Evaluating the role of watershed properties in long-term water balance through a budyko equation based on two-stage partitioning of precipitation. *Water Resources Research*, 53(5):4142–4157.

Tempel, N. (2023). The effect of climate variability on the root zone storage capacity.

Thornton, P. E., Thornton, M., Mayer, B., Wei, Y., Devarakonda, R., Vose, R. S., and Cook, R. B. (2016). Daymet: Daily surface weather data on a 1-km grid for north america, version 3.

van Noppen, E. (2022). Climate-analogy mapping as a tool to develop a temporally-adaptive hydrological model of the meuse basin for more reliable predictions under change.

Velde, Y., Vercauteren, N., Jaramillo, F., Dekker, S. C., Destouni, G., and Lyon, S. W. (2014). Exploring hydroclimatic change disparity via the budyko framework. *Hydrological Processes*, 28.

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., van Dijk, A. I. J. M., Guerschman, J. P., Keys, P. W., Gordon, L. J., and Savenije, H. H. G. (2016). Global root zone storage capacity from satellite-based evaporation. *Hydrology and Earth System Sciences*, 20:1459–1481.

Winkler, K., Fuchs, R., Rounsevell, M., and Herold, M. (2021). Global land use changes are four times greater than previously estimated. *Nature Communications*, 12:2501.

Winkler, K., Fuchs, R., Rounsevell, M. D. A., and Herold, M. (2020). HILDA+ Global Land Use Change between 1960 and 2019.

Wit, M. J., den Hurk, B. V., Warmerdam, P. M. M., Torfs, P., Roulin, E., and van Deursen, W. P. A. (2007). Impact of climate change on low-flows in the river meuse. *Climatic Change*, 82:351–372.

Wit, M. J. M. D., Warmerdam, P. M. M., Torfs, P. J. J. F., Uijlenhoet, R., Roulin, E., Cheymol, A., Deursen, W. P. A. V., Walsum, P. E. V. V., Ververs, M., Kwadijk, J. C. J., and Buiteveld, H. (2001). Effect of climate change on the hydrology of the river meuse.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P. S., Vermote, E. F., Vogelmann, J., Wulder, M. A., and Wynne, R. (2008). Free access to landsat imagery. *Science*, 320(5879):1011–1011.

Wu, Q. (2020). geemap: A python package for interactive mapping with google earth engine. *Journal of Open Source Software*, 5:2305.

Xu, D. (2014). Compare ndvi extracted from landsat 8 imagery with that from landsat 7 imagery. *American Journal of Remote Sensing*, 2:10.

Young, D. (2014). Catchment streamflow response to climate change conditioned by historic alterations of land-use: Forest harvest, succession, and stand conversion.

Zhang, D., Cong, Z., Ni, G., Yang, D., and Hu, S. (2015). Effects of snow ratio on annual runoff within the budyko framework. *Hydrology and Earth System Sciences*, 19(4):1977–1992.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J. (2021). $\mathrm{Glc}_{fcs}30 : globalland - coverproductwithfineclassificationsystemat30musingtime - serieslandsatimagery. EarthSystemScienceData, 13 : 2753 - -2776.$

Zhao, J., Xu, Z., and Singh, V. P. (2016). Estimation of root zone storage capacity at the catchment scale using improved mass curve technique. *Journal of Hydrology*, 540:959–972.

Zhu, Z., Gallant, A. L., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., Jin, S., Dahal, D., Yang, L., and Auch, R. F. (2016). Optimizing selection of training and auxiliary data for operational land cover classification for the lcmap initiative. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122:206–221.

A Appendix

A.1 Potential Evaporation Equations

Included in this section are the various methods of evaporation employed by van Noppen (2022); Tempel (2023); Bouaziz et al. (2022) to build the datasets

A.1.1 Penman-Monteith Equation

$$\lambda_v E_a = \frac{\Delta(R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})} \tag{20}$$

The Penman-Monteith approach, as described, encompasses all the factors that influence energy exchange and the resulting latent heat flux (evapotranspiration) from uniform vegetated areas. These parameters can be measured or easily derived from weather data. This equation allows for the direct estimation of evapotranspiration for any specific crop by accounting for crop-specific surface and aerodynamic resistances Allen et al. (1998). This formulation is standard within the CAMELS-GB data but replaced by the Makkink formulation for E_p (equation 27).

A.1.2 Priestly-Taylor Formulation

The Priestly-Taylor equation was derived in 1972 to estimate potential evaporation over large, well-watered areas B and J (1972). It evolved from the Penmann-Monteith method (equation 20) to reduce reliance on observations, only radiation and α (experimentally $\alpha \approx 1.34 \pm 0.05$) are required:

$$E_P = \alpha \frac{\Delta}{\Delta + \gamma} - (R_n - G) \tag{21}$$

This formulation is employed in CAMELS-US as standard but is replaced by the Makkink formulation (22).

A.1.3 The Makkink Formulation

The Makkink formulation was derived in 1957 to estimate potential evaporation from temperature and radiation:

$$E_P = 0.65 R_n \cdot \frac{\Delta}{\Delta + \gamma} \frac{Rn}{\lambda \cdot \rho_w} \tag{22}$$

This formulation is used to calculate E_P in the Meuse data in the form detailed in Hiemstra and Sluiter (2011). For the sake of standardisation and remaining consistent with van Noppen (2022) and Tempel (2023) all datasets are re-calculated with the Makkink equation. For the CAMELS-US we replace the 21 by applying the equation 22 where Δ is the slope of the saturation water vapor pressure curve:

$$\Delta = \frac{7.5 \cdot 273.3}{(273.3 + T)^2} \cdot log_{10}(e_s) \tag{23}$$

The saturated vapor pressure in $[Pa]$:

$$e_s = 0.6107 \cdot 10^{\frac{7.5 \cdot T}{273.3 + T}} \tag{24}$$

The latent heat of vaporisation in $[Jg^{-1}]$:

$$\lambda_v = (2501 - 2375 * T) * 1000 \tag{25}$$

The psychrometeric constant in $[kPaK^{-1}]$:

$$\gamma = 0.0646 + 6e - 5 \cdot T \tag{26}$$

An alternative method is derived as elaborated in van Noppen (2022) to compensate for the unit difference in the CAMELS-GB set, which is provided in $Jm^{-2}d^{-1}$, the modified form of 22:

$$E_P = 0.65R_n \cdot \frac{\Delta}{\Delta + \gamma} \frac{K}{\lambda \cdot \rho_w} \tag{27}$$

To satisfy the units of 27, the slope of the saturated vapor pressure curve must be calculated as follows:

$$\Delta = \frac{abc}{(c+T)^2} \cdot exp(\frac{bT}{c+T}) \tag{28}$$

where a is found experimentally to be $6.1078[mbar]$, $b = 17.294$, $c = 237.3[^oC]$.

A.2 Github Link for Python Scripts

`https://github.com/manrahan/MOH_MSc_Thesis.git`

The above link contains all Python scripts used in this analysis.

A.3 Hydroclimatic and Budyko Analysis Supplement

A.3.1 Dataset Grouping

The following images display the trajectory analysis, with frequency, direction and magnitude. Each image is a dataset, GB and US where the Meuse image is attached in the main text.

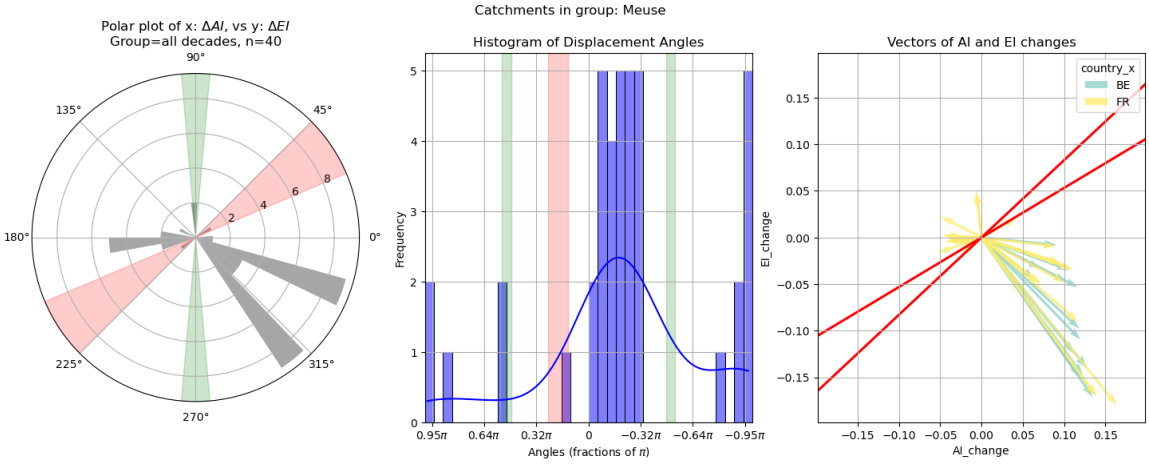


Figure 18: CAMELS GB trajectory histograms and vectors

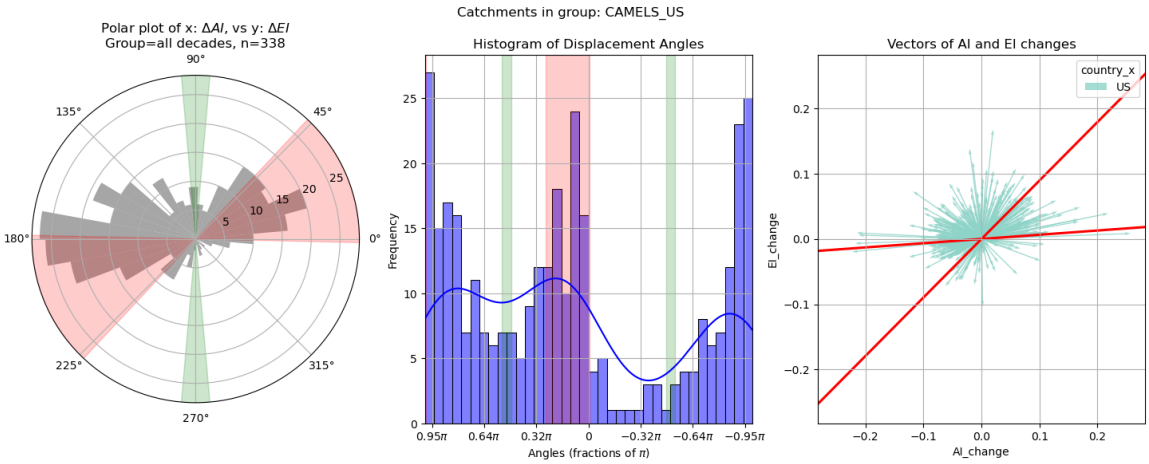


Figure 19: CAMELS US trajectory histograms and vectors

A.3.2 Aridity Grouping

The following data displays two tables for the proportion of each aridity group that is increasing or decreasing in either the aridity index or the evaporative index.

Table 9: Mean AI Change by Group

Group	**Mean AI Change**	**Percentage of $\Delta AI > 0$**	**Percentage of $\Delta AI < 0$**
0.2-0.4	-0.0098	23.86%	51.14%
0.4-0.6	-0.0018	27.86%	44.29%
0.6-0.8	-0.0055	27.79%	40.69%
0.8-1.0	-0.0087	26.03%	42.47%
1.0-1.2	0.0253	40.00%	26.67%
1.2-1.4	-0.0040	22.22%	44.44%
1.4-1.6	0.0353	55.56%	11.11%
1.6-1.8	-0.0325	20.00%	46.67%
1.8-2.0	NaN	NaN%	NaN%

Table 10: Mean EI Change by Group

Group	**Mean EI Change**	**Percentage of $\Delta EI > 0$**	**Percentage of $\Delta EI < 0$**
0.2-0.4	0.0003	34.09%	40.91%
0.4-0.6	-0.0092	35.00%	37.14%
0.6-0.8	0.0057	39.26%	29.23%
0.8-1.0	0.0139	44.86%	23.63%
1.0-1.2	0.0193	46.67%	20.00%
1.2-1.4	0.0162	55.56%	11.11%
1.4-1.6	0.0013	33.33%	33.33%
1.6-1.8	0.0038	40.00%	26.67%
1.8-2.0	NaN	NaN%	NaN%

The following images are the boxplots and trajectory plots to analyse tendencies within the data. data are grouped by aridity as in the previous table.

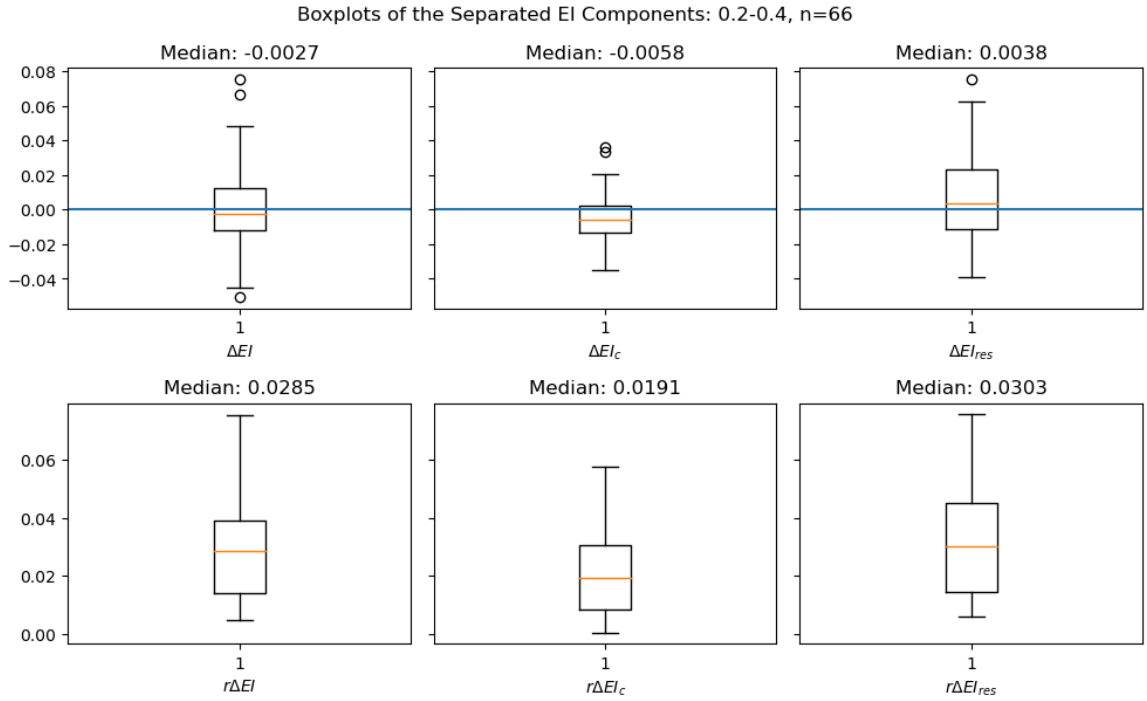


Figure 20: Aridity group 1 $0.2 > AI < 0.4$ boxplots of EI separations and overall magnitude of movements

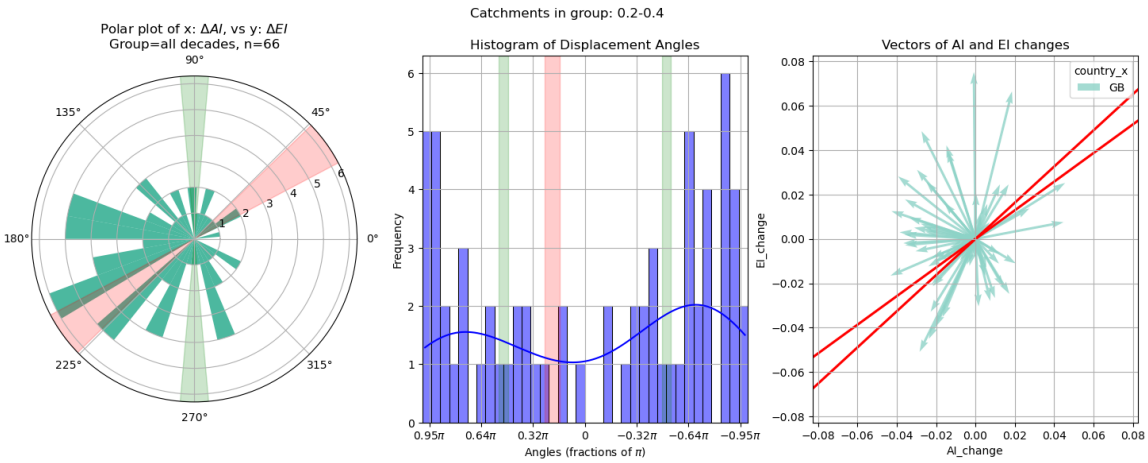


Figure 21: Aridity group 1 trajectory histograms and vectors

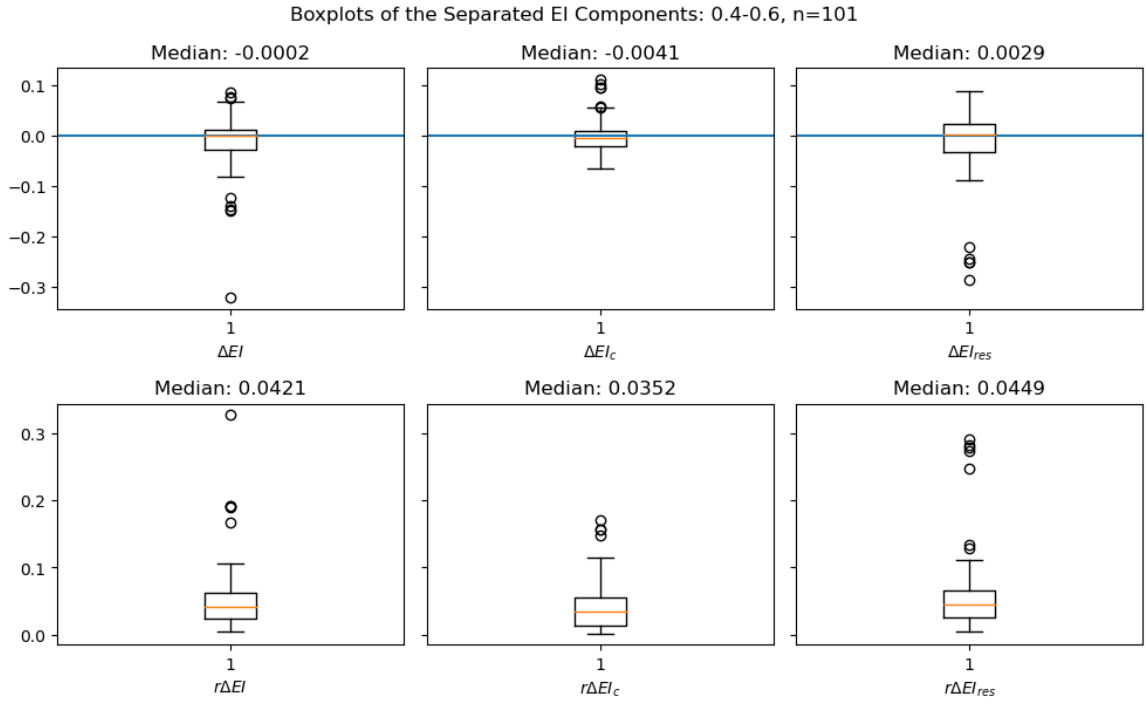


Figure 22: Aridity group 2 boxplots of EI separations and overall magnitude of movements

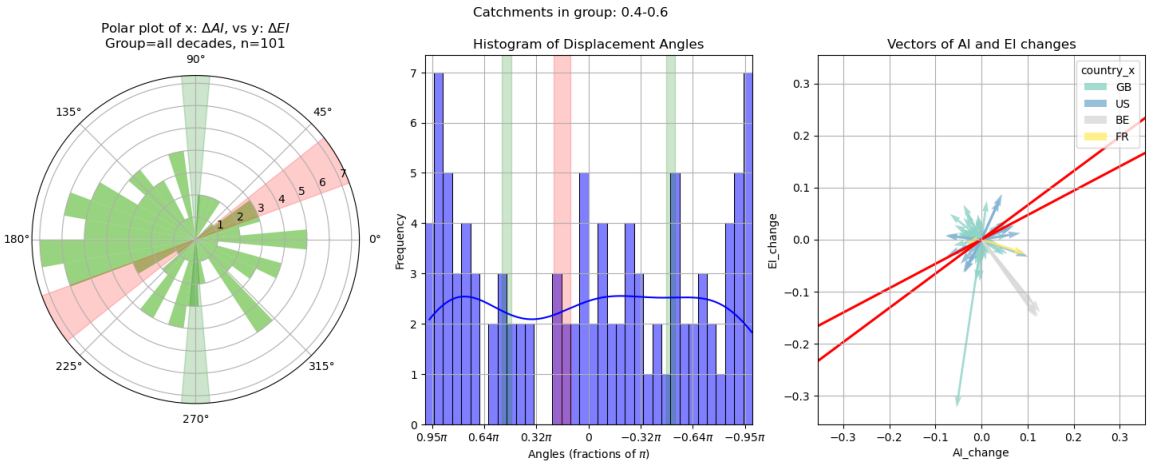


Figure 23: Aridity group 2 trajectory histograms and vectors

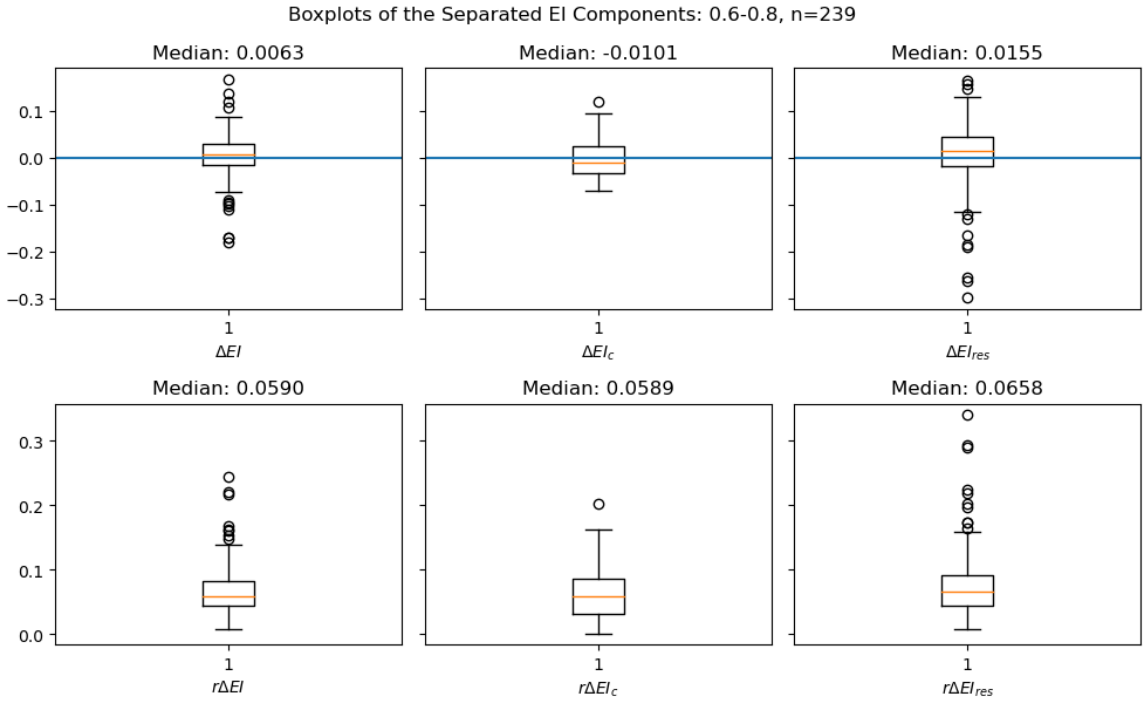


Figure 24: Aridity group 3 boxplots of EI separations and overall magnitude of movements

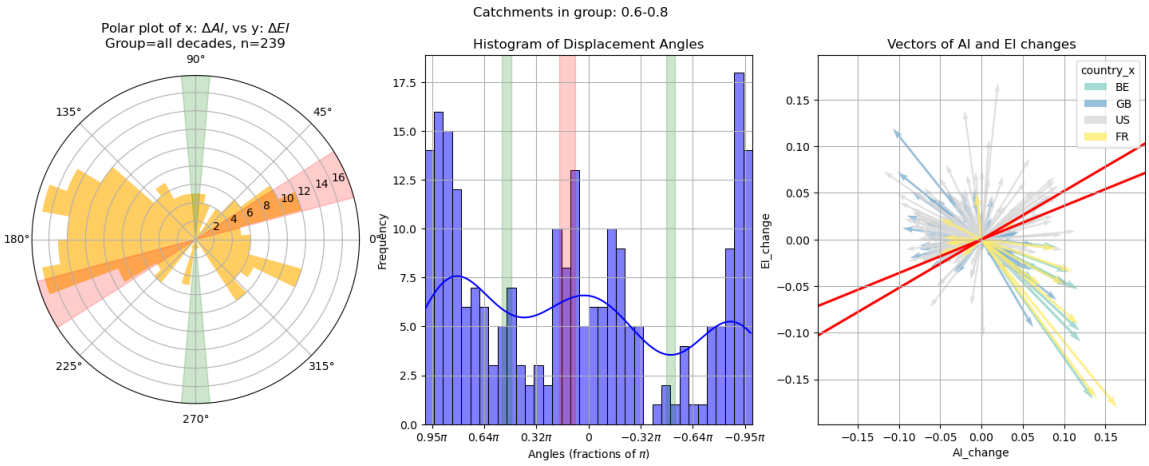


Figure 25: Aridity group 3 trajectory histograms and vectors

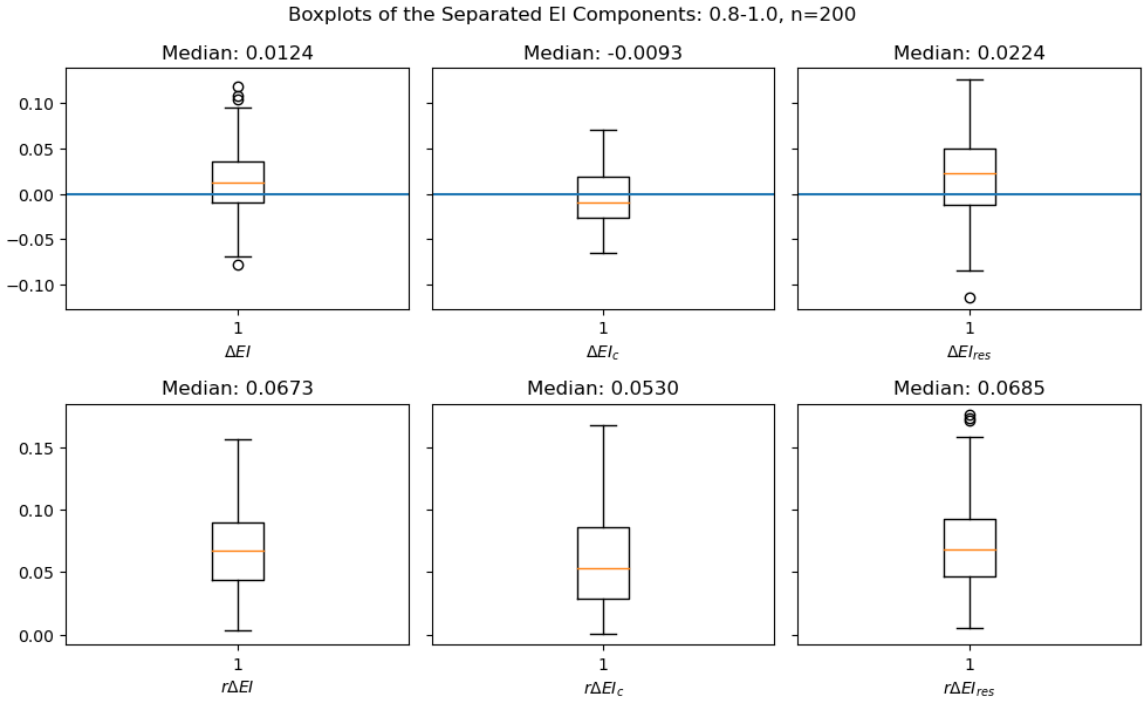


Figure 26: Aridity group 4 boxplots of EI separations and overall magnitude of movements

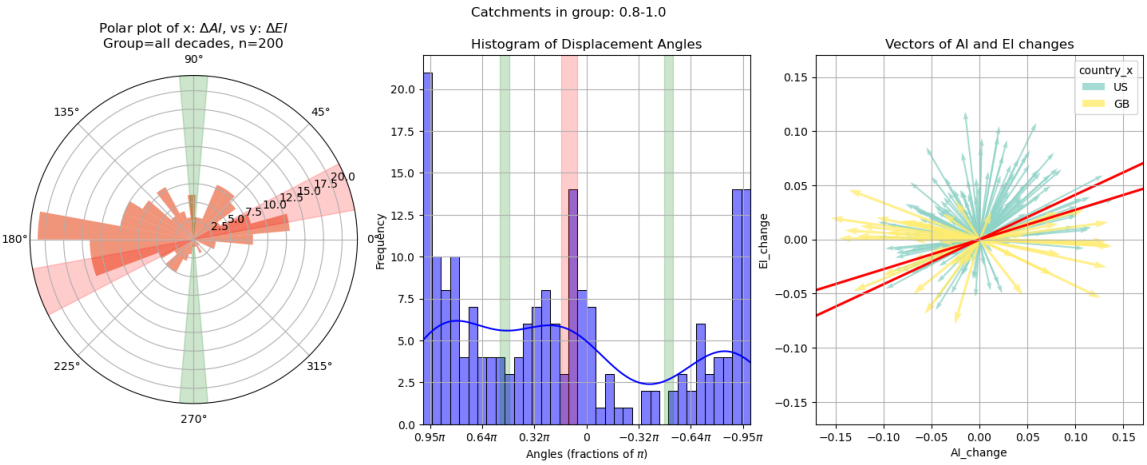


Figure 27: Aridity group 4 trajectory histograms and vectors

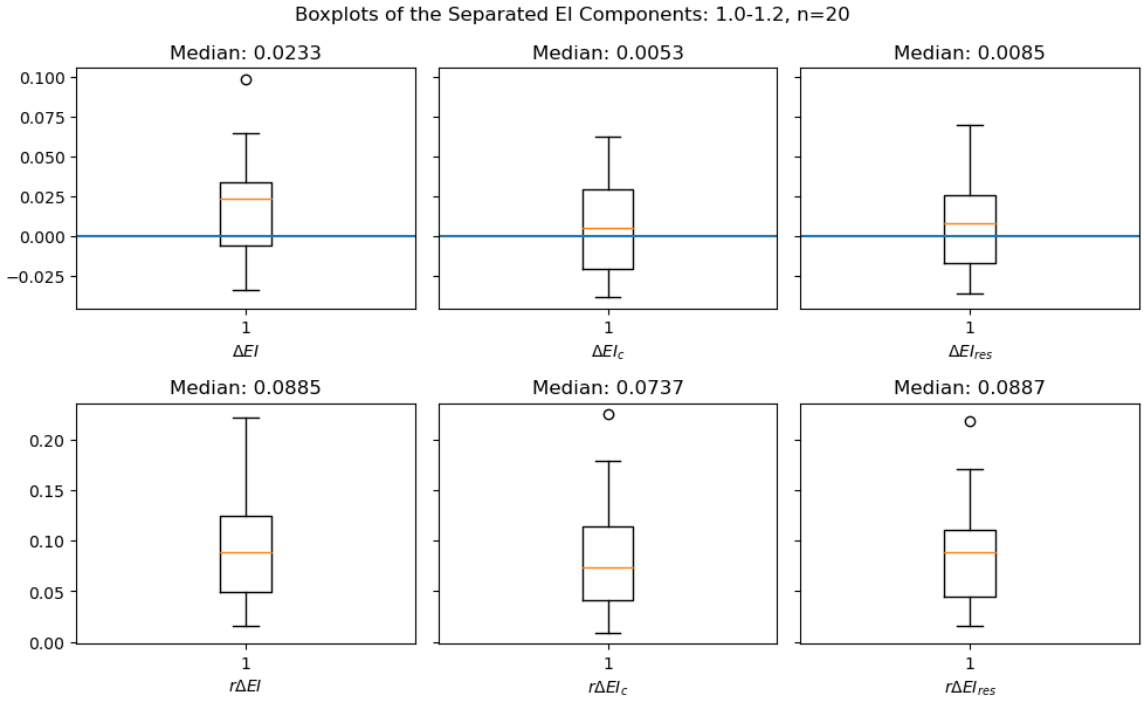


Figure 28: Aridity group 5 boxplots of EI separations and overall magnitude of movements

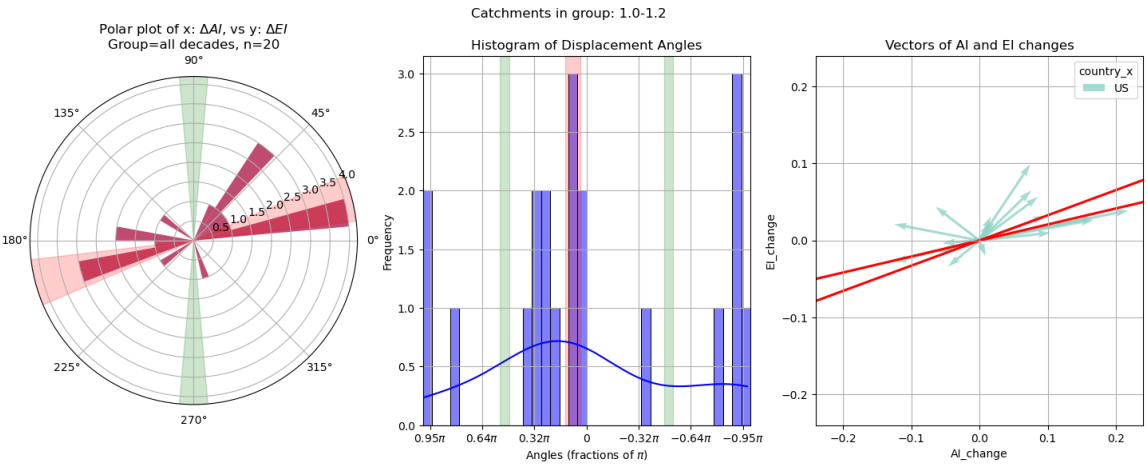


Figure 29: Aridity group 5 trajectory histograms and vectors

A.3.3 Error Propagation

Section 5.1.2 uses the law of error propagation to derive the following and estimate the error in relative contribution e.g in table 7:

$$\sigma \frac{|\Delta EI_r|}{|\Delta EI_c|} = \left| \frac{|a|}{|\mathrm{b}| \cdot a} \right| \cdot \delta a + \left| -\frac{|a|}{\mathrm{b} \cdot |\mathrm{b}|} \right| \cdot \delta b \tag{29}$$

A.3.4 Residual Change Values per Aridity Class

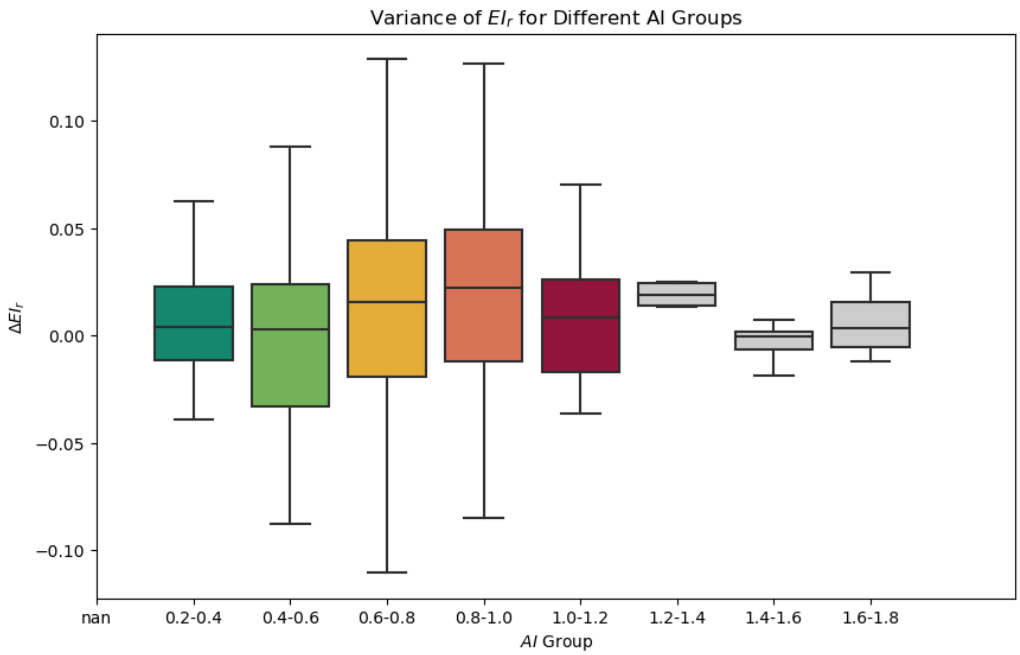


Figure 30: Variance of the residual component with increasing aridity

A.3.5 Additional Maps

Maps supporting the main maps included in the main text.

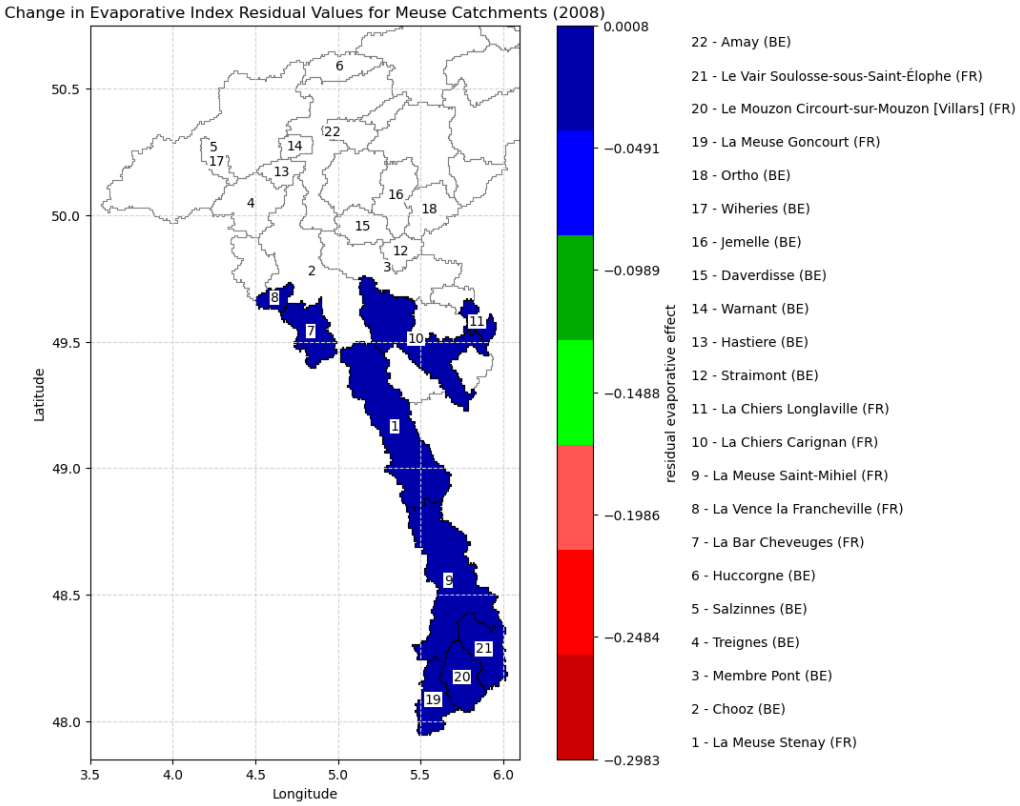


Figure 31: Meuse basin 2008 preceeding the residual anomaly

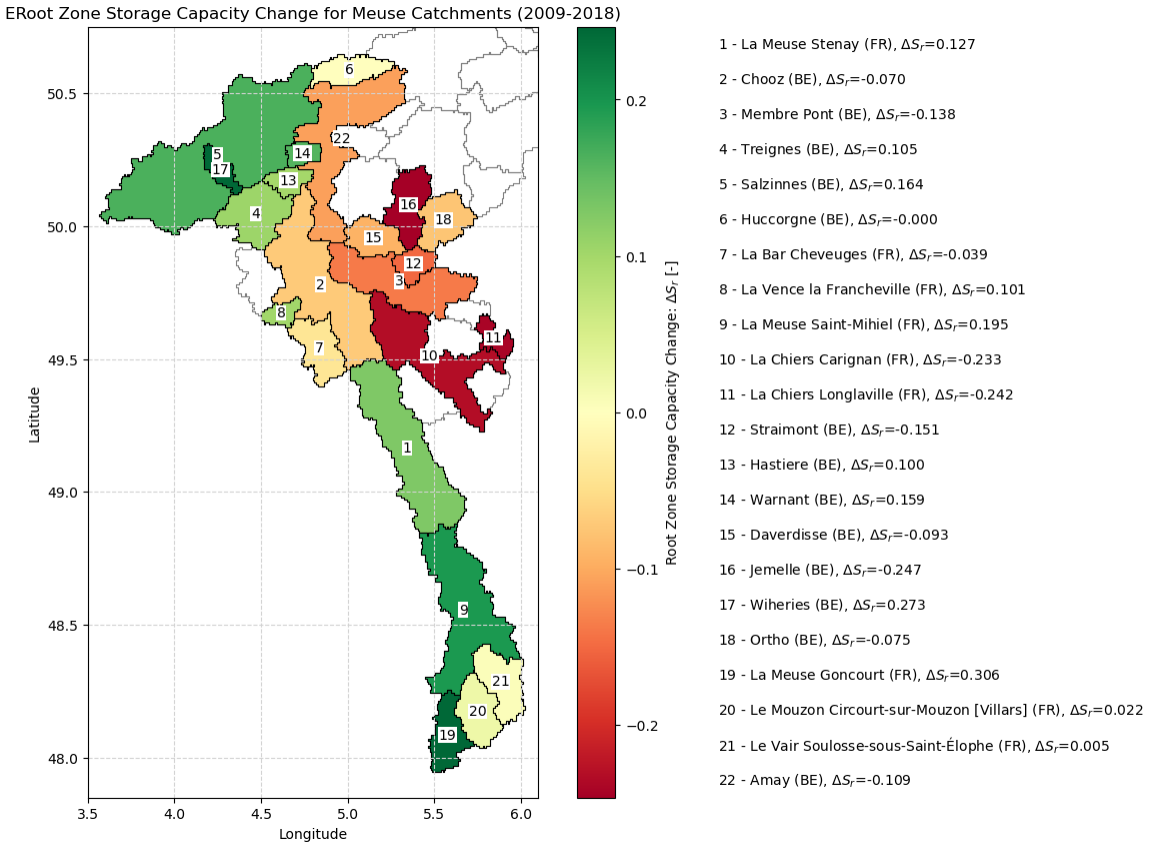


Figure 32: A map displaying the proportional change in root zone storage capacity ($\Delta S_{r,20y}$) component across the Meuse basin from the interdecadal period t_1:1999-2008 to t_2: 2009-2018

A.4 Low Flow Variability in The Meuse

Decade Ending	Q_{50_sum}	Q_{50_sum}	Q_{50_90}	Q_{50_90}	Q_{90_50}	Q_{90_50}	n catchments
Metric	Median	IQR	Median	IQR	Median	IQR	Sum
Unit	(m^3/decade)	(m^3/decade)	[-]	[-]	[-]	[-]	[-]
1999	2.17×10^{16}	9.04×10^{16}	3.046	1.840	0.328	0.000	9
2009	3.53×10^{13}	2.51×10^{13}	2.943	1.654	0.340	0.000	22
2018	2.78×10^{13}	1.24×10^{13}	3.390	1.812	0.295	0.000	22

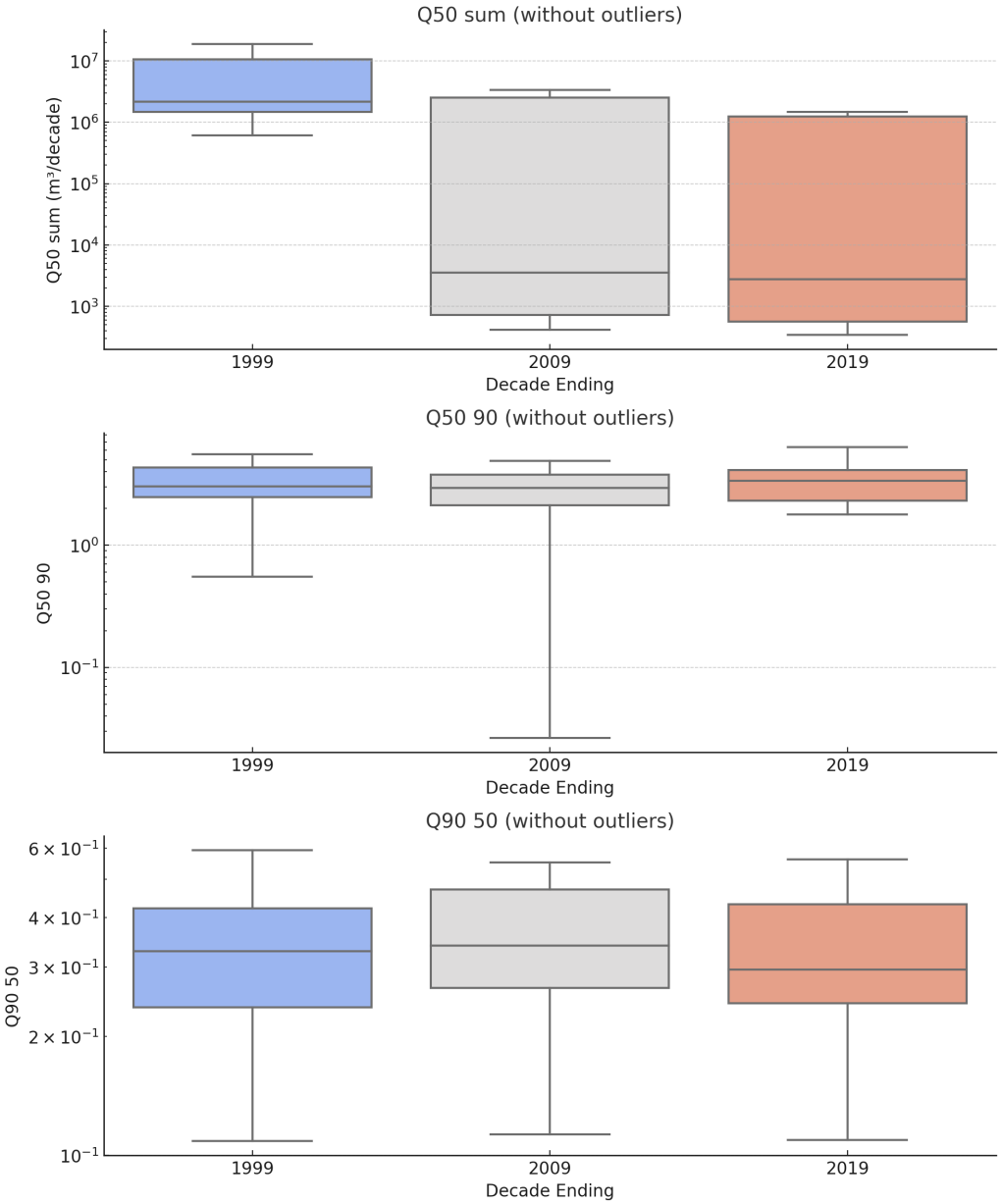


Figure 33: Boxplots for three decadal periods showing the range and variability of the low flow values over time in the Meuse

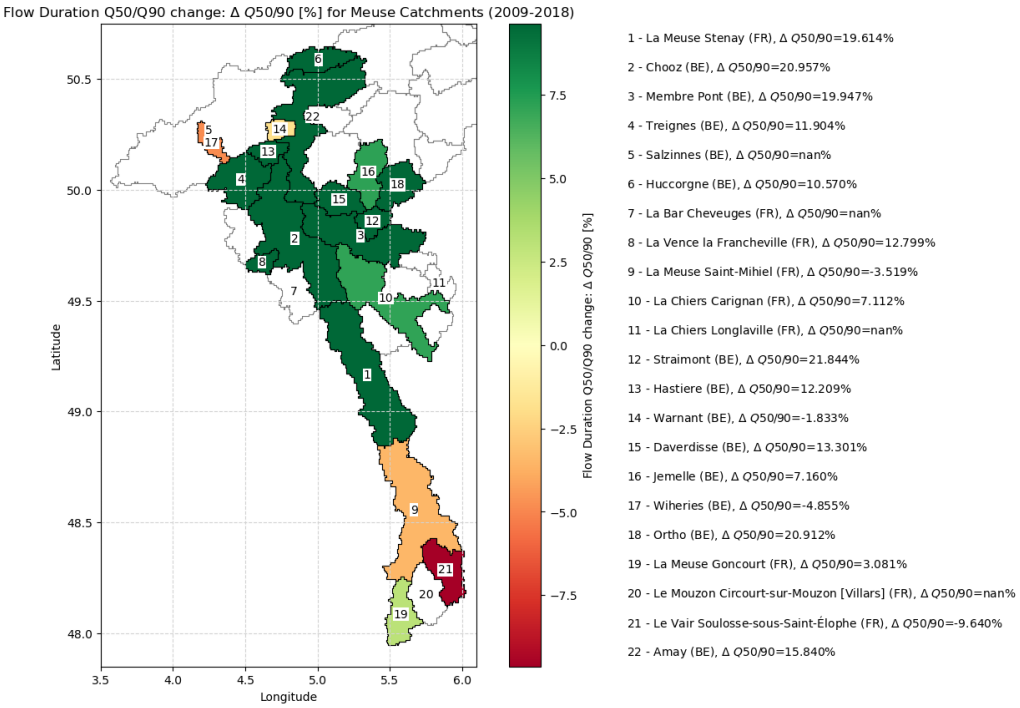


Figure 34: Meuse basin 2018 change values for the Q50 to Q90 ratio.

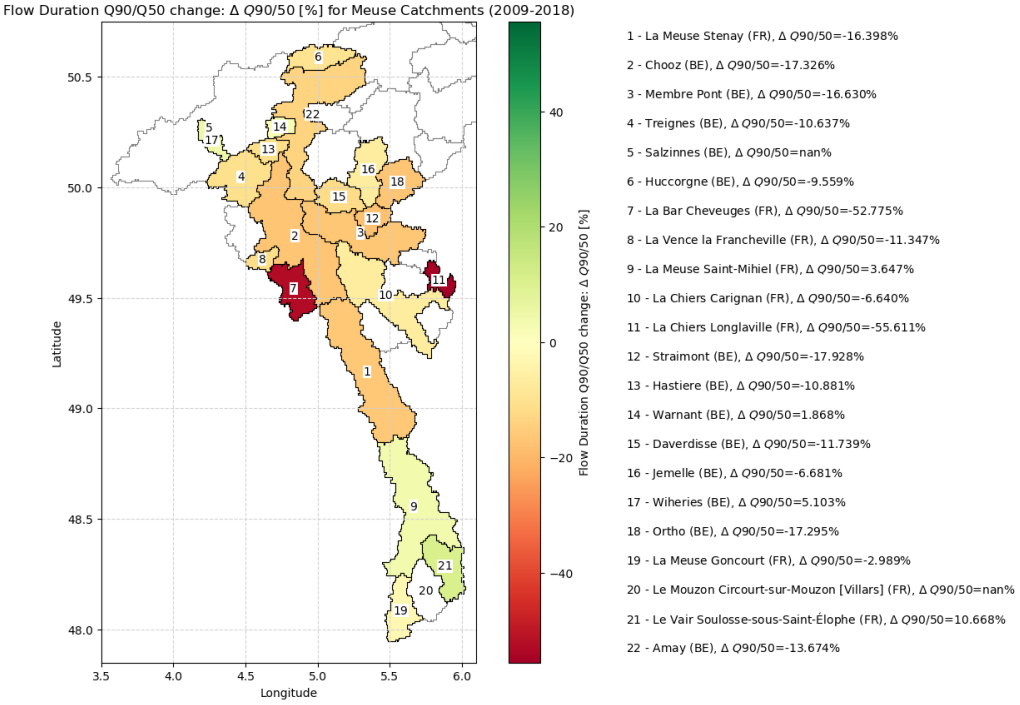


Figure 35: Meuse basin 2018 change values for the Q90 to Q50 ratio.

A.5 Landcover Change and Random Forest Classification

A.5.1 Zonal Statistic Bias Assessment

Table 11: Surface reflectance index biases calculated by the mann-kendall test for trend ($p_{sig} < 0.05$. The tested timeseries were for the full time period, the 199 total catchments tested is less than the total catchments in the overall study data. Data collection for this phase was halted as it was computationally expensive and not useable.

Index	Increasing	Decreasing	No Trend	Total
B1 mean	0	181	18	199
B2 mean	0	181	18	199
B3 mean	0	166	33	199
B4 mean	6	36	157	199
B5 mean	2	100	97	199
B7 mean	10	69	120	199
EVI mean	15	30	154	199
GNDVI mean	169	0	30	199
NBR mean	43	22	134	199
NDMI mean	40	12	147	199
NDVI mean	154	0	45	199
TCA mean	143	1	55	199
TCB mean	0	140	59	199
TCG mean	55	4	140	199
TCW mean	20	33	146	199

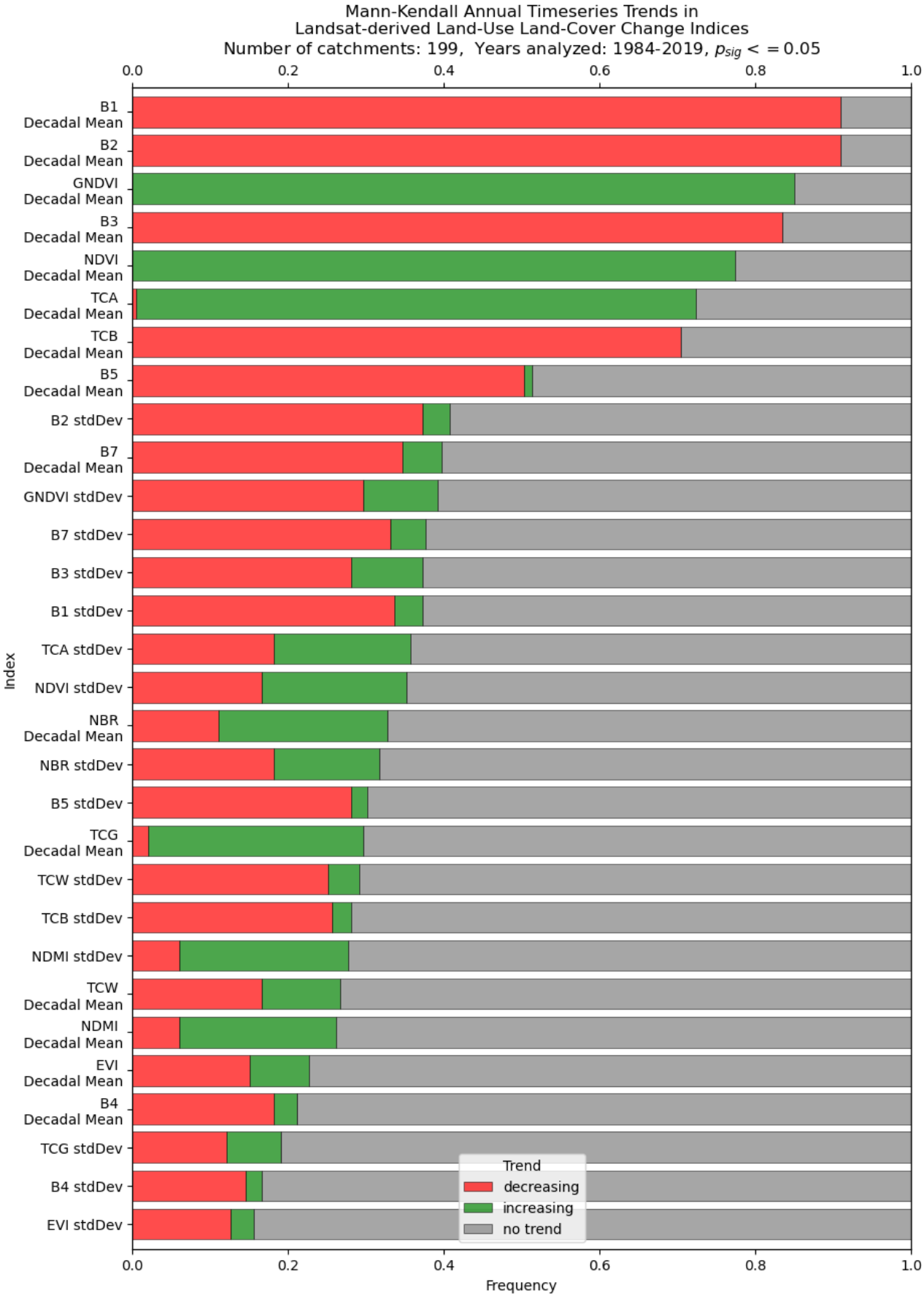


Figure 36: Proportion of Landsat band timeseries showing significant trends

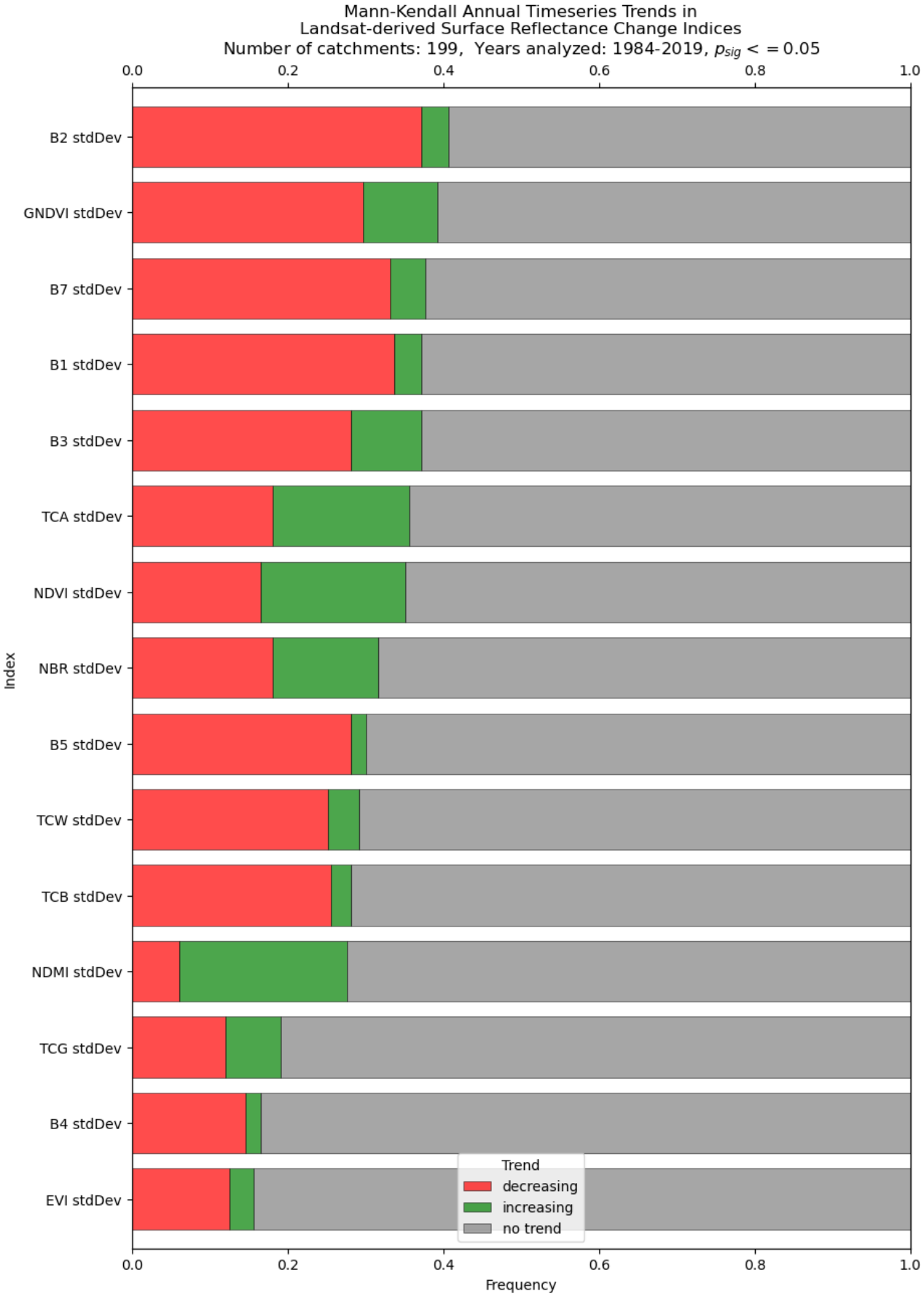


Figure 37: Stacked bar plot with proportion of significant trends identified in the all bands, including trends in standard deviations.

A.5.2 Tables of land cover proportions within the Meuse Basin over time

Table 12: Corine Meuse basin above Maastricht (Borgharen) with area reported in square kilometers where change is calculated in kilometers and percent ($t_1 = 1990$, $t_2 = 2018$)

	1990	**2000**	**2006**	**2012**	**2018**	**Change**	**Change %**	**Unit**
1: Artificial	1228.8	1253.4	1288.9	1334.0	1343.8	115.0	9.4	km^2
2: Agricultural	8701.2	8674.3	8647.1	8600.8	8592.1	-109.0	-1.3	km^2
3: Forest and Semi Natural	5966.2	5968.7	5959.2	5957.1	5955.6	-10.5	-0.2	km^2
4: Wetlands	11.5	8.3	8.5	9.9	9.9	-1.7	-14.5	km^2
5: Waterbodies	48.0	51.0	51.9	53.9	54.3	6.3	13.1	km^2

Table 13: HILDA+ Meuse basin above Borgharen Area and accurate change from 1990 to 2018

	1990	**2000**	**2006**	**2012**	**2018**	**Change %**	**Change**	
Urban	1194.6	1311.7	1356.3	1437.4	1498.6	25.5	304.0	
Agriculture	8123.2	8115.6	8036.6	7888.4	7891.4	-2.9	-231.8	Km^2
Forest and Semi Natural	6633.2	6522.1	6552.7	6620.3	6557.7	-1.2	-75.5	Km^2
Sparse/no vegetation	0.2	2.4	3.2	5.6	4.0	1900.0	3.8	
Water	1.6	1.6	1.6	1.6	1.6	0.0	0.0	

Table 14: Meuse basin above Borgharen Landsat Classification Area and accurate change from 1990 to 2018

	1990	**2000**	**2006**	**2012**	**2018**	**Change**	**Change %**
1: Artificial	826.87	756.203	871.75	930	1032.8	205.93	24.90
2: Agricultural	9259.27	8772.176	9215.86	9079	9108.8	-150.47	-1.63
3: Forest and Semi Natural	5812.435	5365.39	5803.77	5844	5752.4	-60.03	-1.03
4: Wetlands	0.1929	0.09521	1.13	1.08	0.7	0.51	237.21
5: Waterbodies	24.828	32.481	36.76	38.68	39.8	15.97	60.31
6: Masked	32.1	1029.3	26.4	62.9	21.2	-10.9	-33.93

Table 15: Meuse Basin above Borgharen All CORINE classes with classification levels and corresponding areas in square kilometers in 2018.

Area [km^2]	**Level 1**	**Level 2**	**Level 3**
4451.83	Agriculture	Pastures	Pastures
4258.65	Forest and semi natural areas	Forests	Broad-leaved forest
4176.05	Agriculture	Arable land	Non-irrigated arable land
2219.48	Forest and semi natural areas	Forests	Mixed forest
1841.35	Agriculture	Heterogeneous agricultural areas	Complex cultivation patterns
1777.83	Artificial surfaces	Urban fabric	Discontinuous urban fabric
1079.44	Forest and semi natural areas	Forests	Coniferous forest
724.59	Agriculture	Heterogeneous agricultural areas	Agriculture, with natural vegetation
247.35	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Transitional woodland-shrub
171.06	Artificial surfaces	Industrial, commercial, and transport units	Industrial or commercial units
61.42	Artificial surfaces	Artificial, non-agricultural vegetated areas	Sport and leisure facilities
51.81	Artificial surfaces	Mine, dump, and construction sites	Mineral extraction sites
43.39	Wetlands	Inland wetlands	Peat bogs
38.41	Water bodies	Inland waters	Water courses
37.87	Water bodies	Inland waters	Water bodies
28.16	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Moors and heathland
26.51	Artificial surfaces	Industrial, commercial, and transport units	Road and rail networks
18.79	Artificial surfaces	Industrial, commercial, and transport units	Airports
13.54	Agriculture	Permanent crops	Fruit trees and berry plantations
11.07	Wetlands	Inland wetlands	Inland marshes
10.87	Artificial surfaces	Urban fabric	Continuous urban fabric
7.42	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Natural grasslands
5.98	Artificial surfaces	Artificial, non-agricultural vegetated areas	Green urban areas
3.81	Artificial surfaces	Mine, dump, and construction sites	Construction sites
3.73	Artificial surfaces	Mine, dump, and construction sites	Dump sites
1.40	Artificial surfaces	Industrial, commercial, and transport units	Port areas
0.41	Forest and semi natural areas	Open spaces with little or no vegetation	Sparsely vegetated areas
0.27	Forest and semi natural areas	Open spaces with little or no vegetation	Bare rocks
0.02	Agriculture	Permanent crops	Vineyards

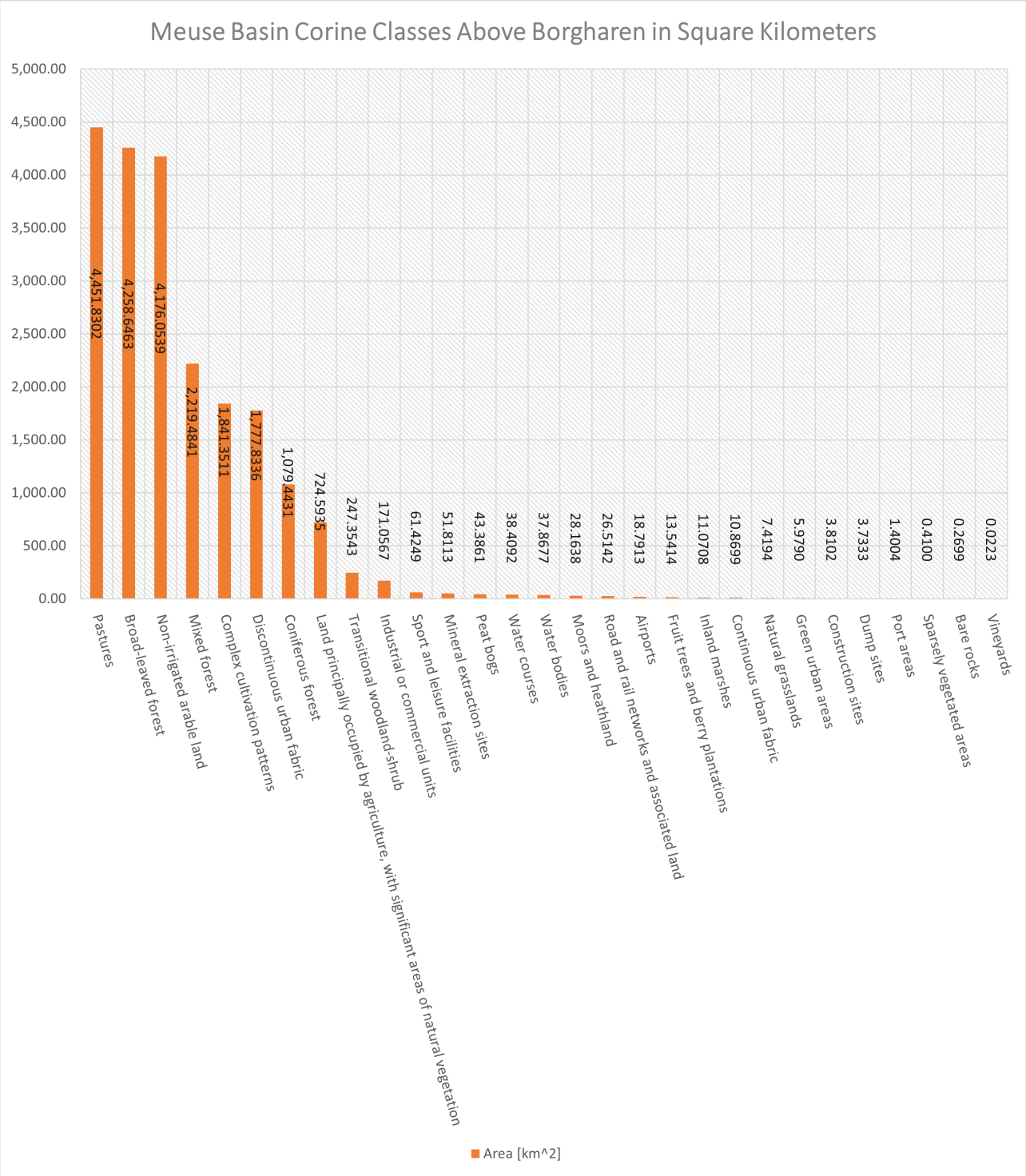


Figure 38: Sorted bar plot with all level 3 classes from CORINE 2012 data. Area is displayed in square kilometers

A.5.3 CAMELS GB Land Cover Changes

Median values for changes in all catchments considering the whole upstream area.

Table 16: Upstream Classification Normalized Change Landsat (CAMELS_GB)

Landsat Classification	Median Normalized Change (%)
Artificial	-0.052%
Agricultural	2.735%
Forest	0.772%
Wetlands	0.179%
Waterbodies	0.142%

Table 17: Normalized Change (CAMELS_GB)

HILDA+	Normalized Change (%)
HILDA + (Artificial)	0.237%
HILDA + (Cropland)	-0.442%
HILDA + (Pasture/Rangeland)	0.617%
HILDA + (Forest)	-0.218%
HILDA + (Grass/Shrubland)	-0.018%
HILDA + (Sparse/No vegetation)	0.000%

A.5.4 CAMELS US Land Cover Changes

Table 18: HILDA+ Normalized Change by Upstream Classification (CAMELS_US)

Upstream Classification	Normalized Change (%)
HILDA + (Artificial)	0.063%
HILDA + (Cropland)	-0.015%
HILDA + (Pasture/Rangeland)	-0.140%
HILDA + (Forest)	0.276%
HILDA + (Grass/Shrubland)	0.000%
HILDA + (Sparse/No vegetation)	0.000%

A.5.5 Meuse Land Cover Changes

Table 19: Upstream Classification Normalized Change Landsat (Meuse)

Upstream Classification	Normalized Change (%)
Classified (Artificial)	0.677%
Classified (Agricultural)	5.799%
Classified (Forest)	-4.435%
Classified (Wetlands)	0.004%
Classified (Waterbodies)	0.004%
Classified (Masked cloud)	-0.602%

Table 20: HILDA+ Upstream Classification Normalized Change (Meuse)

Upstream Classification	Normalized Change (%)
HILDA + (Artificial)	0.181%
HILDA + (Cropland)	0.531%
HILDA + (Pasture/Rangeland)	-0.940%
HILDA + (Forest)	0.263%
HILDA + (Grass/Shrubland)	-0.237%

A.5.6 Hyperparameter Tuning

Figure 39: Calibrating the number of trees hyperparameter

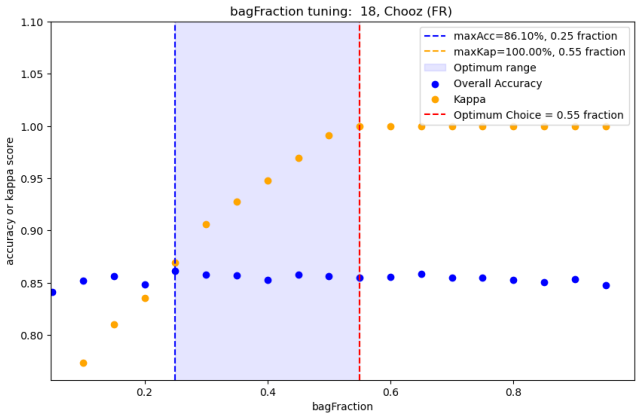


Figure 40: Calibrating the bag fraction hyperparameter

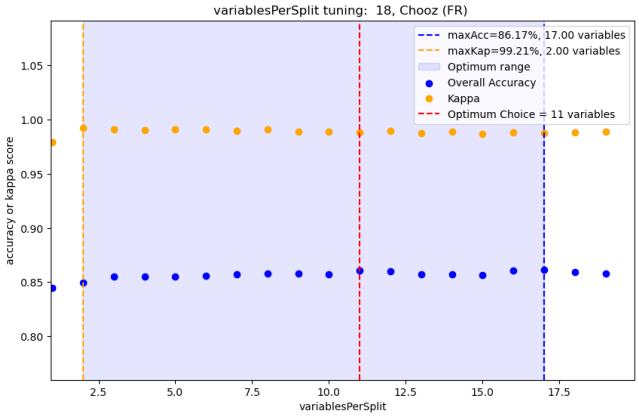


Figure 41: Calibrating the variables-per-split hyperparameter

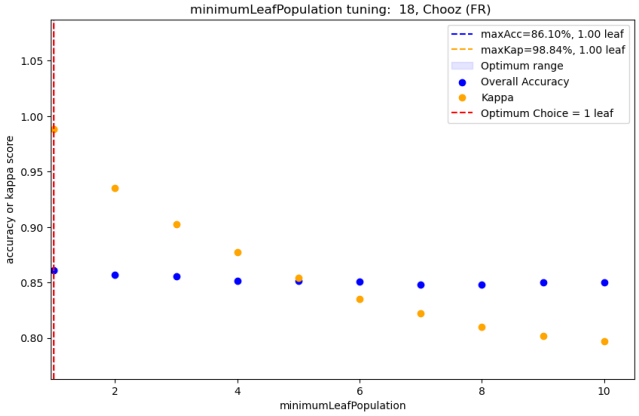


Figure 42: Calibrating the minimum leaf hyperparameter

A.5.7 Unbalanced Classification F1-Score Timeseries

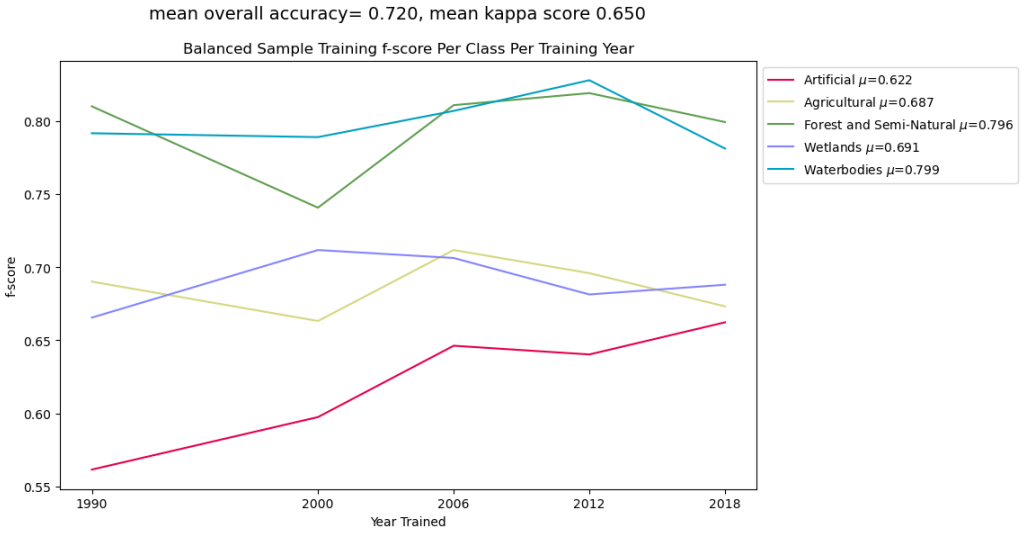
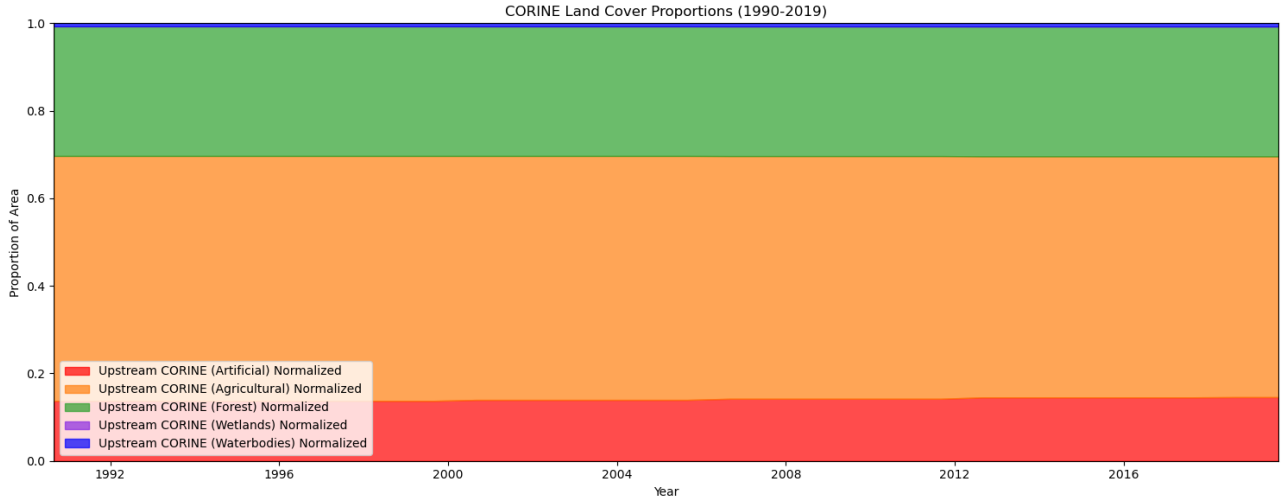
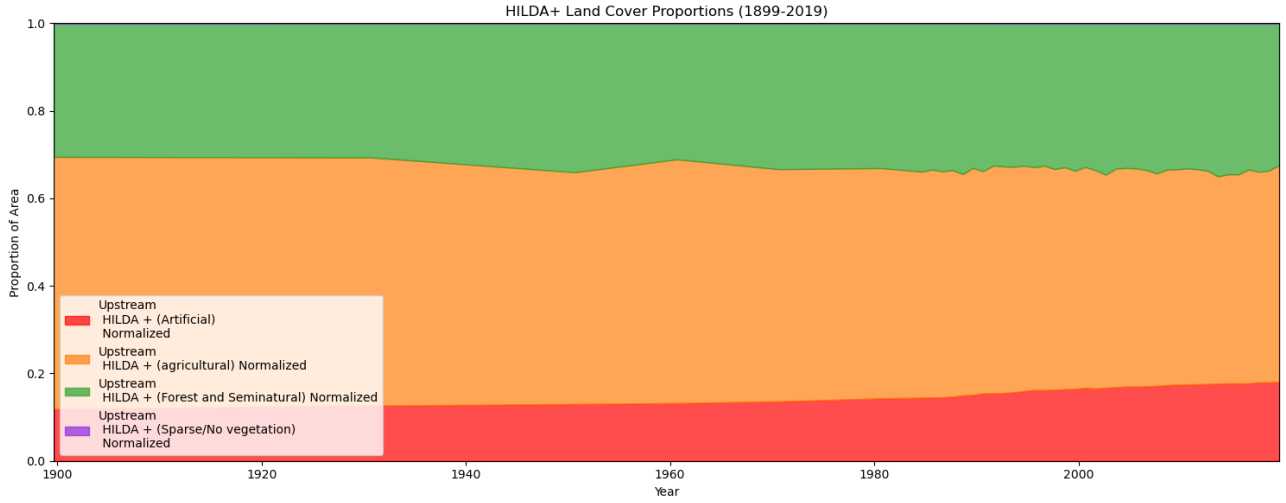


Figure 43: Training and classification for all years in the CORINE dataset with model performance assesed by F1-score

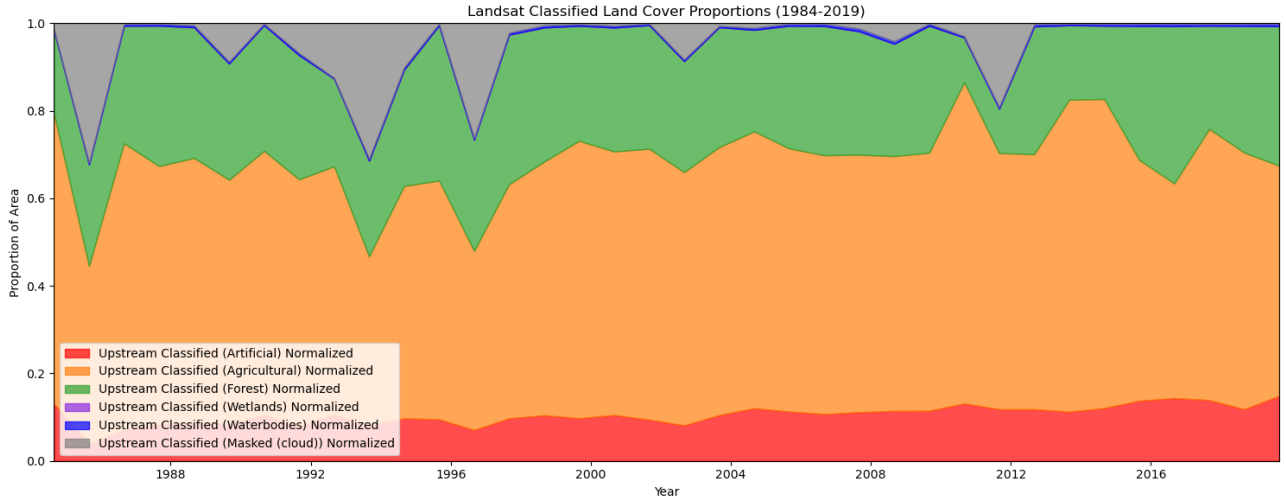
A.5.8 Land Cover Class Timeseries



(a)



(b)



(c)

Figure 44: Three Dataset timelines illustrating the landcover change in the Meuse basin above Borgharen. Red indicates artificial surfaces the classes of which are detailed in 12 (a) is the Copernicus CORINE dataset from 1990 (first available year in mainland Europe, see 2.

A.5.9 Model Feature Importance

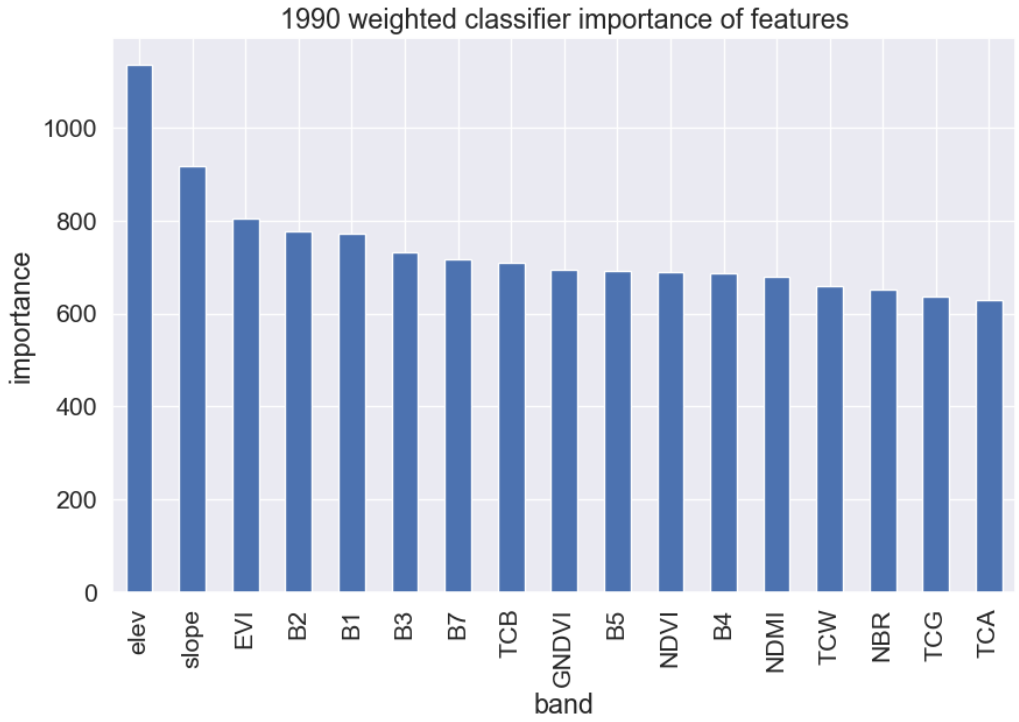


Figure 45: The random forest ranked classification importance of bands in contribution to determining classes.

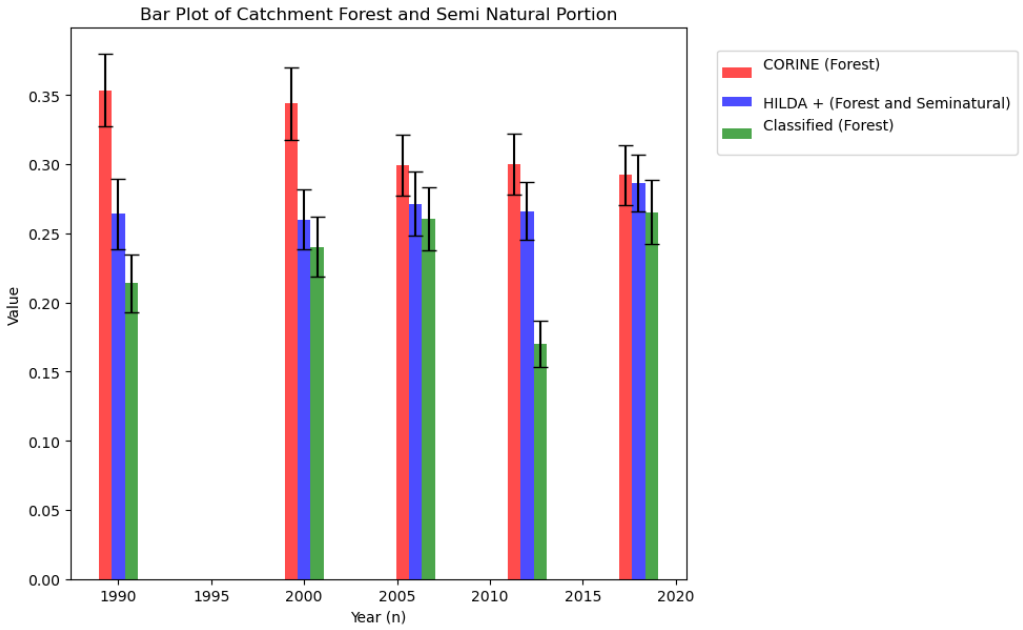


Figure 46: Three models are presented for the Meuse basin and CAMELS GB ($n_{catchments} = 122$). HILDA+ model is in blue, the CORINE data in red, the classified Landsat classified product is in green. Each bar is composed of the a range of area proportions estimated to be forest and semi-natural land, in the case of the Meuse, each bar and standard error represents 22 catchments upstream of Borgharen.

A.5.10 Confusion Matrices for Each Year Classified

For each training year, a model and associated classification confusion matrix is calculated.

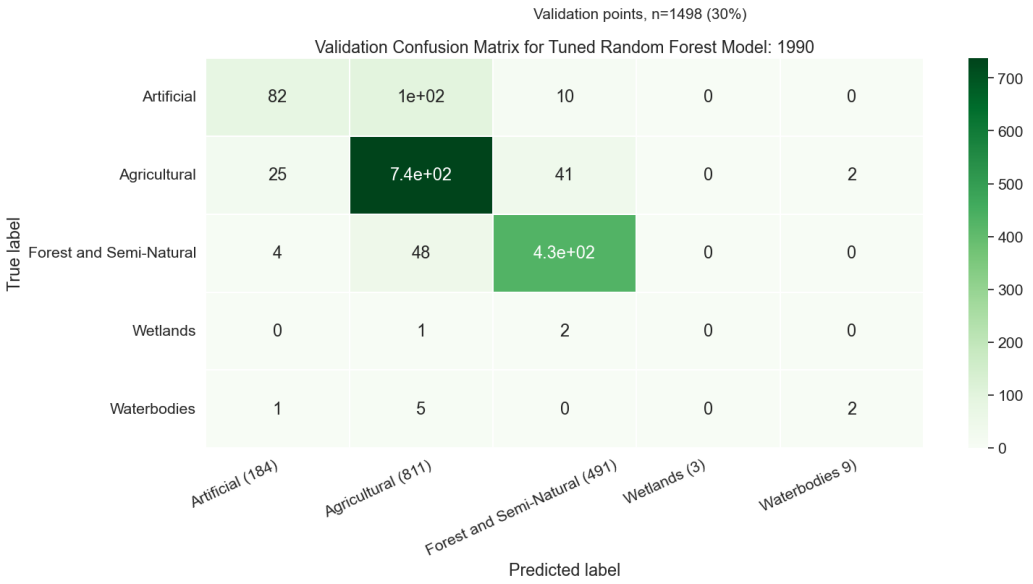


Figure 47: Confusion matrix for the un-balanced, tuned classifier used for data around 1990

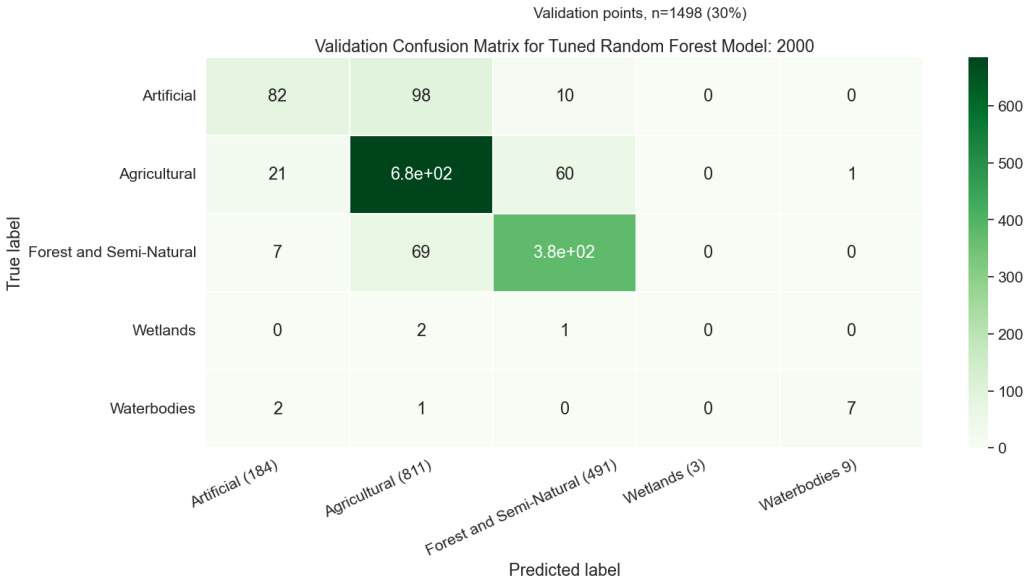


Figure 48: Confusion matrix for the un-balanced, tuned classifier used for data around 2000

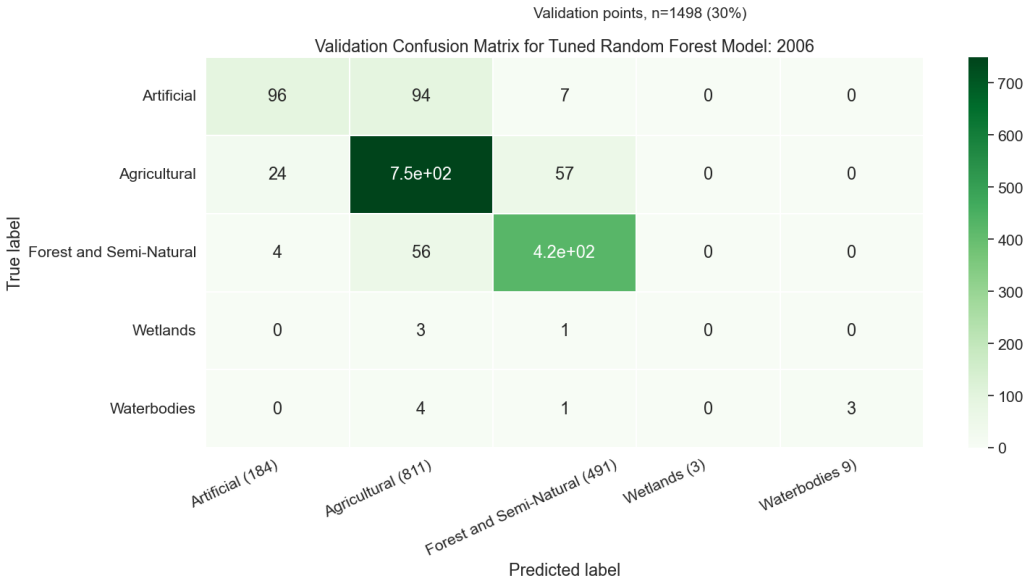


Figure 49: Confusion matrix for the un-balanced, tuned classifier used for data around 2006

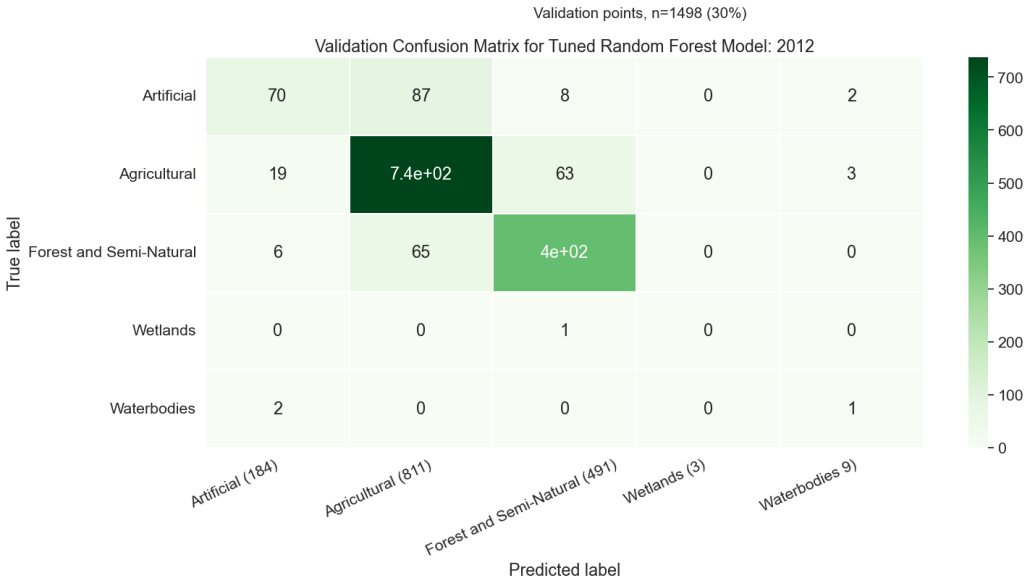


Figure 50: Confusion matrix for the un-balanced, tuned classifier used for data around 2012

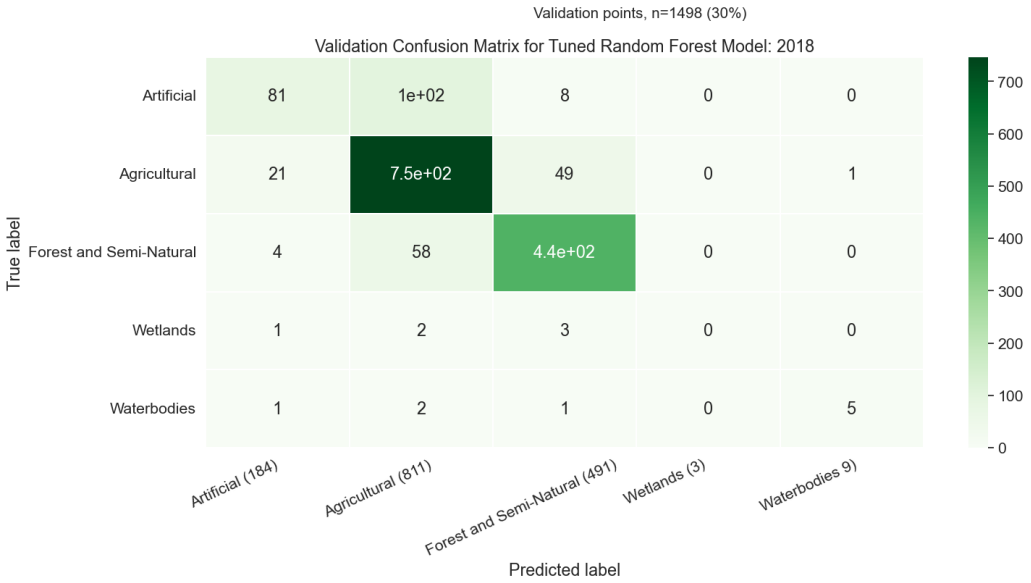


Figure 51: Confusion matrix for the un-balanced, tuned classifier used for data around 2018

A.6 Regression Results

Explanatory	n	Response Var	R-squared	p
HILDA + (Artificial)	333	dy_res	0.0121	1.3160E-18
HILDA + (Cropland)	432	dy_res	0.0285	2.6082E-05
HILDA + (Pasture/Rangeland)	494	dy_res	0.0030	0.0267
HILDA + (Forest)	518	dy_res	0.0188	3.5361E-05
HILDA + (Grass/Shrubland)	356	dy_res	0.0022	0.4838
HILDA + (Sparse/No vegetation)	654	dy_res	0.0033	0.2633
Upstream Classified (Artificial)	175	dy_res	0.1366	0.1501
Upstream Classified (Agricultural)	213	dy_res	0.0247	0.0012
Upstream Classified (Forest)	208	dy_res	0.0656	0.0009
Upstream Classified (Waterbodies)	19	dy_res	0.0681	0.2417
Upstream Classified (Masked cloud)	221	dy_res	0.0041	0.0000
Aridity Index	654	dy_res	0.3883	0.4052
Evaporative Index	654	dy_res	0.0352	0.0758
Potential Evaporation	654	dy_res	0.0955	2.1911E-20
Precipitation	654	dy_res	0.3121	0.0003
Seasonality Index	654	dy_res	0.0549	0.0029
Seasonality Timing	654	dy_res	0.0134	0.2939
Root-zone Storage Capacity	654	dy_res	0.0306	0.0014
Temperature	654	dy_res	0.0845	4.5452E-80
Interstorm Duration	654	dy_res	0.0445	0.1851
Budyko Curve (Fu) Parameter	654	dy_res	0.0985	0.0002

A.7 LAI Timeseries Issues

The leaf area index is a measure of vegetation density. Coupled with areal extent of vegetation, powerful insights into ecosystem scale physiology can be employed. The timescale and spatial resolution of this investigation prohibited the complete integration of LAI products available. It was attempted to use a regression model to join the old FAPAR (AVHRR sensor) which is now defunct to the modis (VIIRS sensor) to establish a sufficiently long timeseries (albeit with a coarse resolution. The result was substandard and determined to be unusable. Of particular concern was the discontinuity in the region of the join at 2004. An example of some select Meuse catchments is in the figure below.

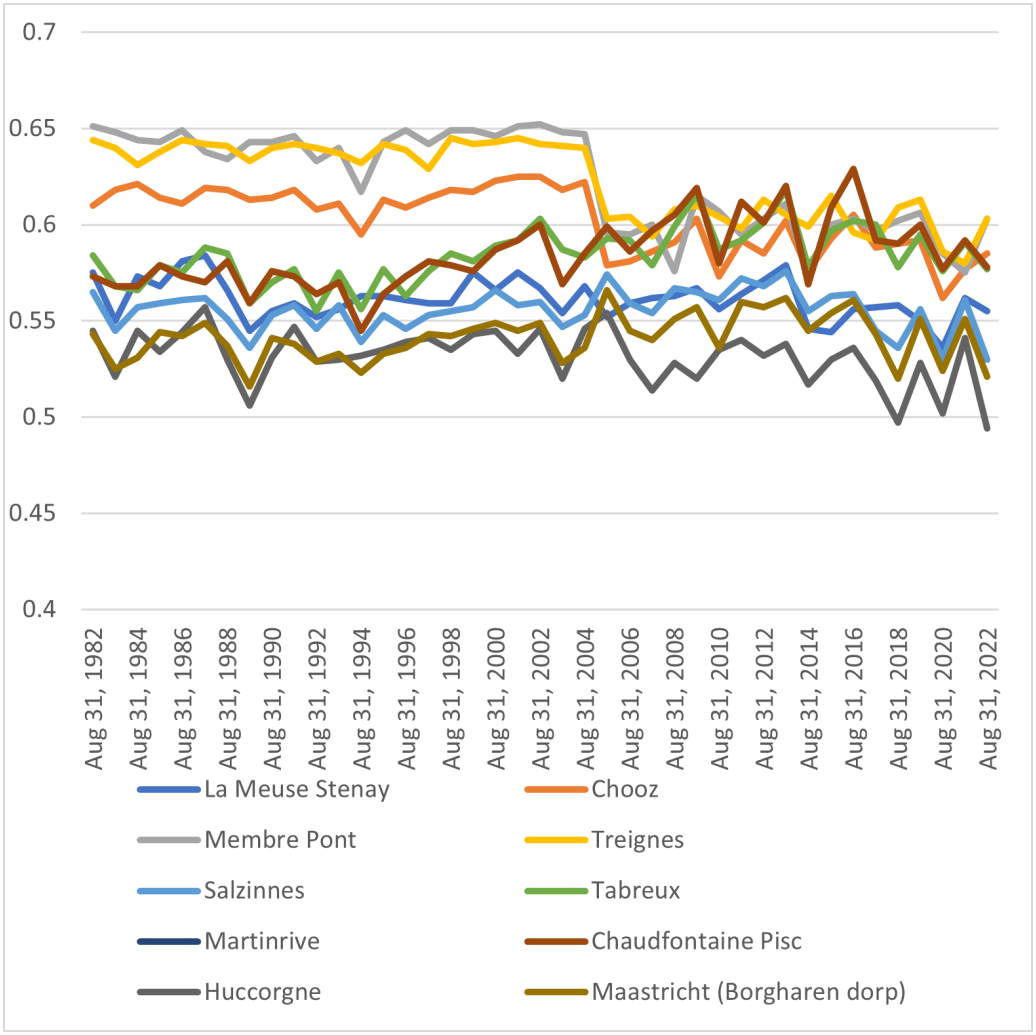


Figure 52: LAI joined timeseries