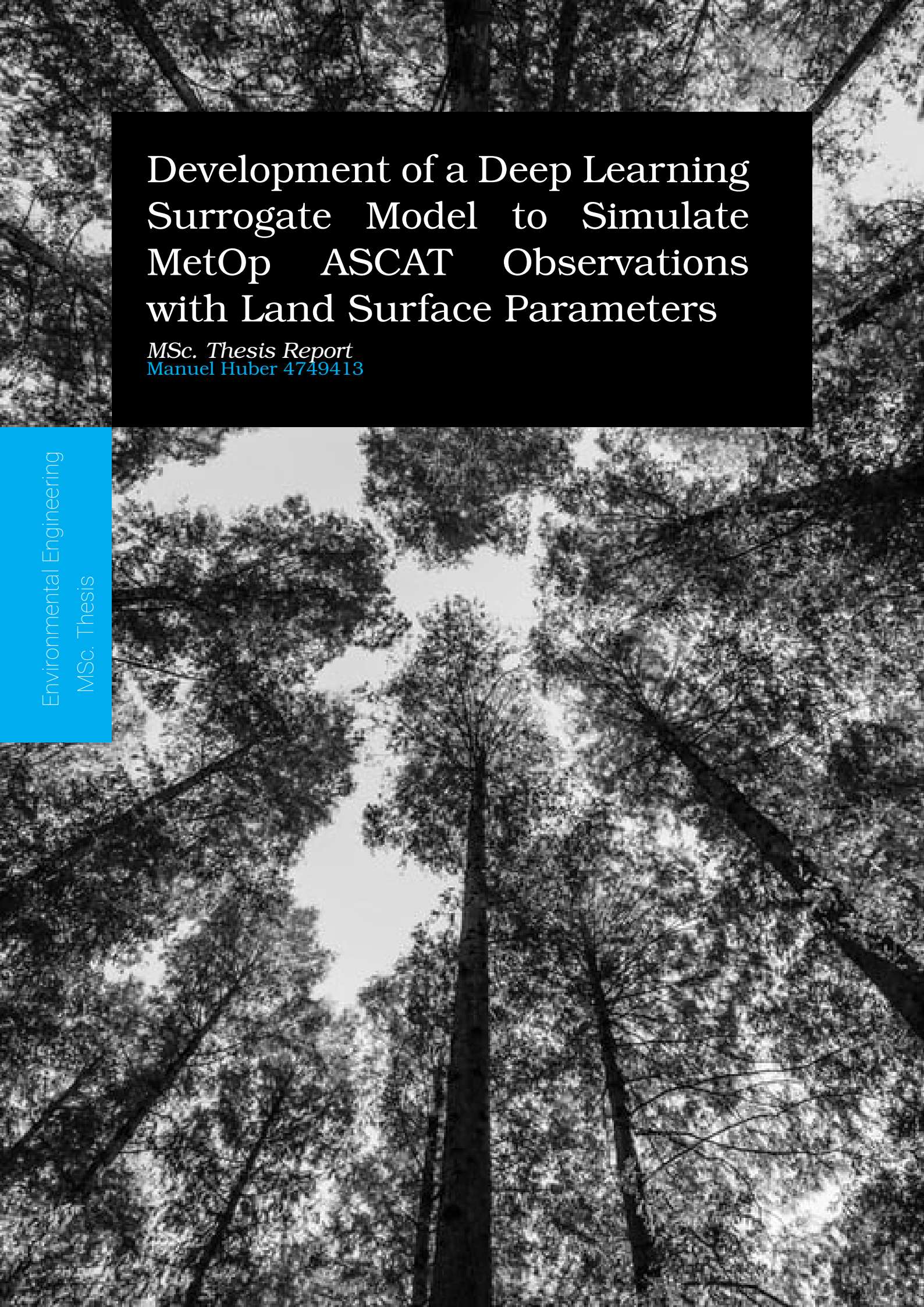

Development of a Deep Learning Surrogate Model to Simulate MetOp ASCAT Observations with Land Surface Parameters

MSc. Thesis Report

Manuel Huber 4749413

Environmental Engineering
MSc. Thesis

Graduation Report

by

Manuel Huber 4749413

Committee

Chair Prof. Dr. ir. Susan Steele-Dunne
Dr. I.V. (Ihor) Smal
Dr.ir. Miriam Coenders (-Gerrits)

TU Delft

Preface

"Learning never exhausts the mind" is a famous quote of Leonardo da Vinci. He also said that experience is the daughter of wisdom. And the daughter definitely tells me that I was and am very exhausted by the vast amount of learning I had to do during my thesis. But I genuinely enjoyed it and I feel very grateful to had such an amazing thesis committee - Miriam, Ihor and Susan. Thank you very much for all your help, motivation and guidance throughout the whole process, especially Susan. It has been a blast to work with you and learn from you. I really enjoyed the ride with all of you.

I could not have enjoyed it without all the great friends and fellow students I met during my stay at TU Delft. Thanks guys! I also want to thank my family at home in Germany and my second family here in The Netherlands. It would not have been possible without all your support, food and love. I specifically want to thank my life partner - Nurxia Dhimat Kinanthi Sutopo. Thank you for all the support, understanding, discussions and laughter. I am endlessly grateful to have you in my life. Now it is time to move forward.

Manuel Huber
Delft, November 2019

Abstract

In this study, deep neural networks are employed to act as a surrogate model between the Meteo France land surface model and Advanced SCATterometer (ASCAT) satellite observations. This provides a measurement operator for the assimilation of ASCAT satellite derivations into this model. Currently, TU Wien uses ASCAT measurements to retrieve soil moisture from backscatter. Next to backscatter signal, two additional vegetation parameters are extracted from the TU Wien Soil Moisture Retrieval Approach. These parameters are slope and curvature and describe the second order Taylor polynomial, which explains the incidence dependency of backscatter. A recent study showed that slope and curvature could contain valuable information about vegetation water dynamics and biomass phenology. The new explored relationship gives an unique opportunity to relate land surface variables with these observation parameters. This is significant as it could be used to create a climatological data set of high quality and temporal consistency.

The surrogate model avoids the need to use a Radiative Transfer Model (RTM) to relate the land surface model to the ASCAT observations. RTM's require complex input variables such as size, shape, height, thickness and orientation of the canopy but also the dielectric properties. Additionally, RTM's are not based on the actual output of the land surface model (LSM), as the LSMs simulate vegetation parameters such as leaf area index, soil moisture, gross primary production, temperature and respiration. This makes RTMs less suitable to act as a measurement operator. The suggested method to simulate ASCAT observations are deep neural networks. Deep neural networks are able to capture every highly non-linear relationship by using only the outputs from the LSMs.

For this study a regular feed forward deep neural network is used to simulate the backscatter signal of the ASCAT instrument, whereas slope and curvature are simulated by a deep convolutional neural network. The results show that the deep neural network is able to simulate the seasonal and inter-seasonal variation of backscatter. Concerning slope the model was capable to capture the seasonal trends and some of the inter-seasonal variations. Curvature shows the worst model performance, as the model is only able to capture the timing of the seasonal changes but not the right magnitudes. In general, the performance depends on the variation of the observation and land surface data. This suggests that the model structure needs to be adapted according to the complexity of the investigated grid point.

A black-box interpretation model, called DeepSHAP, is used to extract the most important features for each observation simulation. This feature importance allows a physical interpretation of the ASCAT observations. The most relevant feature for backscatter is soil moisture, which is consistent with previous research and gives confidence to the feature importance extraction method. The slope signal is mostly related to the gross primary production and therefore biomass assimilation. Curvature shows the highest correlation for LAI. Both results confirm the previous assumptions that curvature and slope are, respectively, related to structural and phenology changes.

The results of this research are substantial as they allow to the first time to relate actual vegetation parameters to the slope and curvature signals. It additionally proves that deep neural networks are a possible choice to act as a surrogate model between ASCAT observations and a land surface model.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
AMSR	Advanced Microwave Scanning Radiometer
ANN	Artifical Neural Network
ASCAT	Advanced Scatterometer
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
DNN	Deep Neural Network
ESCAT	ERS Scatterometer
EUMESAT	Exploration of Meteorological Satellites
GPI	Grid Point
GRU	Gated Recurrent Unit
ISBA	Interactions between soil biosphere and atmosphere
KS	Kernel Smoother
LAI	Leaf Area Index
LIME	Local interpretable model agnostics explanation
LSM	Land Surface Model
LSTM	Long Short Term Memory
METOP	Meteorological Opertaional
NASA	National Aeronautics Space Administration
NWP	Numerical Weather Prediction
RE	Ecosystem Respiration
RELU	Rectified linear Unit
RNN	Recurrent Neural Network
RTM	Radiative Transfer Model
SAR	Synthetic Aperture Radar
SCA	Second generation C-Band Scatterometer
SHAP	Shapley additive explanation
SIG	Sigmoid function
SLAR	Side looking airborne Radar
SMAP	Soil moisture Active Passive
SMOS	Soil moisture and Ocean Salinity
SMR	Soil moisture Retrieval
SURFEX	Externalized Surface
SVM	Support Vector Maschine
T-SNE	t distributed Stochastic Neighbour Embedding
TANH	Hyperbolic tangent function
WG2	Soil moisture

1 Introduction

In this chapter the research motivation, problem statement, research objectives and research question of the project are present. Here, the fundamental background information and specific problems are addressed to provide a general overview for the context.

1.1. Research Motivation

Disciplines such as hydrology, climate science, agriculture, water management, weather forecasting and early warnings are strongly contingent on the understanding of the land-atmosphere interactions [17]. Vegetation plays a important role in the mesoscale processes between land and surface and is dependent on several conditions [10] [13]. For example vegetation density and internal water vegetation water dynamics have great influence on several properties such as the surface albedo, infrared emission, transpiration, diurnal temperature range, rainfall, runoff and infiltration and surface energy and moisture budgets [10] [53].

Radar backscatter has proven in many studies to be highly related to soil moisture, vegetation structure and vegetation water content [44][50][51]. Compared to optical waves radar penetrates deeper into the vegetation and soil and therefore reveals information about deeper soil layers and the ground beneath vegetation [51]. Moreover, microwaves are able to penetrate through clouds and are not dependent on the illumination of the sun, which means that the observations do not rely on day or night conditions [51]. This leads to radar observation as the leading technique for many different civilian applications.

In 1991 the satellite mission ERS-1 was launched and generated together with ERS-2 20 years of global active microwave C-band scatterometry data (ESCAT) [53]. During these decades an exceedingly amount of research was carried out to investigate this large data-set, which shows that the seasonal vegetation dynamics as well as soil moisture are strongly correlate to the C-band backscatter data[19]. Many of the ESCAT soil moisture validation studies show good results independent from each other [18][54][55][57]. Next to the soil moisture retrieval, V. Wismann et al.,1995 shows a strong correlation between the normalized radar cross section, incidence angle and precipitation which gives a good indication for the onset of vegetation growth and decay [57]. Friesen et al., 1998 show that the backscatter information at an incidence angle of 45° displays the seasonality of the vegetation cycles[19]. They conclude that the backscatter information at 45° incidence includes two main parameters: soil volumetric water content and vegetation biomass. Still, this simplified model contains an error of 33 percent for vegetation biomass estimations. Friesen et al., 2008 and Steel-Dunne et al., 2019 highlight that the distanglement of soil and vegetation parameters remains a challenge and is very important to improve the accuracy of satellite derived soil moisture and vegetation state estimates [17][43].

Following the success of the ERS-1/2 mission a new series of Meteorological Operational (METOP) satellites were launched. Here a new Advanced Scatterometer (ASCAT) was installed on Metop-A (launched in October 2006), Metop-B (launched in September 2012) and Metop-C (launched in November 2018) and are operated by the European Organization for the Exploitation of Meteorological Satellites (EUMESAT) [53]. The Advanced Scatterometer (ASCAT) runs at the same wavelength of 5.7cm (C-Band) as the ESCAT with a frequency of 5.255GHz with VV polarisation [43][53].

A second-generation of C-band scatterometer called SCA is planned to continue the latest ASCAT mission in 2022 [45]. All three satellite missions (ESCAT, ASCAT, SCA) combined will provide a potential climate record covering at least 40 years of data.

Since the late 1990s TU Wien developed an algorithm to retrieve soil moisture from backscatter observations by first using ESCAT data and currently ASCAT data [52][54][53]. The ability to detect backscatter dependent on range of different incidence angels creates the base of this soil moisture retrieval method [54]. This backscatter dependency on incidence angles can be approximated by a second order Taylor polynomial. TU Wien uses those approximated slope and curvature values to normalize the backscatter measurements to a reference incidence angle and to correct for vegetation [52][54][55].

The ASCAT instrument contains six antennas with three independent viewing directions covering 550 km of observational area, which creates a unique viewing geometry. This geometry allows to extract the local slope values from the three different backscatter measurements. The noisy signal is further processed using recently developed Kernel Smoother (KS) approach to estimate the slope and curvature values dynamically [47].

Slope and curvature values are commonly used to normalize and correct backscatter observations but not as a source of information about vegetation phenology and water dynamics. For example, Vreugdenhil et al., 2017 extracts the vegetation optical depth from backscatter observations to analyse the inter annual variability in vegetation dynamics [52]. Steele-Dunne et al., 2019, however suggests to investigate slope and curvature as an information source to independently extract the vegetation water dynamics [43]. Their results show a clear correlation between slope and vegetation density and a connection between vegetation phenology and curvature. Still, the vegetation parameters slope and curvature are not fully understood yet.

Understanding and exploiting the ASCAT observations is therefore very important 1) to get a unique insight about water dynamics from soil and vegetation 2) to generate a climatological data set of vegetation states related to the observations. This data can then be further used in many different land applications such as rainfall estimation, landslides and soil erosion and drought and flood monitoring but also for meteorological and climate applications to estimate the water fluxes between surface and atmosphere, evapotranspiration and air temperature [5][53].

1.2. Problem Statement

Directly quantifying vegetation and soil variables from satellite observations on a global scale is rather complex and requires a measurement operator, which is able to translate observations to vegetation states [15][14][16][39]. This translation between variable states of the LSM's into the corresponding observational space is done by a surrogate model. A sophisticated and applicable surrogate model is a Radiative Transfer Model (RTM), for example the Michigan Microwave Canopy Scattering model (MIMICS) [44][51].

RTM's rely on very detailed description of the soil and vegetation such as size, shape, height, thickness and orientation of the canopy but also the dielectric properties [51]. For simulating the dynamic vegetation parameters those required variables are too scarce or not available and would need expensive field campaigns to generate. Another issue for using RTM's is that the relationship of the parameters and vegetation water dynamics is not well understood yet and therefore not simulated in LSM's. Furthermore, the connection between RTM's and vegetation parameters is rather static and therefore is not able to represent the complex and dynamic reality.

Recent research has shown great potential using Artificial Neural Networks (ANN) to overcome the complexities and necessary inputs required from RTM's [15][16]. Forman et al. uses ANN's and Support Vector Machines (SVM) to simulate passive microwave brightness temperature from a land surface model [15][16]. ANN's are very useful because, they are able to capture any non-linear relationship without giving physical explanations and relations. Especially deep learning architectures, which are a sub-field of ANN's, have proven to be state-of-the-art algorithms to explore highly-non-linear data [22].

The suggested approach in this thesis to overcome the implementation issues of radiative transfer models, is to use a deep neural network as a surrogate model. This method will allow to capture the non-linear relationships

between vegetation dynamics and ASCAT observables and could therefore directly relate vegetation states with observations.

1.3. Research Objective

The aim of this research is to exploit the ASCAT observations, namely backscatter, slope and curvature, to improve the understanding of vegetation phenology and vegetation water dynamics and follows up on the most recent work from Steele-Dunne et al., 2019 [43].

The overall goal of this thesis is to directly link vegetation water processes to the ASCAT backscatter, slope and curvature. 10 years of ASCAT data from the Metop-A satellite and 10 years of LSM data provided by Meteo-France are used to establish this relationship. The hypothesis here is that machine learning can provide the mean to connect LSM states to the ASCAT observations.

The first objective is to find the best Deep Neural Network (DNN) architecture fitting to the input data. Here, it is important to take into account the physical interpretation of the results but also the logic of the input to train the neural network.

The next objective is to understand and validate the results provided by the DNN, by comparing the variations of the ASCAT and LSM data with the performance of the developed surrogate model and the land surface model vegetation parameters. Additionally, the feature importance of the deep learning models could create a direct link between the LSM and ASCAT parameters.

The overarching goal is to work towards an operational measurement operator for vegetation monitoring. Here, the surrogate model is used to assimilate ASCAT observations to constrain the LSM vegetation variables. This will not only help to exploit the ASCAT measurements in the sense of a new perspective on water dynamics of vegetation but also to explore new insights of microwave scattering from living vegetation.

1.4. Research Question

Can deep neural networks be used to establish a surrogate model to reconcile the land surface model and the ASCAT observations?

The following research sub questions aim to help answering the research question:

- RQ1: How does the model performance vary in space and time?
- RQ2: How does the feature importance vary in space and time?

Both research sub questions are used to discuss both feature importance and model performance in space an time. Space refers to the vegetation type distribution throughout the study area, which means that both metrics are evaluated and compared regarding the land cover types. Time is a representative for seasonal and inter-seasonal variations. Specific time series comparison reveal if the model is able to simulate small time variations and how the feature importance is changing in time.

1.5. Research Guide

The remainder of this report is structured as follows: the "Introduction", which presents the research motivation, problem statement, research objective and research question in Chapter 1; the "Theoretical Background" section, which gives a more detailed description of remote sensing, ASCAT instrument and observation parameters, Land Surface Model and Deep Neural Networks in Chapter 2; the "Methodology" section, which describes the study domain, in - and output data, implementation, training and validation process of Deep Neural Networks in Chapter 3.1; the "Results and Discussion" section, which visualizes and describes the achieved results in order to answer the research questions in Chapter 4; the "Conclusions and Recommendations" section, which gives concluding remarks and outlook as well as suggestions for future research and applications in Chapter 5.

2

Theoretical Background

In this Chapter the theoretical background needed to understand the results, discussion and conclusion is discussed. It provides a detailed review of the basics of microwave remote sensing, the ASCAT instrument, the TU Wien Soil Moisture Retrieval Algorithm, the used Land-Surface-Model and deep learning.

2.1. Microwave Remote Sensing

1978 the first operational microwave observing satellite Seasat was launched by the National Aeronautics and Space Administration (NASA) [34]. Microwave remote sensing has several upsides compared to other observational wavelength ranges. First it penetrates through clouds and is not depended on the illumination of the sun, which means it can provide a consistent data set without any interruption in any condition [51]. Next to consistency it achieves a deeper penetration into the soil and vegetation compared to optical and thermal waves [51]. How deep it the waves penetrate into the soil and vegetation is dependent on several properties, such as vegetation density, geometric properties, water content, volumetric water content, soil type and soil moisture [48]. In general longer waves reach deeper into the vegetation-soil layers, which makes choice of wavelength and sensor not arbitrary rather a deliberate considering which parameter are important to measure. Commonly the frequency range reaches from 0.225 to 100 GHz and is divided into 9 sub-ranges, called bands, see Table 2.1.

Band	Frequency Range in GHz
P	0.225-0.0390
L	0.390-1.550
S	1.550-4.20
C	4.20-5.75
X	5.75-10.9
K	10.9-36
Q	36-46
V	46-56
W	56-100

Table 2.1: Common microwave frequency ranges.

Microwave sensors can be distinguish in two main classes 1) Passive (Radiometers) and 2) Active (Radar). Both sensors act like receivers but radars additionally transmit radar waves. Radiometers are mainly used to measure the thermal emission of microwave transmissions, which is depending on the physical temperature and electrical properties [51]. Active microwave sensors send out pulses and use the properties of backscatter applying Doppler processing to extract information. Active microwave sensors are divided into five groups: Altimeters, Scatterometers, Synthetic-Aperature Radar (SAR), side looking ariborne radar (SLAR) and meteorological radars. Radars can be applied either on a ground-based mission or spaceborne and are mainly used to create high quality images, altimetry and wind measurements. The difference between SAR and the other radar sensors are the antenna processing techniques of which SAR employs a single synthetic-aperture antenna and the rest usually multiple real aperture antennas. The single antenna of a SAR returns multiple

measurements per time-period which with the right signal processing allows to increase the beam-width and therefore synthetically increases the antenna length [51]. This technique is mainly used for producing images. Non-imaging radars such as scatterometers and altimeters measure backscatter at the same location from different angles. One important application of scatterometers are wind measurements. The ERS-1/2, National Aeronautics and Space Administration (NASA) Scatterometers, ASCAT, Indian Oceansat-2 and SeaWinds scatterometers are all operational sensors used for near surface wind observations, but differ in the geometry of antennas as well as spatial resolution, daily coverage and swath width [51]. Next to wind-measurements scatterometer can be used to detect vegetation density, soil moisture, sea-ice type, ice extent and detection of icebergs [51]. As an answer to the fast development and increasing capability of scatterometers, especially in the field of soil moisture measurements, NASA launched a satellite mission called "Soil Moisture Active Passive" (SMAP). The combination of a L-band radar and a radiometer instrument gives a new angle and opportunity to improve the accuracy of soil moisture retrievals [11].

Radiometers are highly sensitive sensors and have been used in many fields of environmental remote sensing, e.g. hydrology, agriculture, oceanography and weather related applications [51]. One of the reasons is the ability of multi-wavelength observations, which means more then one geophysical parameter can be estimated [51]. Still one downside of passive microwave sensors is the relatively coarse resolution of 25 to 50km. This limitation was solved by introducing the synthetic-aperture to the radiometers system, called interferometer. The Soil Moisture and Ocean Salinity (SMOS) mission was the first to implement such a system and was launched 2009.

Considering the wide range of microwave based satellite missions many are interesting to exploit vegetation variables. However, ESCAT and ASCAT are unique in several ways, 1) the real-aperture antennas allow several observation angles of one observation point, which can not be provided from a SAR sensor, 2) the ERS 1/2 and ASCAT are operationally used, which guarantees a consistent but also high quality data supply [51], 3) ESCAT and ASCAT observing at a C-band frequency, which proved sufficient sensitivity to soil moisture and vegetation changes [44][43][53], 4) ESCAT, ASCAT and SCA provide a climatological dataset of at least 40 years, which none of the most advanced radiometer and interferometer missions can provide [51].

2.1.1. Basic Operations of Radar

Scatterometers are used, as mentioned before, for active microwave remote sensing. Even though there are several groups within the remote-sensing radars, the basic elements of the operation is the same. Radar stands for radio detection ranging, which stems from the estimation of distance between aircraft's and ships. The basic idea is to transmit microwave signals onto the observational area, at which the waves are scattered and reflected backwards. This reflected signal, also called backscatter, contains information about the properties of the surface. Therefore, it can detect 1) the distance 2) direction, 3) speed and 4) scattering cross section. The challenge is to translate those properties into vegetation and soil variables. Figure 2.1 shows the basic components of remote-sensing radar, including the transmitter, receiver, transmit-receive switch (T/R switch), frequency synthesizer, signal processor and antenna.

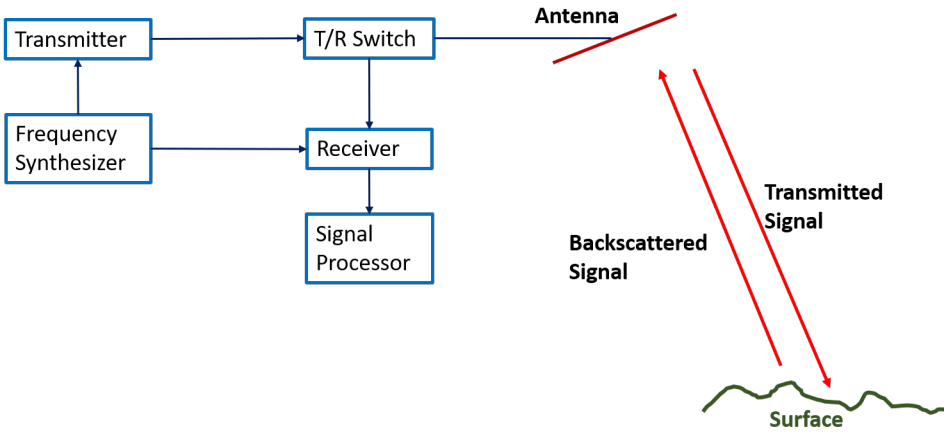

Figure 2.1: Basic components of a remote-sensing radar [51]

The transmitter sends out a signal using the antenna, once the impulse is reflected/scattered by the observational target the receiver amplifies and processes the backscatter. Hereby, the T/R switch controls the incoming and out going signals and frequency synthesizer controls the comparability and communication between the received and transmitted signals [51].

2.1.2. Microwave Scatter Mechanisms

The echoed signal received from the satellite antennas is a result of two major scatter mechanisms, 1) surface scattering and 2) volume scattering [51]. In general, backscatter is a result of both scatter mechanisms due to a very heterogeneous surface composition. Plain soil surfaces can cause either a single-scattering, which means all reflected backscatter oriented towards the satellite or a multi-scattering, in which case the incidence signal causes multiple reflections on the surface, see Figure 2.2(a). Volume scattering is referred to take place within the vegetation canopy and is illustrated in Figure 2.2(b). This medium encloses all scatters between the top of the canopy and soil surface.

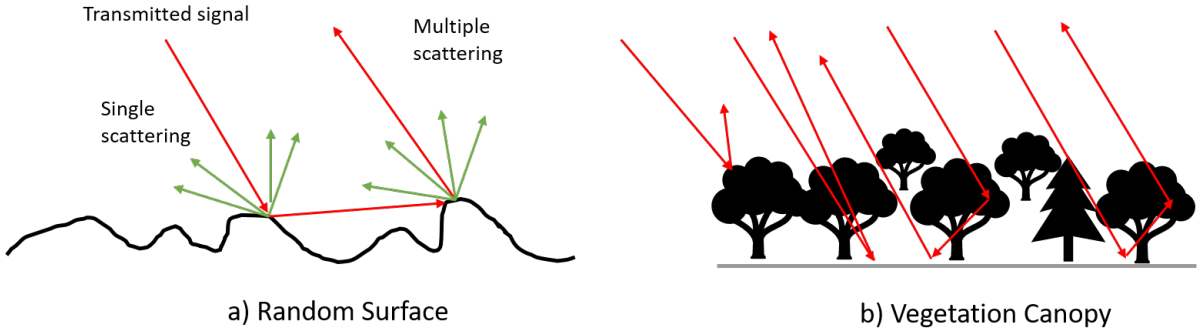

Figure 2.2: Illustration of basic scatter mechanisms on a a) Random surface and b) vegetation canopy [51]

The attenuation as well as the direction of backscatter signals is very complex and difficult to model, due to the high variation of surface properties. Surface properties are described in surface-scattering-models as a statistical distribution of size, soil composition, dielectric constants, orientation and volume scatters etc. [51]. This resolves at best as a good approximation of the actual scatter processes.

Scatterometers retrieve backscattering coefficients, which are a function of frequency, polarisation, incidence angle of the emitted signal as well as surface parameters such as geometric shapes, roughness and dielectric properties - also referred as normalized scattering cross section per unit area. Whereas the roughness is function of incidence angle and wavelength and describes a relative concept. For example water bodies act as a smooth and land as a rough surface [51]. The dielectric properties are dependent on soil and vegetation volumetric moisture content as well as soil moisture, for example loamy soil shows an increase of the dielectric constant by increasing volumetric soil moisture. The volume backscatter coefficient is defined per unite volume and consists of four major scattering contributions, considering a vegetated surface: 1) Direct scattering from the plants, 2) Plant - Ground - Plant Scattering, 3) Ground - Plant - Ground Scattering and 4) Direct scattering from the soil.

Chapter 1 mentioned the capability of radiative transfer models to model these multiple scattering processes for a in-homogeneous surface. Still, the above described scatter mechanisms give only a small glimpse into of complexity to mathematically describe scattering mechanisms of complex surfaces, whereas machine learning algorithms require far less input information.

2.2. Advanced SCATterometer (ASCAT)

The ASCAT instrument is mounted on the operational Metop(A/B/C) missions on a altitude of 822km and orbits in a sun-synchronous 29 day cycle in which it crosses the equator at 9:30 (descending) and 21:30 pm (ascending) [13]. A global coverage is achieved after 3 days, however, phasing the Metop A mission with B results in global observation already in 2,5 days [9]. The daily coverage is around 82 % of the worlds area which causes a irregular spatial and temporal resolution, which needs to be considered for simulating these obser-

vation but also interpreting them [53]. ESCAT and ASCAT scatterometers use C-band at 5.355GHz microwaves with a VV polarization, which means the antenna transmits and receives vertical polarized electromagnetic waves. Next to the similarity one big difference is that the ASCAT antenna consists of two sets of three fan-beam antennas, whereas ESCAT only mounts one set [13]. Figure 2.3 illustrates the unique ASCAT antenna geometry, look angles and ground track.

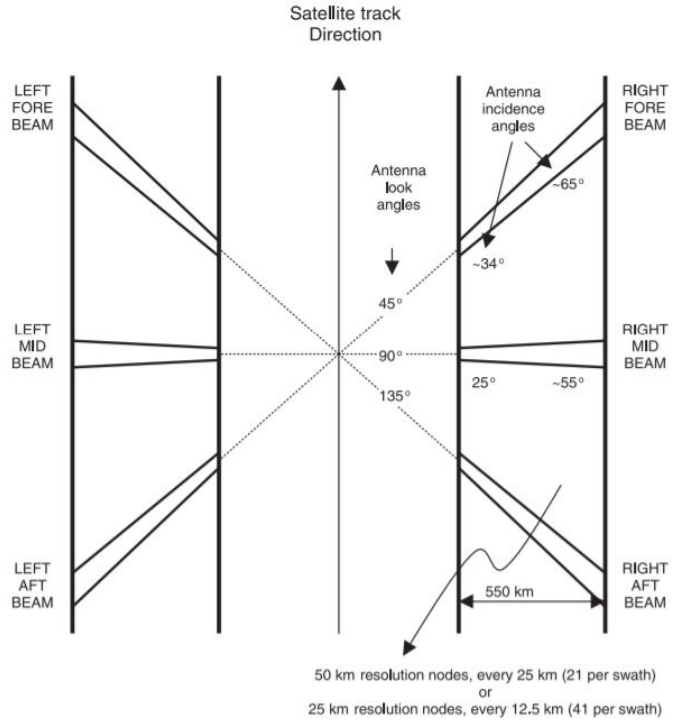

Figure 2.3: ASCAT antenna geometry, showing the symmetric set-up of the satellite antennas to respect of the ground track direction. ; Source: [13]

The ground track of the satellite is covering 550km per swath for each set of antennas and 360km apart from the satellite ground track, which is referred in the Figure 2.3 along the satellite track direction. Both sets of fan-beam antennas are mounted each in 3 different angles in respect to the satellite ground track in a 45°,90° and 135° angle, whereas the incidence angles ranges between 25 and 65° [13]. The measurement uses echo power detection, which means that the backscattered electromagnetic wave gets amplified and processed via radar equations to the so-called backscattering coefficient $\sigma°$ expressed in units [m2m-2] or decibels [dB] [53]. With this measurement geometry, a triplet of backscatter coefficients $\sigma°$ can be observed for each overpass for either a 50km (25km sampling) or 25km (12.5km sampling) resolution [13].

The good and stable performance of the ASCAT instrument has already been established and proven by the ERS scatterometer instruments and further investigated by several studies, which have shown a maximum calibration error of 0.15-0.25 dB, considering a very a very stable annual change over the amazon rain forest of around 0.02 dB [2][53]. The observation frequency 5.3 GHz allows great sensitivity estimate liquid water to soil relationship, which means very small changes in soil moisture are detectable because the change of the soil dielectric constant to liquid water is ten times more compared to frequencies over 10 GHz [51][53]. Together with the multiple observation angles provided by geometry of the antennas and the high spatial and temporal resolution ASCAT is a high performing instrument to monitor soil moisture and vegetation processes. Other missions such as the SMAP, the Advanced Microwave Scanning Radiometer 2 (AMSR2) and the Soil Moisture and Ocean Salinity (SMOS) mission act so far only as quasi-operational soil moisture estimates and provide less stable data or performing on a lower temporal/spatial resolution [5][51]. ASCAT is therefore the most suitable instrument/mission to support operational hydrological, climatological and meteorological applications [5][53].

2.3. TU Wien Soil Moisture Retrieval Approach

TU Wien developed a Soil Moisture Retrieval (SMR) algorithm using ASCAT data to generate a global multi-year dataset for soil moisture and achieved together with the EUMETSAT a continuous near-real-time soil moisture product to operationally support Nummerical Weather Predictions (NWP) [53]. The algorithm was originally applied on the ESCAT data-set but now transferred to the ASCAT product and is the most advanced and widely used technique to retrieve soil moisture from satellite observations [5][53]. The core of this algorithm has not changed since its first development and implementation by Wagner et. al 1999 and carries along following assumptions [55][5]:

- Soil moisture and backscatter are in a linear relationship
- The dependency of backscatter on incidence angle can be empirically expressed
- Vegetation effects cancel each other out for a specific incidence angle
- Land cover and surface roughness are constant over time

It is shown by Wager et al. the algorithm is able to compensate for static components of soil and vegetation but seasonal cycles are not accounted for, e.g. the start of the growing season [53]. This change in canopy has an increase or decreasing effect on the observed backscatter. Wagner et al. showed that the dependency of σ° and incidence angle is controlled by the dynamic vegetation component and soil moisture [55][53]. Hence, the attenuation of σ° can be distinguished between vegetation and soil moisture depending on importance. This provides an opportunity to account for vegetation dynamics by choosing the incidence angles at which the two physical effects cancel out. Finally, in order to derive soil moisture from the multi-angular ASCAT observations two incidence angles θ need to be known to estimate backscatter for 1) the dry conditions θ_{dry} and 2) wet conditions θ_{wet}. Dry conditions determine the influence on backscatter caused by seasonal variation in vegetation density, whereas the wet condition represent the highest value of backscatter and is dominated by soil moisture [43][53]. All backscatter coefficients σ° are normalized to a reference angle of 40 ° θ_r. After applying several data processing steps including noise estimations, normalization, accountance for azimuth- and incidence angle etc. following equation 2.1 is applied to calculate soil moisture from ASCAT backscatter [53]:

$$m_s = \frac{\sigma^\circ - \sigma^\circ_{dry}}{\sigma^\circ_{wet} - \sigma^\circ_{dry}} \tag{2.1}$$

In equation 2.1 σ° represents the backscatter observation inverted to the dry state σ°_{dry} and the moist state of the soil σ°_{wet}, all normalized to the reference angle θ. The whole retrieval process is implemented as a near-real-time operator with a software package called WAter Retrieval Package (WARP) [47].

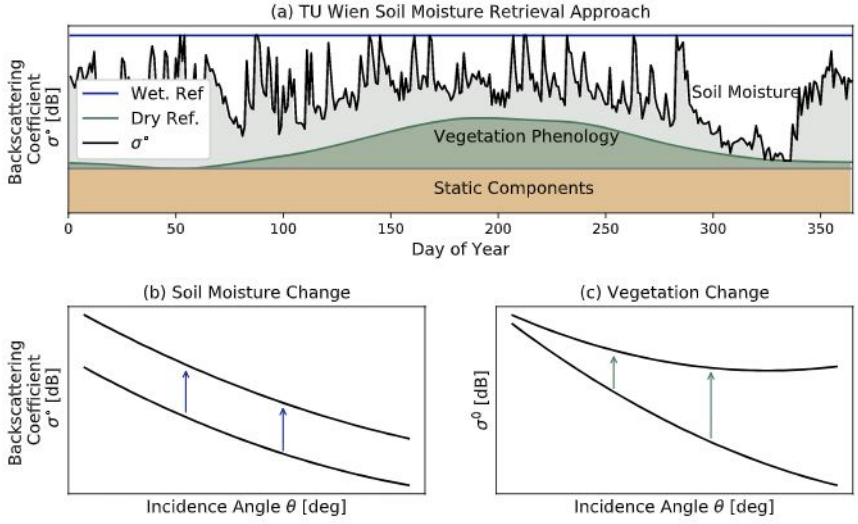

Figure 2.4: Backscatter Coefficients as a function of time and incidence angle [44]

Figure 2.4 (a) illustrates the dependency of the normalized backscatter coefficient σ° on the dynamic vegetation phenology and soil moisture as well as the baseline of static components [43]. It clearly shows the importance of vegetation to backscatter as a smooth and seasonal whereas soil moisture is noisy and based on events. Following up on this Wagner et al. proved a linear relationship between the slope ($\sigma^{'}$) [dB/deg] and incidence angle, see equation 2.2 with θ_r representing the reference angle [54]:

$$\sigma^{'}(\theta) = \sigma^{'}(\theta_r) + \sigma^{''}(\theta_r) \cdot (\theta - \theta_r) \tag{2.2}$$

This leads to a further approximation of the backscatter - incidence angle relationship described as a second order Taylor polynomial, see equation 2.3 :

$$\sigma^\circ(\theta) = \sigma(\theta_r) + \sigma^{'}(\theta_r) \cdot (\theta - \theta_r) + \frac{1}{2}\sigma^{''}(\theta_r) \cdot (\theta - \theta_r)^2 \tag{2.3}$$

This expression allows to calculate backscatter for any incidence angle once slope $\sigma^{'}$, curvature $\sigma^{''}$ and $\sigma^\circ(\theta_r)$ are known. A further transformation allows to express backscatter as a function of reference angle, which removes the incidence angle dependency and makes it comparable to different backscatter observations, see equation 2.4:

$$\sigma^\circ(\theta_r) = \sigma(\theta) - \sigma^{'}(\theta_r) \cdot (\theta - \theta_r) - \frac{1}{2}\sigma^{''}(\theta_r) \cdot (\theta - \theta_r)^2 \tag{2.4}$$

Slope and curvature coefficients can be estimated using the geometry of the unique ASCAT antenna set-up. The fore (f), mid (m) and aft (a) beam observations enable to observe three independent backscatter measurements of three different incidence angles. This information can be directly used to estimate the backscatter dependence on incident angles by computing the "local slope" , see equation 2.5 [23][43] :

$$\sigma^{'}\left(\frac{\theta_{mid} - \theta_{a/f}}{2}\right) = \frac{\sigma^\circ_m(\theta_m) - \sigma^\circ_{a/f}(\theta_{a/f})}{\theta_m - \sigma^\circ_{a/f}(\theta_{a/f}} \tag{2.5}$$

where m indicates the mid-beam antenna, f the fore beam and the a the aft beam antenna. The result is often a very noise signal, which can be further processed averaging several local slope values over the entire range of incidence angles or over a time window [23]. The ERS 1/2 used several years of local slope products to create seasonal climatology slope and curvature parameters to provide robust coefficient estimates [35]. A new approach proposed by Thomas Melzer introduced a new technique to calculate slope and curvature dynamically using a Kernel Smoother (KS) approach [47]. This gives the unique opportunity to account for inter-annual variations, rather then seasonal parameterisation.

This processing is done using the Epanechinkov kernel by integrating to one with an finite sample range [-λ,λ]. λ is defining the kernel width and is an important parameter, because it controls how many observations in time are affecting the final result [23]. Hahn et al. established a window size of 42 days (λ = 21) as a sufficient balance between variance and bias. This means that slope and curvature are a product based on local slope information of 42 days. The magnitude of influence of the local slope values are dependent on the distance to the estimated time-step, whereas the influence decreases with increasing distance [23].

Figure 2.4 (b) and (c) show the impact of soil moisture and vegetation change on the incidence angle dependence of backscatter respectively. In sub-figure (b) the observed normalized backscatter curve shifts upwards for all incidence angles with increasing soil moisture, which was to expect as the relationship is defined linearly. Sub-figure (c) shows that with an increase of vegetation the σ°-θ curve is rotating and therefore changes its slope $\sigma^{'}$ and curvature $\sigma^{''}$ [43].

This new KS approach allows to estimate $\sigma^{'}$ and $\sigma^{''}$ on a daily bases which means it improves the SMR algorithm significantly but also gives the opportunity to investigate the parameters itself. Vreugdenhil et al. used this development to assess the vegetation dynamics over Australia, by retrieving vegetation optical depth from backscatter [52]. This is significant, because it proves the ability of this new methodology to detect inter-annual variability of vegetation dynamics.

Steele-Dunne et al. suggested not only using backscatter as a source of information but utilizing slope and curvature itself [43]. In this study the inter-annual variability of water dynamics over the north american grassland was investigated to exploit the relationship between slope and curvature with vegetation and soil dynamics.

So far, slope is used as a measure of vegetation density but is not directly linked to biomass or water content in vegetation. Expanding this knowledge, Steele-Dunne et al. showed a correlation between vegetation types and slope but also established a significant interrelationship of curvature and vegetation phenology [44]. This interaction can be explained by the dominance of direct scattering caused by the vertical structure of the grass land. However, the dominance of direct scattering over ground-bounced scattering is suggested to be dependent on the total vegetation water content and geometry of the canopy [44]. The results show that slope and curvature time-series contain valuable information about the water dynamics within vegetation. Still, a clear correlation between the dynamic vegetation parameters and actual land-surface variables has not been established yet.

2.4. Land-Surface-Model

The foundation of this thesis is based on combining high-end land-surface-modeled data with ASCAT observations. Hereby, Meteo-France, the national meteorological service, is chosen to provide the data needed for this research project. The following explains the background and variables of the operational land-surface model.

Land surface models are a mathematical description of the real interactions between soil, biosphere and atmosphere (ISBA) and are generally used for numerical weather predictions and global climate models [4][29]. Over the last decades LSMs have been improved and extended, by including transpiration, photosynthesis, stomates controlled evaporation and the assimilation of CO_2 [29]. Especially, in the field of numerical weather forecasting surface modelling is most demanded and developed [29]. One special effort was done to create a stand-alone surface model called EXternalized SURFace (SURFEX), which can act by it self or as a physical surface model coupled model to any atmospheric models [29]. This means that for each time step and grid point the atmospheric model provides the SURFEX with specific parameters such as wind components, pressure, long-short wave radiation, upper air temperature etc., which then be returned as fluxes of latent, sensible heat and momentum [29]. Figure 2.5 shows the basic description of the processes and interactions between the atmosphere and the SURFEX model platform [29].

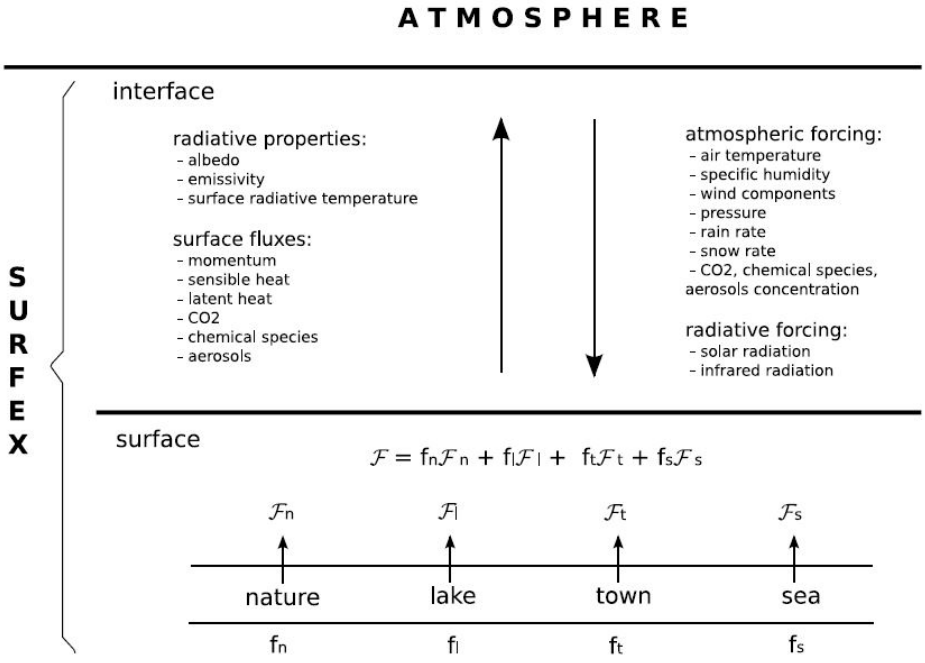

Figure 2.5: Basic description of the SURFEX model; F represents the variables whereas f is the cover-fraction for a specific grid-box [29]

The SURFEX modulates per grid point and is described by four major tiles such as nature, urban areas, sea/ocean and lakes. Every of these surfaces is described by the ECOCLIMAP database [29]. This description contains information about the land cover maps as well as satellite observations.

The SURFEX parameterization has been established over years of research and combines all different developments and improvements made in the field of surface schemes [29]. Parameterization has been done for water surfaces (e.g. 1D Turbulent Kinetic Energy Oceanic Model and Lake Model FLake), urban and artificial areas (e.g. Town Energy Budget Scheme), soil and vegetation (e.g. ISBA surface scheme, ISBA-A-gs surface scheme and ISBA-CC model), chemistry and aerosols and surface boundary layers [29]. A detailed description of all physical processes and parameters can be found at the "Surfex Scientific Documentation" [29].

Considering established correlation between slope and curvature with vegetation phenology and water dynamics only the most relevant ISBA vegetation and soil parameters are concerned in this study. The ISBA surface scheme is a subsection of the SURFEX model and provides several vegetation states and variables, which might be related to the dynamic vegetation parameters observed with the ASCAT instrument. Most relevant vegetation variables are leaf area index, soil moisture, ecosystem respiration and gross primary production, which are explained in following subsections.

2.4.1. Leaf Area Index

Leaf Area index (LAI) is defined as the ratio of leaf area surface divided by the total are of consideration and is expressed in $[m^2 m^{-2}]$ [29]. The LAI used in the ISBA is not the same as the one observed by satellites, because it referrers to the whole thickness of the vegetation whereas satellite deduced LAIs are based solely on the top of canopy [29].

2.4.2. Soil Moisture

Soil moisture is defined as the volumetric soil moisture concentration in $[m^3 m^{-3}]$. The ISBA LSM provides soil moisture for three different dephts: WG1 is the surface soil moisture up to 1cm; WG2 is the root-zone soil moisture up to 0.8m including WG1 and WG3 is soil moisture between 0.8 and 1m [12]. However, in the context of this thesis the most valuable soil moisture sequence is WG2, because it contains information aligning with the depth of penetration from C-band electromagnetic waves.

2.4.3. Ecosystem Respiration

Respiration (RE) is the opposite process to photosynthesis, which means it releases CO_2 instead of removing and transforming it from the air by decomposing soils organic matter. The process of respiration and its rate is dependent on the surrounding temperature and soil moisture [1]. The ISBA-CC model can successfully represent the heterotrophic respiration and soil moisture, which is then expressed as the CO_2 ecosystem respiration [29].

2.4.4. Gross Primary Production

The Gross Primary production is defined as uptake of carbon by photosynthesis processes, also defined as carbon assimilation or carbon fixation [29]. The ISBA model uses GPP and RE to estimate the net ecosystem CO_2 exchange, by simply subtracting the ecosystem respiration form the gross primary production [29].

2.4.5. Vegetation Type

The ISAB vegetation types are based on a global land cover map at 1/120° resolution and contains 19 different classes. In the overall context, the SURFEX model defines four major land covers, which are then interpreted with the 19 fractions provided by the ISAB vegetation types. Therefore, each grid points vegetation type is expressed as a combination of 19 different fractions of different classes [29].

2.5. Deep Neural Networks

Artificial Intelligence (AI) is a technique and mathematical concept to create machines that think. This desire of developing such machines reaches all the way back to the ancient Greece and has not lost its importance until today. The original purpose of AI was to solve and describe problems which are hard to describe by human, e.g. speech or face recognition [22]. Those tasks feel natural and intuitive for people but describing it within a mathematical formal framework is very difficult. This intuition of our brain is based on repetition and experience. Those experiences and knowledge can be transferred to a hierarchy of concepts describing complex problems with simple concepts [22]. Stacking up those concepts creates a "deep" (multi) layered AI, which can be called Deep Learning. Self-sufficiently extracting patterns from raw data is referred as machine learning, which is a sub-field of AI, whereas logistic regression is the simplest from of it [22].

Every piece of information given to the machine learning algorithm is called feature and the representation is crucial for its performance. When the importance of features for a certain prediction are unknown simple machine learning algorithms do not know which information is negligible or not which results in a bad model outcome. A solution to this problem is representation learning, which is able to discover a good features quickly compared to manually creating features. Still, high variations in the data remains a problem, for example a speaker's accent or different angles of photographs etc. [22]. This disentanglement of variations can be achieved by deep learning. This algorithm is able to break down very complex concepts in several simpler ones, for example in edges, corners and contours for images. The complexity representation of these networks can be controlled by the depth and extend of the network. Figure 2.6 illustrates the hierarchy of the different concepts of artificial intelligence. In the following sub-section the basic mathematical concept of deep learning is explained, which is transferable to all other possible architectures such as the regular feedforward neural network, convolutional neural network and recurrent neural network [22].

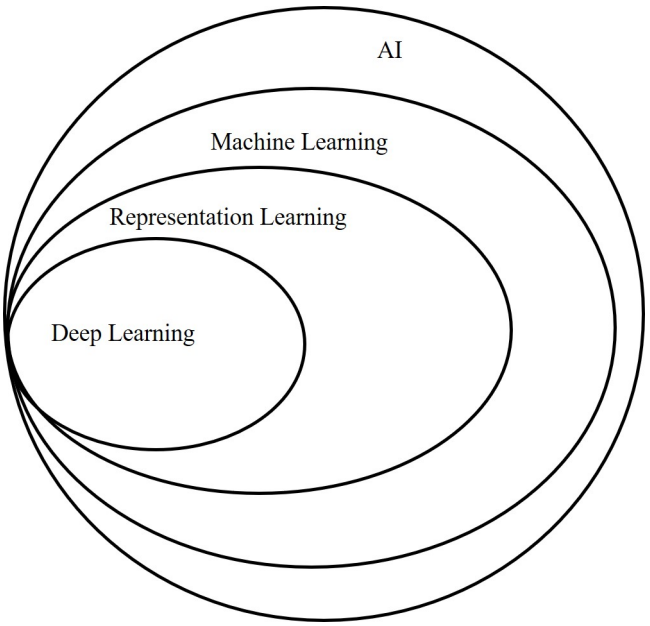

Figure 2.6: Venn diagram of all fields of Artificial Intelligence; Source: Goodfellow et. al page 9 [22]

2.5.1. Mathematical Concept of Deep Learning

Deep feedforward Networks are the core of deep learning algorithms and are able to approximate any possible mathematical function [22]. The terminology is "neural" stems form neuroscience, which these networks are inspired on. Feedforward in this context means that the model starts at the input, computes the function and predicts an outcome. When including feedbacks from previous predictions the deep learning model is called Recurrent Neural Networks (RNN). Still, RNN's are based on the same "learning" process as regular deep neural networks (DNN).

Figure 2.7 illustrates the basic architecture of a feedforward neural networks. The basic structure contains a 1) Input layer, 2) n amount of Hidden Layers and 3) Output Layer. The input layer is defined by the structure and amount of date fed to the network, whereas the hidden layers are designed by the user. The hidden layers do not output their estimated values. Every hidden layer is made out of nodes, which are represented as "H" in Figure 2.7. The amount of nodes per hidden layer as well as the number of hidden layers define the complexity of the DNN. The larger the network is the more complex it is, which means the capacity increases to approximate very complex data patterns. Still, there is a trade off between complexity of the network and a reliable estimate, which means that the DNN is endangered for false predictions when it is to complex or too simplistic[31]. The complexity choice is unique for the data itself and therefore needs be adapted for individual problems [31].

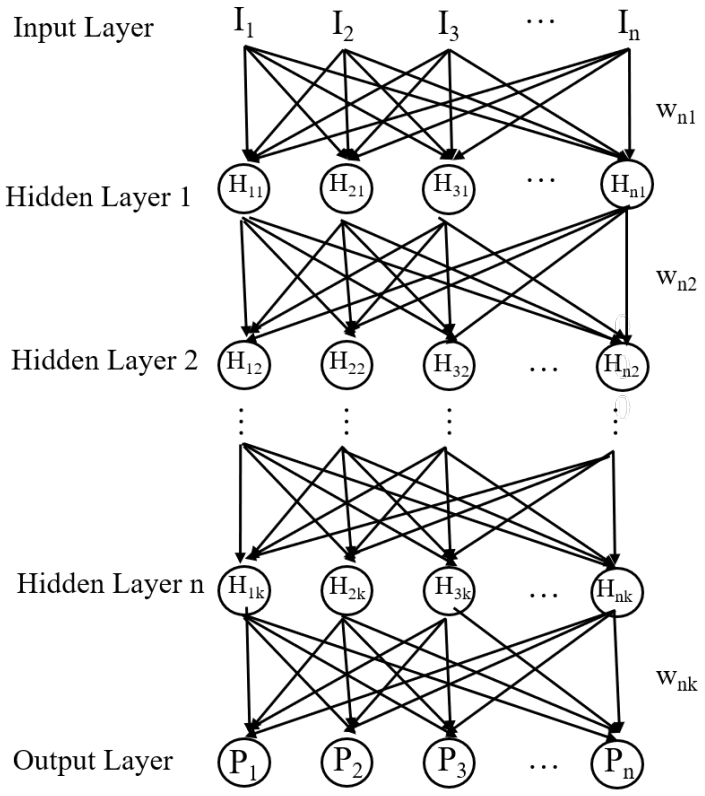

Figure 2.7: Schematic of a regular feedforward deep neural network; I represents the input features, H a single node of a hidden layer and P represents the output features. This schematic is considered as a fully connected neural network, because all nodes are connected through out the whole network.

The interconnection between input as well as hidden and output layers are described by weights and are displayed as black arrows in Figure 2.7. Weights can be seen as the strength of connection between neurons, this neurons connected with high weight are more activated and more important to the prediction than connections with lower weights. Every neuron in the hidden layer has its own activation and is fully connected to the layer before and after. The output of each interconnection is expressed in following equation 2.6:

$$a_j^k = f\left(\sum_{i=1}^{n} w_{i,j}^k a_i^{k-1} + b_j^k\right) \tag{2.6}$$

Hereby, a represents the activation for a neuron with subscript k for the layer number and j as well as i the position of the neuron, w is gives the weight of connection from the previous neuron, b is the bias of the neuron and f represents the activation function. The bias is an expression of how easy or likely a neuron is to fire or output [36]. An activation function is a non-linear function and is needed to introduce non-linearity to network, so it can approximate any function. Most common activation functions are the sigmoid function (sig), rectified linear unit (ReLU) and hyperbolic tangent function (tanh) [22].

This means that the accuracy of the output is dependent on all weights and biases of each neuron throughout the network. Every little change of any bias or weight has influence on the prediction and optimizing them to achieve the correct result is called learning or training. Training can either be done with an unsupervised or supervised learning algorithm. Unsupervised means that no outcome, or labels, are known for the input, which is useful for clustering or dividing data according to similarity [22]. Supervised learning is associated with labels for each sample and therefore needs an instructor whereas unsupervised learning can be used without guidance [22].

First step to train the system by updating weights and biases in the network is to quantify the performance of the DNN. This is done by applying the cost function, which calculates all square differences between the outputs of the network and the actual values. This is done for all training examples and averaged over all samples to calculate the "average cost function", see equation 2.7 [36]:

$$C(w,b) = \frac{1}{2n}\sum_{x}\left\|y(x)-a\right\|^2 \tag{2.7}$$

In this equation w denotes the weights, b the biases, n the number of samples,y represents the function 2.6 and a is the vector resulting form the network with x as input. The cost function calculates therefore the mean squared error, which means the a high value indicates a big discrepancy between the output and the truth, whereas a low value indicates a good prediction. This means in order to optimize the DNN the cost function C(w,b) needs to be minimized.

The mathematical principle to minimize the average cost function is called gradient decent [36]. Gradient decent is a method to find the direction of the steepest increase in a multi dimensional space. In the case of minimization the direction of the steepest decrease is desired expressed as - $\nabla C(w,b)$. This process is computed multiple times with small steps to finally reach the global minimum. Still, in most cases the global minimum is never reached because the negative gradient computation converges in local minimum of the cost function [36]. Considering that C is a function of all variables within the network (e.g. v = v_1,....,v_1). The gradient decent vector can be expressed as:

$$\nabla C(v) = \left(\frac{\partial C}{\partial v_1},....,\frac{\partial C}{\partial v_m}\right)^T \tag{2.8}$$

The heart of deep neural network is backporpagation, which computes the negative gradient of the cost function throughout the network by updating all weights and biases. Every layer is represented by a vector of weights and bias and the length of this vector is represented by the number of neurons in the layer. This means that output of a neuron, also called activation, is basically a simple matrix multiplication. As mentioned before the basic forward mechanism of a DNN is 1) apply a weight matrix to an activation, 2) add the bias vector and 3) apply the activation function. Equation 2.8 explains that the gradient of the cost-function computes all partial derivatives to any variable, which are in the case the weights and biases, using the chain rule. The partial derivatives of $\frac{\partial C}{\partial w_{n,k}}$ and $\frac{\partial C}{\partial w_{n,k}}$ are then used to update all weights and biases in order to minimize the cost function [22]. This updating process from the cost-function back to the last hidden layer is called backpropagation [36].

2.5.2. Regular Deep Neural Networks

Many different deep neural network architectures are available and used for many different applications such as voice recognition, image classification, time-series prediction, parameterization, object detection and motion modelling [22][36]. The simplest form of a deep neural network is the regular forward propagating deep neural network, which includes an input layer, x-amount of hidden layers and an output layer as explained in the previous subsection. It is used for regression but also time-series forecasting, air quality predictions and simulating satellite observations [6][15][30][38].

2.5.3. Deep Convolutional Neural Networks

Deep Convolutional Neural Networks (DCNN) are a subsection of deep learning and inspired by the biological visual cortex and mainly used for time-series analysis and image processing [22]. CNN‘s or DCNN‘s are used in many applications frameworks such as prediction of plant stress, Soil moisture retrieval, Time Series Classification, Surface Water Mapping [20][21][28][56].

Regular DNNs consist of fully connected layers, DCNNs however connect neurons only over a local region at which weight and biases are shared. This means the convolutional layers are not fully connected, which is done to reduce the dimensionality of weight and bias matrices [28]. The advantages of DCNNs are 1) sparse interaction, which means fewer parameters are stored hence less operations are required, 2) parameter sharing, which increases the computational efficiency because sets of weights are trained rather each weight separately and 3) equivariance to translation, which means due to the parameter sharing layers are equivariant hence changes in the input change the output in the same way - expressed mathematically as f(g(x)) = g(f(x)) [22].

The reduction of dimensionalty is defined by the kernel size, which represents the local region at which weights and biases are shared. Figure 2.8 shows the basic components with arbitrary dimensions of a DCNN architecture, containing the input layer, convolutional layers, pooling layer and the fully connected layer. Pooling is used to reduce the complexity of the convolutional layer with a summary statistics of the previous layer. The most common pooling technique is max-pooling which propagates only the maximum activation of the input region [36].

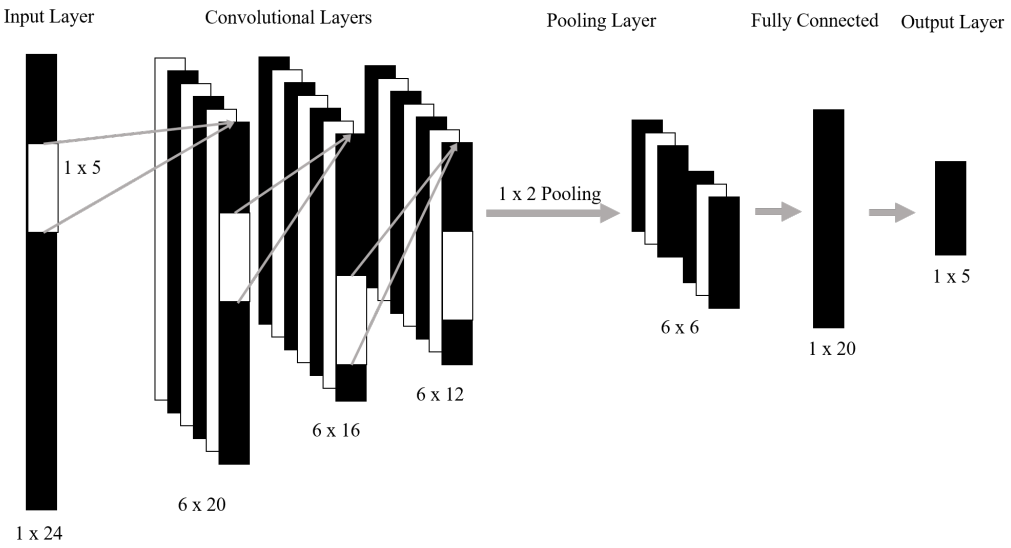

Figure 2.8: Schematic of a deep convoluational neural network

2.5.4. Recurrent Neural Network

Recurrent Neural Networks (RNN) are architectures designed for sequential data with applications in speech processing, music composition and time-series predictions [26]. CNNs and regular DNNs are based on the feed-forward network dynamic at which all activation's are based on the previous layer. RNN's however include a dynamic component which allows to additionally change weights and biases based on previous states [36]. This mechanic allows to establish a memory, which is valid throughout the whole time sequence. Similar to CNN parameters are shared, but in the case of RNN every activation is a function of the previous time-step. Figure 2.9 illustrates the schematic of a recurrent neural network.

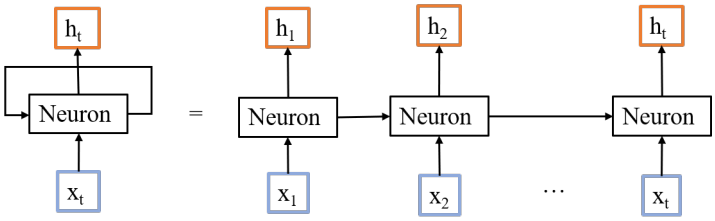

Figure 2.9: Schematic of a recurrent neural network

Hereby, x denotes the input and h the output of a piece of neural network and the subscripts indicate time-steps of the input sequence. Hence, every piece of an RNN has an loop implemented which feeds the previous states of weights and biases to the next one [22].

The biggest issue of the regular RNN is the vanishing or exploding gradient problem. This means that during the backporpagation the errors of weights and biases exponentially increase or decrease during the temporal evolution [26]. The most advanced and state-of-the-art algorithm is the Long-Short-Term Memory (LSTM) overcomes this issue by employing an "appropriate gradient-based learning algorithm" [26]. This was achieved by utilizing a constant error flow through the network.

The network unit of a RNN is shown in Figure 2.9 and consists of an updating loop feeding the outputs back into the activation function. LSTM's network units are build with several different neural networks called gates. First, the cell state, which carries information throughout all time-steps, but it is updated every time by adding or removing information. The "forget-gate" tells the cell state what states to remove or keep hereby information from the input and the output from the previous time-step is used. The "input-gate-layer" is the part of the network which keeps track what information should be stored in the cell state. Finally, the updated cell state as well as the new input determine the output of the network unit [26][8].

Anther version of an advanced recurrent network is the Gated Recurrent Unit (GRU) [8]. It also solves the vanishing gradient problem by using an update and reset gate, but contrary to the LSTM it does not incorporate the cell state. Both methods are strong methods to keep long and short term memories [8][26][36].

2.5.5. Model interpretation

The complexity of a machine learning model contradicts with interpretability, which means highly complex models such as deep learning neural networks might predict very accurately but the output does not explain how the solution was generated. Machine learning is widely used in economics, decision making and scientific applications but rather not fully understand, especially if prediction is done with as so called "black-box" model. This means decisions and interpretation are made without considering the origin of this prediction, which can lead false or even dangerous and discriminating conclusions [24]. Therefore, interpretability is a very important and active study area, in which many approaches have been developed to interpret a complex model, e.g. LIME, Guided backprop, DeepLIFT, Shapley reg. values, integrated gradients, relevance prop and QII [32][40]. In the context of this thesis only the two most important feature importance models are discussed: 1) Shapley Additive explanation and 2) Local interpretable model-agnostic explanations (LIME)

The most recent model to achieve interpretability for a "black-box" model is SHAP (Shapley Additive explanation). SHAP assigns for each feature a specific importance for a particular prediction. Here, a simpler model is used to explain the complex model, called the "explanation model", which gives interpretable approximation of the "black-box" model. The explanation model uses a simplified input to track the original inputs by using a mapping function. Figure 2.10 illustrates the basic principle behind the SHAP algorithm.

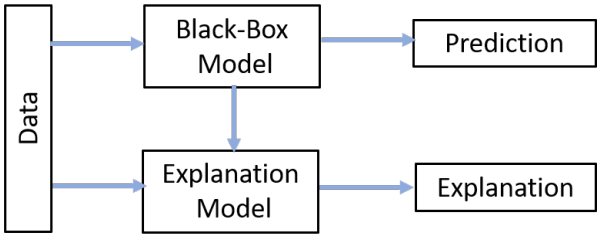

Figure 2.10: Schematic of SHAP interpreting Model for black-box algorithms

The principle is based on additive feature attribution methods which are summing the effects of all feature's approximation to the black-box model output and shapley values. Shapley values originated from coalitions game theory by explaining/approximating a fair pay-out among players [32]. As in difference to LIME shapley values allow contrastive explanations rather than comparing the average of predictions. The SHAP interpretation should guarantee the three following properties, at which f represents the original prediction model, g

denotes the explanation model, x is the original input, x' the simplified input, $\phi \in \mathbb{R}$ is the effect on a feature and h is the mapping function [32]:

1) Local Accuracy: The simplified input needs to results in the same output for the explanation model and the original model.

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x_i^{'} \tag{2.9}$$

2) Missingness: Missingness states that features missing in the original data do not have an impact.

$$x_i = 0 \longmapsto \phi_i = 0 \tag{2.10}$$

3) Consistency: If A change in the model occurs which leads to either an increase or no change in the simplified input contribution should not cause a decrease on the inputs attribution. $f_x(z') = f(h_x(z')$ for all input $z' \in \{0,1\}^M$ with M as the number of simplified input features and *z'/i* denotes $z'_i = 0$.

$$f_x^{'}(z') - f_x^{'}(z'/i) \geq f_x(z') - f_x(z'/i) \tag{2.11}$$

The combination of the properties 1-3 with the following theorem:

$$\phi_i(f,x) = \sum_{z' \subseteq x'} \frac{|\, z' \,|!(M - |\, z' \,| - 1)!}{M!} [f_x^{(}z') - f_x(z'/i)] \tag{2.12}$$

implies only one solution for the feature importance attribution method, which proposes a unified approach to improve previous interpretability models [32].

The Local interpretable model-agnostic explanation (LIME) model was developed by Riberio et al. aiming to develop an explanation model for any classifier [40]. LIME is based on the additive feature attribution method, which means that summing up all effects from all features is approximately the output of the prediction model. However, it does not include the above theorem and properties but tries to minimize following objective function with $g \in G$, where G is a class of potential models [40]:

$$\xi = argminL(f,g,\pi_{x'}) + \Omega(g) \tag{2.13}$$

Ω penalizes the complexity of the explanation model, whereas L represents square loss.

3
Methodology

The methodology proposed to answer the research question is illustrated in form of a flow chart in Figure 3.1. The green box is a given circumstance for this thesis, which defines the extent of the study domain as well as data provider namely, ASCAT and Meteo France. The blue boxes represent individual tasks, whereas the yellow arrows show the milestones during this process. Finally, the dark yellow boxes indicate the evaluation process in order to answer the research question, which will be discussed in the Chapter 4. In the following sections each of the tasks as well as milestones are discussed.

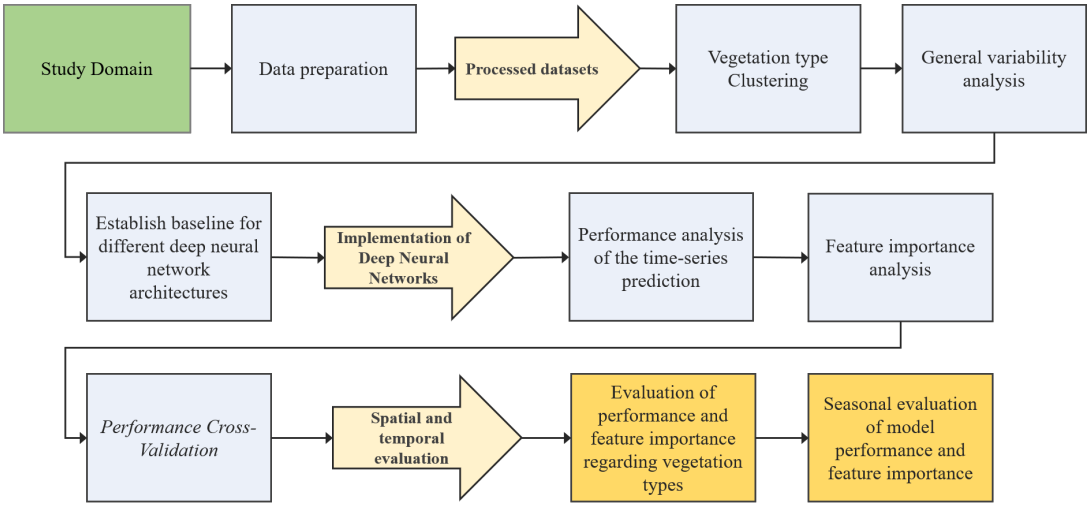

Figure 3.1: Schematic overview of the methodology applied

3.1. Study Domain

The study domain covers the main part of West-Europe extending from 40°N to 54°N and 5°W to 10°E. The grid-size is defined to 0.25° x 0.25°, which can be translated to approximately 25km x 25km grid, which leads to 2096 grid points. All data-sets are provided with the same grid configuration. The following subsections explain in more detail the data-sets within the study domain as well as the basic climatology.

3.1.1. Climatology

Figure 3.2 illustrates the study area with a red square on a Koppen Geiger Climate Classification map [37]. This allows simultaneously to illustrate the climatology of this area. The climatology classification is based on long-term monthly precipitation and temperature data sets. The main climate types in the study domain are temperate C[$T_{hot} > 100 < T_{cold} < 18$], cold D[$T_{hot} > 10 T_{cold} 0$] and polar E[$T_{hot} < 10$], of which T_{hot} represents the hottest month and T_{cold} the coldest month. This shows that the climate of the study area has a dominant temperate climate only the mountains areas such as the Alps and the Pyrenees are polar. This consistency in climate is beneficial to compare the model performances on their robustness as it limits the amount of variables which could cause inconsistencies and anomalies.

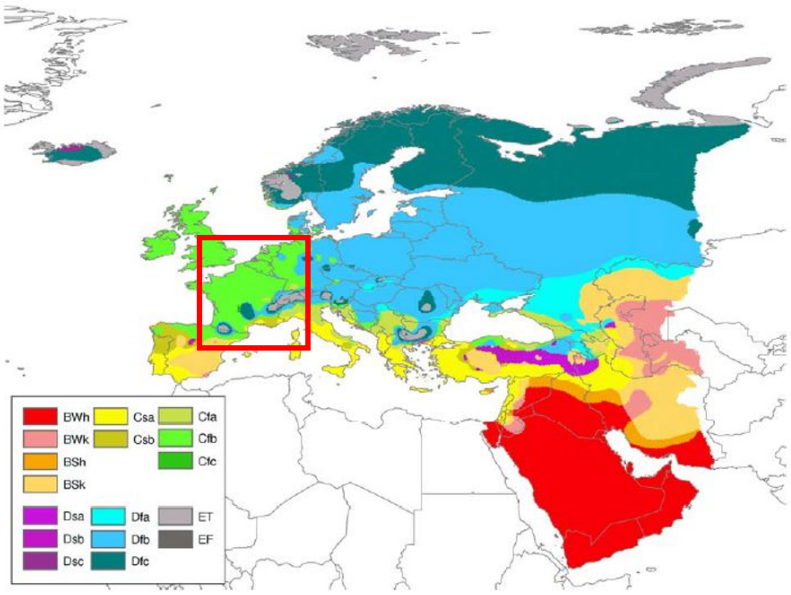

Figure 3.2: Extend of the study area illustrated within the red square; The color schemes on the map based on the Koppen Geiger Climate classification; The abbreviation B[...] stands for arid areas, C[...] for temperate areas, D[...] for cold areas and E[...] for polar [37]

3.1.2. Land-Surface Data

The land surface data is provided by Meteo France ISBA land-surface-model and contains four different surface variables: Leaf Area Index, Soil moisture (WG2), Gross Primary Production (GPP) and Ecosystem Respiration (RE). Those parameters are explained more in detail in Chapter 2. All of the variables have a daily resolution for a 25km x 25km grid. The time period covers 8 years is from 2010 to 2018. Next to the changing parameters Meteo France provided 19 different vegetation type fractions for each Grid-Point (GPI), see Table 3.1. This means each GPI consists of a combination of these land cover types.

Land Cover Type	Abbreviation
1: no vegetation (smooth)	NO
2: no vegetation (rocks)	ROCK
3: permanent snow and ice	SNOW
4: temperate broadleaf cold-deciduous summergreen	TEBD (TREE)
5: boreal needleleaf evergreen	BONE (CONI)
6: tropical broadleaf evergreen	EVER
7: C3 cultures types	C3
8: C4 cultures types	C4
9: irrigated crops	IRR
10: grassland (C3)	GRAS
11: tropical grassland (C4)	TROG
12: peat bogs, parks and gardens (irrigated grass)	PARK
13: tropical broadleaf deciduous	TRBD (TREE)
14: temperate broadleaf evergreen	TEBE (TREE)
15: temperate needleleaf evergreen	TENE (CONI)
16: boreal broadleaf cold-deciduous summergreen	BOBD (TREE)
17: boreal needleleaf cold-deciduous summergreen	BOND (CONI)
18: boreal grass	BOGR (GRASS)
19: shrub	SHRB (TREE)

Table 3.1: ISBA Vegetation Types [29]

3.1.3. ASCAT Data

The ASCAT data is provided by the EUMETSAT Data centre from the Metop-A satellite mission. Hereby, the backscatter measurements are re-sampled to a earth grid and further processed to account for azimuthal effects [35][43]. As mentioned in Chapter 2 the Metop satellites orbit with a 29 day cycle around the earth which provides overpass times at 9:30am and 9:30pm. For this study only the descending data-set is used. Naeimi et al. shows that there are no significant differences between these two different observations by showing very similar statistics between ASCAT and ESCAT data [35]. Still, Steele-Dunne et al. suggests that the use of descending observation is more practical because the plant-physiological processes are less influenced by dew [43]. Considering this, each grid point has three ASCAT parameters namely: Normalized (40°) backscatter coefficient, slope and curvature. Slope and curvature are provided on a daily bases whereas for backscatter the temporal resolution is dictated by the revisit time. All three data sets are available for a 10-year time-series from 2007 to 2018.

3.2. Data Preparation

In order to implement the data into a deep neural network both labels (ASCAT data) and inputs (Land Surface Model data) need to have the same spatial and temporal resolution. As mentioned before the data sets are provided on the same earth grid. The land surface model data is only available from 2010 to mid of 2018, which means the first two years of the ASCAT data can not be used for applying the deep learning model. Slope and Curvature as well as the LSM data have a daily resolution and need no further processing. Combining the backscatter data set with the LSM data sets requires to neglect several time-steps due to the temporal inconsistency of the backscatter observations. This inconsistency occurs because the satellite has different over pass times depending on the location.

3.3. Vegetation Type Clustering

Many studies show that the land cover and vegetation type is a crucial factor for the backscatter mechanisms and therefore observations of the ASCAT instrument [43][51][52]. Steele-Dunne et al. shows variability in slope and curvature values due to different vegetation phenologies such as grass types as well as the influence of forest. This has to do with the internal water dynamics but also vertical structure of the vegetation. Therefore, the vegetation land covers are suggested to establish clusters within the study domain to create a baseline to inter compare the results.

The vegetation types for this study domain are given in 19 fractions. This high dimensionality makes it very difficult to form similar clusters. Additionally, the fraction types are varying spatially in a small radius, which makes it even more difficult to establish robust cluster types.

In order to reduce the high-dimensional data into a low-dimensional space the t-distributed Stochastic Neighbor Embedding (t-SNE) machine learning algorithm is suggested. t-SNE is a state-of-the-art visualization tool developed for high-dimensional data by Van der Maaten et. al [33]. Another commonly used dimensionality reduction approach is Principal Component Analysis (PCA). However, it is only able to capture linear projections and therefore the t-SNE is used for this research project.

The basic concept is based on a t-distributed probability distribution to express the relationships between neighbouring data-points with the goal to create a low dimensional output of either 3D, 2D or 1D. In this case a reduction to 1D is chosen, hence every GPI has a specific value which representative for a specific combination of the given 19 fractions. Once the values per GPI are established a threshold is set to combine similar values to one clusters. The choice of threshold is based on the distribution of the t-SNE values.

The library used to implement the t-SNE algorithm is a module from the sklearn called "sklearn.manifold.TSNE" [1].

[1] `https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html`

3.4. General Variability Analysis

The general variability analysis is important to understand the seasonality of all input data but also to see the variation in space. To illustrate the spatial variability of the data temporal variance of the time series is suggested. Here, the time-series data are averaged over a grid point using 10 years of data. This gives a mean value for each point of time per GPI. The representative variance is then calculated over the one year time frame. Furthermore, the time-series can be split into three signals trend, seasonal and noise. This give an opportunity to identify the inter-seasonal variations and anomalies diverging from the seasonal development. This is done with a seasonal decomposition model using moving averages. The assumption hereby is that the signal can be represented by an additive model consisting of three signals: Trend (T), Season (S) and Residual (R). This simple model structure can be expressed in following equation - t represents a time step 3.4:

$$Model(t) = T(t) + S(t) + R(t) \tag{3.1}$$

Here the "$statsmodels.tsa.seasonal.seasonal\_decompose$" library in python 3.6 is applied [2]. This algorithm first extracts the seasonal cycle by applying an convolution filter and returns its average as a seasonal component. This approach, however, is not the most sophisticated method, because it does not capture non-linear behaviour between the three signals. Still, it can provide a very good indication of anomalies in the data time series and is therefore used as a first order analysis to identify if the model is able simulate anomalies and not only the seasonal trend. An anomaly in the context of this thesis is defined as following:

Anomaly Time Series = Orignal Data Time Series - Seasonal Trend of the Time Series

3.5. Establishing a Baseline for Different Deep Neural Network Architectures

The observational parameter normalized backscatter is provided as an instantaneous measure of the ASCAT observation and is utilized as a daily value. Still, the temporal resolution per grid point is different considering swath frequency and coverage of the Metop satellite. Slope and curvature however are due to the high noise in the raw signal further on processed to a continuous signal. This processing is done by applying the Epanechinkov kernel by integrating to one with an finite sample range [-y,y] [47]. The data processing is done with a kernel window of 42 days whereas the influence of the neighbouring measurement decreases with increasing distance to the estimated time-step [23].

Considering this the input for simulating backscatter is an instant estimation, which means given values of day x should also simulate observation (backscatter) for date x, because only this state has impact on the instantaneous measurement. A long-term dependency of the surface properties should be reflected in the current state of the input data. Slope and curvature as explained before contain information from 42 days (x -21 days and x+21 days). This results in a different architecture of the deep neural network compared to the backscatter simulation. Therefore, a different architecture is suggested, which can estimate slope/curvature from a time-sequence (42days) of the input values. Every time-sequence should be treated as a single sample, which means that it can be fed randomly to the model without influencing the training results. This format can be achieved by applying a sliding window to the input/output data and mix this outcome randomly. Hence, long-term dependencies are avoided.

Many deep neural network architectures are available of which the most important are: Regular Deep Neural Networks, Convolutional Neural Networks and Recurrent Neural networks. The simplest form of a deep neural network is a multi-layered neural network, which includes an input layer, x-amount of hidden layers and an output layer. For backscatter this is the most appropriate one, considering the one-dimensional input of only the four land surface variables. Slope and curvature needs to be processed differently because of time as a second dimension. This means slope and curvature can be treated as a time-series-prediction problem.

State-of-the art networks for time-series predictions are Recurrent Neural Networks (RNN) such as Long-Short-Term Memory and Gated Recurrent Neural Network [26][8]. The problem with RNN is the long-term dependency, which is useful for general time-series prediction but in the case of slope and curvature not

[2] `https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.seasonal_decompose.html`

wanted, because the mathematical dependency of the signal processing should solely involve a specific time window (kernel width of 42 days). In general, the goal is not to solely create an accurate time-series prediction it is rather a trade of between keeping the results physical explainable and achieving an accurate prediction. Hence, an architecture is needed, which considers sequential data but at the same time does not keep long-term dependencies. This can be achieved by Deep Convolutional Networks. Here, the date is processed sequentially using a time-window without keeping long-term dependencies.

3.6. Implementation of Deep Neural Networks

As a conclusion from the previous subsection, for backscatter a Regular Forward DNN (RFDNN) is suggested and for slope as well as curvature a Convolutional DNN (CDNN). For implementing deep neural networks several libraries in Python are available, e.g. TensorFlow, Keras and Pytorch to name the most common one. TensorFlow is the high-end library, giving a lot of flexibility in adjusting the code. Keras and Pytorch are more modular based, which means most functions are pre-defined, hence easier to implement. As this thesis is a prove of concept it does not require self written neural network concepts. Therefore, the user friendly library Keras is suggested, as it is also widely used in the scientific community. Keras is a open source neural network library and runs on top of TensorFlow. This allows to use functions applicable to TensorFlow and therefore gives Keras and advantage over Pytorch.

Both, RFDNN and CDNN are sequential models, which means the layers are linearly stacked [3]. This allows to build a model layer by layer. The basic set-up of a sequential model is a core layer, a compiler to calculate the loss and backpropagate through the network and fit. Fit is the part in which the model is trained and the model design is stored.

The RFDNN contains multiple core layers in form of so called dense layers[4]. Those layers are fully connected. The complexity of the system can be increased by implementing more layers or/and adding more neurons to the layers.

The deep structure of a CDNN is due to multiple convolutional layers (Conv1D[5]) and several dense layers stacked on top of each other. As for the RFDNN the complexity can be increased by stacking more layers on top of each other and increasing the number of neurons within the layers. Additionally, the kernel size can be changed. By increasing the kernel size more data points are taken into account at once and by decreasing less data at once, which means the layer following layer size is either decreased or increased respectively.

Next to the amount of layers and neurons several other parameters can be changed and adjusted within a deep neural network, e.g. the activation function, loss function, learning rate, optimizer, dropout, amount of epochs and batch size. All these variables are called "Hyper-parameters" and are briefly explained in table 3.2. By changing these parameters the deep neural network can be optimized to achieve the best result. Further details about the training and optimization procedures are elaborated in following subsections.

[3] https://keras.io/models/sequential/
[4] https://keras.io/layers/core/
[5] https://keras.io/layers/convolutional/

Hyper-Parameter	Definition
Activation Function	An activation function introduces non-linearity to the model applying it onto the activation's per neuron, weights and bias. The most common activation functions are [22]: 1) Sigmoid $$\sigma(y) = \frac{1}{1+e^{-y}}, \quad (3.2)$$ which squishes the values between 0 and 1; 2) Hyperbolic tangent activation function (Tanh) $$f(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}, \quad (3.3)$$ which transforms the values between -1 and 1; 3) Rectified linear unit (ReLU) $$g(z) = max\{0, z\}, \quad (3.4)$$ which neglects all negative values and scales the other values between 0 to infinity
Loss Function	The loss function gives an equation to find the loss between the label and the outcome of a simulation. It is the key driver to distill the "loss" to one number which is used to backpropagate through the network. Keras offers several different functions for example mean squared error, mean absolute error and mean squared logarithmic error etc.[6].
Optimizer	Combined with the loss function, the optimizer tries is responsible to update the weight and biases in the network. This can be done with a mathematical concept called gradient descent, as explained in Chapter 2. The stochastic gradient descent is the most basic concept, but Keras offers several more optimizer. The following are the most common ones: Stochastic Gradient Descent (SGD), Adagrad, Adam and Nadam [7].
Learning Rate	The learning rate expresses how fast the neural network is learning, or how big the steps are within the optimization process. If the steps are too big the optimizer is more likely to end up in a local minimum instead of absolute minimum. Every optimizer has a recommended learning rate, but still can be adjusted and therefore be used to optimize the model performance.
Dropout	Dropout is a tool to increase the regularization of the model by removing randomly non-output units. The amount of ignored units can be set as an fraction between 0 and 1.
Hidden Layers and Neurons	With the amount of hidden layers and neurons the complexity of the network increases. However, too complex or too simplistic models can result in wrong estimate . Therefore, the choice of the number of hidden layers and neurons are important to achieve regularization and an accurate result.

Table 3.2: Hyper-parameters, which have been used to optimize the deep neural networks

[6] https://keras.io/losses/
[7] https://keras.io/optimizers/

3.6.1. Training - Inputs and Outputs

The data is provided for 2097 grid points over Europe. In general, the results of the DNN's should be used to one simulate the observation parameters and second to give a physical insight. There are two possible ways to train the model: 1) Per multiple grid points and 2) per individual GPI. By training the model using data from all or multiple grid points the interpretation will be limited because every grid point is expected to have different feature importance depending on their vegetation type, soil composition, climatology etc.. Therefore, every observation parameter is predicted per grid point, which means every GPI has its own model or even two considering that backscatter and slope/curvature will be predicted separately. Still, training on more data resolves in better accuracy of the prediction, hence interpretability is a trade off with model performance.

For this research the project the model is trained per GPI, because interpretability has a higher priority than an increased model performance. As the ASCAT observation parameters are not fully understood yet it is from a high scientific significance to assign properties to them in order to explore unknown properties and patterns.

The inputs for both DNNs are the surface variables mentioned in subsection 5.1.2 (LAI, WG2, GPP and RE) and the outputs or labels contain the three ASCAT observations (backscatter, slope and curvature). Table 3.3 gives a brief summary of all input and output variables. Before training the model the data needs to be normalized. This is necessary, because the ranges of the data are given in different magnitudes and the model might be intrinsically be more influenced by values with higher magnitude then by data with small magnitudes. Therefore, the "MinMaxScaler"[8] from the sklearn.preprocessing is applied on the data, squeezing it between 0 and 1 . In following the input-output structure is elaborated separately for the RFDNN and CDNN model.

Inputs	Symbol
Leaf Area Index	LAI
Soilmoisture	WG2
Gross Primary Production	GPP
Ecosystem Respiration	RE
Outputs	**Symbol**
Normalized Backscatter	σ
Slope	σ'
Curvature	σ''

Table 3.3: DNN Inputs and Outputs [29]

1) RFDNN backscatter simulation:

The set-up of the the RFDNN requires daily sets of inputs and labels, see Figure 3.3 (a) with t representing a daily time step. This can be expressed in two vectors: 1) 1x4 as input and 2) 1x1 as output.

2) CDNN slope and curvature simulation:

The set-up of the the CDNN are created with a sliding window, taking always 42 days consecutive input observations into account for one tuple of outputs at time step t+21, illustrated in Figure 3.3 (b). This leads to following input and output vectors:1) 42x4 as input and 2) 1x2 as output.

[8] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

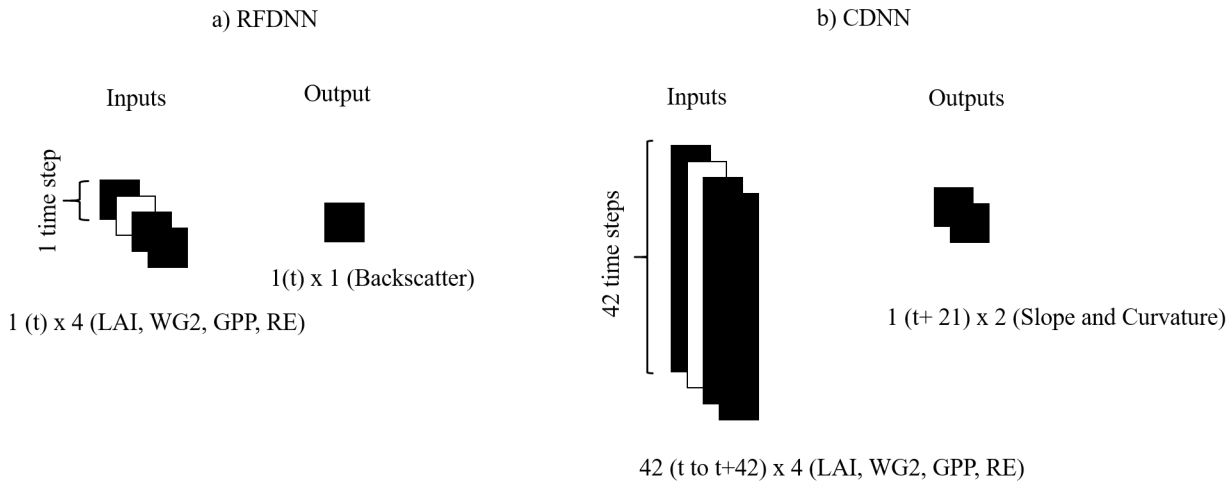

Figure 3.3: A conceptual plot to illustrate the training loss/error and testing loss/generalization error development; Source [22] Page 112, Figure 5.3

In total 8 years of data is utilized to train and test both models. Hereby, one year is reserved as a test dataset and the remaining data is used to train the model. The reference case uses 2017 as a testing year and the other years to train the model. However, two more cases with different testing years are evaluated in from of a cross validation to test the robustness of the model.

3.6.2. DNN Optimization

Optimizing a deep neural network is often refereed to as the "black-art" [3][22][42]. The optimization process involves tuning every single hyper-parameter by constantly comparing the outcome. Here, the goal is to create a model which has the ability to perform accurate simulations for unseen inputs, also referred to as generalization [22]. This means the model performance is judged on the accuracy or error given unseen data. The training and optimization procedure consist of four steps: 1) Set arbitrarily the hyper-parameters, 2) Analyzing the training error, 3) Analyzing the test error or generalization error and 4) Update the hyper-parameters.

The general practice in machine learning is to pick extreme parameter values and compare the results manually. However, there are other possibilities to find the best parameter set-up for example the Bayesian optimization and Grid Search [3][42]. The Grid Search technique allows to automatically estimate the best hyper parameters, by trying every possible combination of a given parameter space. The parameter space is pre-defined, which means the user sets up all parameter combinations and grid search returns the best combination[3]. This is done in a nested for-loop and computational very expensive considering the high dimensionality of the hyper parameter sets. The Bayesian optimization however uses prior evaluated parameters and then searches through the hyper-parameter space by updating the prior belief whilst training. This makes the optimization process more efficient then grid search but still has a great demand of computational power.

This thesis uses the manual approach of selecting parameters by constraining it with extreme parameter values. As mentioned before this research is a proof of concept rather then finding the most optimum hyper-parameter combination. A scientific proof optimization process needs more time and is suggested for further research as the investigated method proves to be successful.

A powerful tool to judge the performance and training progress is to plot the training as well as the testing loss development per epoch. An epoch is a one whole forward and backprogagation process through the network. Several epochs means therefore that the network is trained several times with the same data by constantly updating the weights and biases. Evaluating the training and testing loss can result in three cases: 1) Underfitting when the model is not able to perform well on the training data, 2) Overfitting, which means the model performs well on the training data but not on the test data and 3) Optimal Capacity, which gives the same model performance for training and testing data.

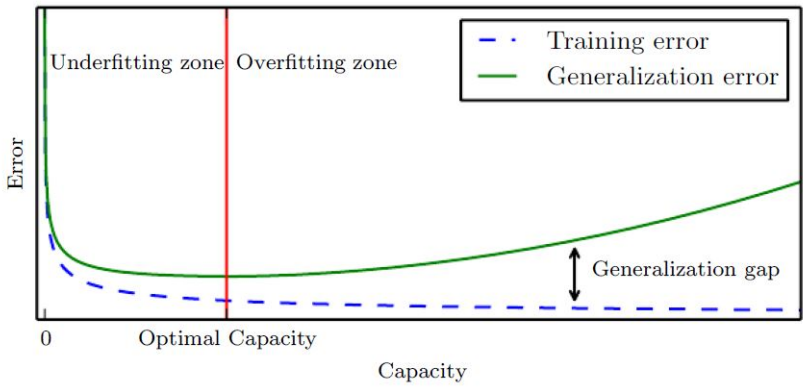

Figure 3.4: A conceptual plot to illustrate the training loss/error and testing loss/generalization error development; Source [22] Page 112, Figure 5.3

Figure 3.4 shows a typical training loss and testing loss relationship. In case the training loss diverges from the testing curve the model is overfitting. Underfitting occurs if the error is too big in both cases the parameters need to tuned. The final result should aim to reach the optional capacity of the model which is indicated with a red line in Figure 3.4. This is done by adjusting the hyper-parameters.

3.6.3. Regular Forward DNN - Architecture

The architecture was developed based on the manual optimization processes explained in the previous subsection 5.5.2. A three layered deep architecture with 128, 64 and 42 neurons is selected. As an optimizer "Adam" and as a loss function the mean squared logarithmic errors with a learning rate of 0.001 is applied. For the first two layers the tanh activation function performed best and for the last layer Relu is implemented. This following python code snapshot illustrates the implementation sequence:

```
model = Sequential()
model.add(Dense(128,activation=tanh, input_shape = (4,)))
model.add(Dense(64,activation=tanh))
model.add(Dense(32,activation=relu))
model.add(Dense(units = 1))
model.compile(optimizer = 'Adam', loss = mean_squared_logarithmic_error, metrics=['accuracy'])
```

3.6.4. Convolutional DNN - Architecture

The architecture was developed based on the manual optimization processes explained in the previous subsection 5.5.2. Two convolutional layers are suggested with a filter size of 64 and a kernel size of 12 for the first layer and 5 for the second layer. After this a MaxPooling layer is implemented, which is then flattened and fully connected to a dense layer of 32 neurons. This dense layer is then full connected to the output layer. As an optimizer "Adam" and as a loss function the mean squared logarithmic errors was chosen with a learning rate of 0.001. All weights and biases are updated using the Relu activation function. The following python code snapshot illustrates the implementation sequence:

```
model = Sequential()
model.add(Conv1D(filters=64, kernel_size = 12, activation =' relu', input_shape = (42,4)))
model.add(Conv1D(filters=64, kernel_size=5, activation='relu'))
model.add(MaxPooling1D(2))
model.add(Flatten())
model.add(Dense(32, activation='relu'))
model.add(Dense(2) )
model.compile(optimizer = 'Adam', loss = mean_squared_logarithmic_error, metrics=['accuracy'])
```

3.7. Performance Analysis

The performance analysis is split into two sections:

1. The general performance analysis, which describes the spatial and temporal model accuracy.
2. The model Performance regarding Vegetation types, which means that performance is evaluated per specific vegetation type and vegetation type cluster

The validation of the DNN derived estimates are only done with the testing year of the data set. This guarantees an unbiased result and shows the models ability to generalize. This analysis concept is separately applied for each ASCAT simulation - Backscatter, Slope and Curvature.

3.7.1. General performance analysis

For the general performance analysis each GPI is evaluated with several validation metrics per year as well as per season. A season is defined as a time-window of three months, hence four seasons per year, namely December-January-February (DJF), March-April-Mai (MAM), June-July-August (JJA) and September-October-November (SON). This seasonal and annual analysis is done for the original data as well as for the anomaly time series. The performance of the model is evaluated by employing four validation metrics:

1) Root Mean Squared Error (RMSE)
Root Mean Square Error is expressed as the standard deviation of the prediction errors. It is a well established measure of the model performance and applied in many environmental fields such as air quality, climate research and meteorology [7]. It expresses how much spread the data is around the line of best fit regarding truth. Mathematically it is expressed as following[7]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{1,i} - x_{2,i})^2}, \tag{3.5}$$

where n is the number of samples within the time series and x_1, x_2 represent individual samples from the observation truth and simulation respectively. In general, the smaller the RMSE value the better is the prediction.

2) Mean Forecast Error (Forecast Bias)
Mean Forecast Error (MFE or Bias) gives a direct indication how much the values are overshooting or undershooting the truth value. This is also called the forecast bias. It is calculated as following:

$$MFE = \frac{1}{n}\sum_{i=1}^{n}(x_{1,i} - x_{2,i}), \tag{3.6}$$

3) Mean Absolute Scaled Error (MASE)
Another method to evaluate the error is the Mean Absolute Scaled Error [27]. Hyndman et al. (2005) suggested this from of performance metrics especially for inter comparison of time series analysis [27]. MASE is an approach to compensate for short-coming of the Mean Absolute Error (MAE) by scaling the data to compare it with forecasts across data sets, adjusting for behavior issues once values are zero, penalizing errors in small and large forecasts equally and by approximating the diebold-mariano test [27]. The scaling is done with a naive MAE* estimation using one-step naive forecasting, see equation 3.8. Y represents the observation value, which means the MAE* is expressed as the mean difference between two observations for time step *i-1* and *i*. e_t represents the error between the observation (x_1) at time step i and forecast (x_2) at time step i [27]. Low values indicate better performance and high values a worse performance compared to one-step naive forecasting. MASE can be expressed as in equation 3.9.

$$e_t = x_{1,i} - x_{2,i} \tag{3.7}$$

$$MAE* = \frac{1}{n-1}\sum_{i=1}^{n}\left|Y_i - Y_{i-1}\right| \tag{3.8}$$

$$MASE = mean\left(\frac{e_t}{MAE*}\right) \tag{3.9}$$

4) Spearman Correlation Coefficient (ρ)
The Spearman correlation coefficient is a measure of monotonicity between two datasets. This gives an advantage over the classical Pearson Correlation Coefficent, because the dataset do not have to be normal distributed [41]. Here, +1 implies a strong monotonic relationship and -1 suggests a negative correlated monotonic relationship. It is a very valuable statistical metric because it gives information about the magnitude of the correlation but also the direction (negative or postive) [9]. It can be expressed in following equation:

$$\rho = 1 - \frac{6\sum d_i)^2}{n(n^2-1)} \tag{3.10}$$

where ρ is the Spearman rank correlation, di the difference between the ranks of related variables and n is the number of measurements.

3.7.2. Model performance regarding vegetation types

The evaluation of the model performance regarding vegetation types allows to correlate performance to land cover types but also to investigate the inter-annual variations in more detail. In order to establish a correlation, first the grid points with the most dominant vegetation fraction are investigated. Secondly, the vegetation type cluster calculated with the t-SNE are suggested to explore further correlations. This can be done with time-series plots for each dominant vegetation type fraction and clusters using either the original data and the anomaly time-series data.

3.8. Feature Importance Analysis

The feature importance analysis is split into the same sections as the performance analysis: 1) The feature important analysis, which describes the spatial and temporal feature importance and 2) The feature important regarding vegetation types, which means that the feature importance is evaluated per specific vegetation type and vegetation type cluster. This analysis concept separately applied for each ASCAT simulation - Backscatter, Slope and Curvature.

The feature importance analysis is done using the DeepSHAP[10] algorithm explained in Chapter 2. This algorithm allows to extract information about the feature importance per time step in form of an empirical value.

The general feature importance analysis evaluates the data outcomes per season as well as an annual/overall importance. Additionally, time-series plots per vegetation type cluster are suggested to investigate possible correlations between vegetation types and land-surface variables.

3.9. Performance Cross Validation

The performance cross validation gives an estimate how well the model is performing or how robust it is using different years as testing data sets. However, doing this for each year requires a high computational effort and is therefore executed only for two additional years. In order to evaluate the robustness all three model simulations are compared to each other. Hereby, for each model run the Root Mean Squared Error is calculated per GPI using the whole time-series data (training + testing). This cross validation is separately applied for each ASCAT simulation - Backscatter, Slope and Curvature.

[9] `https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.spearmanr.html`
[10] `https://github.com/slundberg/shap/blob/master/shap/explainers/deep/__init__.py`

4

Results and Discussions

The results are divided into five parts 1) Vegetation type clustering, 2) Data Variability, 3) Backscatter Analysis, 4) Slope Analysis and 5) Curvature Analysis. The data variability as well as the ASCAT simulations are investigated in the same structure by comparing the results first generally, via extreme vegetation types and per distinctive clusters. The backscatter, slope and curvature sections have additionally a discussion section to summarize and interpret the sub-results.

4.1. Vegetation Type Clustering

Clustering is done applying the t-SNE function and allows a spatial model performance and feature importance comparison based on land cover type fractions. In order to optimize this machine learning function the perplexity of this algorithm can be adjusted. Perplexity gives an indication about effective number of neighbors. The developer of t-SNE algorithm states that the most common values are between 5 and 50 [33]. The performance of the t-SNE was tested for perplexity values between 5 and 70. For further analysis the perplexity value of 70 was chosen as the t-SNE values showed natural breaking points throughout the t-SNE value distribution. Those breaking points are then used to create the bin-size for the cluster types, which resolves in 13 distinctive vegetation classes. This is visualized in Figure 4.1, by indicating the natural breaking points with black vertical lines.

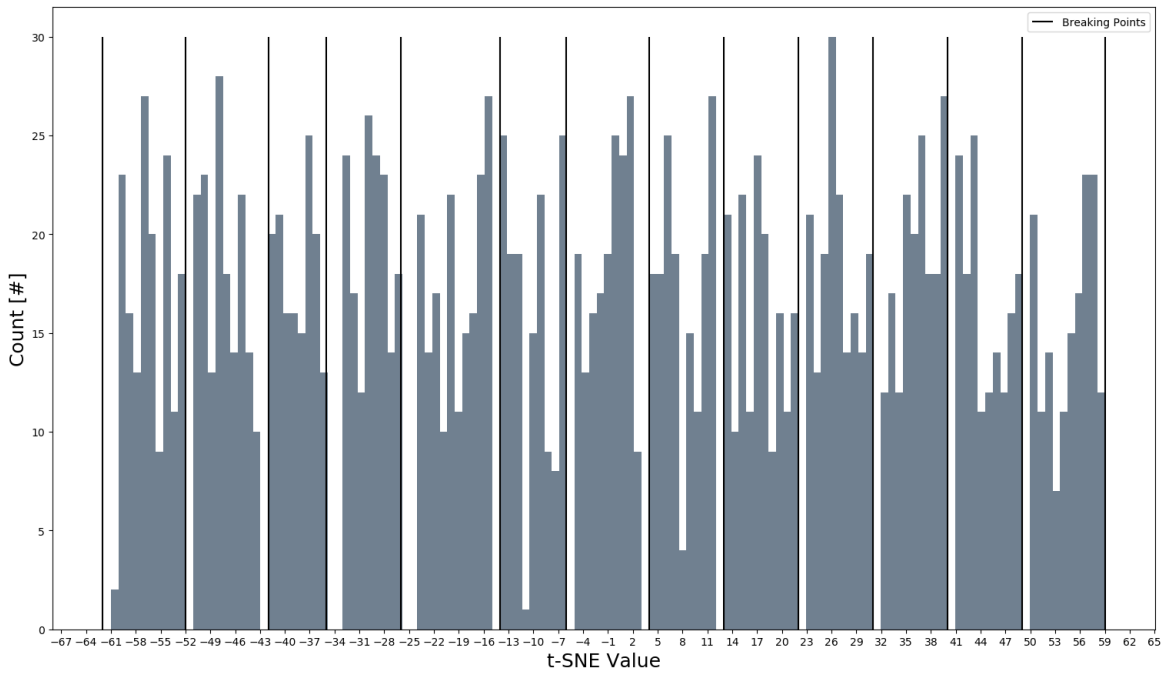

Figure 4.1: The histogram shows the t-SNE value distribution for each GPI. The black lines indicating the natural breaking points, which are used to cluster the data

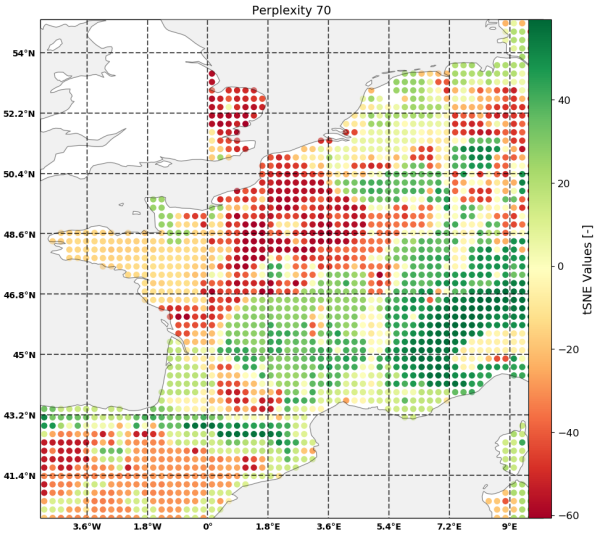

Figure 4.2: t-SNE value distribution over the study domain using perplexity of 70

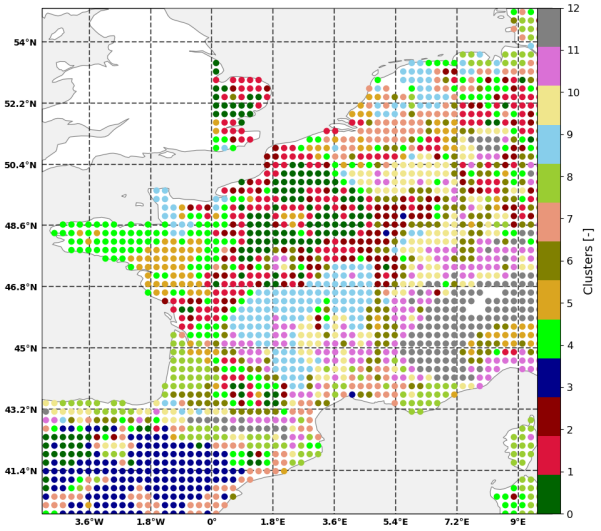

Figure 4.3: Study domain showing the 13 distinct vegetation clusters

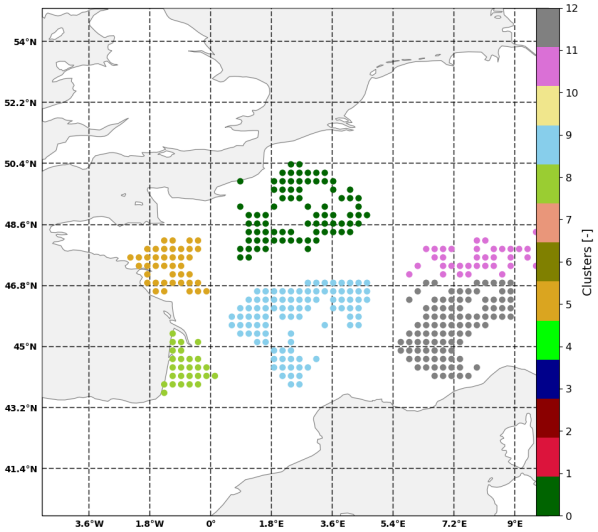

Figure 4.4: Distinctive cluster types within the study domain due to its spatial consistency.

Figure 4.2 shows the distribution of the estimated t-SNE values over the study domain, whereas Figure 4.3 illustrates the 13 distinctive clusters binned by the natural breaking points. These clusters either contain a very mixed land cover type distribution or are distinct for a specific vegetation type.

Table 4.1 shows the mean vegetation type fractions per cluster. Here, the vegetation fraction mean was taken using all GPI's within a cluster. This gives a rough estimation about which land cover types are present and/or dominant for each cluster. Vegetation type fractions above 5 percent are marked red and fractions above 30 percent are illustrated in bold. This indicates that the most present vegetation types within the study domain are 1 (No Vegetation), 4 (Temperate Broadleaf), 7 (C3 Cultures), 10 (Grassland) and 15 (Temperate Needleleaf) as these types are most important for every cluster. The clusters used for further analysis are green high-lighted - namely 0, 5, 8, 9, 11, 12. The combinations of these clusters allow to compare the dominant vegetation type fractions to each other. The chosen clusters are illustrated in Figure 4.4, which highlights the selected areas.

The cluster type 0 is dominated by cultures C3, which indicates that this land-tile is defined by agriculture and therefore is a representative for the C3 Culture types. Cluster 5 has a very mixed land-cover distribution, covering type 1,2,4,7,8 and 10 as fractions above 5 %. This allows to analyse the model performance for more

heterogeneous footprints. Cluster 8 is dominated by grassland and temperate needleleaf and hence used to investigate the model performance on forest. Cluster 9 is mainly dominated by grassland (Veg.Type 10) with smaller fractions of temperate broadleaf and C3 cultures. The high fractions of grassland makes this cluster unique to further discuss the simulations results for grassland. Cluster 11 is characteristic for temperate boradleaf and grassland, which again gives a unique view on the simulation performance regarding temperate boradleaf dominated forests. The mountain area such as the Alps are represented in Cluster 12. Characteristic for Cluster 12 are fractions of boreal grassland and boreal needleleaf but also no vegetation. This allows to critically analyse the deep learning model as the input variables are based on vegetation states, whereas the alps have high percentage of bare rock.

The general vegetation type fraction variability per cluster is illustrated in box plots, see Appendix A.1. Cluster 0 varies mainly in the grassland and C3 Cultures, but shows a high C3 Culture fraction for all grid point within the cluster. Cluster 3 has strong variations of C3 clusters (between 0.5 and 0.1 percent), but shows a high fraction of No-Vegetation for all GPIs. Cluster 5 varies in general a lot for all its main land covers with the highest variation of Grassland. Cluster 9 has a strong consistency in grassland fraction ranging between 40 and 70 percent. Cluster 12 shows in general large variation especially for boreal and C3 grassland, no vegetation, needleleaf and broadleaf.

	Vegetation types													
Cluster	1	2	4	5	7	8	9	10	12	14	15	16	18	19
0	0.09	0.02	0.14	0.0	**0.59**	0.03	0.01	0.1	0.0	0.0	0.02	0.0	0.0	0.0
1	0.08	0.05	0.14	0.0	**0.47**	0.03	0.0	0.17	0.0	0.0	0.05	0.0	0.0	0.01
2	0.07	0.03	0.18	0.01	**0.37**	0.03	0.0	0.23	0.0	0.0	0.07	0.0	0.0	0.01
3	**0.34**	0.03	0.02	0.0	0.28	0.01	0.03	0.17	0.0	0.03	0.06	0.0	0.0	0.02
4	0.09	0.03	0.08	0.01	0.29	0.04	0.01	**0.32**	0.01	0.01	0.07	0.0	0.0	0.04
5	0.08	0.12	0.06	0.01	0.23	0.09	0.02	**0.30**	0.01	0.02	0.04	0.0	0.0	0.03
6	0.07	0.04	0.20	0.02	0.19	0.07	0.02	0.26	0.01	0.0	0.1	0.0	0.0	0.01
7	0.14	0.05	0.11	0.0	0.14	0.04	0.01	**0.30**	0.01	0.03	0.15	0.0	0.0	0.02
8	0.09	0.02	0.08	0.01	0.06	0.02	0.0	**0.30**	0.01	0.09	0.29	0.0	0.0	0.02
9	0.05	0.02	0.16	0.0	0.12	0.02	0.0	**0.55**	0.01	0.0	0.07	0.0	0.0	0.01
10	0.06	0.02	0.23	0.03	0.08	0.02	0.0	**0.33**	0.01	0.01	0.22	0.0	0.0	0.0
11	0.05	0.02	**0.31**	0.06	0.08	0.02	0.0	**0.32**	0.0	0.0	0.13	0.0	0.0	0.0
12	0.1	0.07	0.12	0.20	0.03	0.01	0.0	0.21	0.02	0.0	0.02	0.04	0.16	0.01
Max	0.56	0.63	0.47	0.37	0.72	0.68	0.28	0.73	0.28	0.43	0.75	0.19	0.41	0.23

Table 4.1: Mean vegetation type fractions per cluster. The vegetation type is displayed as a reference number to the description of table 3.1. Red high-lighted are the dominant fractions above 5% and the bold fractions are the fractions above 30%. Mint indicates which vegetation types have a fraction above 45% of the specific vegetation type. Green indicates the most distinctive clusters shown in Figure 4.4. Vegetation types are defined as follows: 1: no vegetation (smooth) , 2: no vegetation (rocks), 3: permanent snow and ice, 4: temperate broadleaf cold-deciduous summergreen, 5: boreal needleleaf evergreen, 6: tropical broadleaf evergreen, 7: C3 cultures types, 8: C4 cultures types, 9: irrigated crops, 10: grassland (C3), 11: tropical grassland (C4), 12: peat bogs, parks and gardens (irrigated grass), 13: tropical broadleaf deciduous, 14: temperate broadleaf evergreen, 15: temperate needleleaf evergreen, 16: boreal broadleaf cold-deciduous summergreen, 17: boreal needleleaf cold-deciduous summergreen, 18: boreal grass, 19: shrub

As mentioned before only the most distinctive and contiguous clusters (0,5,8,9,11 and 12) and main land-over types (1,4,7,10,15,18 and 19) are used for further analysis. Even though vegetation type 19 (Shrub) is not a main contributor it is considered for this analysis as it has a unique structure. These main vegetation type fractions are used to compare single GPI's with the highest fraction, see high-lighted blue in table 4.1. The reference names are used to describe the contiguous clusters and vegetation types shown in table 4.2 and 4.3 respectively.

Cluster	Reference Name
0	Agriculture
5	Pays Mixed Coast
8	Needleleaf Forest
9	French Grassland
11	Broadleaf Forest
12	Boreal Alps

Table 4.2: Overview of clusters and their reference names

Vegetation Type	Reference Name
1	No Vegetation
4	Broadleaf
7	Agriculture
10	Grassland
15	Needleleaf
18	Boreal Grass
19	Shrub

Table 4.3: Overview of vegetation types and their reference names

4.2. Data Variability

This results section is divided into two parts 1) Variability of the Land Surface Parameters and 2) ASCAT Observations. Here, the variability is analyzed first generally (over-all spatially), secondly per cluster and season and thirdly for the main land-cover types. This gives a basic understanding of the input and output data and will further be used to explain the variation in model performance as well as feature importance.

4.2.1. Variability of land surface parameters

Figure 4.5 shows the annual variance per GPI for all Meteo-France land surface parameters. The variance is defined as the squared deviation of a variable from its mean and therefore gives an indication how much spread values are around their mean. In the context of a time-series it shows how strong the seasonal variations are per year. A high value (green) indicates a stronger variation compared to a lower value (red).

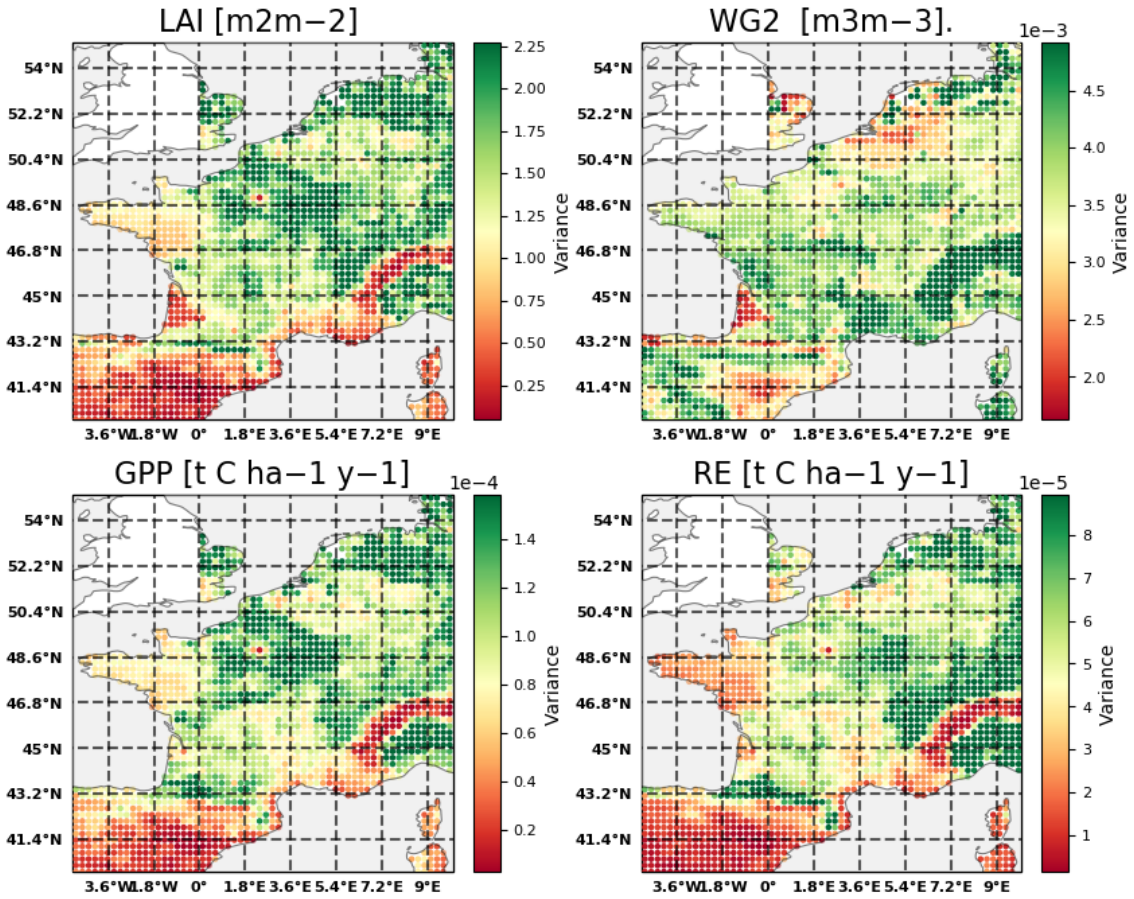

Figure 4.5: Annual variance per grid point for all meteo-france land surface parameters

LAI, GPP and RE show the same spatial variance patterns. Here, the highest variance occurs at the agriculture

land around Paris. This indicates the strong seasonal cycle of the growing and decay seasons of agriculture. The dry rock-land area in Spain as well as the Alps and Pyrenees show low variation, which means the seasonal magnitude is rather low compared to the mean as there is very little vegetation assimilation and phenology change due to a high fraction of no vegetation. Conversely, soil moisture (WG2) shows the highest seasonal variation along the Alps mountain range as well as on the south of the Pyrenees. This is due to a colder and wetter climate, shown in the Koppen Geiger Climate Classification map 3.2. The lowest WG2 variation are in the south-west of Spain as well as in the needleleaf area around Bordeaux.

Table 4.4 shows the mean variance per cluster. Agriculture Land and Broadleaf Forest show very similar extremes in their annual land surface states variance. This counts especially for the LAI, GPP and RE parameters. The lowest seasonal change in LAI can be observed for Needleleaf Forest as in general needleleaf vegetation is consistent throughout the seasons. This cluster additionally shows lowest variance in soil moisture values. The Pays Mixed Coast shows intermediate variance for all land surface states, but lowest for its respiration values.A similar variance is to observe for the French Grassland, however, the magnitude of variance is higher then for the Pays Mixed Coast. The Boreal Alps have the highest variation of soil moisture but rather low changes in LAI, GPP and RE due to the high fractions of no vegetation. This might be due to a colder and wetter climate, as indicated in the Koppen Geiger Climate Classification map 3.2.

	Annual Variance			
Cluster	LAI[m2-m2]	WG2[m3-m3]	GPP[t C ha-1 y-1]	RE[t C ha-1 y-1]
Agriculture Land	2.229234	0.003774	0.000161	0.000066
Pays Mixed Coast	0.892872	0.003980	0.000069	0.000024
Needleleaf Forest	0.459571	0.002330	0.000096	0.000043
French Grassland	1.553287	0.004078	0.000090	0.000048
Broadleaf Forest	2.136135	0.003706	0.000129	0.000092
Boreal Alps	1.148533	0.005662	0.000063	0.000041

Table 4.4: Overview of the mean annual variance for each Cluster and for each land surface variable

Figure 4.6 shows the time series development for seven GPI's with maximum vegetation type fractions of either Vegtype 1 = No Vegetation, 4= Broadleaf, 7 = Agriculture, 10 = Grassland, 15 = Needleleaf 18 = Boreal Grass and 19 = Shurb for LAI, WG2, GPP and RE.

LAI has the strongest magnitude for broadleaf and the lowest for no vegeation. The LAI-Broadleaf (yellow) shows the the broadest annual peaks by increasing in the DJF season of the year and decreasing towards the latest season (SON). The LAI agriculture (green) starts at the same time but has a sharp decrease in the middle of the year and a small growing season around November. Needleleaf (purple) has a small LAI seasonal magnitude by remaining LAI above 1, which means the leaf area remains present throughout the whole year. Grassland has a similar periodicity as broadleaf, however, the magnitude is lower.

Boreal Grass has next to no vegetation the lowest LAI magnitude, with an onset in the middle of the year until September. This is because boreal grass is present in the alpine areas and the cold climate causes a shift in the onset of growing season as the warmer conditions are reached later in the year compared to lower areas. Similar patterns for each max. vegetation type fraction are found for GPP and RE, whereas the signals are more noisy.

WG2 (soil moisture) shows a less periodical cycle for all vegetation types, as it is more dependent on rain events. In general, grassland has the highest magnitude in soil moisture, whereas no vegetation has the lowest. Boreal Grassland shows an exception, by having a very periodic increase and sharp decrease in soil moisture. The decrease occurs at the start of the snow season in the mountain areas (October-November) and remains low until the snow transforms to rain (March-April).

These inter-annual variations are illustrated in Table 4.5. Here, the variance is explained per season (DJF, MAM, JJA, SON) and per land surface variable (LAI, WG2, GPP, RE). The green fields highlight the maximum variance per cluster and variable. LAI has its highest variation throughout all clusters between May and August - only

the Boreal Alps and Broadleaf Forest clusters have their maximum variance in the SON season. This might be due to its geographical close location. Soil moisture is most variant between October and February, which are also the months with the highest precipitation. GPP and RE show most variance for the MAM and JJA season as in those months the growing and harvesting are taken place.

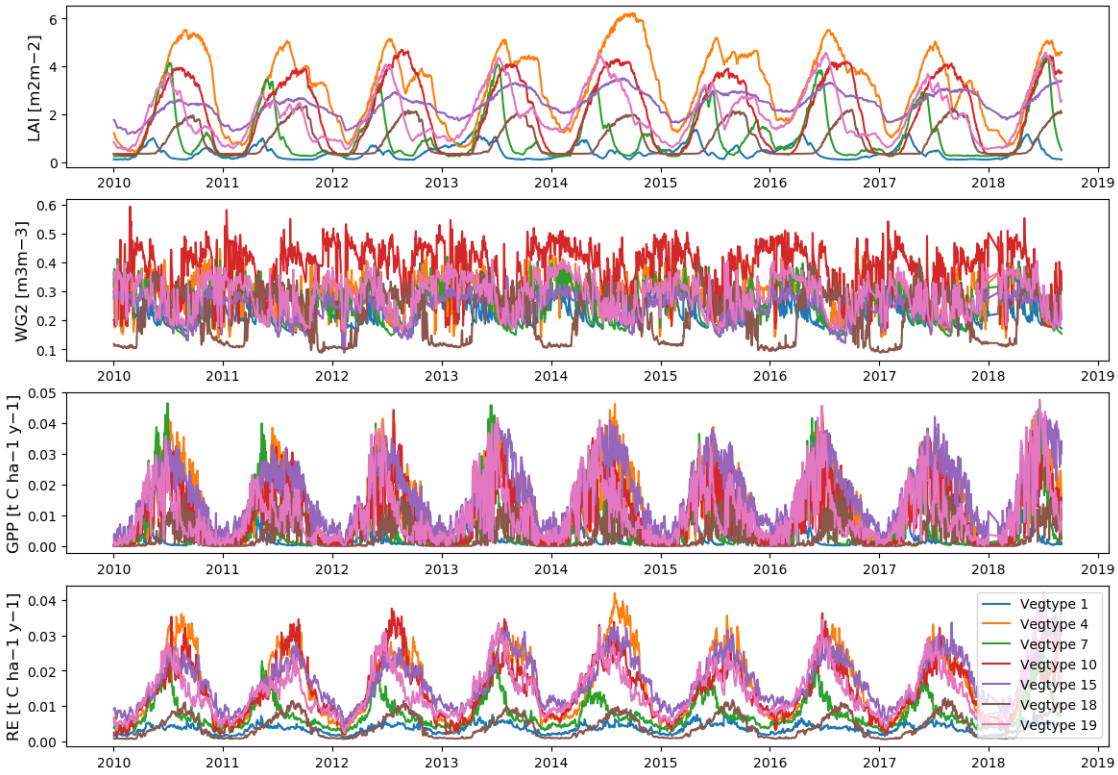

Figure 4.6: Time series plots from 2010 to 2019 showing the strength of the seasonality per land surface variable for the GPI's with maximum vegetation type fraction - Vegtype 1 = No Vegetation, 4= Broadleaf, 7 = Agriculture, 10 = Grassland, 15=Needleleaf, 18 = Boreal Grass and 19 = Shurb

	Variance Agriculture Land	Pays Mixed Coast	Needleleaf Forest	French Grassland	Broadleaf Forest	Boreal Alps
LAI [$m^2 m^{-2}$]						
DJF	0.128473	0.047496	0.122411	0.153101	2.322751e-01	1.068963e-01
MAM	1.182849	0.624901	0.288180	0.646476	4.527089e-01	1.482736e-01
JJA	1.008147	0.799979	0.228394	0.524248	3.770046e-01	3.192324e-01
SON	0.752909	0.294780	0.142568	0.549576	5.731153e-01	3.807104e-01
WG2 [$m^3 m^{-3}$]						
DJF	0.003979	0.003163	0.001366	0.004584	5.955570e-03	2.062045e-03
MAM	0.002170	0.002179	0.001201	0.001937	1.799300e-03	4.833214e-03
JJA	0.003097	0.002263	0.001957	0.003371	2.675631e-03	1.892820e-03
SON	0.002673	0.003263	0.002044	0.003129	1.866586e-03	3.214324e-03
GPP [t C $ha^{-1} y^{-1}$]						
DJF	0.000001	0.000002	0.000005	0.000002	6.414485e-07	1.617941e-07
MAM	0.000140	0.000063	0.000057	0.000064	6.764670e-05	2.237278e-05
JJA	0.000175	0.000085	0.000058	0.000091	8.155641e-05	4.111590e-05
SON	0.000031	0.000016	0.000029	0.000021	5.938111e-05	3.241018e-05
RE[t C $ha^{-1} y^{-1}$]						
DJF	0.000002	0.000003	0.000003	0.000003	2.387306e-06	6.774212e-07
MAM	0.000036	0.000016	0.000014	0.000018	1.793164e-05	6.063931e-06
JJA	0.000054	0.000025	0.000019	0.000024	2.350664e-05	1.173449e-05
SON	0.000012	0.000006	0.000014	0.000014	4.702735e-05	2.371312e-05

Table 4.5: Overview of the seasonal variance for each Cluster and each land surface variable. The green fields indicate the maximum variation per cluster type and variable

Figure 4.7 shows the time series plots from 2010 to 2019 for all GPI's within the Agriculture Land cluster. As for

all land surface variables the the individual lines aggregating around a mean value it can be considered a robust cluster. The rest of the time-series-cluster plots are presented in Appendix A.2. In general, the Agriculture, Pays Mixed Coast, Broadleaf Forest and French Grassland show low variance between the individual cluster GPI's. Needleleaf Forest and Broeal Alps, however show greater variability within the individual grid points. This might be caused by the relative high variation of vegetation type fractions within the clusters. The vegetation type fraction variation is displayed Appendix A.1.

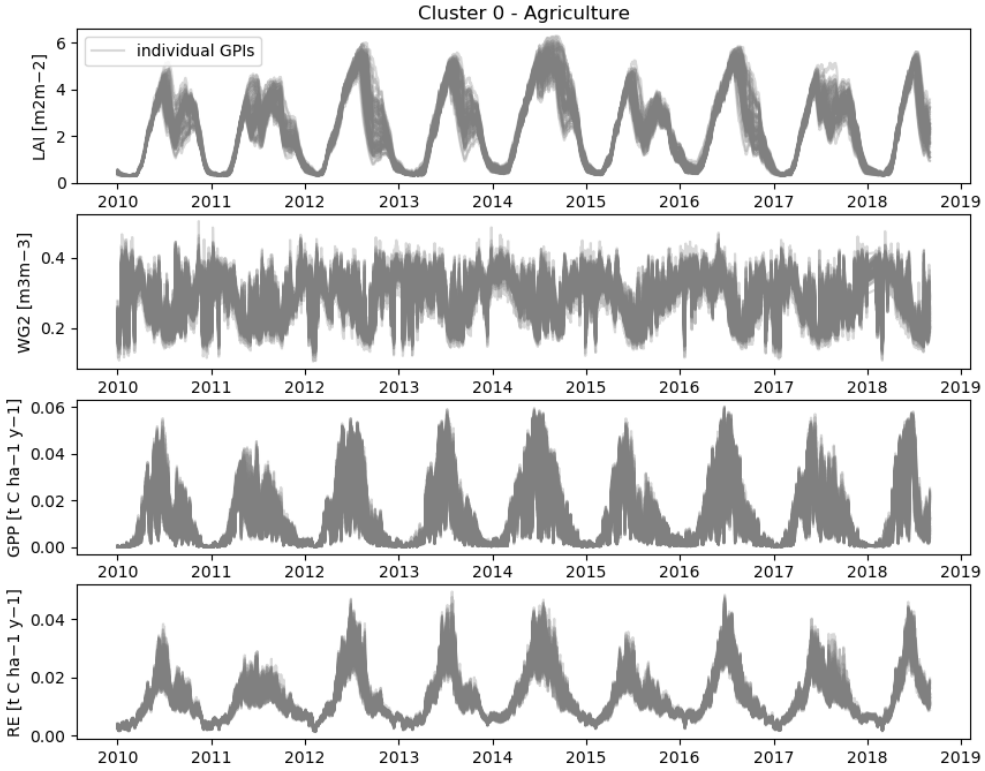

Figure 4.7: Time series plots from 2010 to 2019 showing the strength of the seasonality per land surface variable for all GPI's for cluster "Agriculture".

4.2.2. Variability of ASCAT observations

Figure 4.5 and Figure 4.9 show the annual variance per GPI for backscatter and slope/curvature respectively. For both figures a high value (green) indicates a stronger variation compared to a lower value (red). For all three ASCAT observations the highest variance can be found around the area of Paris as well as the north-east of England as well as the west area of Spain. All those territories are typically dominated by agriculture. Low variability for backscatter, slope and curvature mainly occur in the middle to north east of the study area as well as along the coast lines of France and Spain. Slope and Curvature show in general higher variability along the mountain ranges but lower variance in the north-west of France compared to the backscatter variance.

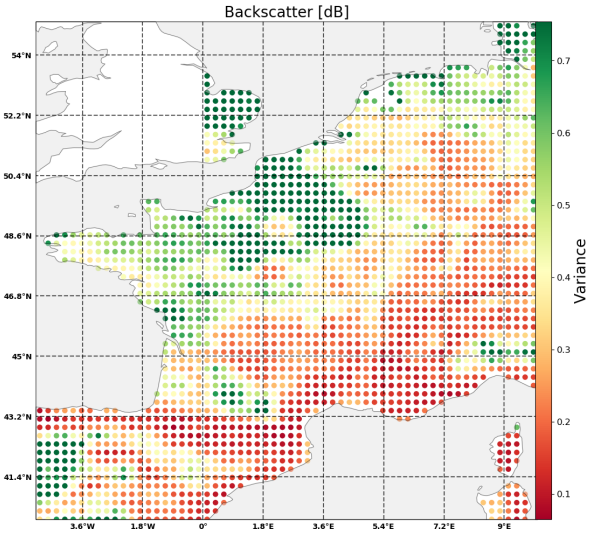

Figure 4.8: Annual variance per grid point for ASCAT backscatter

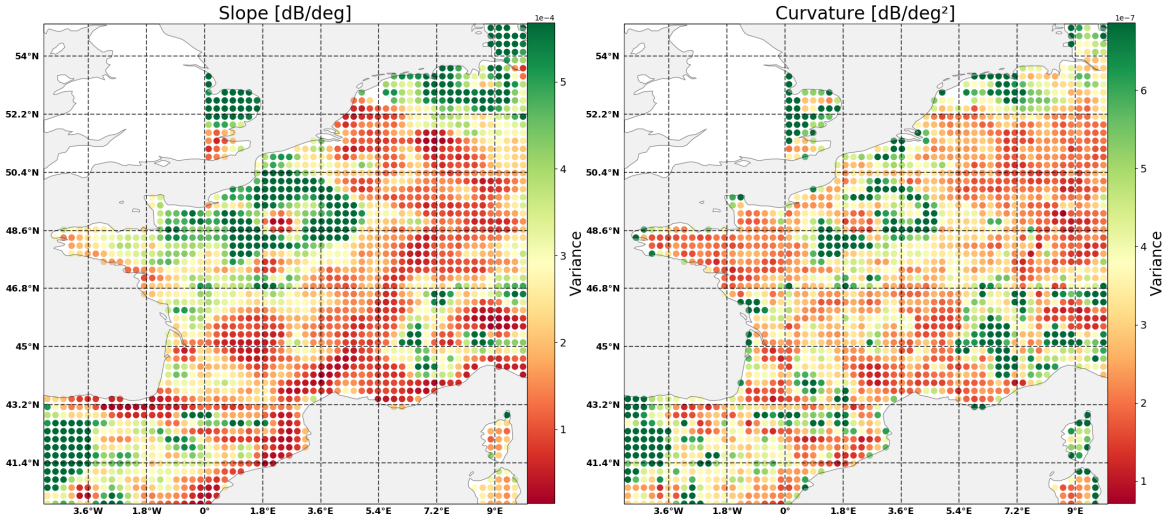

Figure 4.9: Annual variance per grid point for ASCAT slope (left) and curvature (right)

Table 4.6 gives a overview of the mean annual variance for each cluster. The over all results show, that backscatter, slope and curvature change independently from each other. This means, that a high variance of backscatter does not introduce a higher variance in slope or/and curvature. For example, the Boreal Alps show a low backscatter variance but a high variability in slope and curvature. Pays Mixed Coast shows high variance in backscatter and slope but low variance in its curvature values. The Agriculture Land has the highest variance for all three observation parameters. French Grassland generally shows low variance for backscatter, slope and curvature. The smaller annual changes occur for the Broadleaf Forest cluster, whereas Needleleaf Forest shows

intermediate variance for each ASCAT observation.

Cluster	Annual Variance Backscatter[dB]	Slope[dB/deg]	Curvature[dB/deg^2]
Agriculter Land	1.141762	0.000763	7.704335e-07
Pays Mixed Coast	0.521943	0.000303	2.119202e-07
Needleleaf Forest	0.380692	0.000257	3.770216e-07
French Grassland	0.288349	0.000168	2.380026e-07
Broadleaf Forest	0.212221	0.000131	2.522659e-07
Boreal Alps	0.165279	0.000296	5.358971e-07

Table 4.6: Overview of the mean annual variance for each Cluster and for each ASCAT observation

Figure 4.10 shows the seasonality per ASCAT observation for the GPI's with maximum vegetation type fractions - Vegtype 1 = No Vegetation, 4= Broadleaf, 7 = Agriculture, 10 = Needleleaf, 18 = Boreal Grass and 19 = Shurb.

Backscatter is a very noisy signal, which means it is influenced by short time changes but it also generally shows a seasonal trend. Grassland, broadleaf and boreal grass have weak seasonal cycle and low values compared to the other vegetation types. Agriculture on the opposite has a strong seasonal cycle with at the beginning and end of the year, which indicates the growing seasons of crops. The strong seasonality in for no vegetation might be caused by the on and off set of the rainy-season. This is a logical consequence considering that backscatter is highly related to soil moisture and therefore rain events.

Slope and Curvature are smoothed products from the ASCAT observation and are therefore less noisy compared to backscatter. The slope time series for all maximum vegetation types show a similar pattern as for the backscatter. Broadleaf and grassland have a weak seasonal cycle with a low magnitude. However, boreal grass increases in the seasonal magnitude compared to backscatter signal. Agriculture and no vegetation have the same seasonal period but the change over the season is stronger for agriculture. Curvature shows a very irregular seasonality for all vegetation types. The strongest seasonal signal has agriculture, whereas grassland and boreal grass show almost no seasonal trend with continuous negative values.

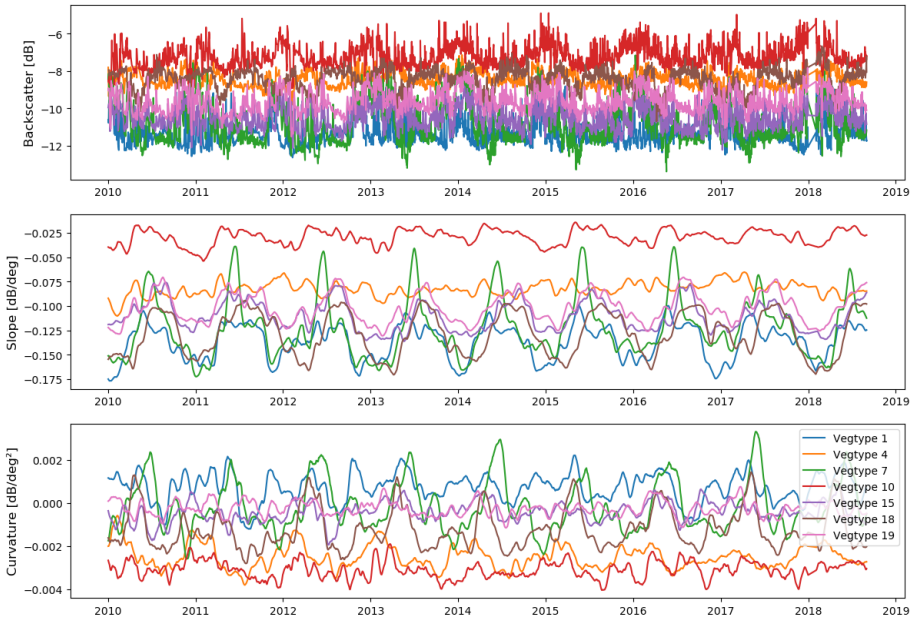

Figure 4.10: Time series plots from 2010 to 2019 showing the strength of the seasonality per land surface variable for the GPI's with maximum vegetation type fraction - Vegtype 1 = No Vegetation, 4= Broadleaf, 7 = Agriculture, 10 = Needleleaf, 18 = Boreal Grass and 19 = Shurb

The inter-annual variations of the ASCAT observations are illustrated in Table 4.7. Here, the seasonal variance is explained per ASCAT observation and cluster. The red fields highlight the maximum variance per cluster.

Backscatter is most variant for the DJF and SON seasons for all clusters and least variant for the JJA months. This is consistent with the strong correlation of backscatter with soil moisture and therefore precipitation. Slope hast the highest variance for the MAM season for all clusters except of Needleleaf Forest. This exception shows highest variance for the SON season, however, the variance is relative consistent between the MAM to SON months. The maximum curvature variance changes between all selected clusters, which indicates more complex correlations between the curvature values and land cover types.

	Variance Agriculture Land	Pays Mixed Coast	Needleleaf Forest	French Grassland	Broadleaf Forest	Boreal Alps
Backscatter [dB]						
DFJ	1.107605e+00	4.281081e-01	4.998181e-01	4.367187e-01	4.231366e-01	1.911949e-0
MAM	1.082654e+00	3.643294e-01	3.439811e-01	2.074273e-01	1.557071e-01	2.149425e-01
JJA	3.643974e-01	2.946439e-01	1.936831e-01	1.796802e-01	9.329256e-02	7.622117e-02
SON	5.905191e-01	4.870571e-01	3.514718e-01	2.286618e-01	1.429387e-01	1.057664e-01
Slope [dB/deg]						
DFJ	1.085524e-04	8.242424e-05	4.708940e-05	4.453916e-05	7.594821e-05	1.593713e-04
MAM	5.928639e-04	1.809378e-04	1.185051e-04	1.324275e-04	1.142876e-04	1.666115e-04
JJA	5.108965e-04	7.724725e-05	1.104768e-04	4.280282e-05	2.388228e-05	4.550347e-05
SON	1.005531e-04	1.547986e-04	1.801615e-04	6.446567e-05	4.040094e-05	8.545619e-05
Curvature [dB/deg^2]						
DFJ	3.543174e-07	1.575744e-07	1.850961e-07	1.459963e-07	2.040661e-07	1.786705e-07
MAM	6.655927e-07	1.384113e-07	2.337702e-07	1.387283e-07	1.751338e-07	3.493319e-07
JJA	3.240100e-07	1.712506e-07	2.618222e-07	1.275089e-07	9.897072e-08	1.734987e-07
SON	2.437597e-07	2.125544e-07	2.533081e-07	1.686494e-07	1.296245e-07	1.830740e-07

Table 4.7: Overview of the seasonal variance for each Cluster and each ASCAT observation.

Figure 4.11 shows the time series plots from 2010 to 2019 for all GPI's within the Agriculture Land cluster 0. The seasonality as well as the magnitude are relatively consistent for all ASCAT observations. Only backscatter shows greater variability, which might be due to local rain events. The rest of the time-series-cluster plots are presented in Appendix A.2. In general, the Agriculture, Pays Mixed Coast and French Grassland show low variance between the individual cluster GPI's. Needleleaf Forest, Broadleaf Forest and Broeal Alps, however show greater variability for the backscatter, slope and curvature measurements. Especially for the backscatter values a high variability occurs. This might have the same reason as for the variability of the land surface parameter, caused by highly variant vegetation type fractions but also very local precipitation events.

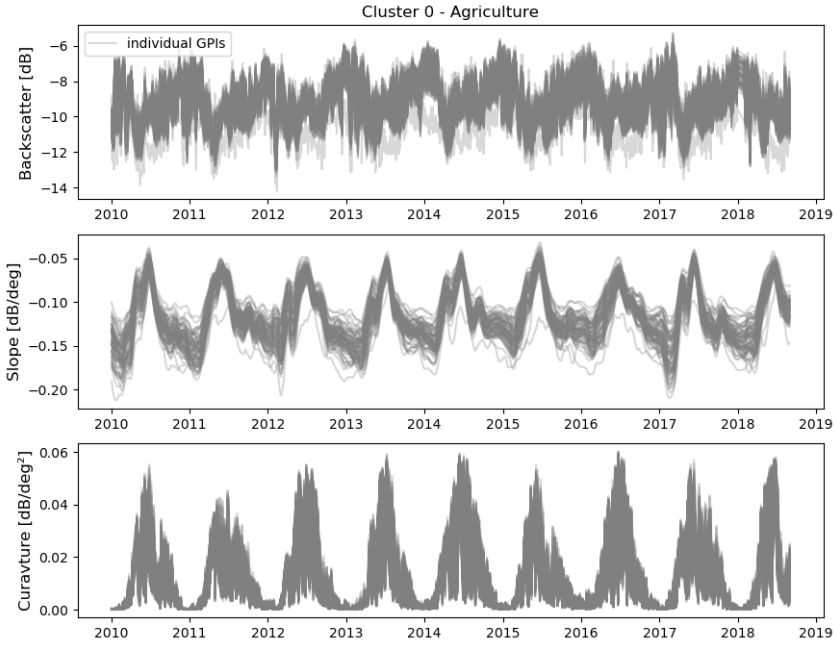

Figure 4.11: Time series plots from 2010 to 2019 showing the strength of the seasonality per ASCAT Observation for all GPI's for cluster "Agriculture".

4.3. Backscatter Evaluation

This section evaluates the backscatter simulations with following structure: 1) Spatial and Seasonal Model Performance, 2) Model Performance regarding Vegetation types, 3) Cross-Validation, 4) Spatial and Seasonal Feature Importance and 5) Feature Importance regarding Vegetation type. The evaluation of the results is followed by a discussion and summary of results, which uses the data variability and land cover types to explain the differences in model performance and feature importance. This will help to finally answer a part of the research question if machine learning can reliable estimate backscatter from ISBA states.

4.3.1. Spatial and seasonal model performance

Figure 4.12 shows RMSE, MASE, Spearman correlation coefficient and Bias for each GPI. Here only the the testing dataset (year 2017) is evaluated. The description normalized is used because, the normalized output of the DNN to is used to calculate the different metrics. This means that the simulated data as well as the observed data is scaled between 0 and 1, which then is further used to calculate the performance metrics. The RMSE and MASE give a indication how well the model performing by calculating the error. However, MASE is normalized by the mean absolute error calculated using native forecasting and is useful to confirm the results achieved with the RMSE but also to compare it with the results of the slope and curvature simulations. RMSE and MASE show the same spatial pattern, which indicates that the RMSE is a good performance metric. For further interpretation the RMSE is used to analyse the performance as it is easier to create interpretation compared to MASE values. The highest error is found at the area of the Alps, agriculture area around Paris and the Brittany. However, the magnitude of error is very similar throughout the study domain, ranging between 0.075 to 0.105 of the normalized RMSE.

The Spearman correlation coefficient shows a high positive correlation between prediction and observation for almost the whole study domain except of the area of Alps, Pyrenees, mountain areas at Corsica and Sardinia and coastlines of Spain. The bias expresses how much the prediction over- (positive) or undershoots (negative) the truth value. In the center of France as well as the Alps underestimates the truth whereas the north of the study domain generally overestimates the observation value.

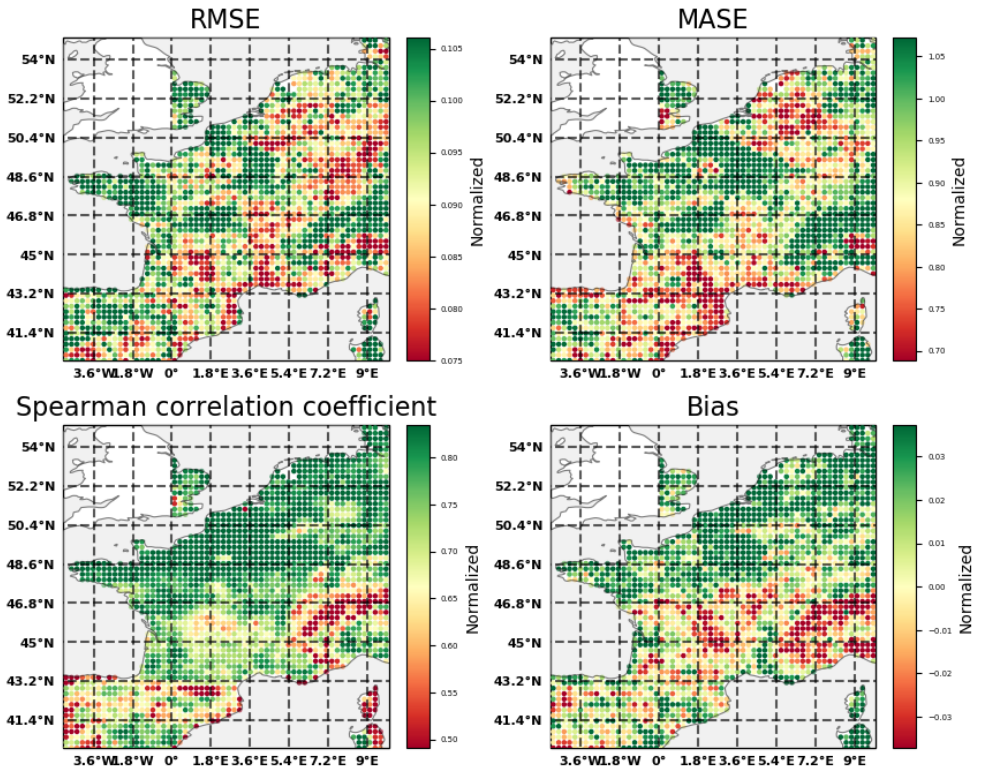

Figure 4.12: RMSE, MASE, Spearman Correlation Coefficient and Bias are calculated for each GPI within the study-domain using only the test year 2017. The description normalized is used because, the normalized output of the DNN to is used to calculate the different metrics.

Figure 4.12 shows the performance evaluated over the whole year. Figure 4.8 on the other hand illustrates the RMSE per season (DJF, MAM, JJA and SON). In the winter season between December and Feburary (DJF) the overall performance has the highest RMSE values compared to the other seasons. Here, the largest errors are found at the alpine areas. In the MAM and JJA season the performance along the mountain area improves, whereas the RMSE for the agriculture area around Paris stay at a similar magnitude as for the DJF season. SON is a transition period, at which the RMSE increases over the whole study domain.

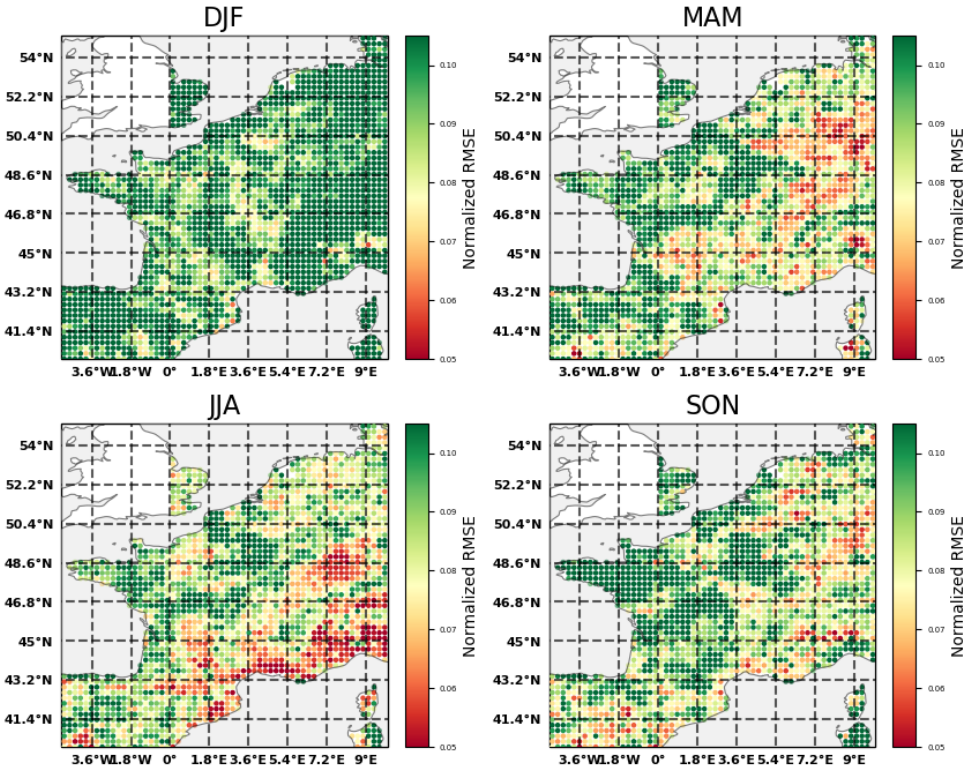

Figure 4.13: Backscatter: RMSE calculated for each Season: December-January-Feburary (DJF), March-April-May (MAM), June-July-August (JJA), September-October-November (SON). Normalized RMSE refers to the RMSE calculated from the min-max scaled output of the DNN and has therefore no unit.

Table 4.8 gives an overview how well the model is performing for each cluster per season. This illustrates that the overall performance for each cluster is very similar, but the seasonal RMSE diverge between the different vegetation type clusters. The boreal alps cluster performs worst during the DJF season and the best in the summer season JJA. This might be because of the presence of vegetation during the summer months. All the other clusters show the same seasonal behavior, however, the magnitude of the RMSE is smaller. In general, the overall seasonal performance is worst in the DJF season and best during the JJA season.

	Normalized RMSE				
Cluster	DJF	MAM	JJA	SON	Over-all
Agriculture Land	0.107142	0.104281	0.096865	0.102508	0.102699
Pays Mixed Coast	0.103270	0.090879	0.086450	0.104304	0.096226
Needleleaf Forest	0.119083	0.094123	0.087744	0.096965	0.099479
French Grassland	0.100489	0.086592	0.086457	0.098945	0.093121
Broadleaf Forest	0.112724	0.071787	0.073981	0.090501	0.087248
Boreal Alps	0.153388	0.088027	0.074565	0.085764	0.100436
Over-All	0.116016	0.089281	0.084344	0.096498	0.096534

Table 4.8: Backsatter: Table to compare the mean RMSE per cluster for Simulation 2017 for different seasons: December-January-Feburary (DJF), March-April-May (MAM), June-July-August (JJA), September-October-November (SON)

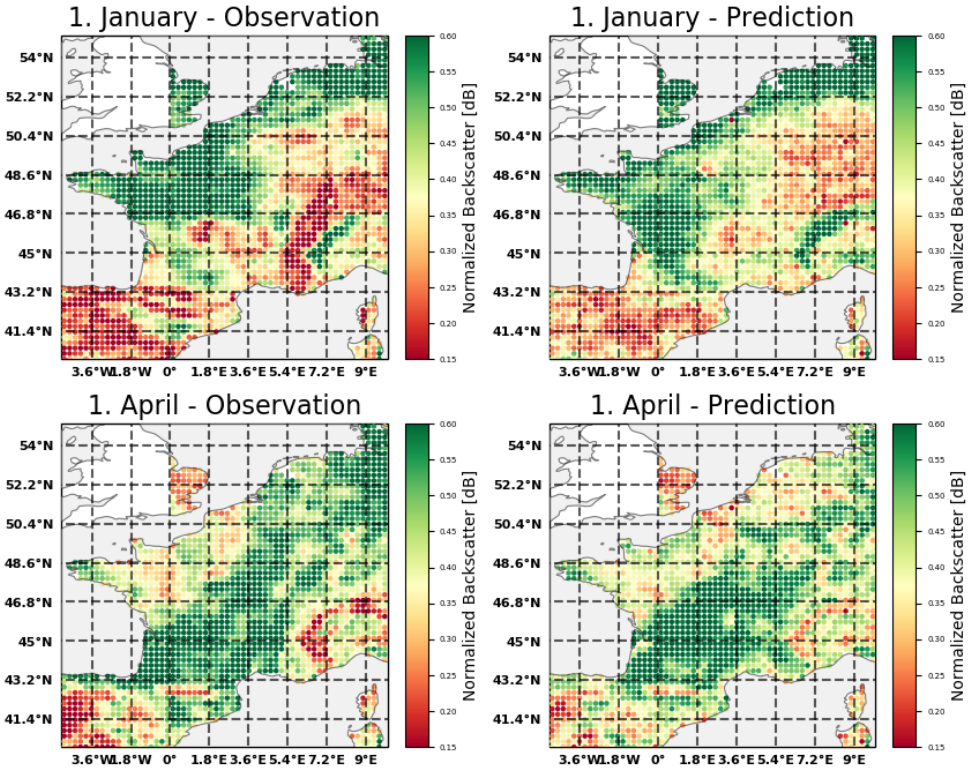

Figure 4.14: Backsatter: 1. January and 1.April observations for each GPI opposed to the prediction for the same date using test year 2017 to illustrate spatial consistency

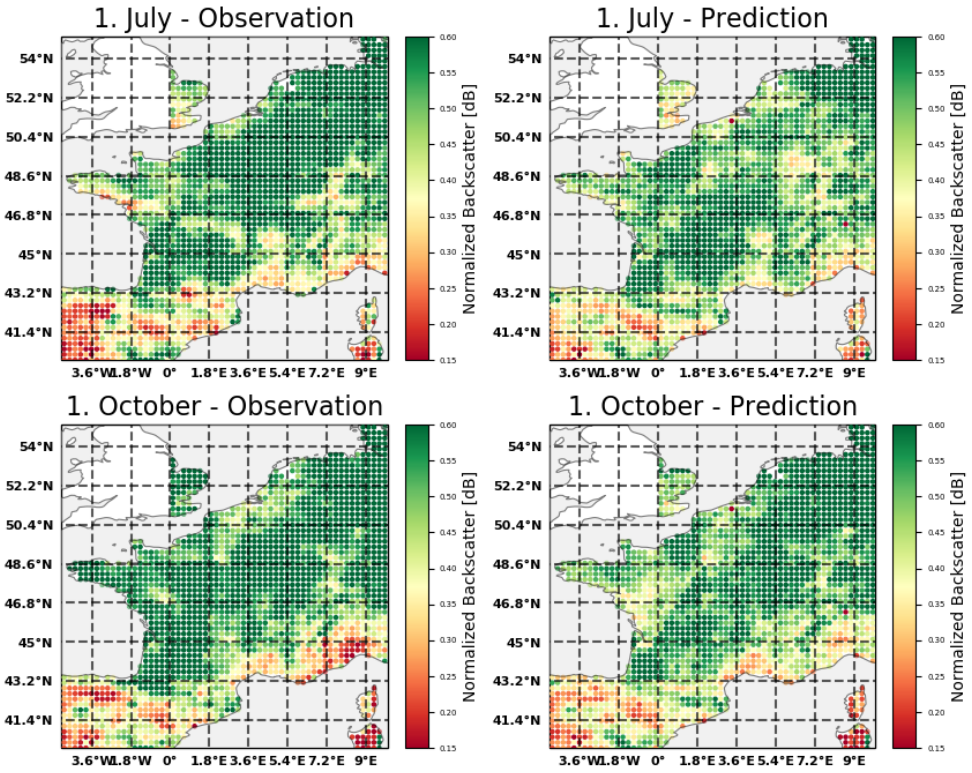

Figure 4.15: Backsatter: 1. July and 1.October observations for each GPI opposed to the prediction for the same date using test year 2017 to illustrate spatial consistency

Figure A.26 in the Appendix A.4 shows the RMSE values per GPI per cluster. This gives an indication of how robust a cluster is in its performance. A high variance of RMSE values indicates a bad robustness whereas a low variance explains a robust cluster. Agriculture land and French Grassland have similar variance with some outliers. Pays Mixed Coast, Needleleaf Forest and Broadleaf Forest are the most robust clusters by showing the lowest variance values. Boreal Alps on the other hand has the highest variance within the cluster. This shows that the t-SNE clustering was not successful to establish a robust cluster for the alpine area.

Figure 4.14 and Figure 4.15 show observed and predicted backscatter for four time steps: 1st January - 1st April and 1st July - 1st October respectively. All four observations/predictions dates show a similar pattern. In general, every prediction is able to capture the spatial distribution, however, it fails to predict the the magnitude of backscatter. This was already indicated by the Bias estimation at which the mountain areas are underestimating the backscatter magnitude.

4.3.2. Model performance regarding vegetation types

Figure 4.16 shows the capability of the model to predict time-series. Seven GPI's are selected to investigate the relationship between vegetation type and model performance. The time series prediction vs. observation for the original signal is presented in the Appendix A.4. Here the model is able to detect the seasonality as well as the general trends.

Figure A.27 allows to compare the observation and prediction time series more critical. The grey line represents the test set prediction, the green line the training set prediction and the red dotted line is the truth (observation).

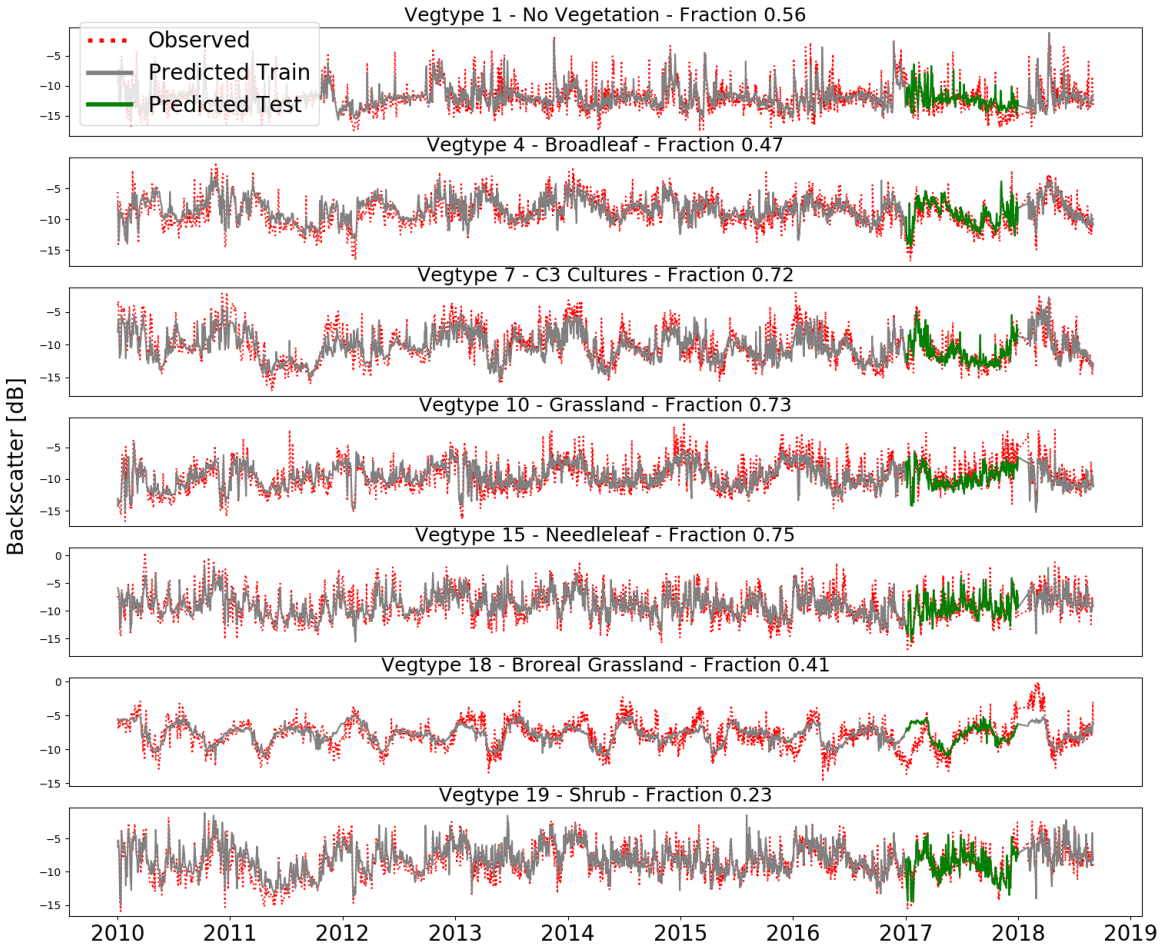

Figure 4.16: Backsatter: Observed (red) and predicted (grey) anomaly time series of backscatter for seven different vegetation types (Type: 1,4,7,10,15,18,19). The green line indicates the prediction using the test data-set.

Figure 4.16 shows that the time series for all tested vegetation types are generally well captured. Table 4.9 illustrates the RMSE for each time series prediction. No Vegetation, Agriculture and Needleleaf have the highest RMSE, whereas Shrub and Boradleaf show the smallest RMSE values. In general, backscatter is a very noisy signal with high frequent time variation - expressed in spikes. As the general trend is well predicted the model fails to predict every spike in the right magnitude. This coincides with the time-step analysis from Figure 4.14 and Figure 4.15, which shows that the model is not able to predict the correct backscatter magnitude. Moreover, the model performs equally well for training and testing data set and is therefore able to generalize.

	No Vegetation	Broadleaf	Agriculture	Grassland	Needleleaf	Boreal Grass	Shrub
RMSE [dB]	1.5922	1.3341	1.564	1.4116	1.5028	1.4482	1.2992

Table 4.9: Backsatter: Table to compare the RMSE per Vegetation type for the whole time series

4.3.3. Cross-Validation

The cross validation is evaluated by comparing three different models, namely Simulation 2012, Simulation 2014 and Simulation 2017. Simulation 2017 is the reference model. Simulation 2012 uses year 2012 as test year, which means all years except of 2012 are used to train the model. The same counts for Simulation 2014 and 2017 of which the years 2014 and 2017 are used to test the model respectively. Figure 4.17 shows a scatter plot with Simulation 2017 on the x-axis and the Simulation 2012 (grey) and 2014 (red) on the y-axis. This shows a clear linear relationship between RMSE of Simulation 2017 to the RMSE of Simulation 2012 and 2014. The Spearman correlation between Simulation 2017 and 2012 is 0.907 and between Simulation 2017 and 2014 it is 0,901. A spatial distribution of the RMSE values for each simulation is illustrated in Appendix A.4. Both illustrations show a clear and consistent correlation between the different simulations.

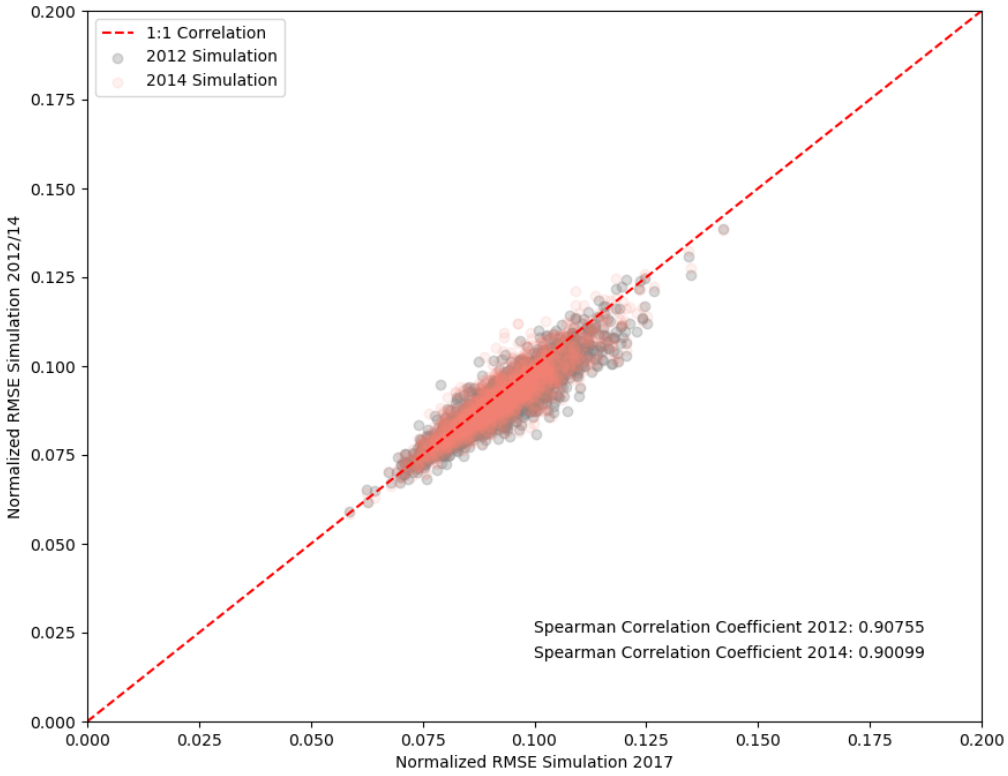

Figure 4.17: Backsatter: Scatter plot showing the linearity of the correlation of three different models using either 2012, 2014 or 2017 as test year. 2017 is used as a reference year, therefore the Spearman Correlation Coefficient is calculated between 2012-2017 and 2014-2017.

Table 4.10 shows the RMSE values for each cluster. This illustrates that the general RMSE is very consistent and varies for most clusters only in a magnitude range of +/- 0.03. Only the "Simulation 2012 - Cluster 12" has the highest error by diverging by 0.1 of the Normalized RMSE compared to the Simulation 2017 - Cluster 12. Furthermore, the linearity of RMSE values is less for higher RMSE values, whereas for lower RMSE values the linearity is higher and therefore more robust.

As a final evaluation Figure 4.18 shows the time-series predictions of Simulation 2012, 2014 and 2017 for seven vegetation types. This illustrates the strong similarity between those three simulations, but also shows that some of the short term variations differ in magnitude. The general performance of the three models show sufficient similarity (Spearman Correlation Coefficient >0.9).

	Normalized RMSE		
	Simulation 2012	Simulation 2014	Simulation 2017
Cluster 0	0.100899	0.098016	0.100986
Cluster 1	0.091938	0.093328	0.095772
Cluster 2	0.091417	0.092093	0.094238
Cluster 3	0.096972	0.097447	0.099178
Cluster 4	0.090921	0.091465	0.093795
Cluster 5	0.090392	0.091173	0.092520
Cluster 6	0.085969	0.086563	0.087085
Cluster 7	0.085645	0.085872	0.090710
Cluster 8	0.092121	0.097554	0.096522
Cluster 9	0.089431	0.090217	0.090873
Cluster 10	0.086490	0.086768	0.087976
Cluster 11	0.086059	0.086608	0.087350
Cluster 12	0.091564	0.099205	0.099380

Table 4.10: Backsatter: Table to compare the mean RMSE per cluster for Simulation 2012, 2014 and 2017. Here the whole time-series from 2010 to 2018.

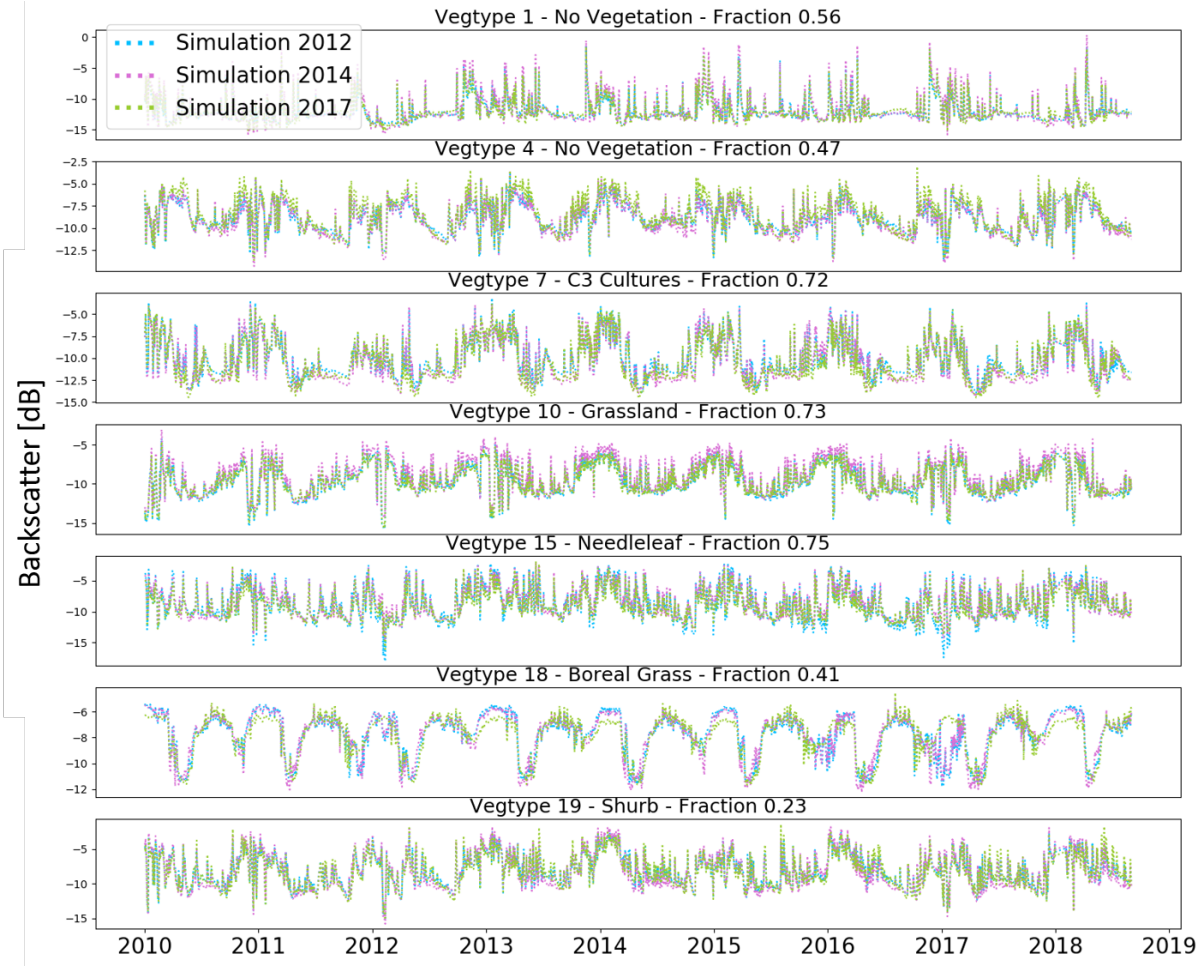

Figure 4.18: Backsatter: Time series prediction from each Simulation (2012 (blue), 2014(pink) and 2017(green) for seven dominant vegetation types (1 = No vegetation, 4 = Broadleaf, 7= C3-Cultures, 10 = Grassland, 15 = Needleleaf, 18 = Boreal Grassland, 19 = Shrub

4.3.4. Spatial and seasonal feature importance

For further investigation the models feature importance is evaluated. This gives information about which variables are used by the deep neural network to establish a relationship between the land surface parameters and ASCAT observation and explains which land-surface variables are relevant to simulate the variations in the ASCAT observable. For the feature analysis only the years between 2016 and 2018 are investigated due to its high computational effort.

Figure A.32, Figure 4.19, Figure 4.20 and Figure A.31 show the fraction of the total feature importance of LAI, WG2, GPP and RE respectively for the backscatter simulation. Every land surface variable is investigated per month. The highest spatial and temporal variation of the feature importance is found for WG2 and GPP. LAI and RE show generally a small feature importance fraction with almost no temporal variation and hence displayed in the Appendix A.4. LAI has only in some parts of the study domain higher feature importance fractions, e.g. in the Alps and Pyrenees. RE only shows significant importance at the western coast line of Spain and France as well as the high Alps at which a high percentage of "no vegetation" and "rock" are present.

Soil moisture (WG2) on the other hand has great importance throughout the center of the study domain but has low importance at the high mountain ranges, the agriculture land around Paris and along the northern coast line of Europe. The low importance for agriculture land is due to the high variation in the vegetation cycle compared to soil moisture. In general, the soil-moisture importance undergoes a very annual cycle. During the May, June and July months the importance is low for almost the whole study area, except of small areas in the south of France. From August to October the importance increases from the west of the study domain towards the east. During the November, December, January months the highest WG2 fractions occur in the mid, east part of the study domain, whereas in February and March the feature importance increases again in the west coast of the study area. This indicates a strong annual cycle of the WG2 feature importance by showing a clear spatial and temporal pattern.

The GPP has in general the highest fraction in the north of the study domain, especially the agriculture land is dominated by the GPP fraction. This is consistent with a small area in the east of Spain, which has an increased GPP feature importance, as it is also mainly defined by agriculture land fractions. Furthermore, GPP shows a strong seasonal importance cycle. Here, April and May as well November and December contain the highest feature importance fractions, which are also the periods when most of the changes in GPP occur. This follows a weaker influence of soil moisture on backscatter. All other months show low importance throughout the whole study domain. In summary, Figure 4.19 and Figure 4.20 clearly show that the feature importance is not arbitrary and has a strong spatial and temporal dependency.

Table 4.11 gives an overview of the importance per feature per season for each cluster. This summarizes the the spatial feature importance figures and indicates the most important features per season as well as area. For Agriculture Land and French Grassland the most important features are GPP and WG2 throughout the whole year, whereas Boreal Alps is dominated by WG2 and RE. Pays Mixed Coast is dominated by WG2 and GPP. This again, shows the small importance fraction of LAI and RE and clearly indicates WG2 and GPP as the most important source of information for the backscatter simulation.

	Fraction [-]															
	LAI				WG2				GPP				RE			
Cluster	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Agriculture	0.19	0.13	0.21	0.21	0.33	0.26	0.23	0.37	0.39	0.47	0.36	0.32	0.09	0.14	0.20	0.10
Pays Mixed Coast	0.05	0.06	0.08	0.07	0.38	0.34	0.39	0.48	0.34	0.39	0.25	0.27	0.23	0.22	0.28	0.19
Needleleaf Forest	0.08	0.08	0.11	0.10	0.45	0.51	0.34	0.50	0.20	0.22	0.20	0.15	0.27	0.20	0.35	0.25
French Grassland	0.08	0.08	0.12	0.09	0.58	0.52	0.39	0.59	0.24	0.27	0.20	0.19	0.10	0.12	0.29	0.12
Broadleaf Forest	0.10	0.12	0.15	0.16	0.61	0.50	0.36	0.45	0.18	0.24	0.24	0.22	0.11	0.14	0.25	0.17
Boreal Alps	0.21	0.27	0.19	0.24	0.30	0.28	0.27	0.32	0.20	0.18	0.25	0.18	0.30	0.28	0.29	0.27

Table 4.11: Table to compare the mean feature importance fraction per cluster for different seasons. Cluster 0 = Agriculture Land, Cluster 5 = Pays Mixed Coast, Cluster 8 = Needleleaf Forest, Cluster 9 = French Grassland, Cluster 11 = Broadleaf Forest and Cluster 12 = Boreal Alps

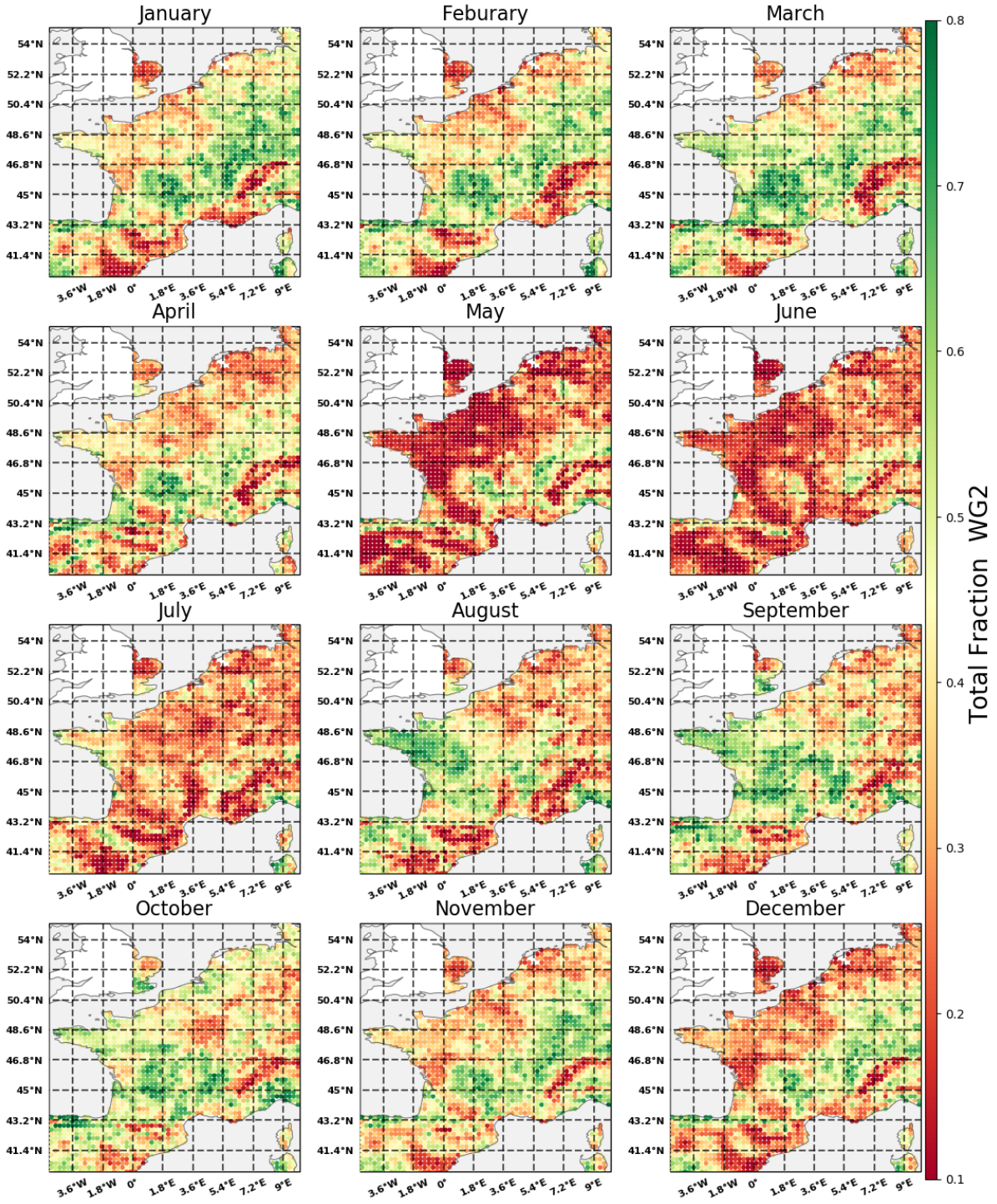

Figure 4.19: Backsatter: WG2 feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

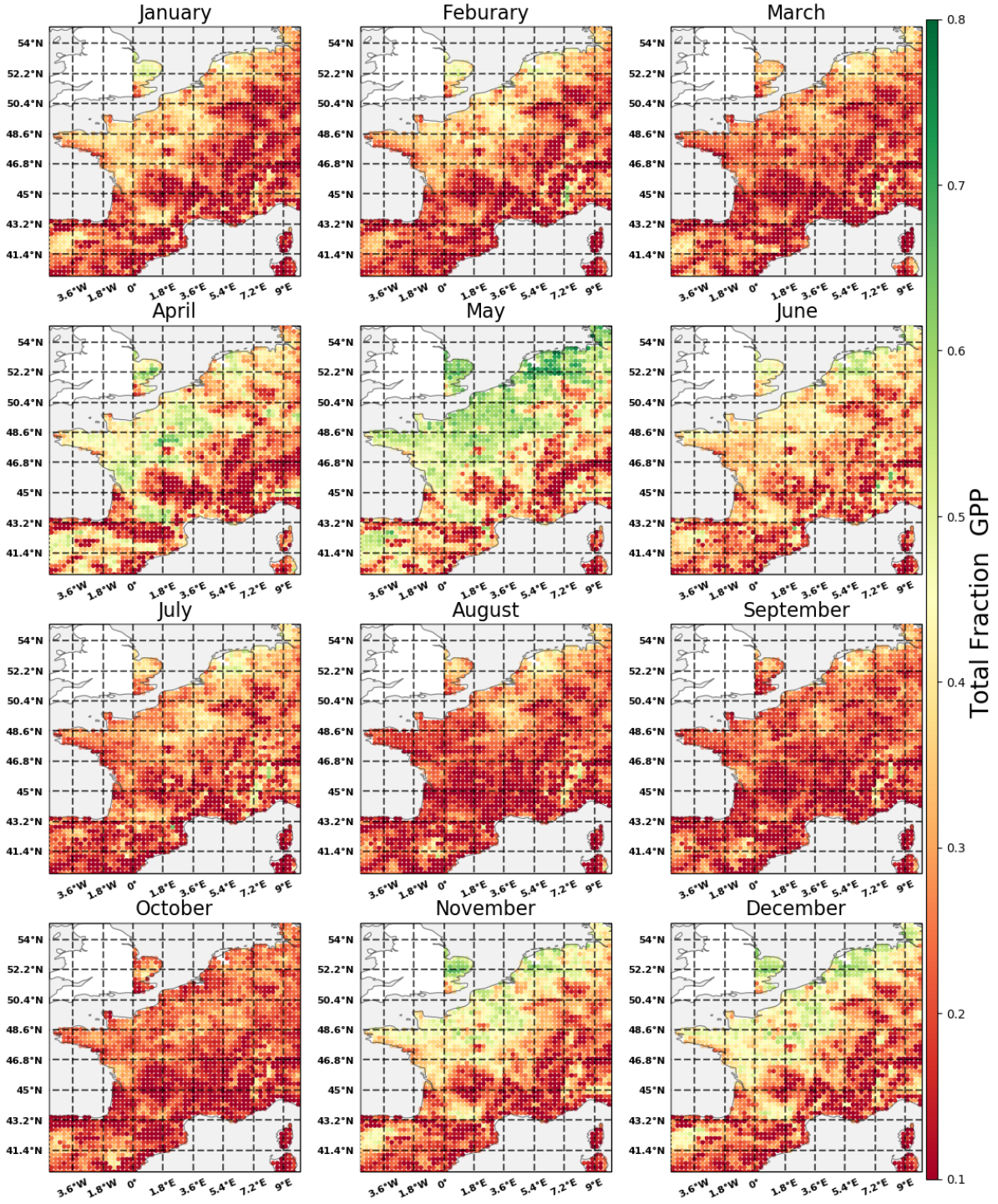

Figure 4.20: Backsatter: GPP feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

For further analysis the feature importance time series are plotted against the actual input variables for each cluster. This allows to see the relationship between the land surface raw input and feature importance. Figure 4.21 shows an example for Cluster 0 (Agriculture Land). The other cluster time series plots are provided in the Appendix A.4.

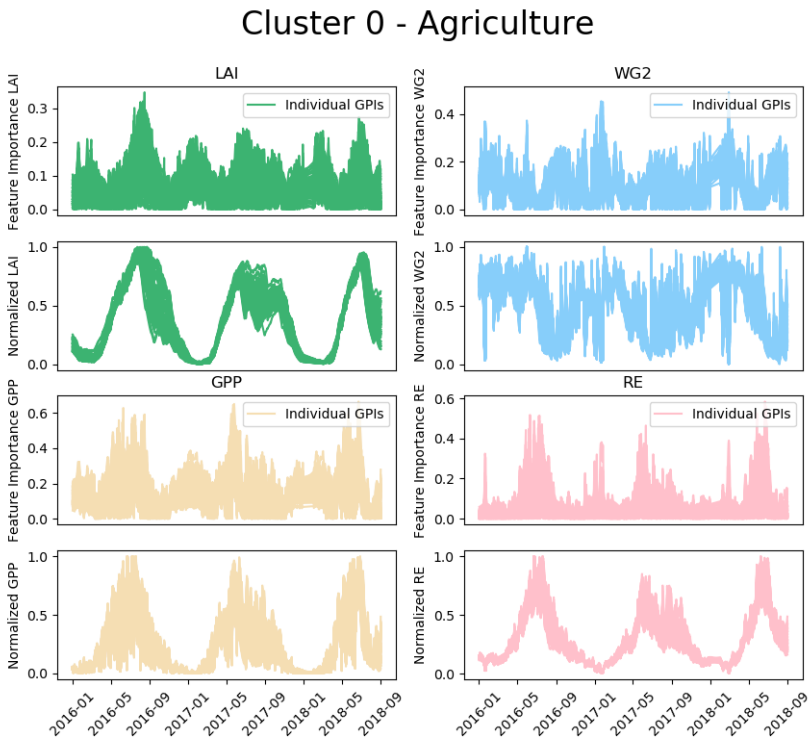

Figure 4.21: Backsatter: Feature importance and land surface parameter time series for the cluster 0 - Agriculture Land

In general, the feature importance varies between each GPI, however, all of them share the same timing when the maximum magnitudes occur. The variations between GPIs are displayed by the spread of lines within the time series. The peaks in the feature importance occur once the land surface variables reach their minimum or maximum values. This means that the feature importance has in general four peaks depending on the change of the input.

This pattern is consistent for each cluster. Still, the maximum feature values are dependent on the over-all importance assigned from the model. For example, the feature importance magnitude of LAI for cluster 5 (Pays Mixed Coast) is very low even-though the input seasonality is strong. This is because the over all importance of LAI for cluster 5 is low. On the other hand cluster 5 shows great dependency on soil moisture (WG2) and therefore has high magnitudes for the soil moisture feature importance. This means the feature importance reflects input seasonality once it is relevant for the prediction for a certain GPI. This counts for all analyzed clusters.

4.3.5. Feature importance regarding vegetation types

The time-series plots per cluster illustrate how the features change importance throughout out the year. However, it does not illustrate which features are important for a specific vegetation type and how those relate to each other. Figure 4.22 shows the feature importance for seven vegetation types per input feature. In year 2018 a straight line can be observed in the time series, because of a data gap in the input data.

The graph shows that LAI has a strong seasonal influence on boreal grassland, no vegetation and needleleaf. Here, all three signals show the highest magnitude during the inter-annual peaks. The seasonal trend of soil moisture has the same timing for all vegetation types, however the magnitudes are different. It also shows several short term peaks throughout the time series, especially for the land cover No Vegetation. The GPP feature

has its strongest influence on the agriculture vegetation type. It reflects a strong seasonality with the highest magnitude at the MAM season. The other land covers follow the same seasonal trend but with a lower magnitude. Respiration is most relevant for needleleaf, agriculture, shurb and boreal grass. Broadleaf, grassland and no vegetation extract very little information from respiration.

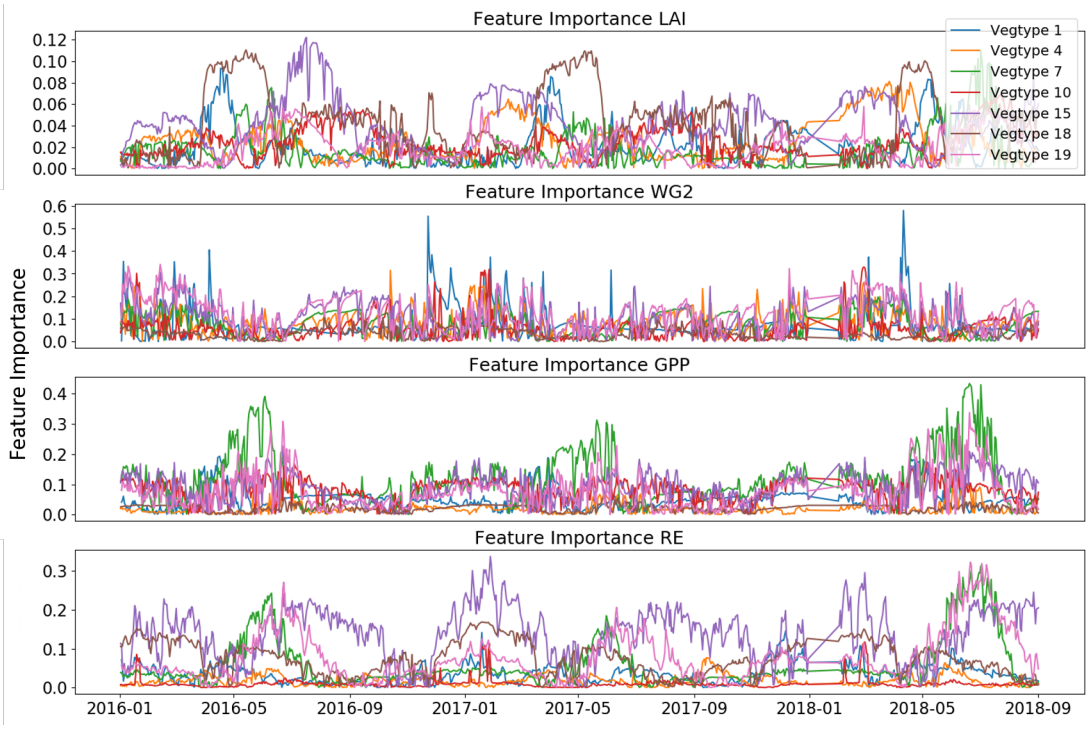

Figure 4.22: Backsatter: Feature importance time series estimations for each feature (LAI, WG2, GPP and RE) for seven vegetation types: Vegtype 1 = No Vegetation, Vegtype 4 = Broadleaf, Vegtype 7 = Agriculture, Vegtype 10 = Grassland, Vegtype 15 = Needleleaf, Vegtype 18 = Boreal Grass and Vegtype 19 = Shrub

The following table 4.12 gives a quick summary of the overall feature importance for the specific vegetation types, by taking the mean over the whole time series. The green high lighted field are representing the main contributors. No Vegetation, Broadleaf and Shrub are dominated by soil moisture, whereas Agriculture and Grassland is explained by gross primary production and Needleleaf and Boreal Grass by respiration.

	Feature Importance Fraction			
Vegetation Type	LAI	WG2	GPP	RE
No Vegetation	0.12	0.40	0.28	0.20
Broadleaf	0.18	0.54	0.13	0.15
Agriculture	0.06	0.33	0.41	0.20
Grassland	0.16	0.30	0.49	0.05
Needleleaf	0.11	0.28	0.24	0.36
Boreal Grass	0.26	0.21	0.12	0.41
Shurb	0.07	0.44	0.25	0.24

Table 4.12: Backsatter: Table to compare the mean feature importance fraction for seven vegetation types. Considering the time frame from 2016 to 2018.

4.3.6. Summary and discussion of backscatter results

The following discussion is based on two research sub-questions, mentioned in Chapter 1, and covers the results of the backscatter analysis. The results are summarized by following Figure A.26, Figure 4.24 and table 4.13. Figure 4.23 shows the summary of results for the performance analysis, by comparing the min-max scaled variance of the input and output variance as well as the RMSE values of the backscatter simulation for every month. The displayed variance per month is calculated by taking the mean of all monthly variances over the 10 years of data. The monthly RMSE values are calculated in the same manner. Figure 4.24 illustrates the total feature importance fractions per cluster between 2016 and 2018. The colored lines represent the vegetation parameters, whereas the red line represent the RMSE backscatter and the black dashed line the variance of backscatter. The dots in both figure 4.23 indicate the individual values, as the lines don't represent continuous data.

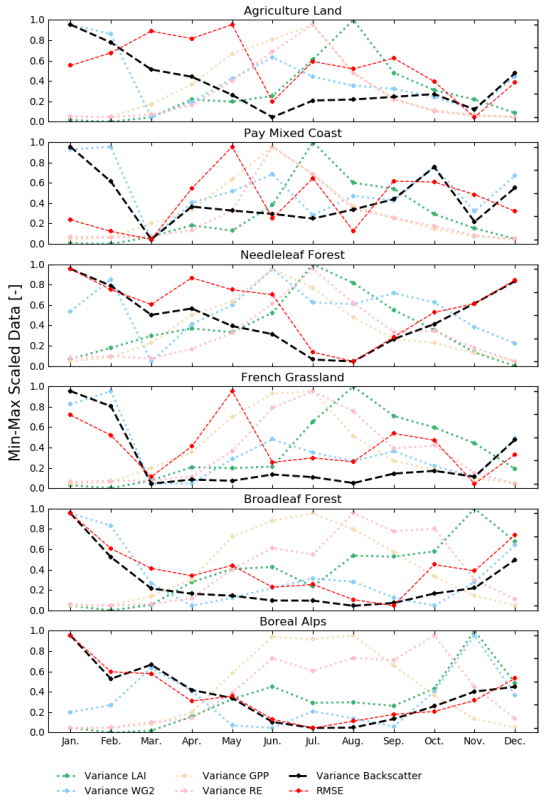

Figure 4.23: Comparison of min-max scaled variance for all input features (LAI, WG2, GPP, RE) as well as backscatter and RMSE for each month and distinctive cluster. The dashed balck line represents the RMSE calculated from the prediction vs observation, the grey dashed line represents the variance of curvature, the green the variance of LAI, the blue the variance of WG2, the yellow the variance of GPP and the pink line the variance of RE.

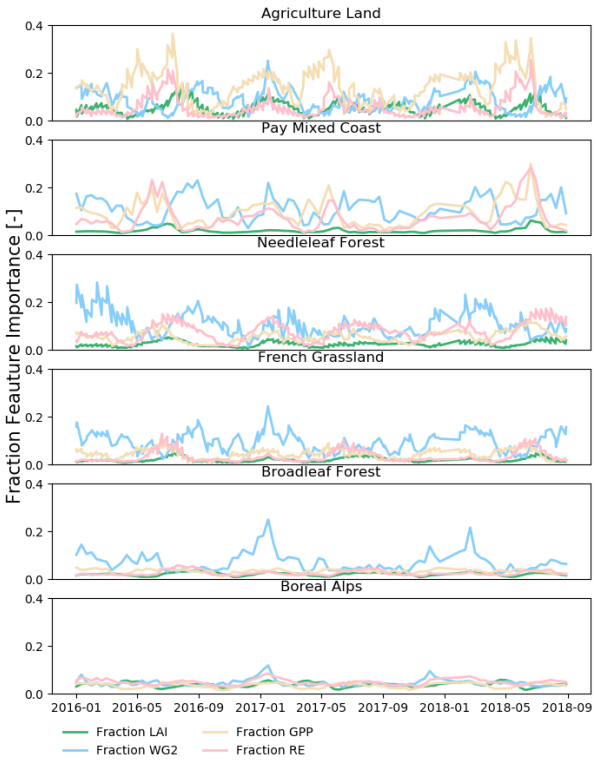

Figure 4.24: The fraction of the total feature importance for all input features (LAI, WG2, GPP, RE) between 2016 and 2018 for all distinctive clusters. The green line represents the feature importance fraction of LAI, the blue the importance of WG2, the yellow tthe importance of GPP and the pink line the importance of RE.

Cluster	Normalized RMSE	Most relevant Feature	No Vegetation	Broadleaf	Agriculture	Grassland	Needleleaf	Boreal Grass	Shurb
Agriculture Land	0.102699	GPP	0.09	0.14	0,59	0.10	0.02	0	0
Pays Mixed Coast	0.096226	WG2	0.08	0.06	0,23	0.30	0.04	0	0.03
Needleleaf Forest	0.099479	WG2	0.09	0.08	0.06	0.30	0.29	0.0	0.02
French Grassland	0.093121	WG2	0.05	0.16	0,12	0.55	0.07	0	0.01
Broadleaf Forest	0.087248	WG2	0.05	0.31	0.08	0.32	0.13	0.0	0.0
Boreal Alps	0.100436	WG2	0.10	0,12	0.03	0.21	0.02	0.16	0.01
RMSE [σ]			1.5922	1.3341	1.564	1.4116	1.5028	1.4482	1.2992
Most relevant Features			WG2/GPP	WG2/LAI	GPP/WG2	GPP/WG2	RE/WG2	RE/WG2	WG2/GPP

Table 4.13: Backsatter: Table to illustrate the over all normalized RMSE and most important feature per cluster

Table 4.13 gives an overview of the over all model performance and most important feature per cluster as well as the feature importance and RMSE for the seven most distinct vegetation types.

How does the model performance vary in space and time?

Figure 4.23 and table 4.13 show a clear correlation between RMSE values and land cover type. The worst performance was found or Agriculture Land and Boreal Alps. The reason for a worse performance might be that exactly those two areas have highest seasonal variance of the input variables, see Figure 4.6 and table 4.5. Agriculture Land has almost double as much variance for all vegetation states compared to the other clusters. Boreal Alps shows the highest variance for soil moisture. Additionally, a big part of the Alps is dominated by rock and no vegetation tiles and are therefore hard to represent by vegetation inputs. The high RMSE values between October and February of the Boreal Alps cluster can be explained by the increased snow land cover fraction, as snow is not considered in the LSM, however is observed by the satellite. This increased complexity makes it hard for the model to capture all variations, which leads to high errors. This behavior is consistent throughout the individual cluster points, which confirms the robustness of this assumption.

Agriculture crops experience fast changes in its vegetation structure due to the accelerated growing and harvesting seasons. This leads to abrupt changes in the surface structure from crops to almost no vegetation or grassland. Therefore, the land cover type is changing, however, the land surface model is not adopting to this change, which means all variables are still computed with the assumption that harvesting did not occur. This might additionally lead to an increased error for agriculture based GPIs.

Table 4.13 shows the RMSE values per vegetation type. Here, no vegetation and agriculture land cover type show the highest RMSE followed by needleleaf. This is consistent with the results of the cluster type Agriculture Land and Boreal Alps as these cluster also have the highest RMSE values. Further evidence, that the vegetation type is correlated with the magnitude of the RMSE values is the consistency between the broadleaf and grassland clusters and land covers. Both, land covers and clusters, show the lowest RMSE values. This superior performance, can be explained by the fact that for the broadleaf and grassland the backscatter signal has a clear seasonal cycle with low inter annual variations, see Figure 4.10.

The perfromance of the DNN is worse in the Needleleaf Forest than in the Boreal Forest and French Grassland. A reason for this could be that in that area a cyclone destroyed a big part of the forest, which caused a switch from a forest land cover to bare soil and grassland. This change was not implemented in the LSM, which again causes high derivations in the output of the LSM as the satellite observes these changes. Therefore, the model might have been trained partly with false data, as the model did not adapt its land tiles, which leads to a mismatch of the LSM inputs and satellite lables in the DNN. Additionally, the neeleleaf vegetation type shows a high inter seasonal variation in the backscatter data compared to broadleaf, which leads to higher RMSE values. This is displayed in Figure 4.10. The relative high RMSE values for Pays Mixed Coast might be because the cluster has a high variation in land cover types but also high inter seasonal variation as it contains large agriculture fractions, see table 4.13.

Figure 4.23 compares the change of the input and output variance with the RMSE values per month. The Spearman correlation matrix shows the monotonic ranks between each of the changes, see Figure A.39. Both figures clearly display and quantify the strong correlation between variance in the backscatter data and RMSE values. This counts especially for the Agriculture Land, French Grassland, Needleleaf Forest, Broadleaf Forest and Boreal Alps clusters. That suggests that the simulation error is mainly dependent on the variance in the output data.

For French Grassland and Broadleaf Forest the variance in backscatter additionally shows the same temporal development as soil moisture. This indicates a relationship between backscatter and soil moisture and suggests that the variance of backscatter is dominated by soil moisture. That strong correlation is consistent with previous research in which the backscatter/soil moisture dependency is used to directly retrieve soil moisture form backscatter observations [53] [52] [47]. Still, the strong correlation between soil moisture and RMSE values is only present for clusters with a high WG2 feature importance. Clusters with a complex combination of different feature importance show a diverging RMSE pattern compared to the soil moisture trend.

Agriculture Land shows a complex combination of feature importance, which then causes a low correlation

between the variance of the input - and output data and RMSE values. The RMSE values of Pay Mixed Coast follow a very similar behavior as the variance in soil moisture in the beginning and end of the year, but diverge during the period between April and June. This might be caused by the sudden increase of the GPP and RE feature importance. The same pattern can be observed for French Grassland. Here the highest divergence between the soil moisture variance and RMSE occurs within the months at which GPP and RE are most important. This strongly suggests that the model error is highly dependent on the variance of the most important feature.

The Boreal Alps on the other hand shows a very different behavior. Here the variance of the performance almost perfectly correlates with the variance of backscatter. This means that the deep learning model is adapting the internal parameters solely according to the variance in the backscatter signal. That also reflects in the feature importance, as none of the input variables are dominantly relevant for the DNN. This is consistent with the fact that the alps have a big part of plain rock and no vegetation and snow in the land tiles. Therefore, the input variables can not reflect the properties of the surface as the input is soil and vegetation related. This again shows that the results of machine learning algorithms need to be treated carefully, as these try to extract patterns and information out of any given data- even the data has no logical or physical correlation.

The seasonal decomposition of the observed and predicted signal shows that the anomaly of the signal is relatively well captured as the seasonality is almost perfectly simulated. It only fails to predict some daily variations in the anomalies. Especially, the capturing of the correct magnitude fails in many cases. This might be because the variation in land cover types as well as the variance in the input and output data differs a lot for each cluster and even GPI. This difference in variance might even be magnified as the normalization and scaling of the input data is done over the whole study domain. This might lead to false estimations as some of the signals show very weak seasonality compared to others, which suppresses the input signal for weak signals drastically. Therefore, a more flexible model structure is needed, which increases or decreases in complexity depending on the complexity of the GPI but also normalizes the data for each GPI.

How does the feature importance vary in space and time?

Figure 4.24 and table 4.13 illustrate that the magnitude of importance is dependent on the land cover types. Agriculture Land is highly dependent on the gross primary production and soil moisture. Especially, during the April/May and November/December months GPP has significant influence on the prediction, whereas during the other months soil moisture plays a big role. This strong seasonal importance might be caused by the growing and harvesting seasons of the cultures. These processes cause high variation in the GPP data, which leads to a strong feature importance.

The dependency of GPP and backscatter is confirmed by the time series analyse of the purest agriculture GPI, see 4.13. This relationship between microwave observations and the gross primary production was already established by previous studies [25][46]. Here the retrieved vegetation optical depth and soil moisture are used to estimate GPP. GPP accounts for the fresh biomass production and therefore phenological development.

Needleleaf Forest, French Grassland and Broadleaf Forest are clusters, which are highly dominated by soil moisture. This could be because the vegetation cover is relatively constant during the year compared to agriculture, which gives the soil moisture a higher importance. Hence the vegetation changes are smaller compared to the soil moisture fluctuations. Table 4.5 shows that French Grassland has the highest soil moisture variance compared to the other clusters. The same counts for Needleleaf Forest at which GPP and WG2 is most variant compared to the other vegetation states.

Pay Mixed Coast is a mixture of agriculture and grassland and neatly demonstrates a transition state between the Agriculture Land cluster and French Grassland cluster. Figure 4.24 shows that the signal for April/May and November/December is explained by the GPP and for the rest of the year soil moisture is most relevant for the prediction. This again shows the sensitivity of backscatter for the growing and harvesting season but also soil moisture variations.

For the Boreal Alps cluster soil moisture has the overall highest importance. However, all features share a similar relevance, which means that none of signals can provide a clear correlation to the observed backscatter signal. This might be because the Boreal Alps cluster has a high variation in its land cover types but also that

the input variables contain too little information about the actual backscatter mechanism. Soil moisture still has the highest importance, which is consistent as the input variables for no vegetation and boreal grass show the highest variations for the WG2 signal, see Figure 4.6.

In summary, the seasonality of the feature importance reflects the variation in the input data, the magnitude of importance depends on the land cover type and the main feature to simulate backscatter is soil moisture. This is a logic consequence as soil moisture is highly influences the dielectric properties and structure of the observation medium and therefore backscatter signal [51]. The results show that the backscatter signal is additionally sensitive to the change in phenology expressed in GPP, especially in areas with a strong seasonal cycle in the GPP input data, such as agriculture areas.

4.4. Slope Evaluation

This section is divided into five parts 1) Spatial and Seasonal Model Performance, 2) Model Performance regarding Vegetation types, 3) Cross-Validation, 4) Spatial and Seasonal Feature Importance and 5) Feature Importance regarding Vegetation type. This allows a detailed description of the spatial and temporal variation of performance and feature importance, which is discussed in the summary and discussion subsection. This will help to finally answer a part of the research question if machine learning can reliable estimate slope from ISBA states.

4.4.1. Spatial and seasonal model performance

Figure 4.26 shows RMSE, MASE, Spearman correlation coefficient and Bias for each GPI. Here, only the testing dataset (year 2017) is evaluated. The RMSE and MASE give an indication how well the model performing by calculating the error. The description normalized is used because, the normalized output of the DNN to is used to calculate the different metrics. RMSE and MASE show the same spatial pattern, whereas MASE is a scaled error comparing the MAE to the MAE calculated using native forecasting. This shows that the native forecasting method performance better over the whole study domain compared to the deep neural network as the values are larger than one. In general, the RMSE and MASE are very noisy throughout the whole study area, by showing only a few spatial clusters. For further interpretation the RMSE, Spearman Correlation Coefficient and Bias are used to analyse the model performance.

The highest errors are found around the area of Paris as well as the south east of France. South east of France contains a big part of the Alps but also the coastal area. Still, both named areas are relatively noisy in its RMSE distribution. The Spearman Correlation Coefficient is over-all very high ranging up to 0.9 for most areas. A lower correlation is found along the France mid-sea coast as swell as the north coast of Spain/Portugal. Moreover, big cities such as Paris, Valencia, Barcelona, Turin, Bordeaux, London and Rotterdam show also a low correlation coefficient. The bias results show that the agriculture land around Paris as well as the Normandy and Brittany are generally underestimating the slope values. Overestimation occurs at the region of the alps as well Spain and the French region Pays de la Laire, which is located below the Brittany.

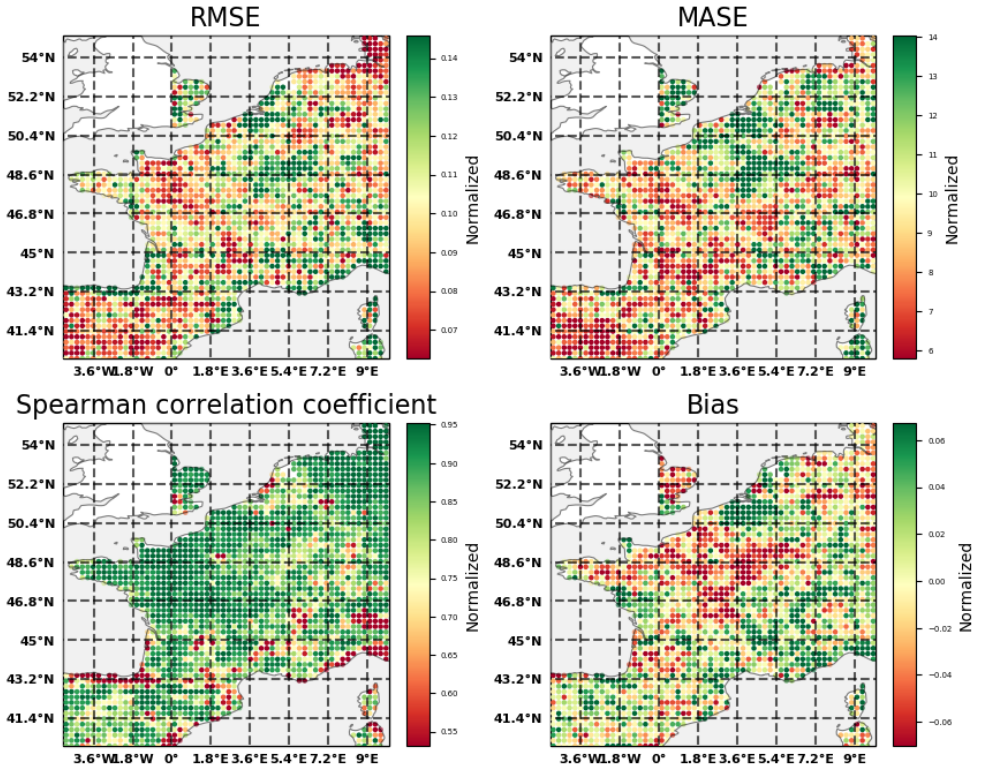

Figure 4.25: Slope: RMSE, MASE, Spearman Correlation Coefficient and Bias are calculated for each GPI within the study-domain using only the test year 2017. The description normalized is used because, the normalized output of the DNN to is used to calculate the different metrics.

Figure 4.26 shows RMSE values per season for the whole study domain. In the DJF season the worst performance is found along the Alps but also in the north-east part of the investigated area. As the JJA season is reached the highest error is found at the agriculture area around Paris and north of London. The Alps show the largest RMSE values during the DJF season and the lowest RMSE between July and August.

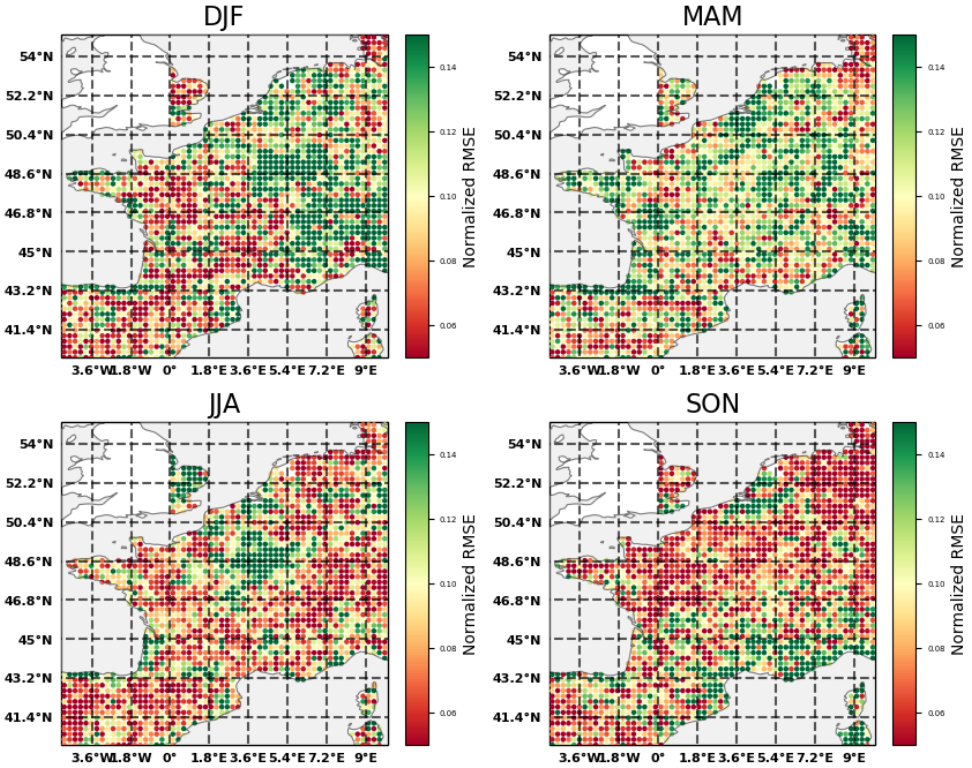

Figure 4.26: Slope: RMSE calculated for each Season: December-January-Feburary (DJF), March-April-May (MAM), June-July-August (JJA), September-October-November (SON). The description normalized RMSE is used because, the normalized output of the DNN to is used to calculate the RMSE.

Table 4.26 gives an overview how well the model is performing for each cluster per season. The lowest RMSE values throughout the study domain are during the JJA and SON season, and the highest are observed in DJF and MAM months. The CDNN performs worst for the Pays Mixed Coast, Needleleaf Forest and French Grassland clusters at the MAM season and for Broadleaf Forest and Boreal Alps during the DJF season. Only for the Agriculture Land cluster the CDNN shows relative high RMSE values in the JJA months. For the Pays mixed Coast and Broadleaf Forest clusters the CDNN has high RMSE values throughout every season. French Grassland and Agriculture show on the opposite the lowest RMSE values. The Boreal Alps and Needleleaf Forest are simulated with intermediate accuracy.

	Normalized RMSE				
Cluster	DJF	MAM	JJA	SON	Over-all
Agriculture Land	0.095907	0.103580	0.116657	0.065847	0.095498
Pays Mixed Coast	0.114095	0.131927	0.102035	0.105350	0.113352
Needleleaf Forest	0.100754	0.115274	0.100550	0.113490	0.107517
French Grassland	0.094521	0.109983	0.083207	0.083551	0.092816
Broadleaf Forest	0.134113	0.120691	0.080899	0.113182	0.112221
Boreal Alps	0.147287	0.094526	0.073841	0.098197	0.103463
Over-All	0.114446	0.1126632	0.092865	0.096603	0.104144

Table 4.14: Slope: Table to compare the mean RMSE per cluster for Simulation 2017 for different seasons: December-January-Feburary (DJF), March-April-May (MAM), June-July-August (JJA), September-October-November (SON)

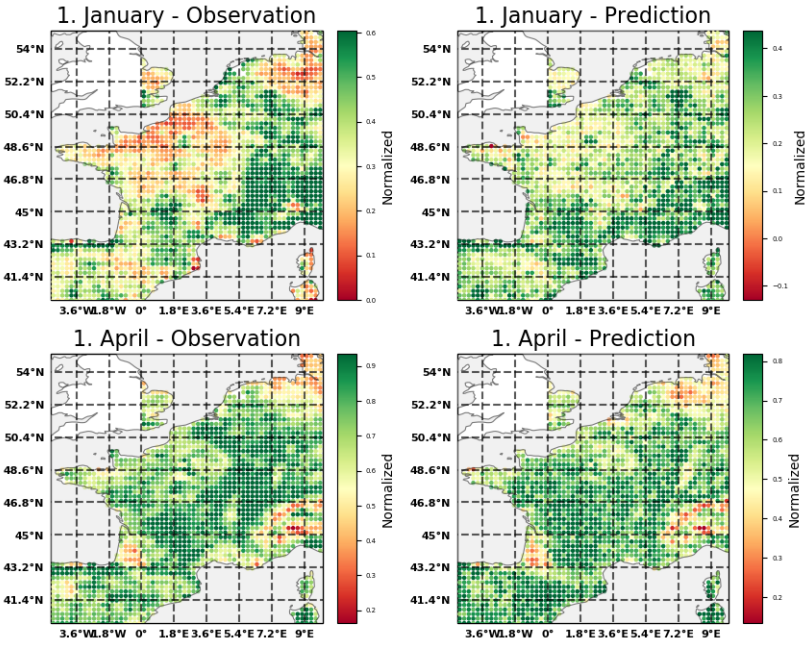

Figure 4.27: Slope: 1st January and 1st April observations for each GPI opposed to the prediction for the same date using test year 2017 to illustrate spatial consistency

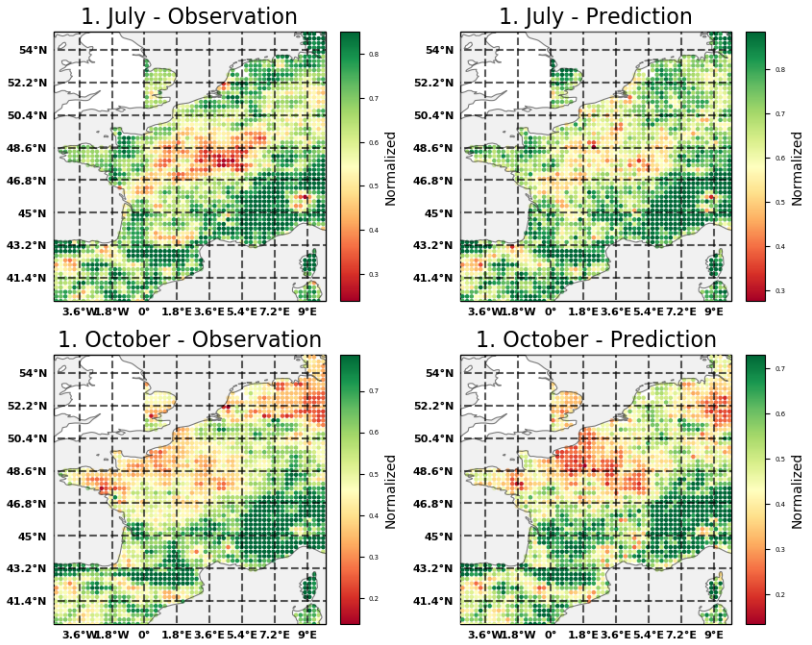

Figure 4.28: Slope: 1st July and 1st October observations for each GPI opposed to the prediction for the same date using test year 2017 to illustrate spatial consistency

Figure A.40 in the Appendix A.5 shows the RMSE values per GPI per cluster. This gives an indication how robust a cluster is in its performance. Agriculture land, Pays Mixed Coast and French Grassland have similar variance with little outliers, whereas Needleleaf Forest, Broadleaf Forest and Boreal Alps have the highest variance in RMSE values.

Figure 4.27 and Figure 4.28 show observed and predicted slope for four time steps: 1st January - 1s April and 1st July - 1st October respectively. All four observations/predictions dates show a very similar spatial pattern. In general, every selected daily prediction is able to capture the spatial distribution, however, it fails to simulate the correct magnitude. Especially, for the agriculture land around Paris the slope values are underestimated during January and July and overestimated during October. This wrong estimations occur because the harvest is not induced in the LSM, which means the input variables forward wrong estimates as these still represent vegetation.

4.4.2. Model performance regarding vegetation types

Figure 4.29 illustrates the simulation performance for seven grid points by comparing the actual observation with the prediction. Here, the grid points represent certain vegetation types, which allows to investigate the relationship between vegetation type and model performance. The grey line illustrates the prediction using the training data set, the green line represents the simulation for the testing data set and the dashed red line illustrates the observation.

The estimated slope follows at least the seasonality for all vegetation types except for boradleaf. For no negetation and boreal grass the model solely able to capture the timing of the seasonality but fails to predict the right magnitude and short term changes. Agriculture, Grassland and no vegetation perform well capturing smaller inter-seasonal changes but still fail to predict the right magnitudes.

For needleleaf the model is able to capture the correct magnitude between the year 2013 and 2019 but not between 2010 and 2013. This shift in performance is only present for this vegetation type. The best simulation performance is for shurb as it captures the right magnitudes as well as inter-seasonal variations.

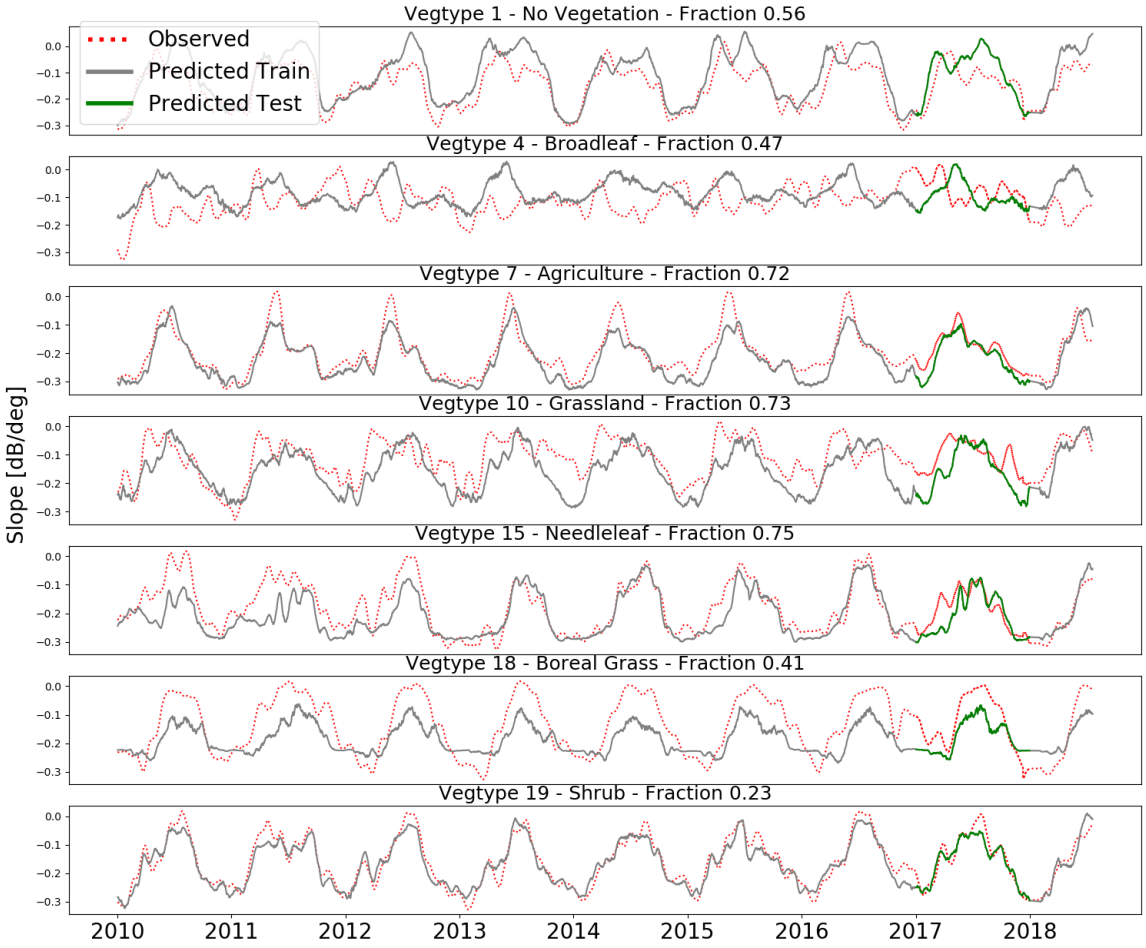

Figure 4.29: Slope: Observed (red) and predicted (grey) time series of slope for seven different vegetation types (Type: 1,4,7,10,15,18,19). The green line indicates the prediction using the test data-set.

For a better performance comparison only the anomalies are compared by subtracting the seasonality signal from the observation signal. Figure 4.30 illustrates the predicted vs. observed anomaly time series. The red dashed line depicts the observation, grey the simulation using the training data and green the prediction using the testing data set.

The anomaly analysis shows that the model is able to detect small inter-seasonal changes for agriculture. It matches very well with the observation anomaly, especially within the years period between 2010 to 2014.

For boreal grass and grassland the model fails to predict the anomalies as neither the timing nor magnitudes are captured. The slope anomalies for no vegetation and needleleaf are able to simulate the timing of inter-seasonal changes but not in the right magnitude. The best model performance was found for the vegetation type shurb.

Table 4.15 illustrates the RMSE for each vegetation type using the whole time-series. This confirms that the model performs best for agriculture and shurb vegetation types.

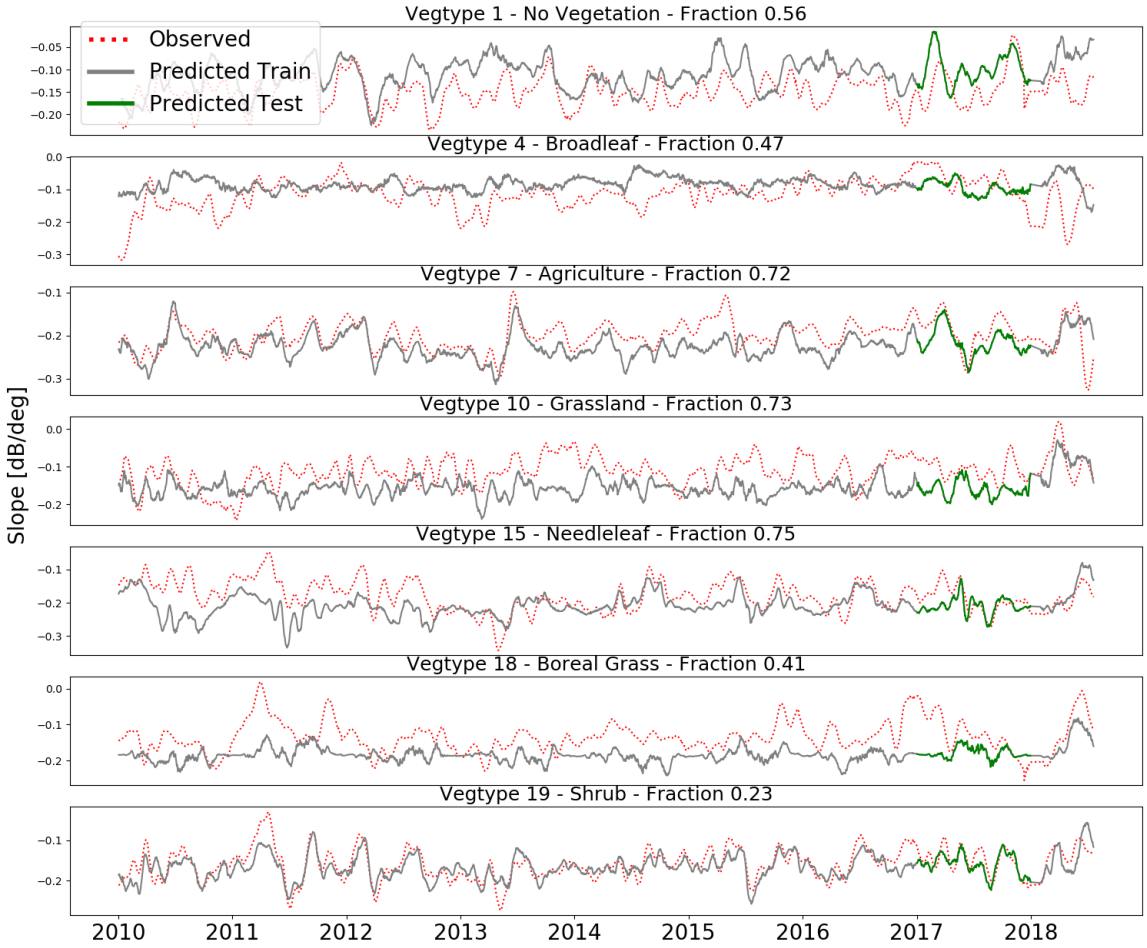

Figure 4.30: Slope: Observed (red) and predicted (grey) anomaly time series of curvature for seven different vegetation types (Type: 1,4,7,10,15,18,19). The green line indicates the prediction using the test data-set.

Vegetation Type	No Vegetation	Broadleaf	Agriculture	Grassland	Needleleaf	Boreal Grass	Shrub
RMSE [dB/deg]	0.053	0.0607	0.036	0.0597	0.0524	0.0645	0.0254

Table 4.15: Slope: Table to compare the RMSE per Vegetation type for the whole time series

4.4.3. Cross-Validation

The cross validation is evaluated by comparing three different models, namely Simulation 2012, Simulation 2014 and Simulation 2017, whereas Simulation 2017 is the reference model. Again, Simulation 2012, 2014 and 2017 are using the year 2012, 2014 and 2017 as test data set respectively. The normalized RMSE refers to the RMSE calculated from the normalized output of the DCNN. Figure 4.31 shows a scatter plot with Simulation 2017 as x-axis and the Simulation 2012 (grey) and 2014 (red) on the y-axis.

This figure shows a cloudy 1:1 linear relationship between RMSE of Simulation 2017 to the RMSE of Simulation 2012 and 2014. The Spearman correlation between Simulation 2017 and 2012 is 0.547 and between Simulation 2017 and 2014 it is 0,567. This indicates a moderate strong correlation between the test and reference simulation. Generally the spread around the 1:1 correlation is larger for higher RMSE values, whereas for lower RMSE values the spread is smaller and therefore more robust. A spatial distribution of the RMSE values for each simulation is illustrated in Appendix A.5. Here, the spatial RMSE patterns are consistent, however the RMSE magnitudes are varying between the simulations.

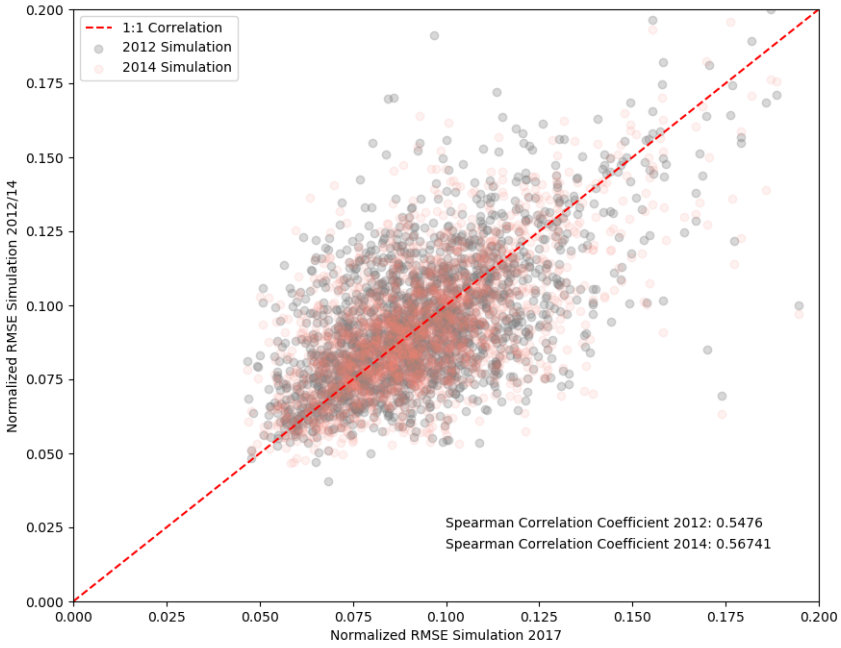

Figure 4.31: Slope: Scatter plot showing the linearity of the correlation of three different models using either 2012, 2014 or 2017 as test year. 2017 is used as a reference year, therefore the Spearman Correlation Coefficient is calculated between 2012-2017 and 2014-2017. The normalized RMSE refers to the RMSE calculated from the normalized output of the DCNN.

Table 4.16 shows the RMSE values for each cluster. This illustrates that the general RMSE is relative consistent and varies for most clusters only in a magnitude range of +/- 0.04.

As a final evaluation, Figure 4.32 is used to show the time-series predictions of Simulation 2012, 2014 and 2017 for seven vegetation types. This illustrates the strong similarity between those three simulations, but also shows that some of the short term variations differ in magnitude. The general performance of the three models show sufficient similarity.

	Normalized RMSE		
	Simulation 2012	Simulation 2014	Simulation 2017
Cluster 0	0.085594	0.084170	0.083179
Cluster 1	0.097726	0.089774	0.093494
Cluster 2	0.098875	0.095503	0.096675
Cluster 3	0.079582	0.075655	0.075449
Cluster 4	0.092066	0.091362	0.090539
Cluster 5	0.106844	0.106123	0.104727
Cluster 6	0.094651	0.094896	0.096198
Cluster 7	0.097504	0.096395	0.096193
Cluster 8	0.103385	0.101534	0.100066
Cluster 9	0.094640	0.095762	0.093308
Cluster 10	0.093037	0.092714	0.092503
Cluster 11	0.100624	0.096527	0.100775
Cluster 12	0.086627	0.087010	0.087589

Table 4.16: Slope: Table to compare the mean RMSE per cluster for Simulation 2012, 2014 and 2017. Here the whole time-series from 2010 to 2018.

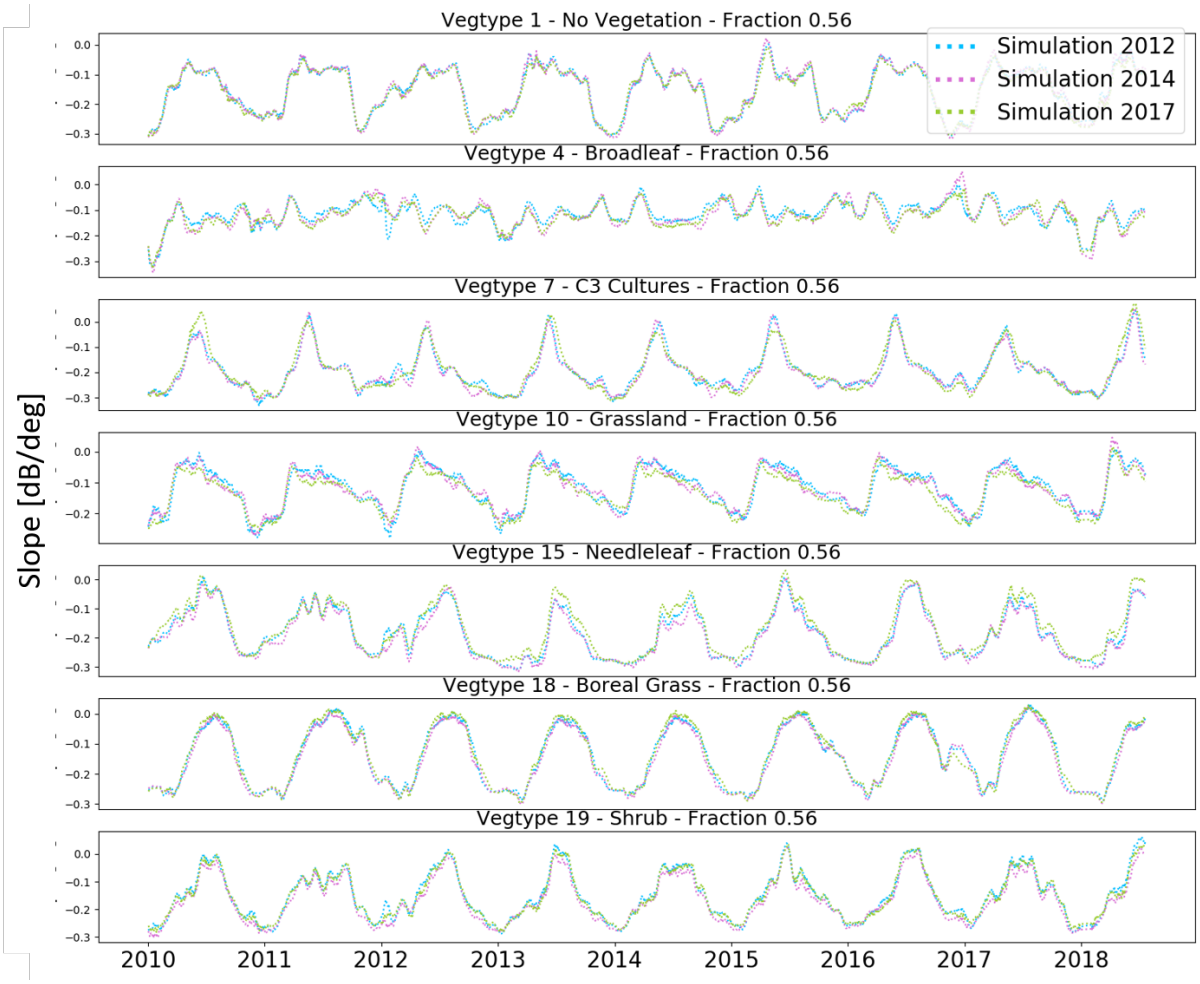

Figure 4.32: Time series prediction from each Simulation (2012 (blue), 2014(pink) and 2017(green) for seven dominant vegetation types (1 = No vegetation, 4 = Broadleaf, 7= C3-Cultures, 10 = Grassland, 15 = Needleleaf, 18 = Boreal Grassland, 19 = Shrub

4.4.4. Spatial and seasonal feature importance

Figure A.44, Figure 4.33, Figure 4.34 and Figure A.45 illustrate the feature importance fractions per month for LAI, WG2, GPP and RE respectively. LAI and RE show a general low importance throughout all seasons for the slope simulation. Only the Alps and Pyrenees show some significant LAI relevance for the prediction. Still, the most dominant features are soil moisture and gross primary production, especially for the vegetation based areas.

The soil moisture feature has two seasonal patterns, in which it has a significant relevance for the prediction. The first occurs between January and February. In this time period the importance fraction reaches up to 0.7 for a small area around Toulouse as well as for the area south of the Pyrenees. The rest of the study domain has a low WG2 fraction. During the spring season from March to June the overall soil moisture importance is very low. Only in some areas especially the alps and Spain it reaches up to 0.4. The second time frame in which soil moisture is relevant is between July and September. Here, Spain and the western part of France reaching the highest values of the feature importance. The Alps, Pyrenees and eastern part of France remain with little soil moisture fractions. October, November and December are transition time frames with low WG2 feature importance.

The GPP parameter is most important for the areas at which WG2 is least important. This covers mainly the centre of the study domain and reaches in two time periods maximum importance. One period of high GPP relevance is between April and June and the other between October and December. For both periods the spatial importance distribution is very similar. In general, whole France has high GPP fractions except of the coast lines, Alps and the forest area below Bordeaux. Between July and September the GPP fraction is overall the lowest. From January to March GPP has a relative high relevance, however, it is rather a transition period as for April the maximum is reached.

Table 4.17 illustrates the feature importance as a mean per cluster and season. This table shows that for Agriculture Land, Pays Mixed Coast, French Grassland and for Broadleaf Forest GPP is the overall most relevant feature. The Boreal Alps and Needleleaf Forest clusters have an almost equal distributed importance fraction for all input variables, which means the slope signal can not be described by a specific vegetation state.

The table again shows that for the feature importance seasonality is different for each cluster. Agriculture and Pays Mixed Coast follow the most similar feature importance trend for each variable, however the feature importance fractions have different magnitudes. GPP shows in general its maximum importance between April and May and its cluster minimum in the summer months. The soil moisture seasonality is not covered well in the clusters, as the selected clusters are not in the Spain. Still, Pays Mixed Coast reflects some of the WG2 importance, especially in the summer month, when the GPP importance decreases.

	Fraction [-]															
	LAI				WG2				GPP				RE			
Cluster	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Agriculture Land	0.26	0.18	0.24	0.22	0.20	0.17	0.24	0.22	0.37	0.47	0.32	0.41	0.16	0.18	0.20	0.16
Pays Mixed Coast	0.26	0.17	0.20	0.20	0.23	0.21	0.41	0.21	0.36	0.47	0.28	0.44	0.15	0.15	0.12	0.15
Needleleaf Forest	0.25	0.24	0.26	0.21	0.26	0.26	0.24	0.29	0.27	0.32	0.24	0.27	0.21	0.17	0.25	0.22
French Grassland	0.20	0.13	0.22	0.17	0.17	0.13	0.18	0.17	0.43	0.52	0.34	0.45	0.20	0.22	0.25	0.21
Broadleaf Forest	0.24	0.19	0.22	0.24	0.16	0.14	0.10	0.15	0.40	0.39	0.37	0.36	0.20	0.28	0.31	0.25
Boreal Alps	0.27	0.26	0.24	0.30	0.21	0.25	0.12	0.23	0.23	0.26	0.33	0.21	0.29	0.23	0.31	0.26

Table 4.17: Slope: Table to compare the mean feature importance fraction per cluster for different seasons. Cluster 0 = Agriculture Land, Cluster 3 = Spanish Rockland, Cluster 5 = Pays Mixed Coast, Cluster 9 = French Grassland and Cluster 12 = Boreal Alps

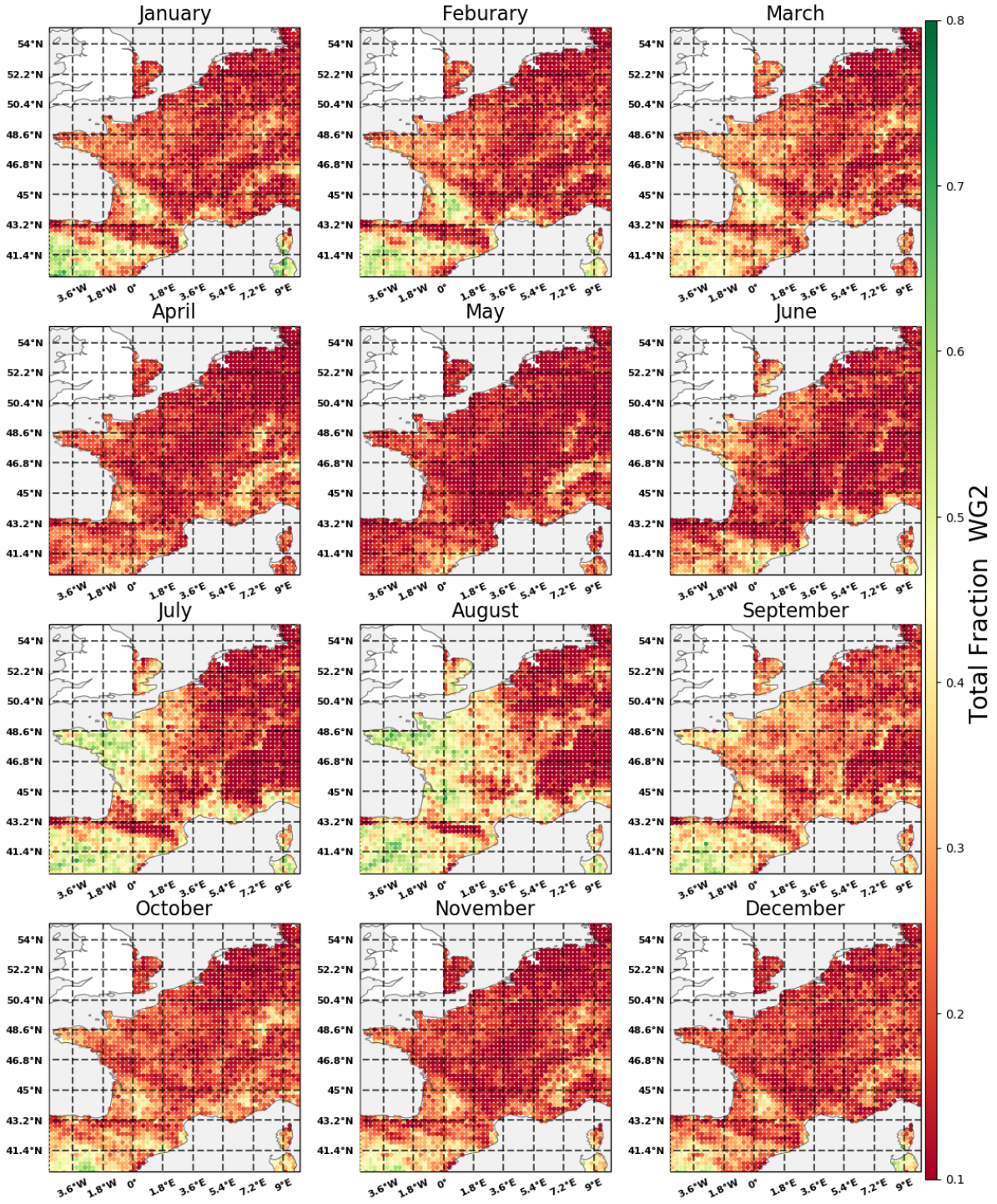

Figure 4.33: Slope: WG2 feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

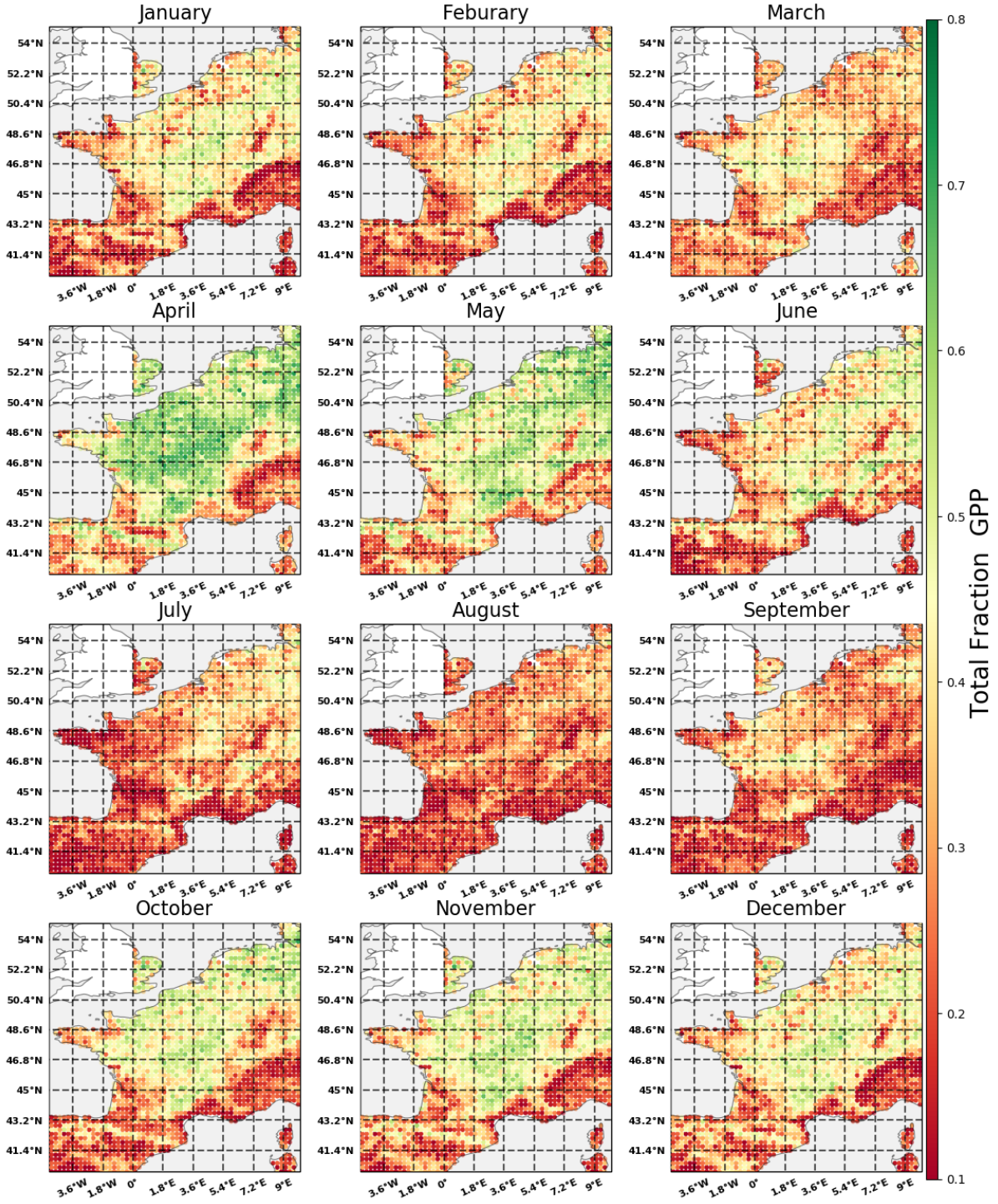

Figure 4.34: Slope: GPP feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

Figure 4.35 further illustrates the time series development of the feature importance between 2016-01 and 2018-07 for the Agriculture Land cluster. Here, the feature importance for each variable as well as the input variable is plotted against time. The green line illustrates LAI for both feature importance and input variable, blue indicates soil moisture (WG2), brown GPP and pink RE.

The table 4.17 shows the mean over the seasons, hence does not give insight how the importance changes through different annual cycles. Figure 4.35 additionally shows how the importance evolves over different years, but also shows how robust the a cluster is in representing itself. This means, cluster with a high variation are less robust then clusters with a low variation. For all clusters the feature importance reaches its local maximum as the input variable is either at its seasonal minimum or maximum. Still, the magnitude of feature importance is highly variable for different years, which means that the maximum feature values can change per season.

As an example the Agriculture Land cluster is displayed to illustrate the general temporal relationship between raw input and feature importance. The example further illustrates a relative strong robustness for the each GPI within the cluster compared to the Boreal Alps cluster, see Appendix A.4. Still, the magnitude of feature importance is highly variable for different years. The example illustrated in Figure 4.35 is relative consistent in its seasonality compared to Pays Mixed Coast, see Appendix A.5. At the Pays Mixed Coast clusters the seasonal importance can vary almost in double of the total magnitude. The clusters French Grassland and Boreal Alps show a similar pattern compared to cluster Agriculture Land by a constant seasonality in the feature importance. In general, all clusters, besides Boreal Alps, show low variance between the GPI, which proofs robustness of the clusters.

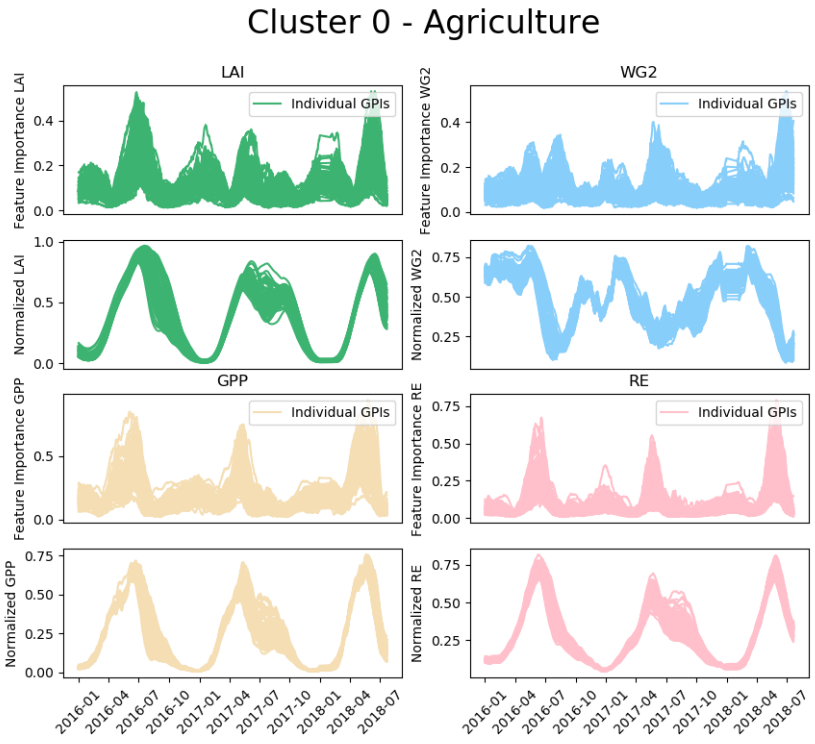

Figure 4.35: Slope: Feature importance and land surface parameter time series for the named cluster 0 - Agriculture Land

4.4.5. Feature importance regarding vegetation types

Figure 4.35 illustrates the feature importance time series for seven different vegetation types. Every GPI is representative for a dominant vegetation type. This again shows the changing magnitude per season. For example the maximum LAI magnitude for Vegtype 7 (Agriculture) has value of 0.4 in the year 2016, 0.25 at 2017 and 0.6 at 2018. This irregularity of magnitudes is found for all features.

Table 4.18 shows the mean annual importance per vegetation type. This gives a general overview how strongly the features are connected to certain vegetation types. For the land covers no vegetation and agriculture LAI is most important, for shurb WG2, for Grassland and needleleaf GPP and for boreal grass respiration (RE). This means, that the slope simulation is mainly influenced by the three vegetation parameters LAI, GPP and RE. Still, the importance magnitude per vegetation type is relative equally distributed as the highest fractions are below 40 %. Therefore, a combination of two main contributors gives a more accurate importance description. GPP has a general low importance except for grassland and needleleaf. The same counts for RE as it is exceptional important for broadleaf, boreal grass and shrub. LAI and WG2 on the other hand have a strong over-all importance.

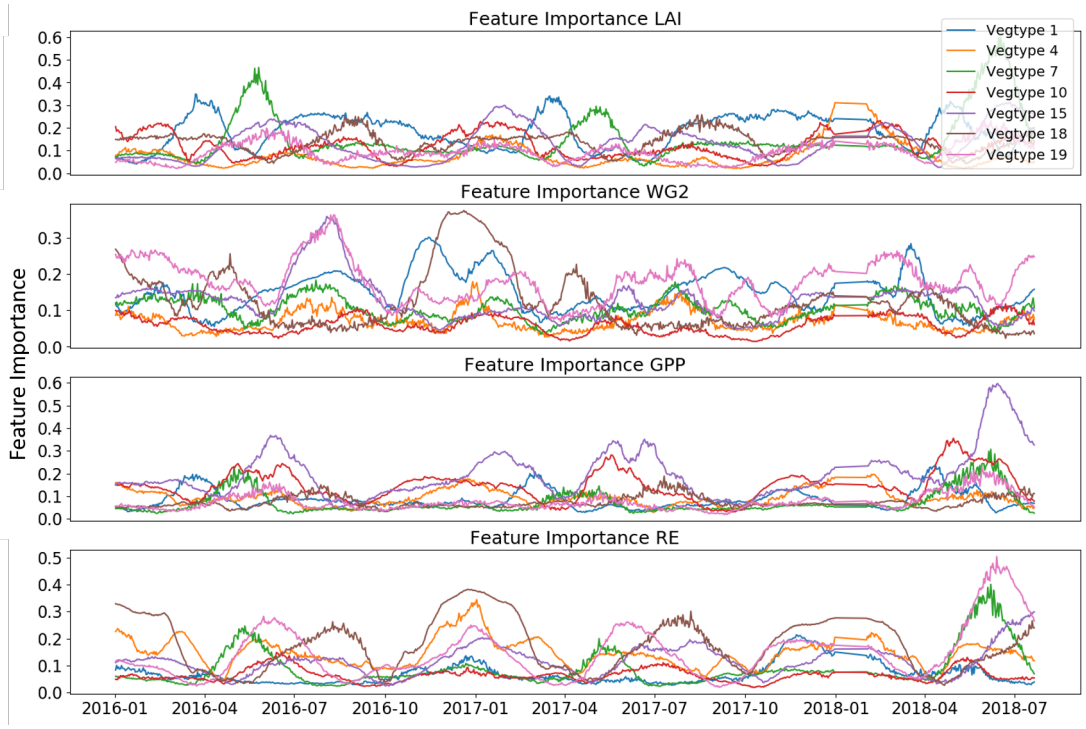

Figure 4.36: Slope: Feature importance time series estimations for each feature (LAI, WG2, GPP and RE) for seven vegetation types: Vegtype 1 = No Vegetation, Vegtype 4 = Broadleaf, Vegtype 7 = Agriculture, Vegtype 10 = Grassland, Vegtype 15 = Needleleaf, Vegtype 18 = Boreal Grass and Vegtype 19 = Shrub

	Feature Importance Fraction			
Vegetation Type	LAI	WG2	GPP	RE
No Vegetation	0.37	0.32	0.17	0.14
Broadleaf	0.18	0.19	0.25	0.39
Agriculture	0.34	0.29	0.17	0.20
Grassland	0.30	0.15	0.37	0.17
Needleleaf	0.22	0.24	0.34	0.20
Boreal Grass	0.28	0.24	0.15	0.33
Shurb	0.19	0.39	0.15	0.27

Table 4.18: Slope: Table to compare the mean feature importance fraction for seven vegetation types. Considering the time frame from 2016 to 2018. LAI = Leaf Area Index, WG2 = Soil moisture, GPP= Gross Primary Production, RE = Respiration

4.4.6. Summary and discussion of slope results

The following discussion is based on two research sub-questions and covers the results of the slope analysis. The summary of the results are presented in Figure 4.37 , Figure 4.38 and Table 4.19. Figure 4.37 compares the min-max scaled variance of the input and output variance as well as the RMSE values of the slope simulation for every month. The displayed variance per month is calculated by taking the mean of all monthly variances over the 10 years of data. The monthly RMSE values are calculated in the same manner. The black dashed line represents the RMSE values, the red line the slope variance and the colored lines the variance for the input variables. Figure 4.38 illustrates the feature importance for all clusters between 2016 and 2017 and table 4.19 depicts the over all performance, most important features as well as the feature importance and RMSE per vegetation type. The green line represents LAI, blue WG2, yellow RE and pink GPP. The dots Figure 4.37 indicate the individual values as the lines don't represent a continuous data set.

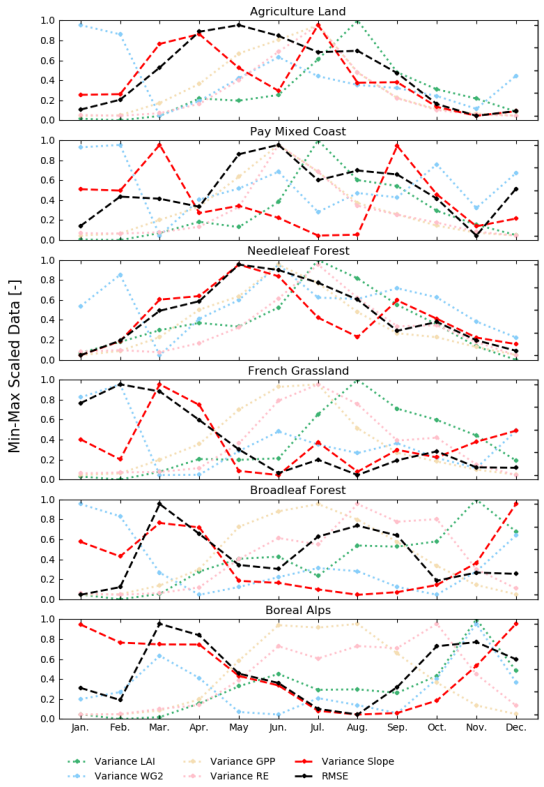

Figure 4.37: Slope: Comparison of min-max scaled variance for all input features (LAI, WG2, GPP, RE) as well as slope and RMSE for each month and distinctive cluster. The dashed balck line represents the RMSE calculated from the prediction vs observation, the grey dashed line represents the variance of curvature, the green the variance of LAI, the blue the variance of WG2, the yellow the variance of GPP and the pink line the variance of RE.

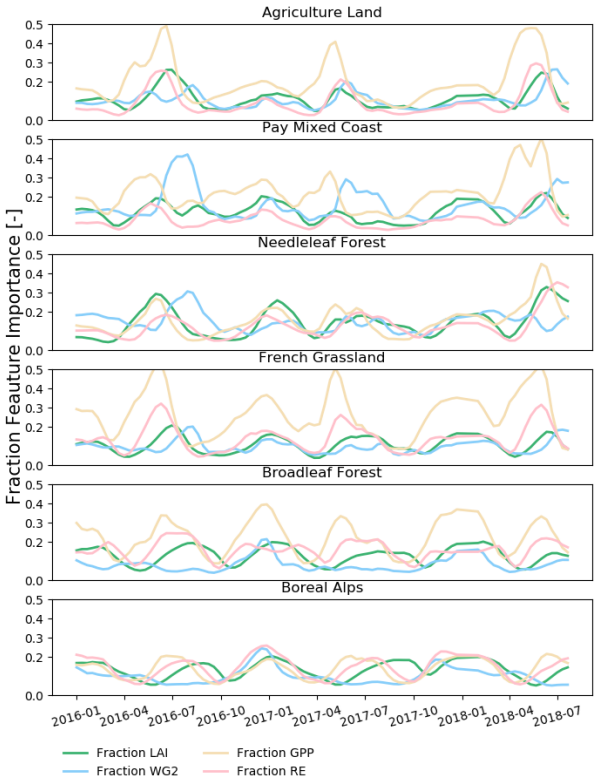

Figure 4.38: Slope: The fraction of the total feature importance for all input features (LAI, WG2, GPP, RE) for every month and distinctive cluster. The greenline represents the feature importance fraction of LAI, the blue the importance of WG2, the yellow tthe importance of GPP and the pink line the importance of RE.

Cluster	Normalized RMSE	Most relevant Feature	No Vegetation	Broadleaf	Agriculture	Grassland	Needleleaf	Boreal Grass	Shurb
Agriculture Land	0.095498	GPP	0.09	0.14	0,59	0.10	0.02	0	0
Pays Mixed Coast	0.113352	GPP	0.08	0.06	0,23	0.30	0.04	0	0.03
Needleleaf Forest	0.107517	WG2/GPP	0.09	0.08	0.06	0.30	0.29	0.0	0.02
French Grassland	0.092816	GPP	0.05	0.16	0,12	0.55	0.07	0	0.01
Broadleaf Forest	0.112221	GPP	0.05	0.31	0.08	0.32	0.13	0.0	0.0
Boreal Alps	0.103463	All	0.10	0,12	0.03	0.21	0.02	0.16	0.01
RMSE [σ]			0.053	0.0607	0.036	0.0597	0.0524	0.0645	0.0254
Most relevant Features			LAI/WG2	RE/GPP	LAI/WG2	GPP/LAI	GPP/WG2	RE/LAI	WG2/RE

Table 4.19: Slope: Table to illustrate the over all normalized RMSE and most important feature per cluster

How does the model performance vary in space and time?
The normalized RMSE values in table 4.19 and Figure 4.25 show that there is a high variation in RMSE values depending on cluster and therefore vegetation type. Here the Pays Mixed Coast and Broadleaf Forest cluster have the highest errors in their predictions. The scaled variance comparison in Figure 4.37 shows that for these two clusters the RMSE values are most correlated to the seasonal trend of the GPP variance. The Spearman Correlation Matrix in Appendix A.52 confirms this correlation between RMSE and GPP for Pay Mixed Coast and for Broadleaf Forest as they have ranking coefficients of 0.69 and 0.48 respectively. The reason for the bad performance of the Broadleaf Forest cluster is probably because the signal is very noisy with almost no seasonality, which means the signal has a high inter-annual variability. This is clearly shown in Figure 4.10, which shows the slope time series for a broadleaf GPI. Pays Mixed Coast consists mainly of agricultural and grassland land tiles but with significant fractions of no vegetation, broadleal and needleleaf. This high variance in land cover fraction could cause errors in the slope simulation as the inter annual variance as well as seasonality are very different for each of the vegetation types. That assumption is consistent with the time series analysis of the distinct vegetation types, see Figure 4.29. Here all vegetation types, except of broadleaf, show a good simulation performance by at least capturing the seasonality.

French Grassland cluster shows the best performance. The reason for this might be because the signal is rather low by showing a clear seasonal cycle, see Figure 4.10. Furthermore, the French Grassland cluster is very robust by showing very little variation in the RMSE distribution Figure 4.37 and land cover fractions A.1.

Agriculture Land shows the second lowest RMSE values. The reason for this good performance could be because the signal has a very strong seasonal cycle with little inter-annual variability, see Figure 4.10. Furthermore, it is a very robust cluster containing mainly agricultural land tiles, see A.1. This suggests that the combination of a strong seasonality and and little inter-annual variation causes a good model performance.

At the Agriculture Land and French Grassland clusters the performance change is most correlated to the variance of the slope data, at which the Spearman Correlation Coefficient is 0.79 for Agriculture Land and 0.41 for French Grassland. For both clusters mainly the maximum magnitudes were underestimated, whereas most of the inter-annual variation are captured, see Figure 4.29.

Intermediate results are found for the Boreal Alps and Needleleaf Forest clusters. Here the RMSE is most correlated with the scaled WG2 variance for the Boreal Alps and for Needleleaf Forest with the scaled slope variance. The intermediate results are consistent with the needleleaf and no vegetation cover results, see Figure 4.29. The reason for higher errors in the prediction could be because the slope signal shows a higher inter-annual variability compared to grassland and agriculture, see Figure 4.10. This counts for the Boreal Alps and Needleleaf Forest clusters.

Overall, Figure 4.25 indicates that the variance in slope is most correlated to the change in the RMSE values independently of the location. Only the Pay Mixed Coast cluster differs from this conclusion as the performance is most related to the GPP variance. This inconsistency might be because of the high variant land covers within this cluster. The RMSE magnitude is dependent on the strength of the seasonal cycle as well as the inter annual variability. Here, a high inter annual variability causes higher errors, whereas a signal with a strong seasonal trend and low inter annual variability is captured well. Furthermore, GPIs with a large variation of land cover types have high RMSE values. This is especially consistent with the Pay Mixed Coast and Boreal Alps clusters.

Next to the spatial model performance dependency time determines the accuracy of the model. For Agriculture Land, Pay Mixed Coast and Needleleaf Forest the highest RMSE values occur between April and June, see table 4.14. A reason for this could be because at this time of the year slope has is highest variance. Theses are also the months at which most of the carbon assimilation takes place. This indicates that the slope signal for these GPIs is highly dependent on the GPP variation and therefore phenology change. This is further confirmed in the high feature importance of GPP for these clusters. French Grassland has the highest RMSE values between January and April, which could indicate the growth of grass within this area as at then the highest variation in the data occurs, see table 4.7.

The Boreal Alps cluster has the worst performance during the winter months. This might be a consequence of an increased fraction of snow in this area, which is not considered in the LSM [29]. This is very cleary il-

lustrated in the time series plots of the input variables, see Figure 4.6. Here all LSM parameters are the lowest with almost no variation within the DFJ months. This means no signal or pattern is forwarded to the DNN and hence creates a large error in the simulation. Additionally the high fraction of rock and no vegetation could lead to an increased wrong interpretation as the model tries to retrieve information from vegetation states.

The seasonal decomposition shows that the model is able to capture the anomaly but solely fails to predict the magnitude of the seasonal cycle. However, Grassland has a bit shifted seasonality, which might be due to a late response of the land surface variables compared to the actual measurements. The vegetation type Shurb shows the best performance for either the original observation as well as the anomaly signal. A reason for this could be the low inter annual variation for this GPI as the anomaly only ranges between -0.1 and -0.2 as the other ranges between 0 and -0.3 or -0.1 and -0.3, see Figure 4.30.

In summary, the general timing of the seasonal cycle is well captured, however the model fails to predict the right magnitude. This means that the model is able to simulate the spatial variance throughout the study domain, but fails to make accurate simulations in time. Here, the model performance depends on the season and land cover types as well as the inter annual variability of the slope signal. The wrong magnitude estimations could be caused by processing procedure of the slope signal at which the signal is smoothed over 42 time steps. This suggests that valuable information might be smoothed out as well, or other kernel widths are better suitable.

How does the feature importance vary in space and time?
Table 4.19 shows the most important features per cluster. This clearly illustrates that GPP is the most important feature for the slope simulation. Especially, Agriculture Land and French Grassland have high GPP importance fractions for all time steps, see Figure 4.20. The high importance of GPP for Agriculture Land might be a consequence as this feature directly represents the assimilation of biomass and therefore harvest seasons. For both clusters the GPP has its highest fraction values during April and May, which again indicates the growing season and therefore phenology change.

Another reason for the high GPP importance for agriculture might because the GPP input data shows the highest variance and seasonality compared to the other features, see Figure 4.6. This high variance could tell the DNN that the GPP is most important compared to the other features. The same counts for French Grassland. French Grassland shows in Figure 4.6 almost no variability for LAI and WG2 but a very high variability for GPP and RE.

RE and GPP have a very similar pattern and correlate highly in their monthly variance, see Spearman Correlation Matrix in the Appendix A.4. This might be a problem for the feature importance interpretation as they can partly represent each other. Still, the model takes GPP as their main source of information. To further investigate this is dependency, an ablation study is suggested by neglecting RE or GPP for the prediction. This allows to see if the signal can still be simulated even with one feature missing.

Another characteristic of the Agriculture Land and French Grassland cluster are the robustness and low variation in land cover types, see Appendix A.1. That suggests that GPIs or clusters with a clear land cover type tend to have a dominate feature, whereas clusters with a mixture of different land covers have lower feature importance fractions. The very distinct feature importance as well as the low RMSE values for Agriculture Land and French Grassland allow a relative confident interpretation.

Pays Mixed Coast shows a high importance of the GPP feature. Only between July and August WG2 is the most important feature for the simulation. This might be because at this point of time most of the vegetation has already developed and therefore the variation in GPP is rather low, which gives soil moisture are larger relevance for the simulation. Therefore, an increased soil moisture could cause higher vegetation water dynamics, which might explains the relationship between soil moisture and slope as slope is assumed to be a vegetation proxy. Still, these interpretation needs to be treated carefully, as the model is solely able to reproduce the seasonality. In general every variable which reflects somehow the seasonality of slope might be considered by the deep neural network as relevant or even correlated.

Boreal Alps as well as Needleleaf Forest show no dominant feature importance as almost all fractions are

around 0.25. This means that the vegetation dynamics and phenology change can not be simulated by the vegetation states inserted to the deep neural network. That can additionally be observed by the low model performance for these clusters. A reason for this feature importance distribution could be that for the Boreal Alps cluster high fractions of no vegetation as well as rock are present. Therefore, no or little correlation can be found with the given vegetation states. This was also observed for the backscatter simulation. The feature distribution of the Needleleaf Forest cluster might be explained by the low variance magnitude of the input variables within this area, see Figure 4.6. Here the LAI signal for needleleaf is almost constant with a very weak seasonality.

In summary, the feature importance seems to be highly dependent on the magnitude of variance of the input data. The most important feature is the gross primary production. Only areas with a high fraction of no vegetation land cover type tend to depend more on soil moisture. This indicates that the slope signal contains valuable information about the vegetation phenology.

4.5. Curvature Evaluation

This section is divided into five parts to evaluate the curvature results: 1) Spatial and Seasonal Model Performance, 2)ModelPerformance regarding Vegetation types, 3) Cross-Validation, 4) Spatial and Seasonal Feature Importance and 5) Feature Importance regarding Vegetation type. In the end of the section the results are summarized and discussed. This will help to finally answer a part of the research question if machine learning can reliable estimate curvature from ISBA states.

4.5.1. Spatial and seasonal model performance

Figure 4.39 shows RMSE, MASE, Spearman correlation coefficient and Bias for each GPI. Here, only the the testing dataset (year 2017) is evaluated. The description normalized is used because, the normalized output of the DNN to is used to calculate the different metrics. Similar to the backscatter and slope results the RMSE and MASE values show the same spatial pattern. The MASE values indicate that the prediction quatilty of the deep neural network is between 8 and 16 times worse then the native prediction method. Especially, the agricultural and grassland areas in the Brittany, Paris and Spain show the highest MASE values. In general, the RMSE and MASE are very noisy throughout the whole study area, by showing only a few spatial clusters. For further interpretation the RMSE, Spearman Correlation Coefficient and Bias are used to analyse the model performance.

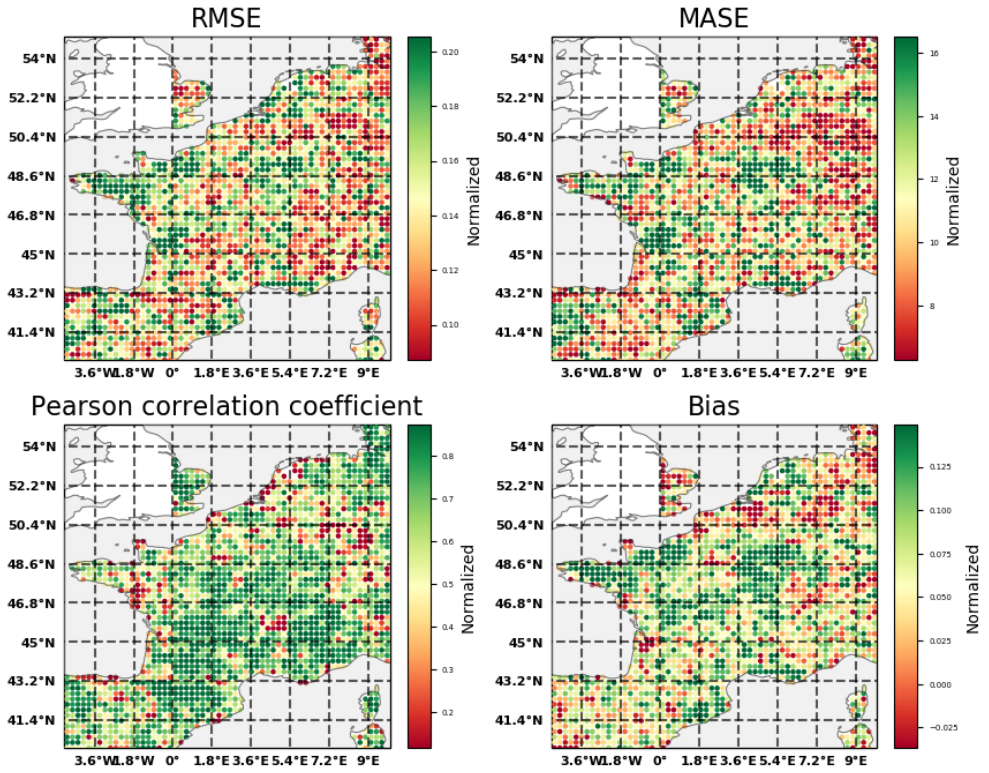

Figure 4.39: Curvature: RMSE, MASE, Spearman Correlation Coefficient and Bias are calculated for each GPI within the study-domain using only the test year 2017. The description normalized is used because, the normalized output of the DNN to is used to calculate the different metrics.

The largest RMSE values occur in the center of the study domain, which is located close to the agriculture area around Paris and the Brittany. Low RMSE values are accumulating mainly at the alpine areas such as the Alps and Pyrenees but also in the south west of France. The Spearman correlation coefficient indicates a general high positive correlation of 0.8. A lower correlation rank is found in the north-east parts of France, the coast line of Belgium and the Netherlands as well as the west coast of France and Portugal. In those areas the bias indicates a general under estimation of curvature values. Over estimation occurs mainly in the parts with high RMSE values, which again is highly accumulated at the Brittany and agricultural sides.

Figure 4.40 gives an overview how the spatial distribution of the model performance is per season. In general, the RMSE distribution remains very noisy without clear spatial patterns throughout the seasons. For a better insight into the spatial-inter-annual variation the clusters are investigated separately. This is illustrated in Table 4.20. Here, the mean RMSEs are taken for each distinctive cluster per season.

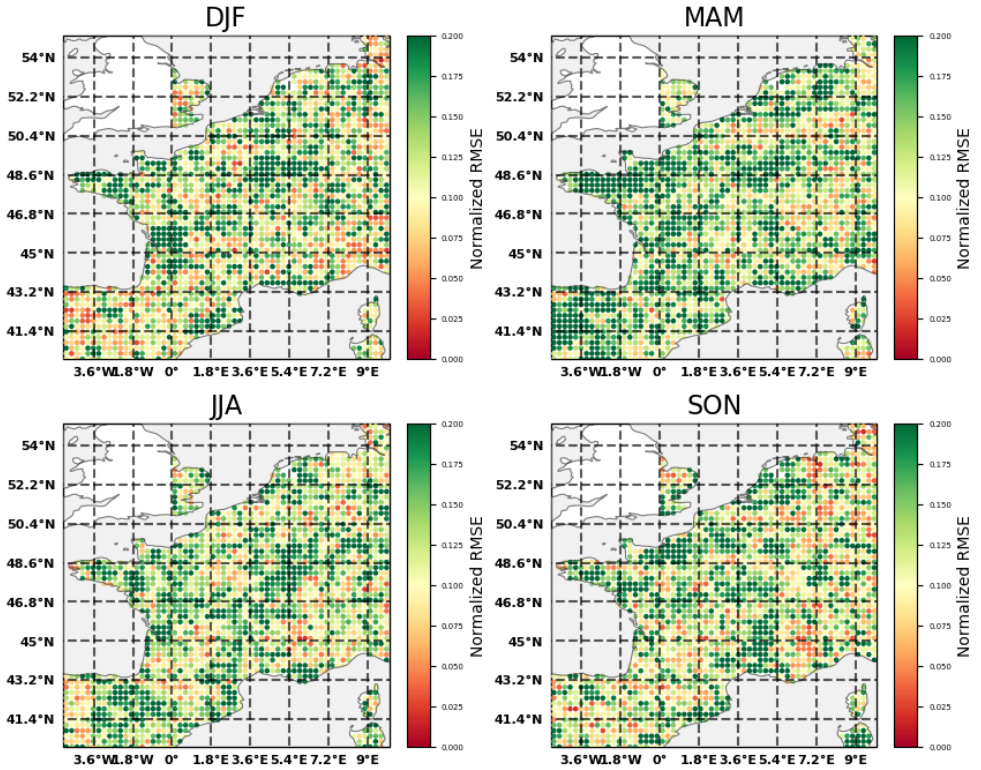

Figure 4.40: Curvature: RMSE calculated for each Season: December-January-Feburary (DJF), March-April-May (MAM), June-July-August (JJA), September-October-November (SON). Normalized RMSE refers to the RMSE calculated from the min-max scaled output of the CDNN and has therefore no unit.

The best performance for each cluster is during the DJF months and the worst is during the MAM. From MAM to SON the simulation accuracy improves. This means that the performance - over-all but also per cluster - follows a seasonal cycle with the lowest RMSE values during the DJF season and the highest RMSE values in the MAM months. Follwoing this pattern the CDNN perfromed best for the Boreal Alps and Broadleaf Forest clusters. For Agriculture Land, Needleleaf Forest and French Grassland the CDNN shows significant worse results. Still, the highest RMSE occurs within the Pays Mixed Coast cluster.

	Normalized RMSE				
Cluster	DJF	MAM	JJA	SON	Over-all
Agriculture Land	0.135109	0.166201	0.125046	0.133831	0.140047
Pays Mixed Coast	0.142723	0.167370	0.162285	0.160446	0.158206
Needleleaf Forest	0.131693	0.156326	0.143611	0.134264	0.141474
French Grassland	0.135512	0.155874	0.154754	0.140274	0.146604
Broadleaf Forest	0.121011	0.127073	0.115625	0.123294	0.121751
Boreal Alps	0.118638	0.126957	0.116834	0.105242	0.116918
Over-All	0.130781	0.149967	0.136359	0.132892	0.137500

Table 4.20: Curvature: Table to compare the mean RMSE per cluster for Simulation 2017 for different seasons: December-January-Feburary (DJF), March-April-May (MAM), June-July-August (JJA), September-October-November (SON)

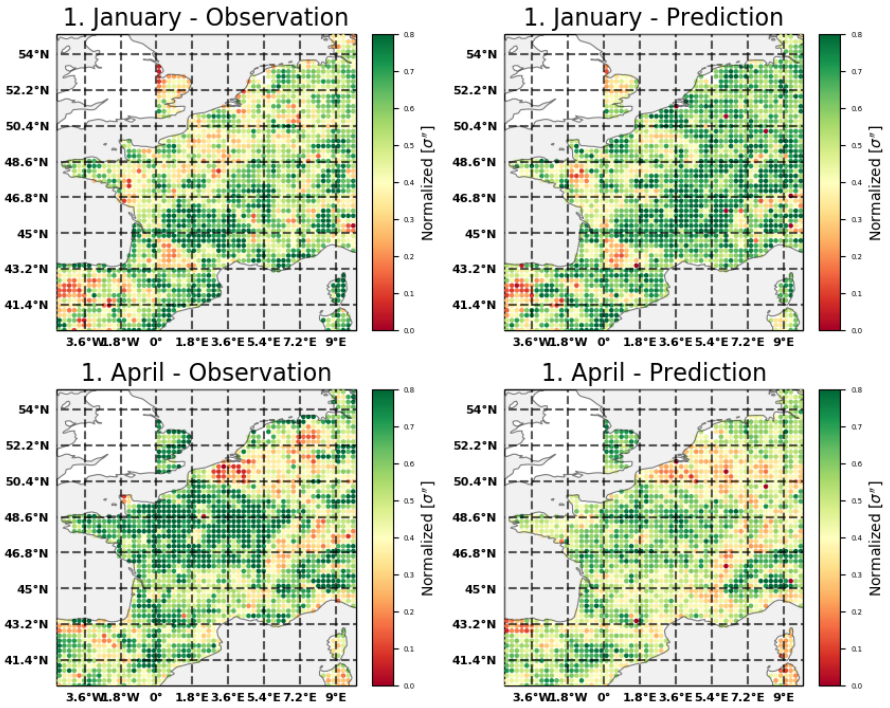

Figure 4.41: Curvature: 1st January and 1st April observations for each GPI opposed to the prediction for the same date using test year 2017 to illustrate spatial consistency

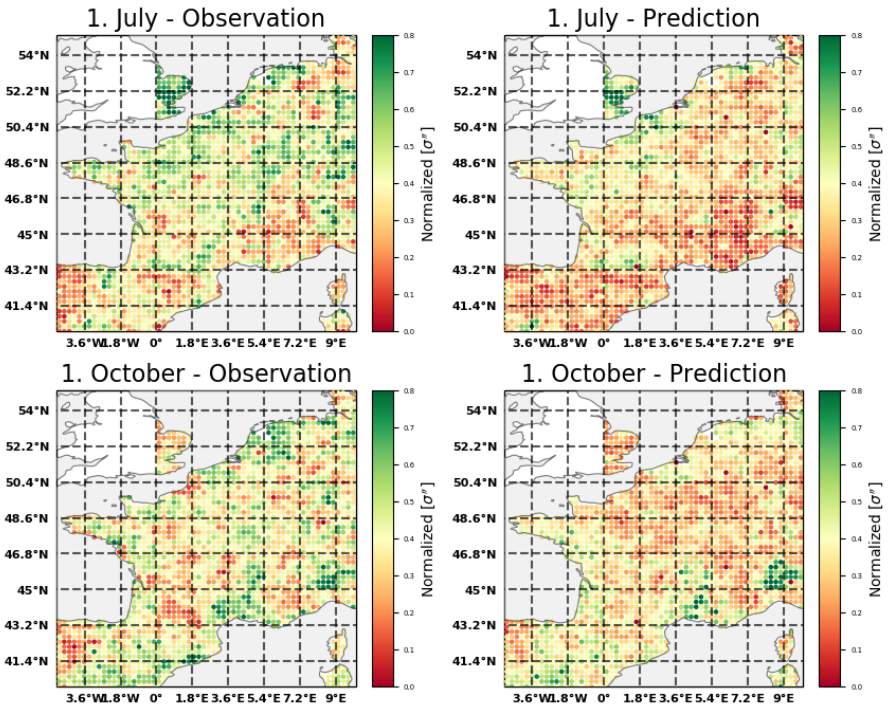

Figure 4.42: Curvature: 1st July and 1st October observations for each GPI opposed to the prediction for the same date using test year 2017 to illustrate spatial consistency

Figure A.53 in the Appendix A.5 shows the RMSE values per GPI per cluster. This gives an indication of how robust a cluster is in its performance. It shows that the RMSE positively correlates with the variance per cluster. This means the cluster with the highest RMSE has also the highest variance.

Figure 4.41 and Figure 4.42 show observed and predicted slope for four time steps: 1st January - 1st April and 1st July - 1st October respectively. The first of January and first of April observations/predictions dates show generally a similar spatial pattern. However, they fail to simulate the correct magnitude of curvature. Within the 1st July and 1st October comparison a poor simulation in magnitude and space can be observed, especially in the centre of the study domain.

4.5.2. Model performance regarding vegetation types

To further investigate the model performance, prediction and observation time series are compared. Figure 4.43 illustrates the performance of the seven distinct vegetation types. The red dashed line represents the observation whereas the grey the prediction using the training data set and the green line the simulation using the testing data set.

The best performance is found for the vegetation type no vegetation as it able to capture the seasonal variations but also most of the inter-seasonal peaks. For agriculture the model achieves to simulate the seasonality, however not in the right magnitude but fails to predict the inter-annual variations. In case of needleleaf, shrub, boreal grass and grassland the model captures roughly the right timing of the main seasonal changes, however is not able to capture the right magnitude nor the inter-annual variations. For broadleaf the model fails to predict either seasonality, magnitude or small term changes.

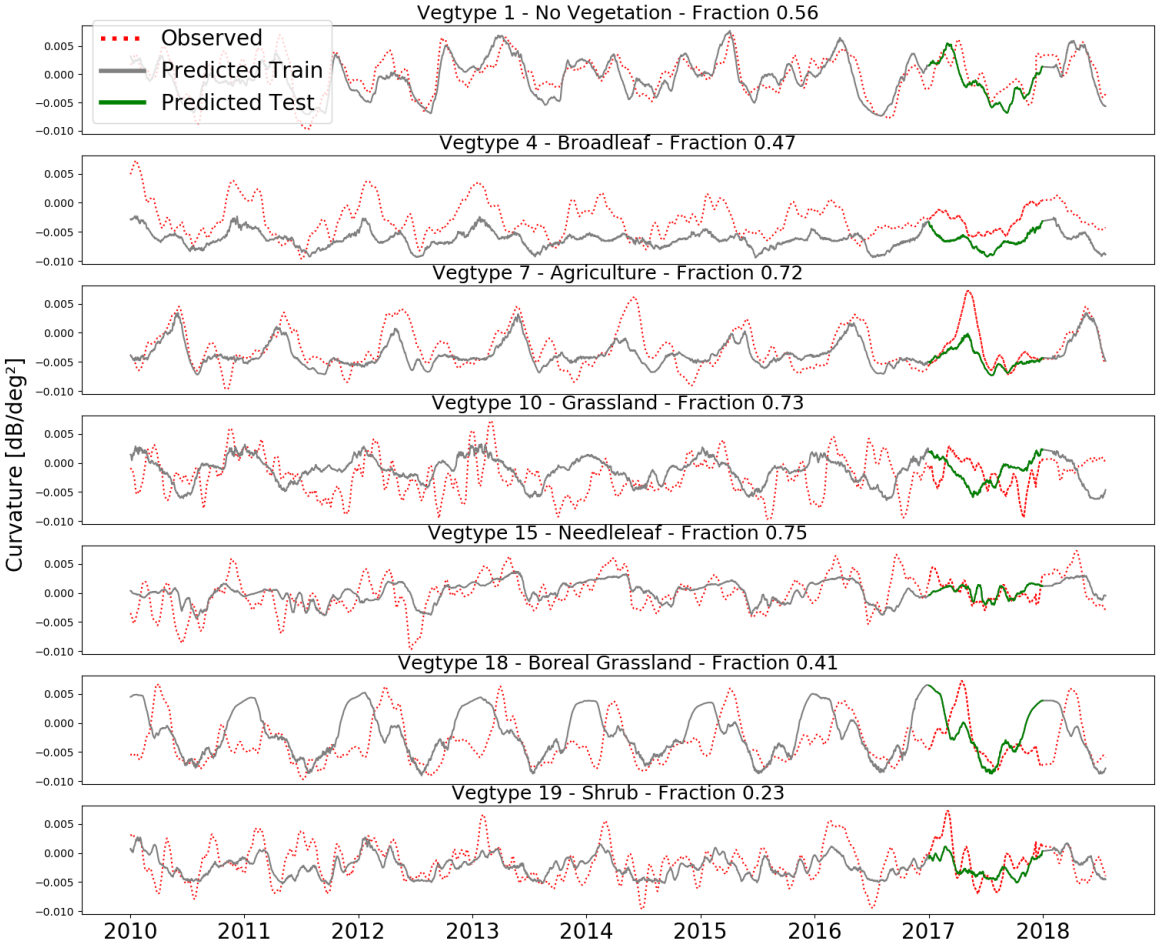

Figure 4.43: Curvature: Observed (red) and predicted (grey) time series of curvature for seven different vegetation types (Type: 1,4,7,10,15,18,19). The green line indicates the prediction using the test data-set.

Figure 4.44 illustrates the anomaly time series analysis for the same vegetation types. This figure shows that the model is not able to capture any of the curvature anomalies with one exception of the grid point representing the no vegetation land cover type. Here, the model simulates almost all of the variations. Table 4.21 shows the RMSE per vegetation type. This again shows that the model for No Vegetation performs the best.

In summary, the model is able to simulate the general behavior of the curvature over time and space. Still, the predicted signal is manly composed of the seasonality of the investigated GPI. This means, it is able to capture the seasonal trend but is not able to capture the right magnitudes nor inter-seasonal variations.

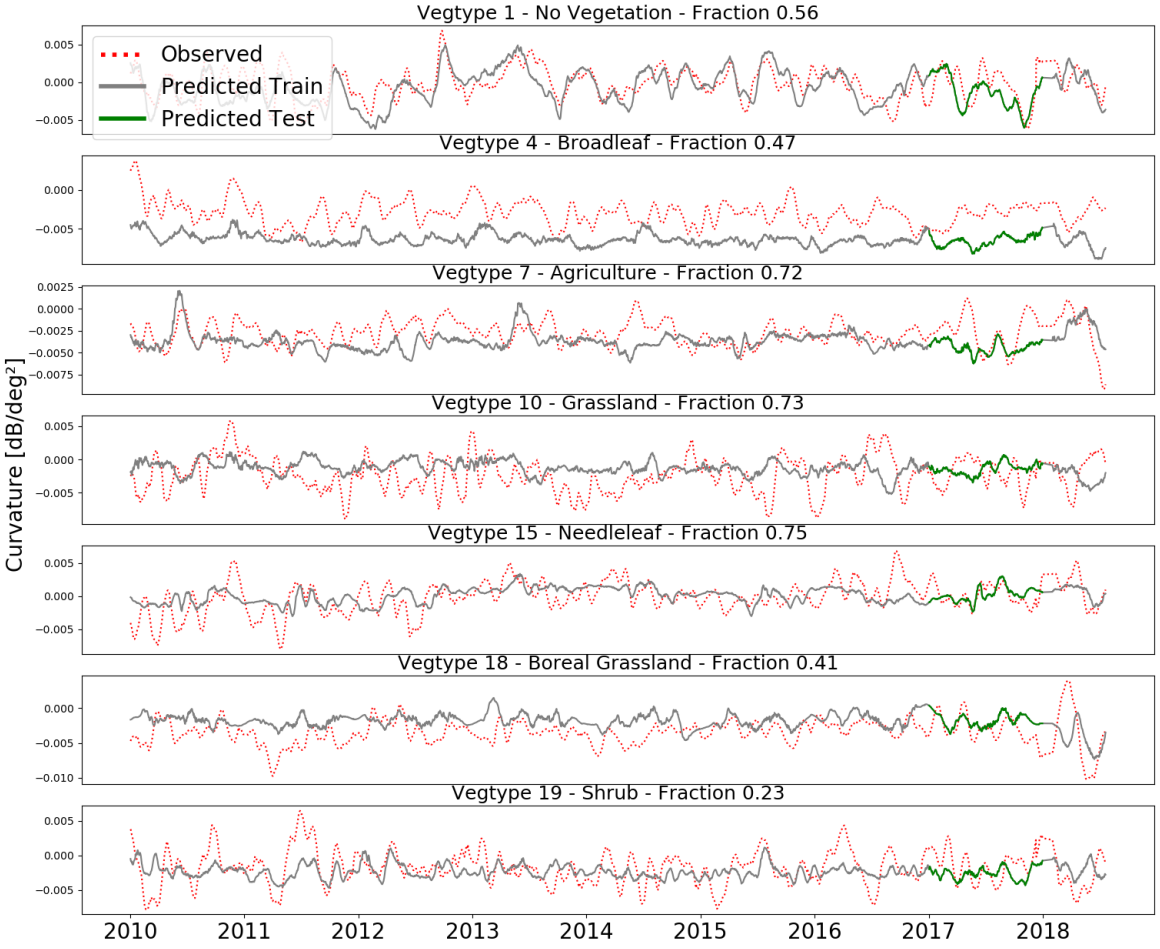

Figure 4.44: Curvature: Observed (red) and predicted (grey) anomaly time series of curvature for seven different vegetation types (Type: 1,4,7,10,15,18,19). The green line indicates the prediction using the test data-set.

Vegetation Type	No Vegetation	Broadleaf	Agriculture	Grassland	Needleleaf	Boreal Grass	Shrub
RMSE [dB/deg^2]	0.0022	0.004	0.0026	0.0038	0.0024	0.0048	0.0025

Table 4.21: Curvature: Table to compare the RMSE per Vegetation type for the whole time series

4.5.3. Cross-Validation

For investigating the robustness of the model a cross validation was performed using 2012 and 2014 instead of 2017 as testing data set. Simulation 2012, 2014 and 2017 are using the year 2012, 2014 and 2017 as test data set respectively. Figure 4.45 illustrates a scatter plot by correlating the normalized RMSE values from Simulation 2012 and 2014 with 2017. The red dots display the values for Simulation 2014 and the grey dots for simulation 2012.

The Spearman Correlation Coefficient is used to show the strength of correlation between the three predictions. Simulation 2012 ranks 0.8 whereas simulation 2014 achieves a coefficient value of 0.78. Both correlation coefficients show a strong positive relationship with the reference simulation 2017. The general distribution around the 1:1 correlation is relative low and maintains relative constant for low and higher RMSE values. Figure A.54, A.55 and A.56 show the spatial distribution of the RMSE values within the study domain. All three simulation show the same spatial pattern and therefore consistency in space.

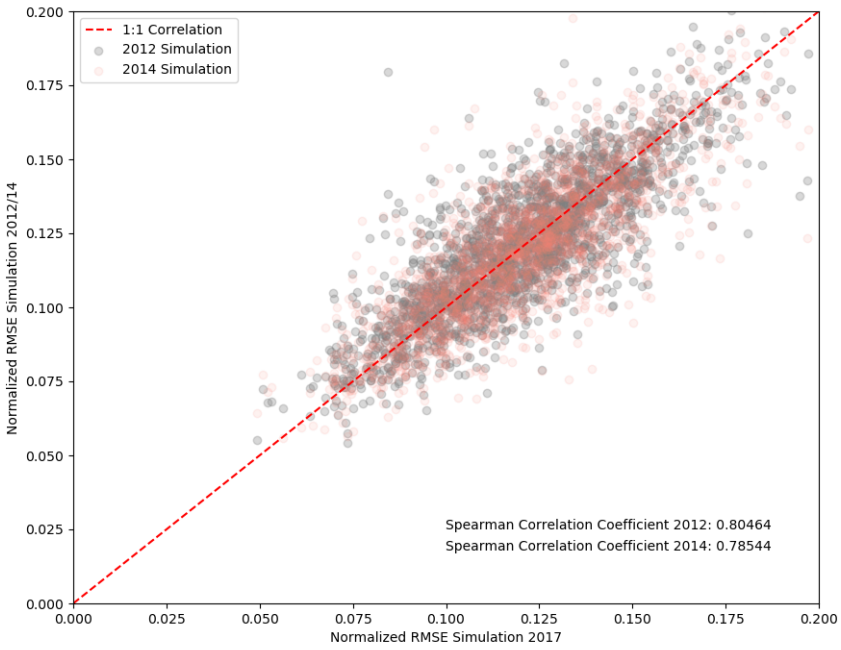

Figure 4.45: Curvature: Scatter plot showing the linearity of the correlation of three different models using either 2012, 2014 or 2017 as test year. 2017 is used as a reference year, therefore the Spearman Correlation Coefficient is calculated between 2012-2017 and 2014-2017.

Table 4.22 shows the RMSE values for each cluster and simulation. This allows to compare the spatial consistency in more detail. All clusters except of cluster 5 show great consistency and the RMSE magnitudes vary maximum in the range of +/- 0.05. Cluster 5, here referred to Pays Mixed Coast, has the highest variation in RMSE between the three simulations. This outlier was also observed in the general performance analyses, as the Pays Mixed Coast performed worst compared to the other spatial clusters. Therefore, the robustness of a model shows to be correlated with the magnitude of the RMSE values.

For a final cross validation check seven time series are directly compared for each simulation, see Figure 4.22. This illustrates the high correlation between the simulation as all three time series for each vegetation type agree for almost all time steps. Still, some of the inter-seasonal variations differ slightly in magnitude but show a good general similarity.

	Normalized RMSE		
	Simulation 2012	Simulation 2014	Simulation 2017
Cluster 0	0.117230	0.113171	0.118108
Cluster 1	0.128454	0.122699	0.134210
Cluster 2	0.124847	0.122578	0.126015
Cluster 3	0.120474	0.122586	0.118746
Cluster 4	0.130526	0.130763	0.130478
Cluster 5	0.136092	0.154532	0.139084
Cluster 6	0.128856	0.124727	0.128466
Cluster 7	0.126936	0.129310	0.127181
Cluster 8	0.127267	0.122798	0.124273
Cluster 9	0.129046	0.127063	0.126341
Cluster 10	0.112807	0.108454	0.107191
Cluster 11	0.105347	0.108703	0.105754
Cluster 12	0.101924	0.101721	0.101107

Table 4.22: Curvature: Table to compare the mean RMSE per cluster for Simulation 2012, 2014 and 2017. Here the whole time-series from 2010 to 2018.

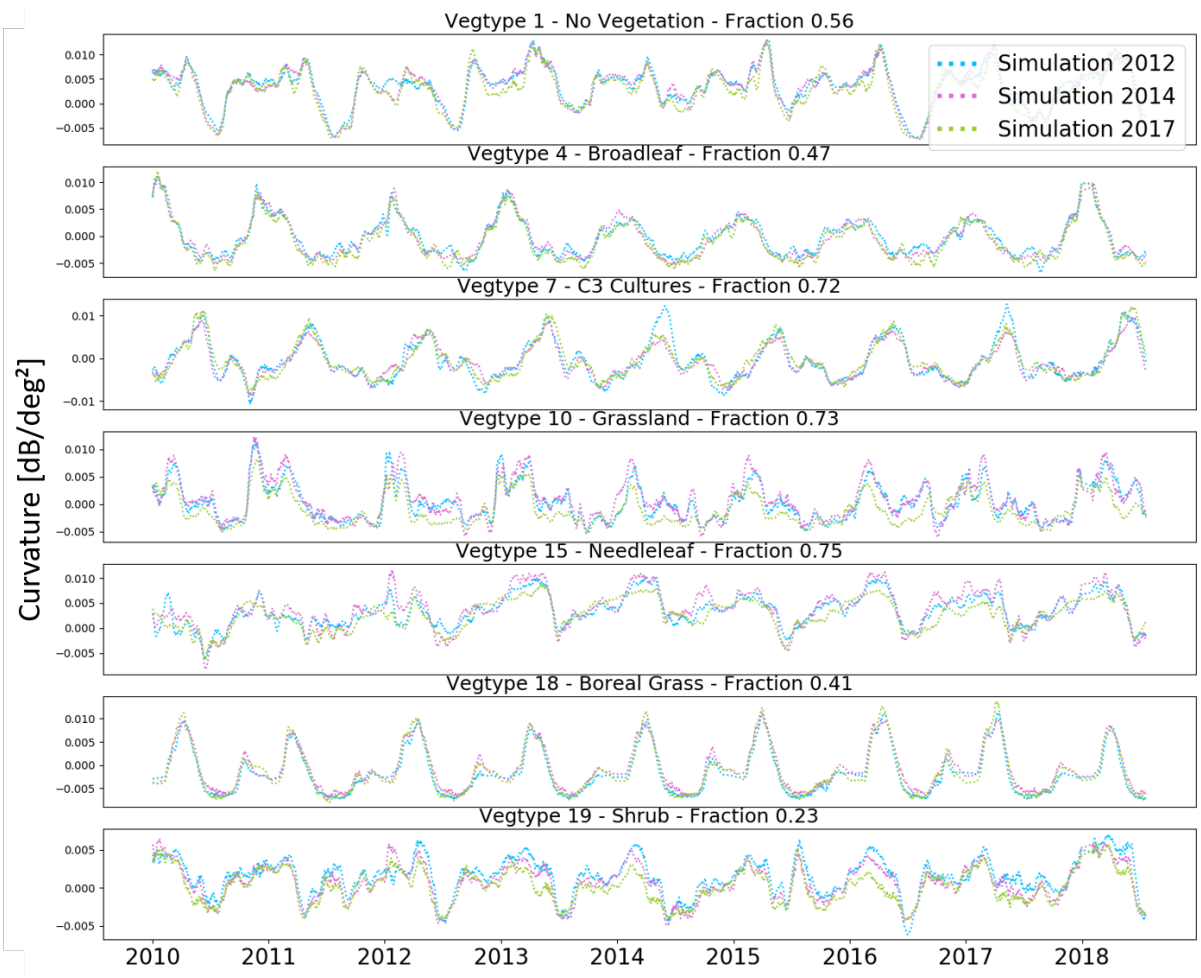

Figure 4.46: Curvature: Time series prediction from each Simulation (2012 (blue), 2014(pink) and 2017(green) for seven dominant vegetation types (1 = No vegetation, 4 = Boradleaf, 7= C3-Cultures, 10 = Grassland, 15 = Needleleaf, 18 = Boreal Grassland, 19 = Shrub

4.5.4. Spatial and seasonal feature importance

The feature importance, as mentioned in Backscatter and Slope Analysis subsections, can give an insight of how relevant features are for a certain prediction. Only the years from 2016 to 2018 are considered due to time limitations as the model runs require high computational effort.

Figure A.57, A.58, 4.47 and A.59 show the monthly averages per each feature LAI, WG2, GPP and RE, respectively. Here, each monthly map represent the average fraction of the total feature importance for 2016, 2017 and 2018. The red areas indicate a weak importance of the feature whereas the green areas show a strong importance of the feature. In this section only the GPP feature importance is displayed, whereas the other features are illustrated in the Appendix A.5. This is because solely GPP shows a significant fraction and distinct clusters within the study domain.

The feature importance of gross primary production has a similar temporal pattern as for the slope simulations. Here, only in April to May and from October to November GPP has a significant relevance. The areas of importance cover mainly the agriculture land around Paris towards the Atlantic coast and UK as well as the agriculture land around Toulouse.

LAI shows a general high importance between July and September. At which the north-east of the study domain contains the highest fractions. During September the importance increases in areas such as the Alps and Pyrenees. WG2 has the smallest feature importance over-all feature importance. Still, it has exceptional high fraction at areas such as the Alps and Spain. During November and May the Alps have a high percentage of WG2 feature importance and Spain has a high relevance for all months except of April, May and June.

The respiration feature importance has a relative noisy distribution over the study domain. For most of the areas RE has a fraction between 0.3 and 0.4. Only the agriculture land and Brittany has very low RE fractions. Still, one area is very distinctive, which is the grassland area in the centre of France. In this tile the fraction of RE is consistently high, except of March and April in which GPP has the highest relevance.

Table 4.23 gives a more clear overview about how the feature importance changes spatially and over time. Hence, gives an opportunity to highlight anomalies and trends more specifically. Here, the mean of feature fractions per cluster and seasons are illustrated. One highlight is that WG2 has no major impact on the prediction, especially for the vegetation covered land tiles. Agriculture Land, Pays Mixed coast and French Grassland are dominated GPP and RE. Boreal Alps, Broadleaf Forest and Needleleaf Forest show a very equally distributed feature importance as the average feature importance is 0.25. Still, the highest fractions (>0.30) are described by respiration.

In summary, for the curvature simulation only GPP has very dominant feature importance. However, solely for agriculture based land covers. Grassland shows a strong correlation to RE but does not contain fractions higher then 0.6. For all other cover types the features share a similar importance and show a weak seasonal trend compared to the slope and backscatter simulation.

	Fraction [-]															
	LAI				WG2				GPP				RE			
Cluster	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Agriculture Land	0.31	0.21	0.31	0.23	0.15	0.14	0.19	0.17	0.38	0.43	0.25	0.42	0.17	0.23	0.25	0.18
Pays Mixed Coast	0.33	0.24	0.28	0.25	0.17	0.15	0.23	0.15	0.28	0.37	0.27	0.34	0.22	0.23	0.22	0.25
Needleleaf Forest	0.25	0.22	0.28	0.22	0.22	0.25	0.22	0.24	0.25	0.28	0.21	0.27	0.28	0.25	0.30	0.26
French Grassland	0.25	0.17	0.24	0.20	0.20	0.17	0.17	0.21	0.24	0.37	0.25	0.29	0.31	0.29	0.34	0.30
Broadleaf Forest	0.27	0.26	0.26	0.30	0.20	0.17	0.12	0.20	0.21	0.30	0.26	0.22	0.31	0.27	0.36	0.29
Boreal Alps	0.25	0.28	0.28	0.29	0.24	0.28	0.12	0.24	0.21	0.22	0.27	0.22	0.30	0.21	0.33	0.26

Table 4.23: Curvature: Table to compare the mean feature importance fraction per cluster for different seasons. Cluster 0 = Agriculture Land, Cluster 5 = Pays Mixed Coast, Cluster 8 = Needleleaf Forest, Cluster 9 = French Grassland, Cluster 11 = Broadleaf Forest and Cluster 12 = Boreal Alps

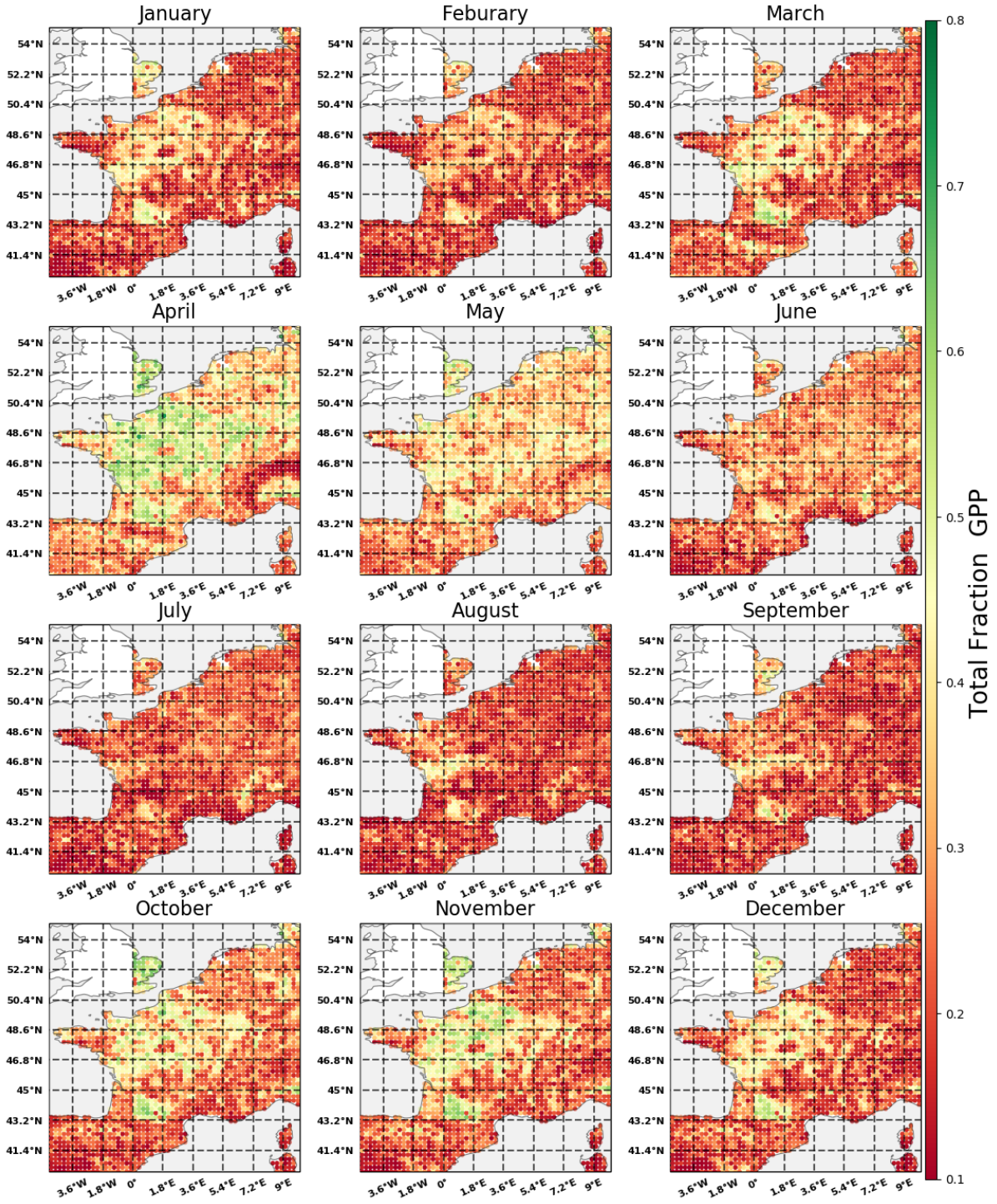

Figure 4.47: Curvature: GPP feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

Figure 4.48 shows the feature importance time series of cluster 0 (Agriculture Land) for each individual GPI and input variable. Again, the seasonality of the input variables drives the feature importance magnitudes. The cluster time series plots are illustrated in the Appendix A.5. In general, the magnitude of the feature importance is defined by the over-all importance of the feature. Here, the relevant features follow a strong seasonal cycle in its importance, whereas least important variables show a very small annual cycle. This pattern was also observed for backscatter and slope. All clusters show good robustness, except of the Boreal Alps cluster. In this exception the variance of the individual time series is rather high compared to the other clusters.

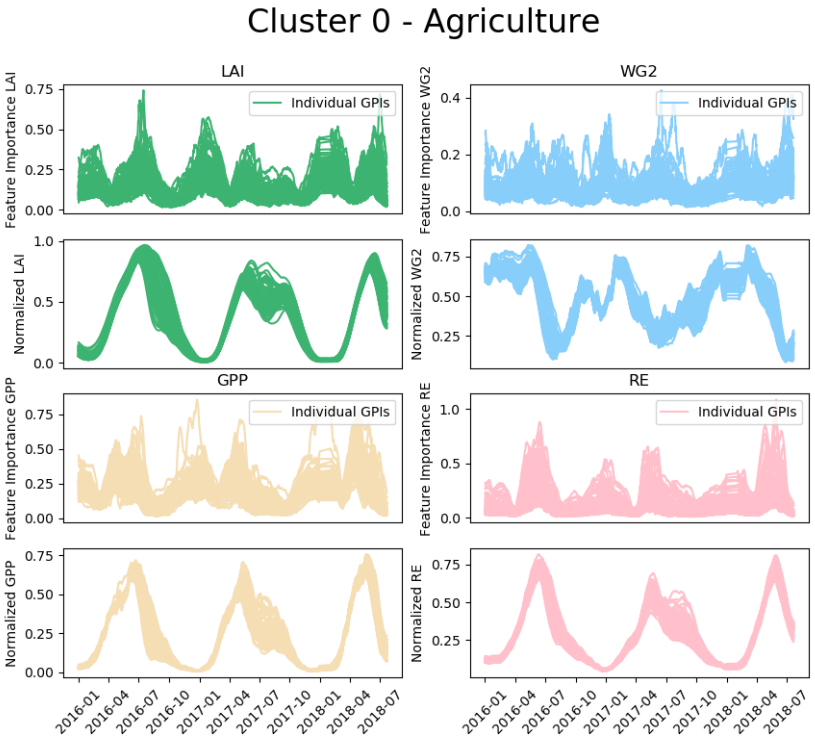

Figure 4.48: Curvature: Feature importance and land surface parameter time series for cluster 0 - Agriculture Land

4.5.5. Feature importance regarding vegetation types

Figure 4.49 shows the feature importance as a time series for seven vegetation types. Here, each feature is displayed as its own figure. This allows to investigate the feature importance dependency for different land cover types.

Leaf Area Index has the most influence for no vegetation land cover type. Here, it has two strong seasonal cycles, one in the MAM months and one from June to November. The second cycle is not as strong in magnitude but has a longer extend. Agriculture shows only one maximum per year and during the months April to July. The remaining vegetation types show a similar pattern as no vegetation, but with a lower magnitude. Soil moisture (WG2) mainly dominates the feature importance for needleleaf and no vegetation land cover types. Needleleaf shows a strong seasonal cycle with two peaks. The first peak occurs between November and February and the second around the MAM season. No Vegetation shows now seasonality in the feature importance of soil moisture. In general, it shows most relevance between April and November.

The GPP time series indicate that the simulations for grassland, no vegetation and agriculture are greatly influenced by GPP. All of them show a strong seasonal cycle. The other land cover types experience little impact from the GPP feature. Broadleaf, grassland and agriculture show a high importance magnitude for RE compared to the other features. All examples have two maximum magnitudes, one during the MAM months and one at the DJF season.

Table 4.24 gives a overview of the over-all feature importance per vegetation type. This high lights that broadleaf is most dependent on respiration, agriculture depends almost equally on all land surface variables, needleleaf and boreal grass is most influenced by soil moisture and no vegetation as well as shurb is mainly defined by LAI. GPP has a general high importance but is never the most relevant feature for a prediction.

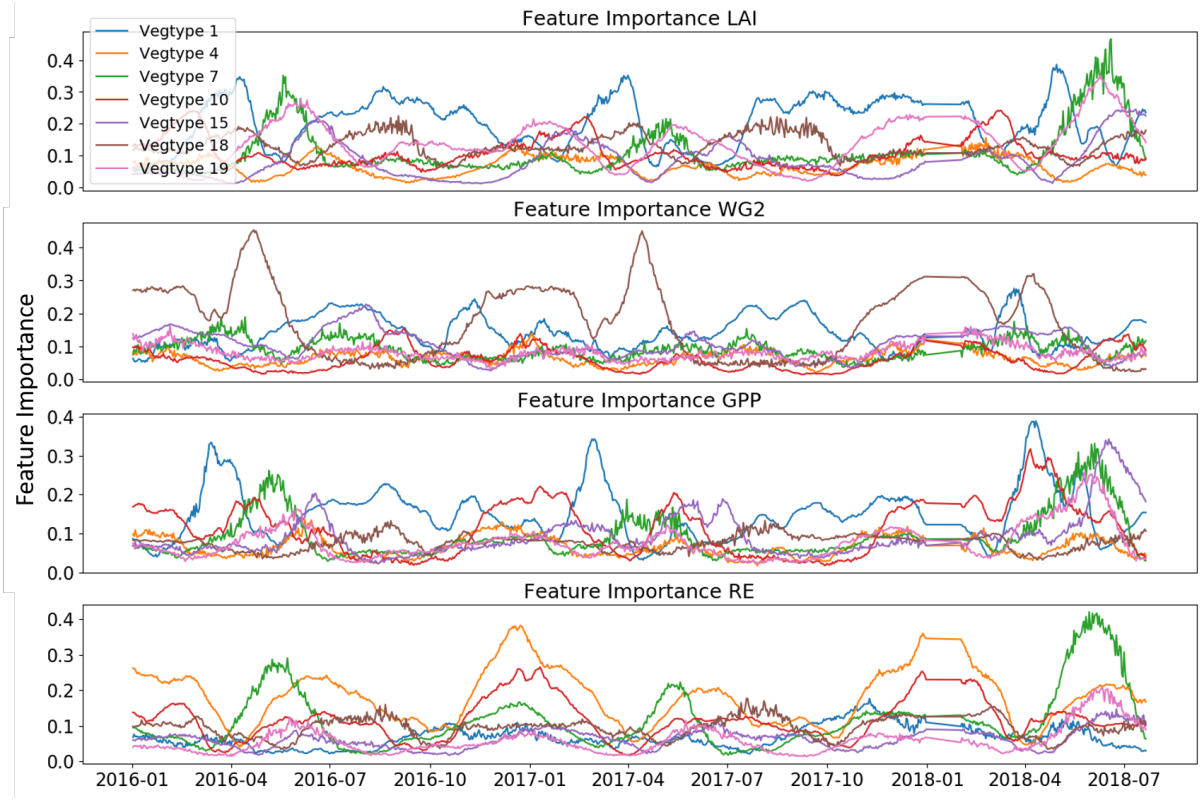

Figure 4.49: Curvature: Feature importance time series estimations for each feature (LAI, WG2, GPP and RE) for seven vegetation types: Vegtype 1 = No Vegetation, Vegtype 4 = Broadleaf, Vegtype 7 = Agriculture, Vegtype 10 = Grassland, Vegtype 15 = Needleleaf, Vegtype 18 = Boreal Grass and Vegtype 19 = Shrub

	Feature Importance Fraction			
Vegetation Type	LAI	WG2	GPP	RE
No Vegetation	0.35	0.26	0.26	0.13
Broadleaf	0.18	0.18	0.18	0.47
Agriculture	0.26	0.25	0.22	0.26
Grassland	0.28	0.15	0.27	0.29
Needleleaf	0.22	0.33	0.27	0.18
Boreal Grass	0.29	0.34	0.17	0.21
Shurb	0.37	0.26	0.22	0.15

Table 4.24: Curvature: Table to compare the mean feature importance fraction for seven vegetation types. Considering the time frame from 2016 to 2018.

4.5.6. Summary and discussion of curvature results

The following discussion is based on two research sub-questions and covers the results of the curvature analysis. Figure 4.50, Figure 4.51 and Table 4.25 are discussed. Figure 4.50 compares the min-max scaled variance of input and output data as well as RMSE values for each month of the year per cluster. The dashed black line represents the scaled RMSE values, the red line the scaled variance of the curvature signal and the colored lines the scaled variance of the input features. Figure 4.51 shows the feature importance per cluster for the prediction between 2016 and 2018. In this figure the green line represents LAI, blue WG2, pink RE and yellow GPP. The dots in Figure 4.50 indicate the individual values, as the lines don't represent continuous data points. Table 4.25 summarizes the over all performance and feature importance per cluster as well as the performance and feature importance per vegetation type.

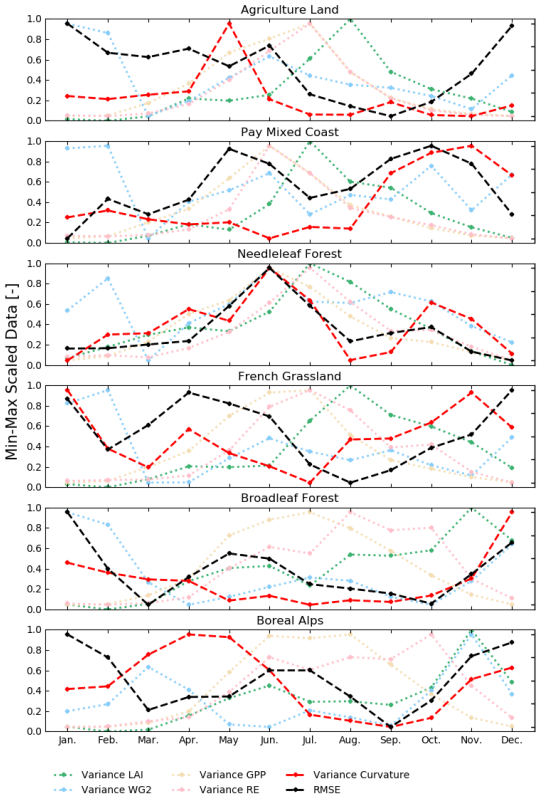

Figure 4.50: Curvature: The relative change in the variance for all input features (LAI, WG2, GPP, RE) as well as curvature and and RMSE plotted for each month and distinctive cluster. The grey line represents the RMSE calculated from the prediction vs observation, the grey dashed line represents the variance of curvature, the green the variance of LAI, the blue the variance of WG2, the yellow the variance of GPP and the pink line the variance of RE.

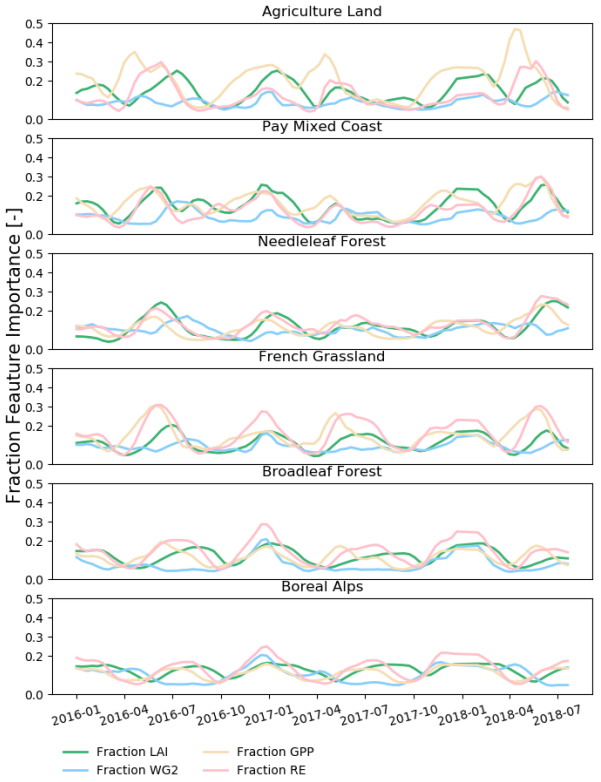

Figure 4.51: Curvature: The fraction of the total feature importance for all input features (LAI, WG2, GPP, RE) between 2016 and 2018 and distinctive cluster. The green line represents the feature importance fraction of LAI, the blue the importance of WG2, the yellow tthe importance of GPP and the pink line the importance of RE.

Cluster	Normalized RMSE	Most relevant Feature	No Vegetation	Broadleaf	Agriculture	Grassland	Needleleaf	Boreal Grass	Shurb
Agriculture Land	0.140047	GPP	0.09	0.14	0,59	0.10	0.02	0	0
Pays Mixed Coast	0.158206	GPP	0.08	0.06	0,23	0.30	0.04	0	0.03
Needleleaf Forest	0.141474	WG2	0.09	0.08	0.06	0.30	0.29	0.0	0.02
French Grassland	0.146604	RE	0.05	0.16	0,12	0.55	0.07	0	0.01
Broadleaf Forest	0.121751	WG2	0.05	0.31	0.08	0.32	0.13	0.0	0.0
Boreal Alps	0.116918	LAI	0.10	0,12	0.03	0.21	0.02	0.16	0.01
RMSE [σ]			0.0022	0.004	0.0026	0.0038	0.0024	0.0048	0.0025
Most relevant Features			LAI/WG2/GPP	RE	LAI/RE	RE/LAI	WG2/GPP	GPP/LAI	LAI/WG2

Table 4.25: Curvature: Table to illustrate the over all normalized RMSE and most important feature per cluster

How does the model performance vary in space and time?
The spatial performance metrics maps in Figure 4.39 of the curvature analysis subsection show a spatially inconsistent performance. However, once averaging the RMSE values over all GPIs within a distinctive clusters, the curvature RMSE values show a spatial consistency. The highest errors are found for the MAM season and the least errors for the winter months (DJF). Within the grassland and agriculture land related tiles the worst performance is to observe. The good performance of Boreal Alps might be due to the fact that curvature has a relative flat and seasonal signal within this area. This is consistent with the curvature time series plot of the vegetation type no vegetation as it fluctuates only between 0.002 and -0.001 db/deg^2, see Figure 4.10. This suggest that the low RMSE values originate from predicting the mean of the signal rather than the actual value. The results of the Boreal Alps need to be viewed critically as the DNN simulates a no vegetation surface with vegetation states.

Agriculture Land has over all very high RMSE values, see Figure 4.39 but the individual agriculture GPI shows a fairly low RMSE value. This inconsistency might be because the model has again a high variation of RMSE values within the cluster - see Figure 4.39. That means that the cluster is not robust and it is difficult to represent by an individual GPI. The over all high RMSE values can be explained by the high variability of the curvature signal for this cluster, see Figure 4.6. Here it has the highest seasonal variability compared to the other vegetation types.

The high RMSE values for French Grassland and Pays Mixed Coast, see table 4.25 are consistent with the high RMSE values for the distinct grassland vegetation type, see Figure 4.43. This can be explained as grassland shows a very flat signal with almost no seasonal cycle. Therefore, the signal is dominated by the inter-annual variability, which is hard to capture by the model. The same counts for the Needleleaf Forest cluster, as it shows a low seasonality and high inter-seasonal variance and therefore high RMSE values. Boradleaf Forest on the other hand has a high seasonality which then leads to lower RMSE values as the inter annual changes are rather low.

Figure 4.50 does not show a clear pattern to explain the performance as the RMSE values barely correlate with the input and output variance. The Spearman correlation matrix between the input and output variance as well as RMSE values per cluster are displayed in A.5. The highest Spearman correlation values are found between the curvature variation and RMSE values of which Needleleaf shows the highest ranking value of 0.68. Still, the RMSE variance is not solely explained by the curvature variance. The RMSE variance also shows strong correlation to the variability of the input data. This changing dependency is rather complex as for the curvature simulation no feature is dominantly important in time and space.

In summary, the inter annual variability and seasonal cycle of the input and output data has a strong influence on the magnitude of the RMSE values. Especially, the variance of the curvature data shows the highest positive correlation to the RMSE values. In best predictions the model is solely able to capture the seasonal cycle but the magnitude of the seasonality as well inter-seasonal variation are not captured. Therefore, the model solely simulates the spatial distribution of curvature in the study domain by representing the seasonal trends. This poor performance can be caused by a too low complexity of the model, too little training data, a weak correlation between the input data and the output data but also due to the fact that curvature is a very noisy signal. For the soil moisture retrieval using backscatter curvature was generally treated as noise rather than a source of information. Only recently the research team from Steele-Dunne found a distinct seasonality in the signal and suggested that curvature might be related to structural changes and therefore scattering mechanisms. The results of this research confirm that there is a seasonal correlation between the vegetation states and curvature, however it also shows that curvature contains a lot of noise and inter annual variability.

How does the feature importance vary in space and time?
Table 4.25 shows the most important features per cluster for the curvature simulation. Here the most important features are very diverse and change for each cluster. The over-all most important feature is LAI as all clusters always have a high feature importance, whereas for example GPP has only high importance for certain periods of the year and for the rest it loses on importance. In general no cluster has a strong domination of a single feature as all features have similar importance. Only Agriculture Land shows high sensitivity to GPP. This is consistent as GPP is a strong indicator of biomass assimilation. Soil moisture has least relevance for the predictions for all clusters, which is consistent with the assumption of previous research that curvature is

correlated to vegetation change [43].

This suggests that LAI creates a baseline of the curvature signal, by introducing the vegetation seasonality component. Whereas the other features are used by the deep neural network to estimate short term changes. However, this is barely achieved by any GPI as none of the models are able to reproduce the anomalies of the curvature data.

The dependency of Boreal Alps on RE is an artefact as there should be no correlation between a vegetation state and bare rock or no vegetation. An explanation why this correlation is captured is that RE is a product of soil moisture and temperature and the simulation might uses the variation in temperature of the area rather the actual scattering mechanisms. This shows again, that deep neural networks need to be treated carefully as they find correlations even though there are none.

In general all feature importance signals have a weak seasonal cycle for all clusters with almost no inter seasonal variability. This again shows that the simulation is based on the seasonal trends of the input data rather then the inter annual variations. For agricultural land the GPP feature importance shows some inter annual variation, which could be related to the very distinct structural change as in April/May the growing season starts. This correlation between structural change and scattering mechanism was already suggested by a previous study [43]. Still, the prediction accuracy is very low, which means that the feature importance interpretation needs to be treated carefully. Furthermore, Figure A.66 shows very high Spearman correlation rank values between LAI, GPP and RE, which means that the feature importance might actually be a mixture of all three components and partly explains why none of the features are dominant for a cluster.

In summary, the curvature simulation is most dependent on the RE, LAI and GPP input features. This clearly indicates that the signal contains information about change in the vegetation phenology. Still, a clear distinction between the features is not possible, as the fraction distribution is very similar but also the features are quite correlated to each other. In general all vegetation states contain information to describe at least the seasonality of the curvature signal but do not contain enough information to explain inter annual variations. This leads to the assumption that the feature importance has a strong dependency on static components rather then the inter-seasonal variation in the curvature data. The over all weak model performance shows, however, that the outcome should be viewed critically.

4.6. Summary of Results

The backscatter simulations show a over all good performance and the MASE values indicate that the deep learning model is performing better than a native forecasting method. All investigated vegetation types show a good accordance between observation and prediction. In general, only the extreme magnitudes are wrongly estimated. This error might stem from the high inter annual variance of the backscatter observations. Therefore, an adaptive model structure is suggested, which increases the complexity when the variance is high. The most important feature is soil moisture of which the magnitude is defined by the variance of the input data. This exploration of WG2 as most important feature is very valuable as it shows consistency with previous research, which show a strong empirical relationship between soil moisture and backscatter [51][53]. The feature importance further reveals that the backscatter also is sensitive to biomass change as GPP is important especially for agriculture land. Here the growing and harvesting season is detectable as the feature importance follows a similar seasonality as GPP.

The cross validation indicates that the model does not strongly depended on the year which is used as a training set. Still, to generate a more robust interpretation, an average over all possible training-test year combinations is suggested.

The slope simulations show a moderate model performance, at which the MASE values indicate a general worse performance compared to a native forecasting method. Here the model is able to capture the spatial and seasonal trends of the slope data. Still, it fails to predict the anomalies and small term variation. In order to improve the simulation a higher complexity depending on the inter annual variance of the slope data is suggested. The variance of the slope data seems to have the highest influence on the model performance.

The most important input feature for the slope prediction is the gross primary production. This indicates that slope contains information about the vegetation phenology as GPP is an indicator for vegetation assimilation and therefore biomass accumulation. The correlation between GPP and hence water dynamics and biomass wtih slope was already suggested by Steele-Dunne et al. [43]. However, this clear relationship is solely established for very distinct cluster such as the Agriculture Land and French Grassland. As for these clusters the feature importance fraction is significantly high.

The cross validation shows that the simulation performance depends strongly on the training-testing year combinations as it has the lowest Spearman correlation coefficent between the different simulations compared to backscatter and curvature. This suggests that the CDNN needs to be trained with a boot-strapping method to average all weights and biases over all different training scenarios.

The curvature simulations show a poor model performance, at which the native forecasting method performs up to 16 times better than the deep neural network. In general, the model is able to capture the spatial distribution of curvature but fails to simulate the right magnitude and inter-seasonal variations. This indicates either a low correlation between the input and output data or that the model structure is not complex enough to capture these variations. Another reason for a low performance could be that the curvature signal is rather noisy which makes it difficult to find a stable correlation.

The most important feature for the curvature simulations is LAI. LAI is related to the vegetation density as well as canopy height and often used as an indicator about for the biomass above the ground [49]. Therefore, this relationship could indicate that curvature contains some valuable information about LAI and vertical structure of the canopy. This correlation between vertical structure and curvature was already suggested by Steele-Dunne et al. in a study published 2019 [43]. Still, the feature importance is fairly equally distributed between all vegetation states, which makes a confident interpretation very difficult. It solely indicates a relationship toward phenology changes rather then soil moisture related properties.

The cross-validation indicates that the three different simulations have strong correlation. Still, in order to generate a more general model an average over all possible training-testing combinations is suggested.

5

Conclusions and Recommendations

The research aims to establish a surrogate model to close the gap between the land surface model and the ASCAT satellite observations, as currently the ASCAT observations are not used to assimilate with land surface models. This is a loss of opportunity as the ASCAT and ESCAT measurements together generate a climatological data set of high quality and temporal consistency. The current possibilities to assimilate the satellite observations with LSM's are complex Radiative Transfer Models, which require complicated input variables of the vegetation such as size, shape, height, thickness and orientation of the canopy but also the dielectric properties. Additionally, RTM's are not based on the actual output of the land surface model, as the LSM's simulate vegetation parameters such as leaf area index, soil moisture, gross primary production, temperature and respiration, which makes RTM's less suitable to act as a surrogate model. The suggested method to simulate ASCAT observations are therefore deep neural networks. Deep neural networks are able to capture every highly non-linear relationship by using solely the inputs from the LSM.

5.1. Conclusions

For this study a regular feed forward deep neural network is used to simulate the backscatter signal of the ASCAT instrument, whereas slope and curvature are simulated by a convolutional deep neural network.

In order to answer the following research question, the model is evaluated in time and space. This allows to correlate the performance to land cover types but also to seasons. Furthermore, a black-box model interpretation analysis, called DeepSHAP, is used to investigate the feature importance for each ASCAT observation.

Can deep neural networks be used to establish a surrogate model to close the gap between the land surface model and the ASCAT observations?

The results show that the backscatter, slope and curvature simulations are able to simulate ASCAT observations using only LSM variables as input. The time and space analysis leads to the conclusion that the magnitude of error highly correlates with the variance of the means of ASCAT data. At which a high inter annual variability leads to high RMSE values. The variation of the data is dependent on the climate zone of the grid point and on the land cover type fractions. It was found that for larger variance in the data and land cover fractions, the GPI was more complex, which leads to the conclusion that the model architecture needs to be adjusted depending on the variance of the in- and output data as well as land cover types. In this research only one architecture was used to simulate backscatter or slope and curvature.

The best results were found in the backscatter simulation, whereas moderate results for slope were found and the weakest results for curvature were found. All three signals are processed, however backscatter is an instantaneous measure whereas slope and curvature are aggregated in time. This indicates that during the kernel smoothing process information is lost. The loss of information for curvature seems more drastic as it is the second derivative of the backscatter signal.

The cross validation showed a high performance dependency on the initialization of the deep neural network but also on the choice of training-testing data sets. Therefore, an average of the trained weights and biases for each training-testing data combination is suggested. Another reason for this dependency is the lack of training data, as DNN's performance highly depends on the amount of data.

The feature importance analysis showed a clear correlation between specific land surface variables and ASCAT observations. The results suggest a high dependency of backscatter on soil moisture and a strong correlation between vegetation parameters with slope and curvature measurements. This confirms the assumption of the latest research regarding slope and curvature, that those two signals are highly correlated to vegetation processes. The results clearly provide evidence that slope contains information about the vegetation biomass, whereas curvature reflects mostly LAI and therefore structural changes.

All feature importance results need to be treated carefully as there are high correlations between GPP and RE. This dependency could potentially falsify the feature importance distribution as these variables could replace each other. Still, feature importance values with fractions above 0.4 can be taken more seriously. Especially the simulation for backscatter shows significant importance for WG2 and GPP, which confirms the current knowledge about backscatter. This gives confidence that machine learning has next to a good performance also physical integrity. The downside of this methodology is that it is rather biased due to the input of the user who tells the algorithm to find a relationship between any given input and output. This means that the DNN possibly has good results but with invalid motivation. For example, areas with no vegetation or rock show good results but the correlation between vegetation states is not logical as there is no vegetation.

The results and conclusions are substantial as they allow for the first time to relate actual vegetation parameters to slope and curvature signals. It additionally proves that deep neural networks are possible choice to act as a surrogate model. The following sub section gives suggestions and recommendations for further research to improve the model interpretation and performance.

5.2. Recommendations

The most common way to improve the performance of a deep learning model is to increase the training data, as the performance is highly dependent on the amount of data. In this study solely the descending data set is used. Including the ascending data set could increase the model performance by doubling the available data. Another way to increase the training data is to train one model with multiple grid points. As the vegetation type fractions as well as local input/output variance is defining the complexity, GPI's with similar attributes are suggested to train the same model. The t-SNE clustering method could be applied to create similar clusters for this method. Cross-Validation can be further used to increase the performance. Here the model is trained for each testing-training combination and subsequently all biases and weights are averaged over each model run. Another solution to increase the model performance is to optimize the hyper-parameters of the deep neural network. Here, the most efficient solution is the Bayesian Optimization.

The simulation performance can further be improved by implementing adaptive land cover tiles for the land surface model. This is necessary as the land cover fractions are constant within the LSM even though there might be changes in the land cover distribution. For example, agriculture has harvesting seasons, which are not considered in the LSM because it is assumed that the crops whiter naturally. This adaptation is very important as the model is only as good as the data provided.

In order to validate the feature importance a constituency check with other "black-box" interpretation models is suggested. These interpretation models need to be treated carefully as they are a very recently developed and not yet fully evaluated. LIME is suggested in previous research as a sophisticated alternative to DeepSHAP and is recommended as an comparison algorithm [32]. Another way to confirm the robustness and actual importance of features is to implement an ablation study. Here the model performance and feature importance is evaluated by neglecting or adding systematically input features. This dimension reduction allows to neglect all variables which are correlated to each other, which gives the feature importance interpretation more integrity.

Another reason for a bad model performance could be that the LSM states have too little correlation with the ASCAT signals. Including new variables to the DNN would be therefore of high interest. For example interception and temperature are potential LSM states to further include. Additionally, most of the simulations fail to predict the magnitude of the signals, which might be caused by normalization and scaling of the input data using the whole study domain. Therefore, it is recommended to scale the data per GPI rather over the whole study domain.

Bibliography

[1] Albergel, C and Calvet, J-C and Gibelin, A-L and Lafont, S and Roujean, J-L and Berne, C and Traullé, O and Fritz, N. Observed and modelled ecosystem respiration and gross primary production of a grassland in southwestern France. *Biogeosciences*, 7(5):1657–1668, 2010.

[2] Craig Anderson, J Figa, H Bonekamp, JJW Wilson, Jeroen Verspeek, Ad Stoffelen, and Marcos Portabella. Validation of backscatter measurements from the advanced scatterometer on Metop-A.

[3] Bergstra, James and Bengio, Yoshua. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

[4] Boone, Aaron and Samuelsson, Patrick and Gollvik, Stefan and Napoly, Adrien and Jarlan, Lionel and Brun, Eric and Decharme, Bertrand. The interactions between soil–biosphere–atmosphere land surface model with a multi-energy balance (ISBA-MEB) option in SURFEXv8–Part 1: Model description. 2017.

[5] Brocca, Luca and Crow, Wade T and Ciabatta, Luca and Massari, Christian and De Rosnay, Patricia and Enenkel, Markus and Hahn, Sebastian and Amarnath, Giriraj and Camici, Stefania and Tarpanelli, Angelica and others. A review of the applications of ASCAT soil moisture products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):2285–2306, 2017.

[6] Busseti, Enzo and Osband, Ian and Wong, Scott. Deep learning for time series modeling. *Technical report, Stanford University*, pages 1–5, 2012.

[7] Chai, Tianfeng and Draxler, Roland R. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 7(3):1247–1250, 2014.

[8] Chung, Junyoung and Gulcehre, Caglar and Cho, KyungHyun and Bengio, Yoshua. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[9] Jose Maria de Juana Gamo, Dorothee Diebel, Pier Luigi Righetti, and Wael El-Dali. Metop-B launch and phasing strategy in the presence of Metop-A. In *SpaceOps 2010 Conference Delivering on the Dream Hosted by NASA Marshall Space Flight Center and Organized by AIAA*, page 2021, 2010.

[10] Dickinson, Robert E. Land-atmosphere interaction. *Reviews of Geophysics*, 33(S2):917–922, 1995.

[11] Dara Entekhabi, Eni G Njoku, Peggy E O'Neill, Kent H Kellogg, Wade T Crow, Wendy N Edelstein, Jared K Entin, Shawn D Goodman, Thomas J Jackson, Joel Johnson, et al. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, 98(5):704–716, 2010.

[12] Fairbairn, D and Barbu, AL and Mahfouf, J-F and Calvet, J-C and Gelati, E. Comparing the ensemble and extended Kalman filters for in situ soil moisture assimilation with contrasting conditions. *Hydrology and Earth System Sciences*, 19(12):4811–4830, 2015.

[13] Figa-Saldaña, J and Wilson, J JW and Attema, E and Gelsthorpe, R and Drinkwater, MR and Stoffelen, A. The advanced scatterometer (ASCAT) on the meteorological operational (Metop) platform: A follow on for European wind scatterometers. *Canadian Journal of Remote Sensing*, 28(3):404–412, 2002.

[14] Forman, Barton A and Reichle, RH and Rodell, Matthew. Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. *Water Resources Research*, 48(1), 2012.

[15] Forman, Barton A and Reichle, Rolf H. Using a support vector machine and a land surface model to estimate large-scale passive microwave brightness temperatures over snow-covered land in North America. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(9):4431–4441, 2014.

[16] Forman, Barton A and Reichle, Rolf H and Derksen, Chris. Estimating passive microwave brightness temperature over snow-covered land in North America using a land surface model and an artificial neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1):235–248, 2013.

[17] Friesen, Jan Christoph. Regional vegetation water effects on satellite soil moisture estimations for West Africa. *Ecology and Development Series 63*, 2008.

[18] Frison, PL and Mougin, E. Monitoring global vegetation dynamics with ERS-1 wind scatterometer data. *Remote Sensing*, 17(16):3201–3218, 1996.

[19] Frison, PL and Mougin, E and Hiernaux, P. Observations and interpretation of seasonal ERS-1 wind scatterometer data over northern Sahel (Mali). *Remote Sensing of Environment*, 63(3):233–242, 1998.

[20] Ge, Lingling and Hang, Renlong and Liu, Yi and Liu, Qingshan. Comparing the Performance of Neural Network and Deep Convolutional Neural Network in Estimating Soil Moisture from Satellite Observations. *Remote Sensing*, 10(9):1327, 2018.

[21] Ghosal, Sambuddha and Blystone, David and Singh, Asheesh K and Ganapathysubramanian, Baskar and Singh, Arti and Sarkar, Soumik. An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences*, 115(18):4613–4618, 2018.

[22] Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron. *Deep learning*. MIT press, 2016.

[23] Sebastian Hahn, Christoph Reimer, Mariette Vreugdenhil, Thomas Melzer, and Wolfgang Wagner. Dynamic characterization of the incidence angle dependence of backscatter using Metop ASCAT. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):2348–2359, 2017.

[24] Hall, Patrick. On the Art and Science of Machine Learning Explanations. *arXiv preprint arXiv:1810.02909*, 2018.

[25] journal=Journal of Geophysical Research: Biogeosciences He, Liming and Chen, Jing M and Liu, Jane and Bélair, Stéphane and Luo, Xiangzhong. Assessment of SMAP soil moisture for global simulation of gross primary production. 122(7):1549–1563, 2017.

[26] Hochreiter, Sepp and Schmidhuber, Jürgen. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[27] Hyndman, Rob J and Koehler, Anne B. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

[28] Isikdogan, Furkan and Bovik, Alan C and Passalacqua, Paola. Surface water mapping by deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11):4909–4918, 2017.

[29] Le Moigne, P and Boone, A and Calvet, JC and Decharme, B and Faroux, S and Gibelin, AL and Lebeaupin, C and Mahfouf, JF and Martin, E and Masson, V and others. SURFEX scientific documentation. *Note de centre (CNRM/GMME), Météo-France, Toulouse, France*, 2009.

[30] Li, Xiang and Peng, Ling and Hu, Yuan and Shao, Jing and Chi, Tianhe. Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, 23(22):22408–22417, 2016.

[31] Lippmann, Richard P. An introduction to computing with neural nets. *IEEE Assp magazine*, 4(2):4–22, 1987.

[32] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, pages 4765–4774, 2017.

[33] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[34] Moreira, Alberto and Prats-Iraola, Pau and Younis, Marwan and Krieger, Gerhard and Hajnsek, Irena and Papathanassiou, Konstantinos P. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1):6–43, 2013.

[35] Naeimi, Vahid and Scipal, Klaus and Bartalis, Zoltan and Hasenauer, Stefan and Wagner, Wolfgang. An improved soil moisture retrieval algorithm for ERS and Metop scatterometer observations. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):1999–2013, 2009.

[36] Nielsen, Michael A. *Neural networks and deep learning*, volume 25. Determination press San Francisco, CA, USA, 2015.

[37] Peel, Murray C and Finlayson, Brian L and McMahon, Thomas A. Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences discussions*, 4(2):439–473, 2007.

[38] Xueheng Qiu, Le Zhang, Ye Ren, Ponnuthurai N Suganthan, and Gehan Amaratunga. Ensemble deep learning for regression and time series forecasting. In *2014 IEEE symposium on computational intelligence in ensemble learning (CIEL)*, pages 1–6. IEEE, 2014.

[39] Reichle, Rolf H and McLaughlin, Dennis B and Entekhabi, Dara. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1):103–114, 2002.

[40] Ribeiro, Marco Tulio and Singh, Sameer and Guestrin, Carlos. Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144. ACM, 2016.

[41] Schober, Patrick and Boer, Christa and Schwarte, Lothar A. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768, 2018.

[42] Snoek, Jasper and Larochelle, Hugo and Adams, Ryan P. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

[43] Susan C Steele-Dunne, Sebastian Hahn, Wolfgang Wagner, and Mariette Vreugdenhil. Investigating vegetation water dynamics and drought using Metop ASCAT over the North American Grasslands. *Remote Sensing of Environment*.

[44] Steele-Dunne, Susan C and Friesen, Jan and van de Giesen, Nick. Using diurnal variation in backscatter to detect vegetation water stress. *IEEE Transactions on Geoscience and Remote Sensing*, 50(7):2618–2629, 2012.

[45] Stoffelen, Ad and Aaboe, Signe and Calvet, Jean-Christophe and Cotton, James and De Chiara, Giovanna and Saldana, Julia Figa and Mouche, Alexis Aurélien and Portabella, Marcos and Scipal, Klaus and Wagner, Wolfgang. Scientific developments and the EPS-SG scatterometer. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):2086–2097, 2017.

[46] Irene E Teubner, Matthias Forkel, Martin Jung, Yi Y Liu, Diego G Miralles, Robert Parinussa, Robin van der Schalie, Mariette Vreugdenhil, Christopher R Schwalm, Gianluca Tramontana, et al. Assessing the relationship between microwave vegetation optical depth and gross primary production. *International journal of applied earth observation and geoinformation*, 65:79–91, 2018.

[47] Thomas Melzer. Vegetation modelling in WARP 6.0. In: Proc. *EUMETSAT Meteorological Satellite Conf., Vienna, Austria*, pages 1–7, 2013.

[48] F Ulaby. Radar measurement of soil moisture content. *IEEE Transactions on Antennas and propagation*, 22(2):257–265, 1974.

[49] Fawwaz T Ulaby. Microwave response of vegetation. *Advances in Space Research*, 1(10):55–70, 1981.

[50] Ulaby, Fawwaz T and Aslam, Adnan and Dobson, Myron C. Effects of vegetation cover on the radar sensitivity to soil moisture. *IEEE Transactions on Geoscience and Remote Sensing*, (4):476–481, 1982.

[51] Ulaby, Fawwaz Tayssir and Long, David G and Blackwell, William J and Elachi, Charles and Fung, Adrian K and Ruf, Chris and Sarabandi, Kamal and Zebker, Howard A and Van Zyl, Jakob. *Microwave radar and radiometric remote sensing*, volume 4. University of Michigan Press Ann Arbor, 2014.

[52] Mariette Vreugdenhil, Sebastian Hahn, Thomas Melzer, Bernhard Bauer-Marschallinger, Christoph Reimer, Wouter Arnoud Dorigo, and Wolfgang Wagner. Assessing vegetation dynamics over mainland Australia with Metop ASCAT. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):2240–2248, 2017.

[53] Wagner, Wolfgang and Hahn, Sebastian and Kidd, Richard and Melzer, Thomas and Bartalis, Zoltan and Hasenauer, Stefan and Figa-Saldaña, Julia and de Rosnay, Patricia and Jann, Alexander and Schneider, Stefan and others. The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*, 22(1):5–33, 2013.

[54] Wagner, Wolfgang and Lemoine, Guido and Borgeaud, Maurice and Rott, Helmut. A study of vegetation cover effects on ERS scatterometer data. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2): 938–948, 1999.

[55] Wagner, Wolfgang and Lemoine, Guido and Rott, Helmut. A method for estimating soil moisture from ERS scatterometer and soil data. *Remote sensing of environment*, 70(2):191–207, 1999.

[56] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline.

[57] Wismann, Volkmar R and Boehnke, Kai and Schmullius, Christiane. Monitoring ecological dynamics in Africa with the ERS-1 scatterometer. In *1995 International Geoscience and Remote Sensing Symposium, IGARSS'95. Quantitative Remote Sensing for Science and Applications*, volume 2, pages 1523–1525. IEEE, 1995.

A Appendix

A.1. Vegetation Type Clustering

The following figures display the box plots representing the vegetation type distribution for all 12 clusters. Still, only five clusters are used for the analysis: 0 -Agriculture,5 Pays Mixed Coast,8 Needleleaf Forest,9 French Grassland,11 Broadleaf Forest,12 Boreal Alps.

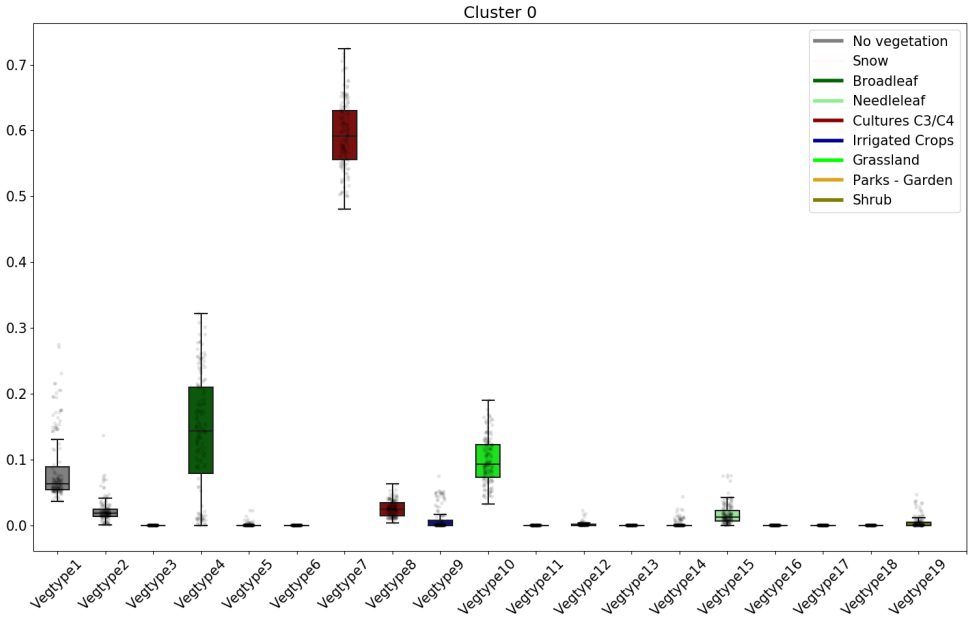

Figure A.1: Vegetation Fraction Box plot for Cluster 0 - Reference name: Agriculture Land

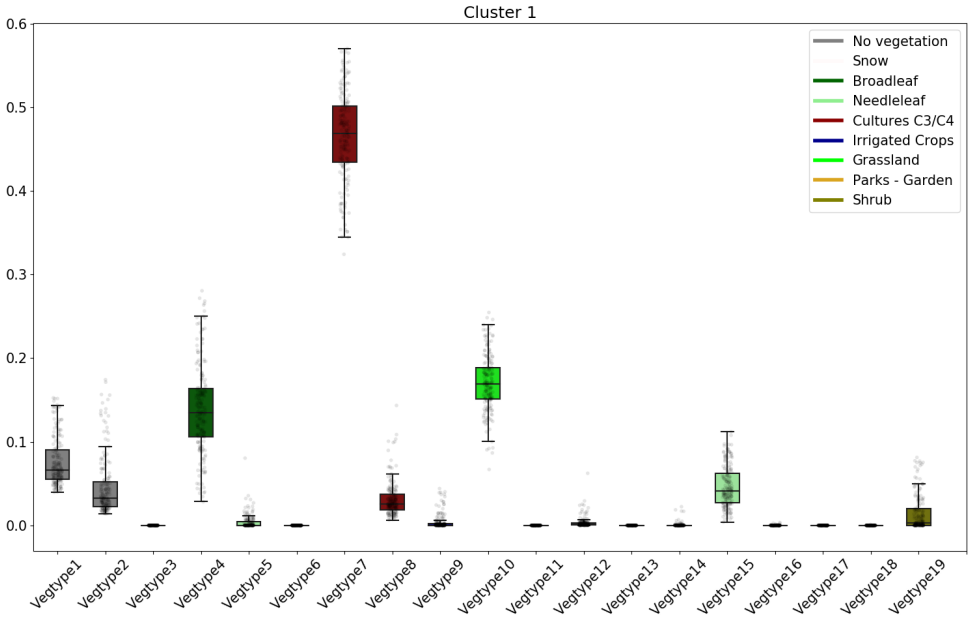

Figure A.2: Vegetation Fraction Box plot for Cluster 1

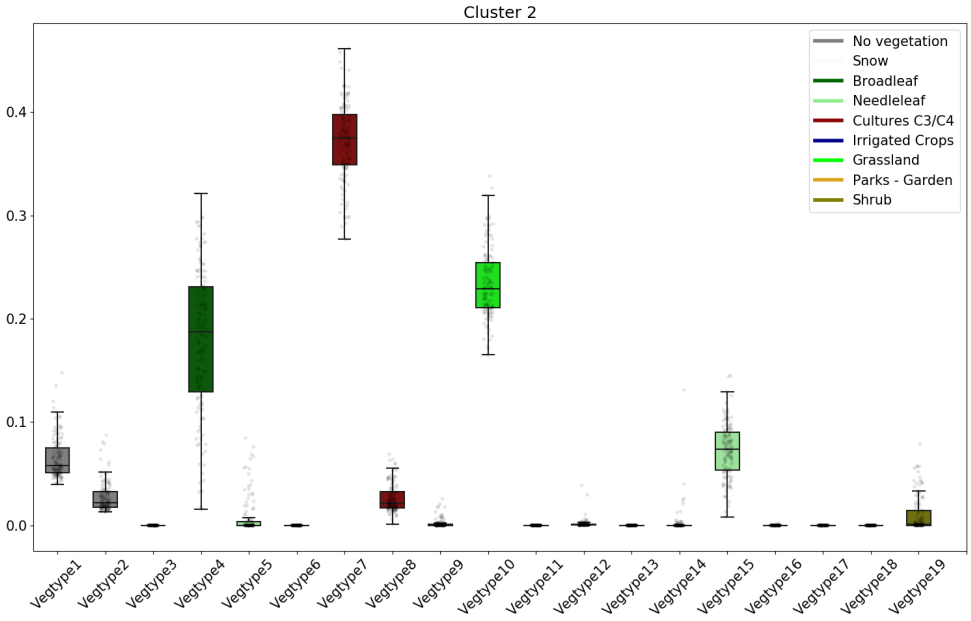

Figure A.3: Vegetation Fraction Box plot for Cluster 2

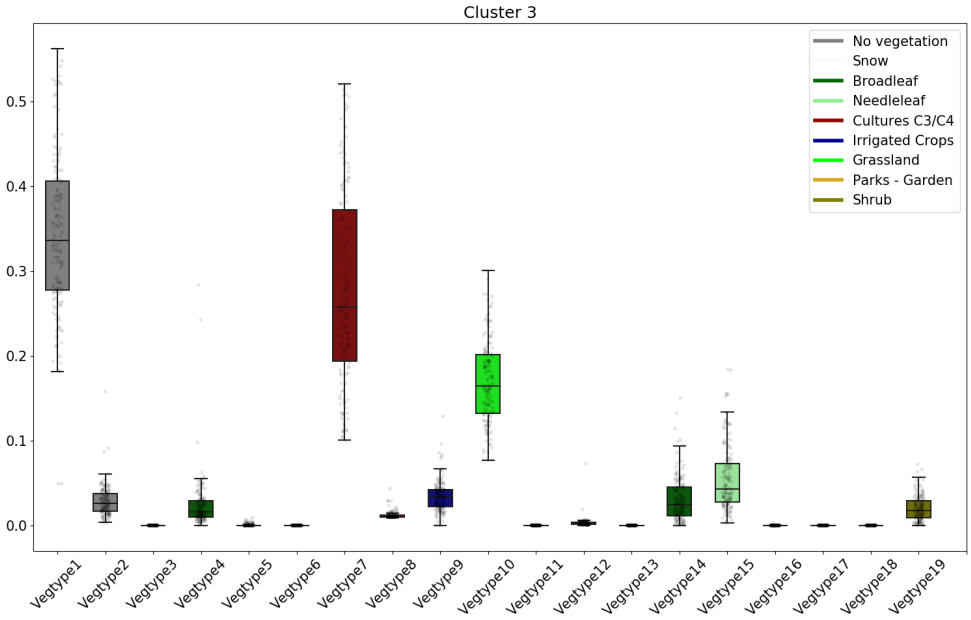

Figure A.4: Vegetation Fraction Box plot for Cluster 3

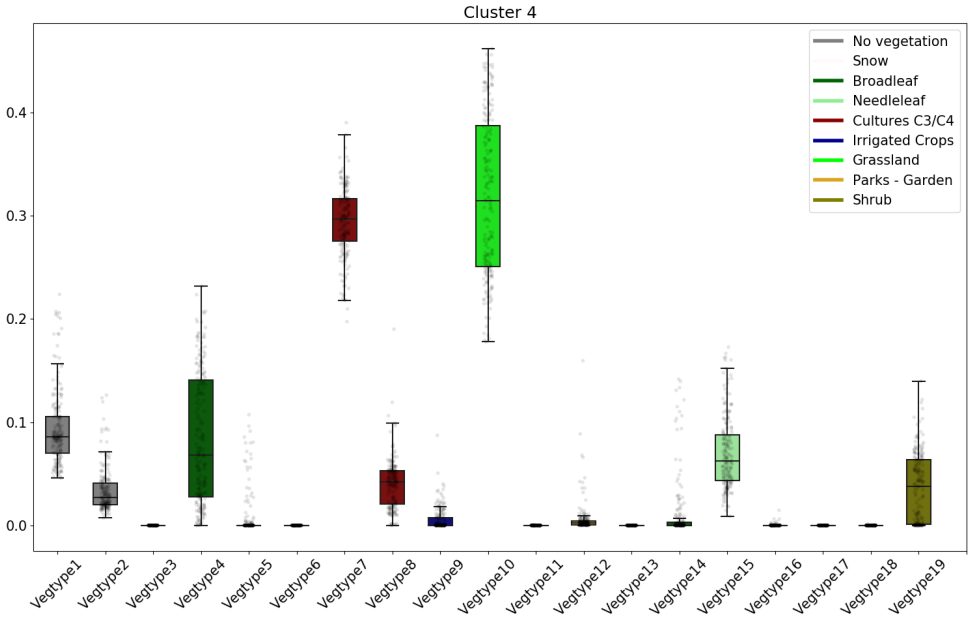

Figure A.5: Vegetation Fraction Box plot for Cluster 4

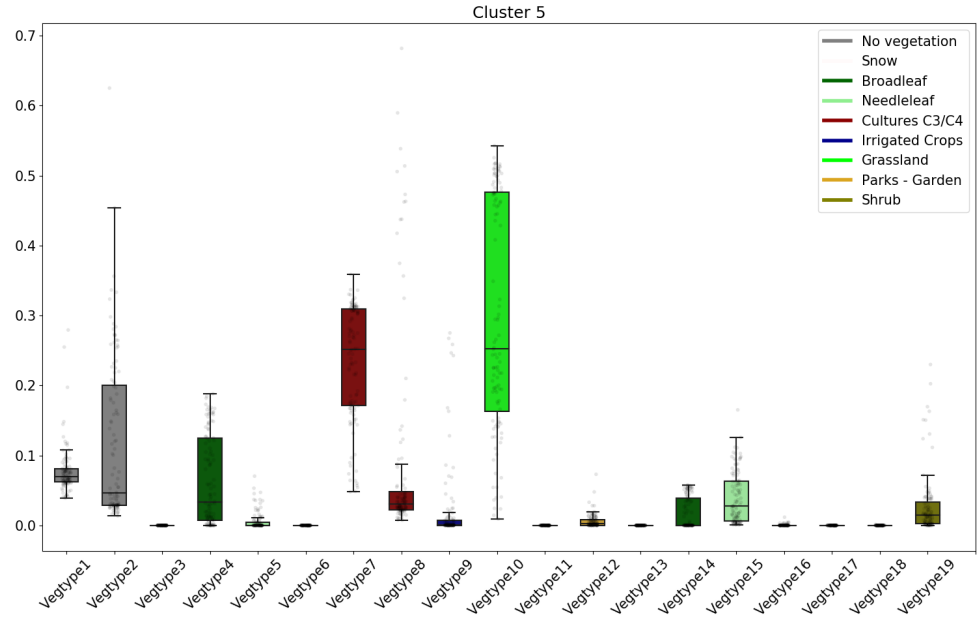

Figure A.6: Vegetation Fraction Box plot for Cluster 5 - Reference name: Pays Mixed Coast

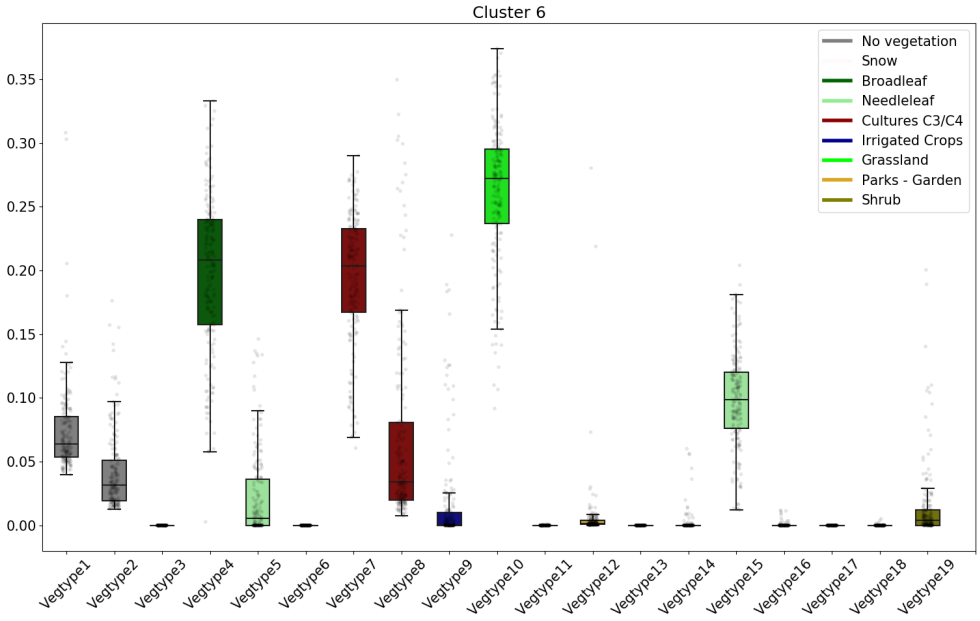

Figure A.7: Vegetation Fraction Box plot for Cluster 6

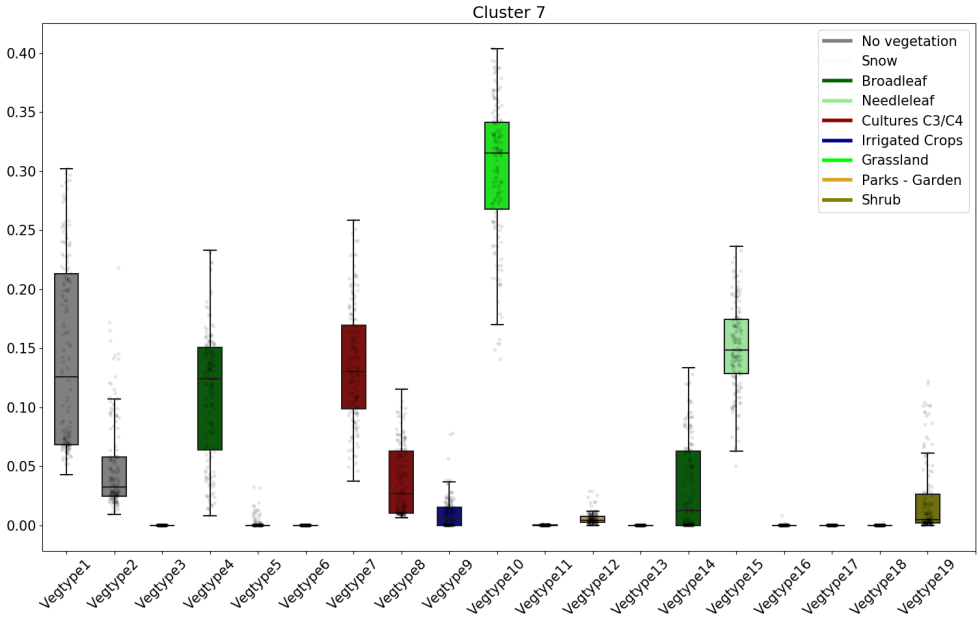

Figure A.8: Vegetation Fraction Box plot for Cluster 7

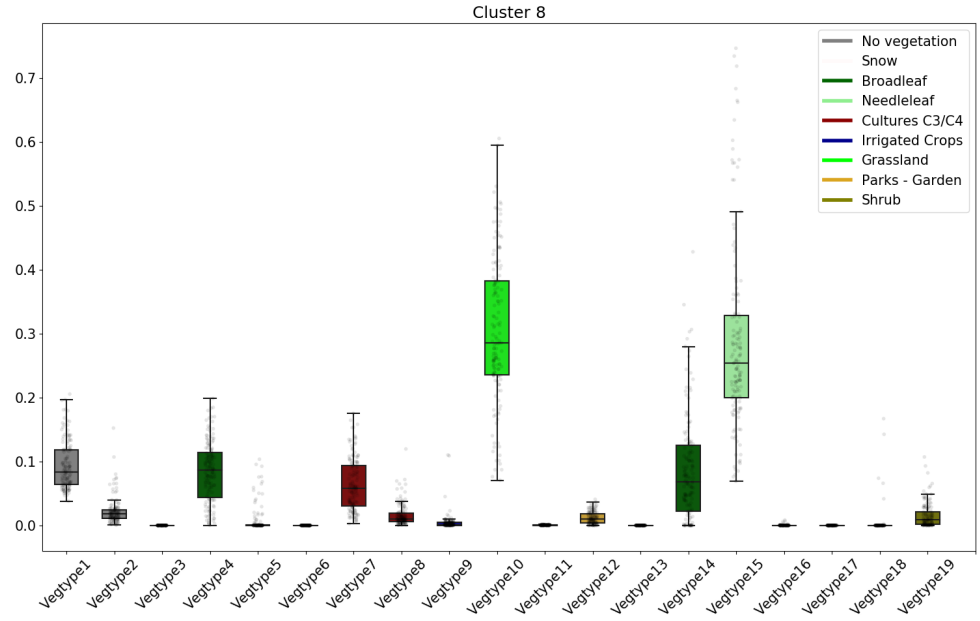

Figure A.9: Vegetation Fraction Box plot for Cluster 8 - Reference name: Needleleaf Forest

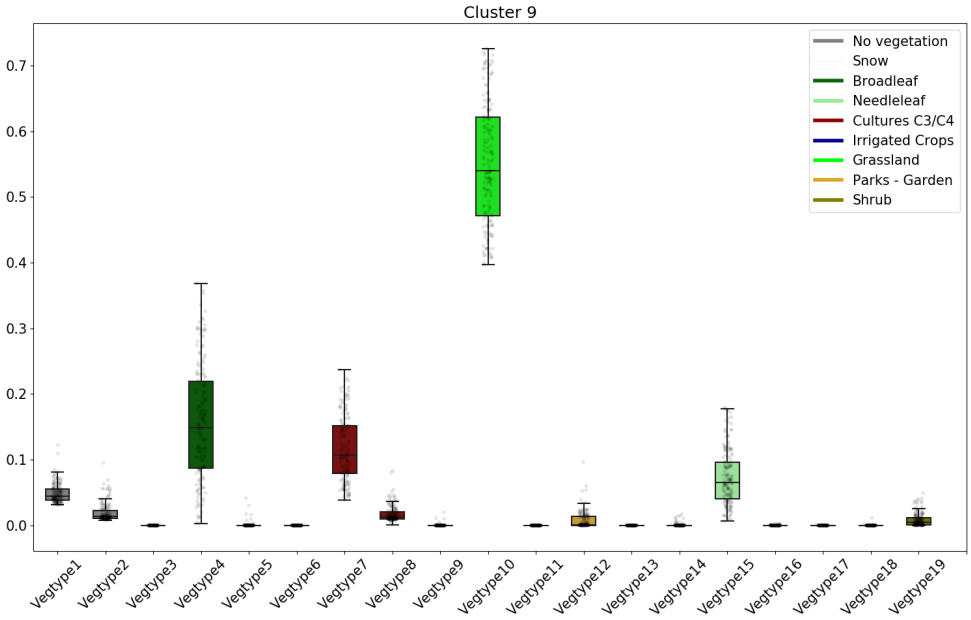

Figure A.10: Vegetation Fraction Box plot for Cluster 9

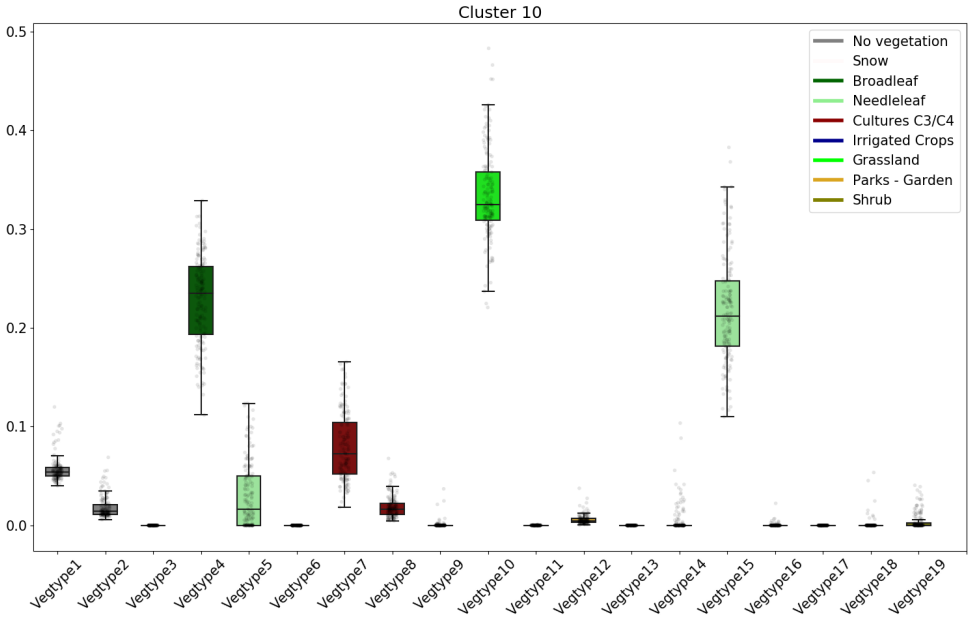

Figure A.11: Vegetation Fraction Box plot for Cluster 10

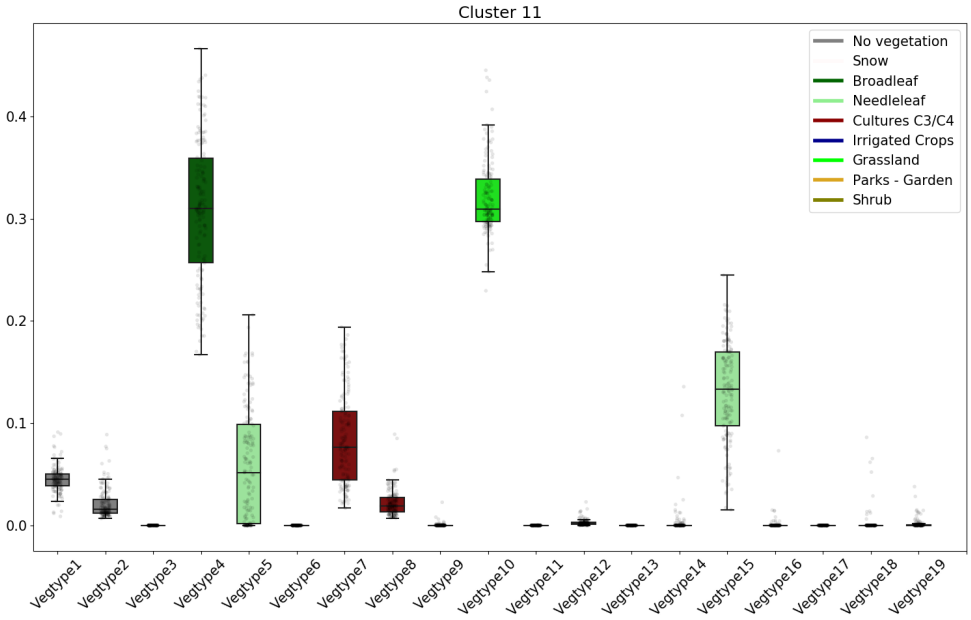

Figure A.12: Vegetation Fraction Box plot for Cluster 11 - Reference name: Boradleaf Forest

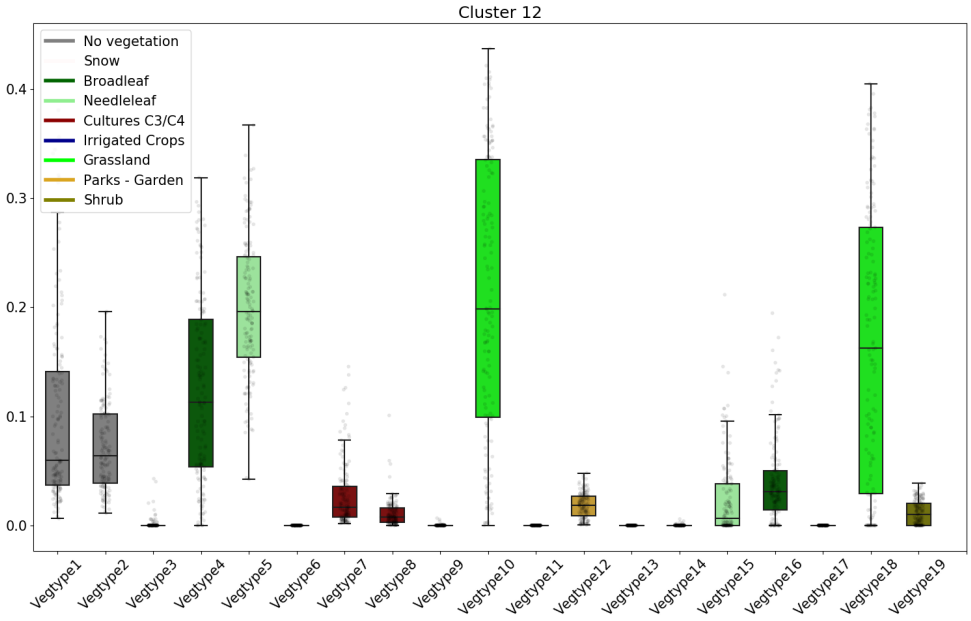

Figure A.13: Vegetation Fraction Box plot for Cluster 12 - Reference name: Boreal Alps

A.2. Data Variability

In this Appendix chapter, all figures for the land surface variability and ASCAT observation variability analysis are provided.

A.2.1. Land surface parameter variability

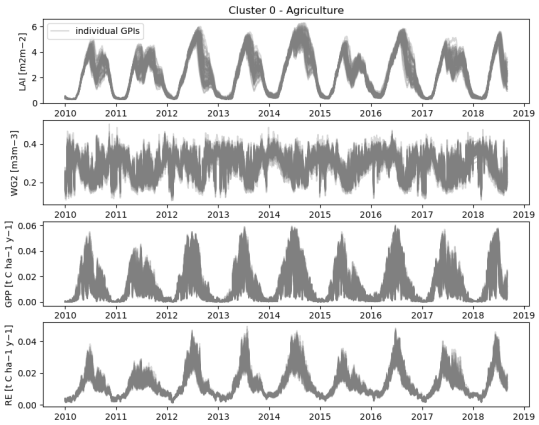

Figure A.14: LAI, WG2, GPP and RE time-series for all individual GPI's in the cluster.

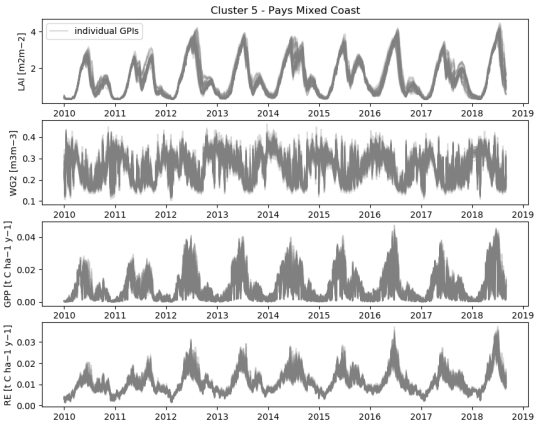

Figure A.15: LAI, WG2, GPP and RE time-series for all individual GPI's in the cluster.

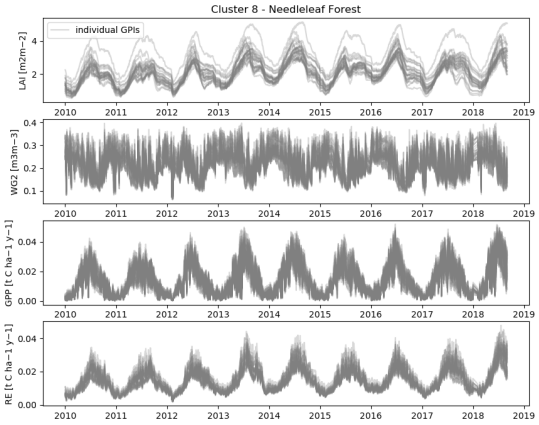

Figure A.16: LAI, WG2, GPP and RE time-series for all individual GPI's in the cluster.

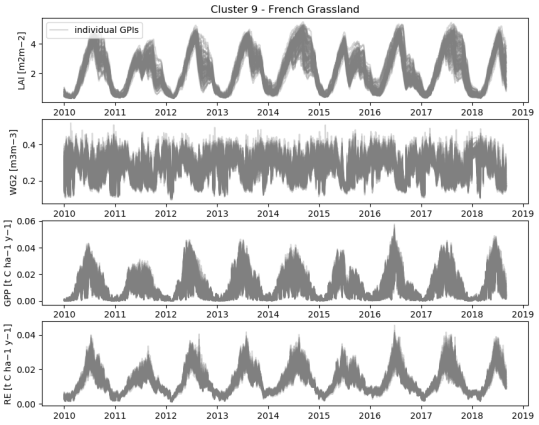

Figure A.17: LAI, WG2, GPP and RE time-series for all individual GPI's in the cluster.

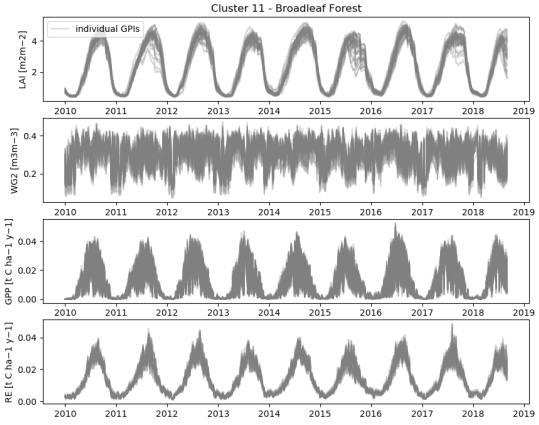

Figure A.18: LAI, WG2, GPP and RE time-series for all individual GPI's in the cluster.

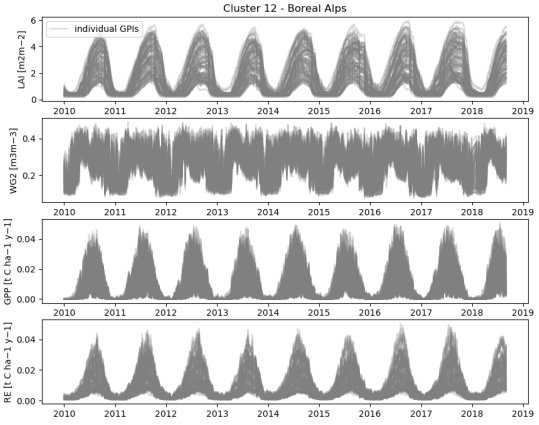

Figure A.19: LAI, WG2, GPP and RE time-series for all individual GPI's in the cluster.

A.2.2. ASCAT observation variability

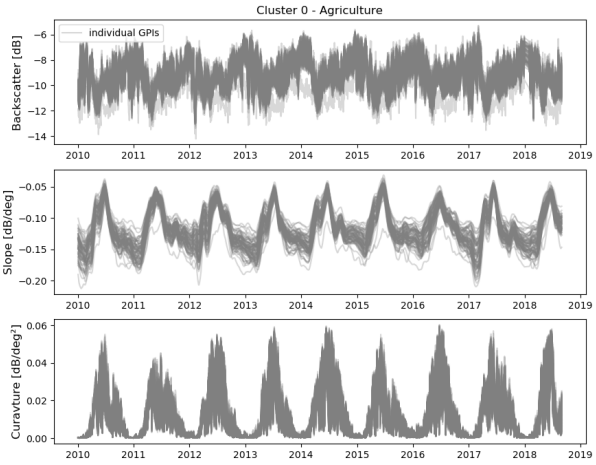

Figure A.20: Backscatter, Slope and Curvature time-series for all individual GPI's in the cluster.

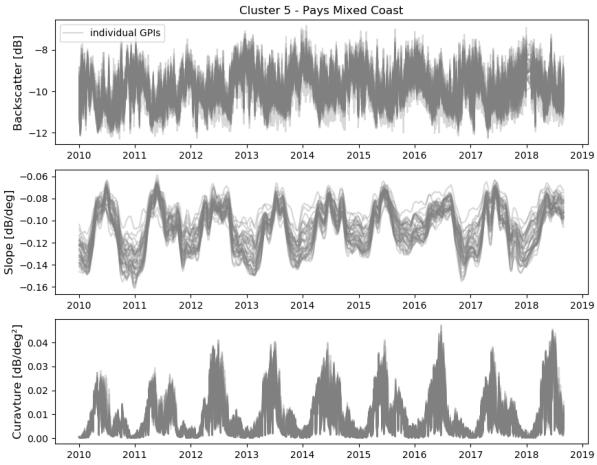

Figure A.21: Backscatter, Slope and Curvature time-series for all individual GPI's in the cluster.

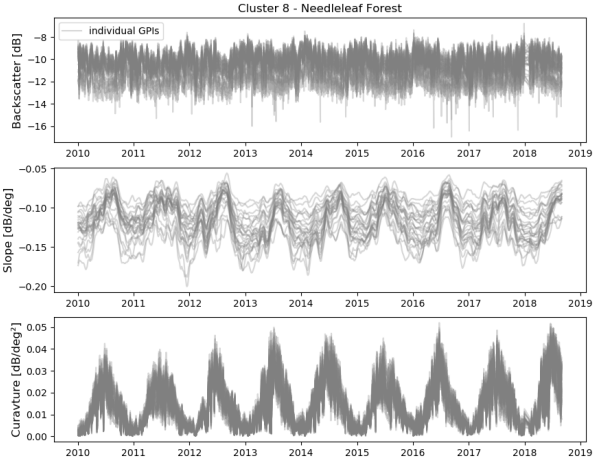

Figure A.22: Backscatter, Slope and Curvature time-series for all individual GPI's in the cluster.

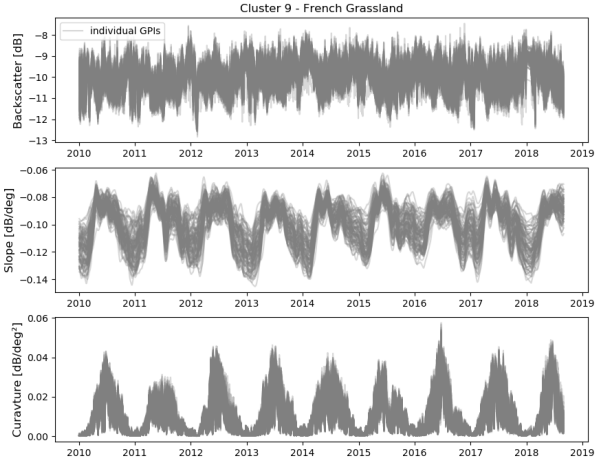

Figure A.23: Backscatter, Slope and Curvature time-series for all individual GPI's in the cluster.

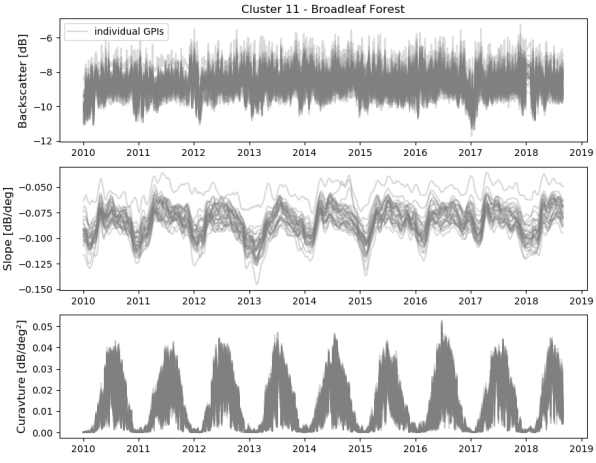

Figure A.24: Backscatter, Slope and Curvature time-series for all individual GPI's in the cluster.

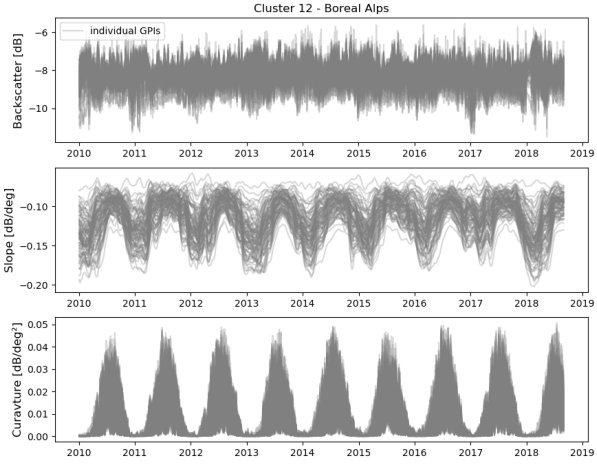

Figure A.25: Backscatter, Slope and Curvature time-series for all individual GPI's in the cluster.

A.3. Backscatter Analysis

In the following section, all extra figures are provided to support the results and discussion section to the backscatter analysis.

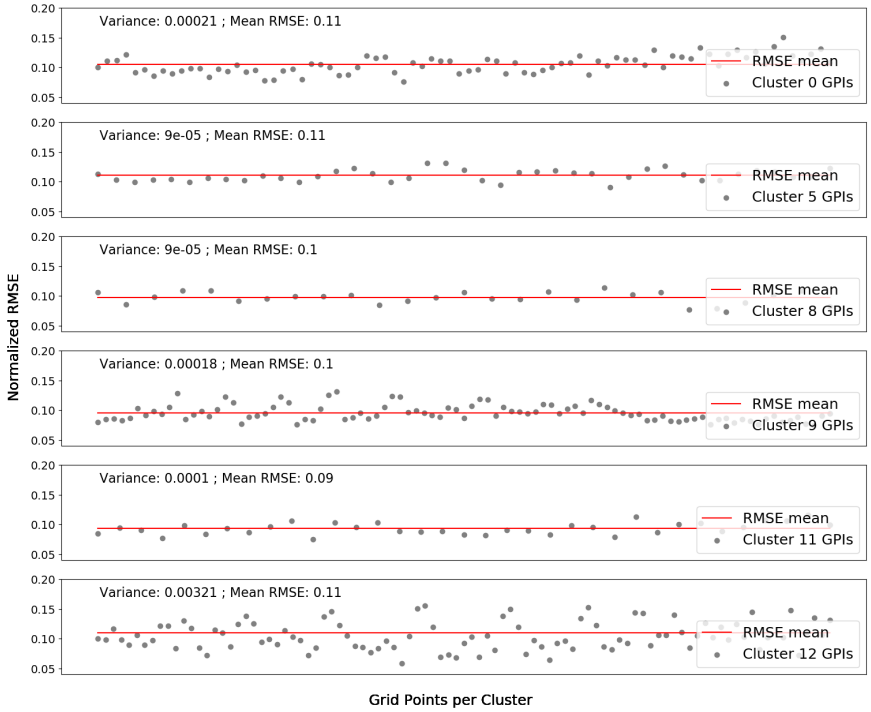

Figure A.26: Backscatter: Normalized RMSE estimations for 5 cluster types (Types: 0,3,5,9,12). The grey dots represent one grid point of the according cluster.

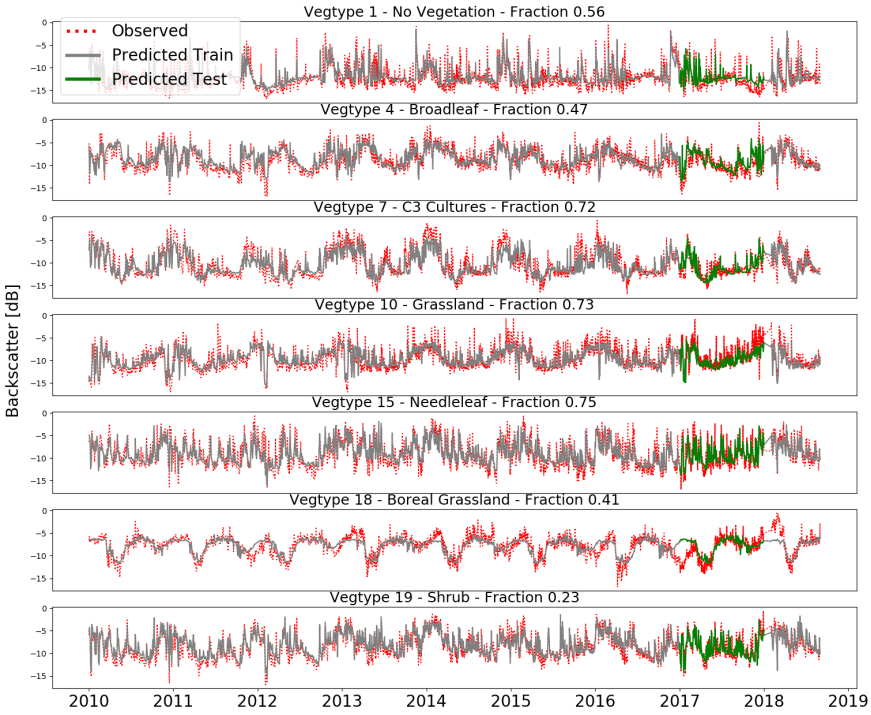

Figure A.27: Backscatter: Observed (red) and predicted (grey) time series of curvature for seven different vegetation types (Type: 1,4,7,10,15,18,19). The green line indicates the prediction using the test data-set.

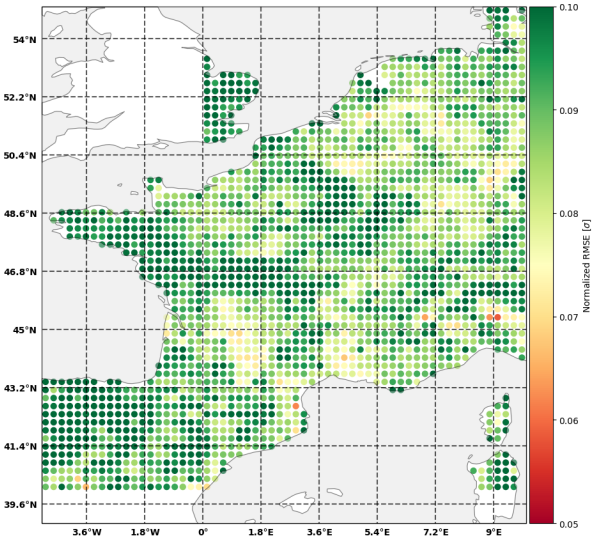

Figure A.28: Backscatter: Cross validation: 2012 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

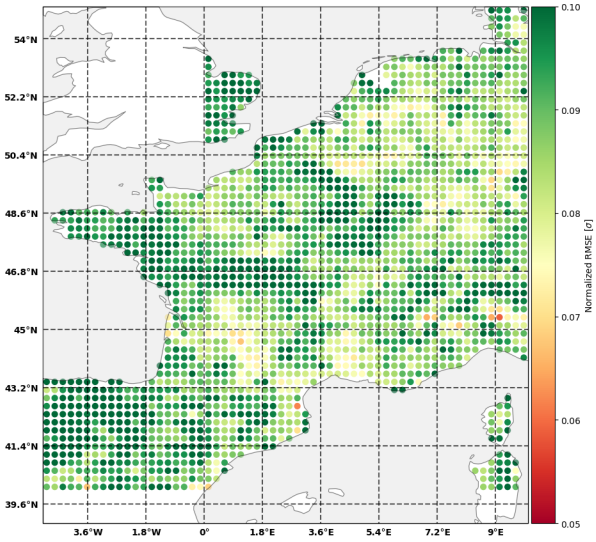

Figure A.29: Backscatter: Cross validation: 2014 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

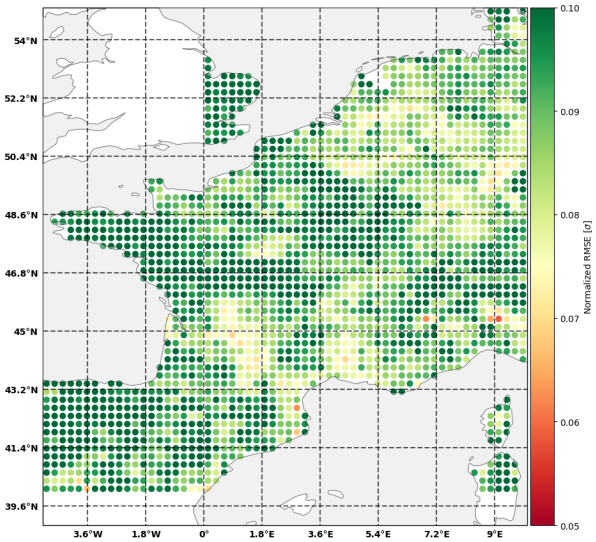

Figure A.30: Backscatter: Cross validation: 2017 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

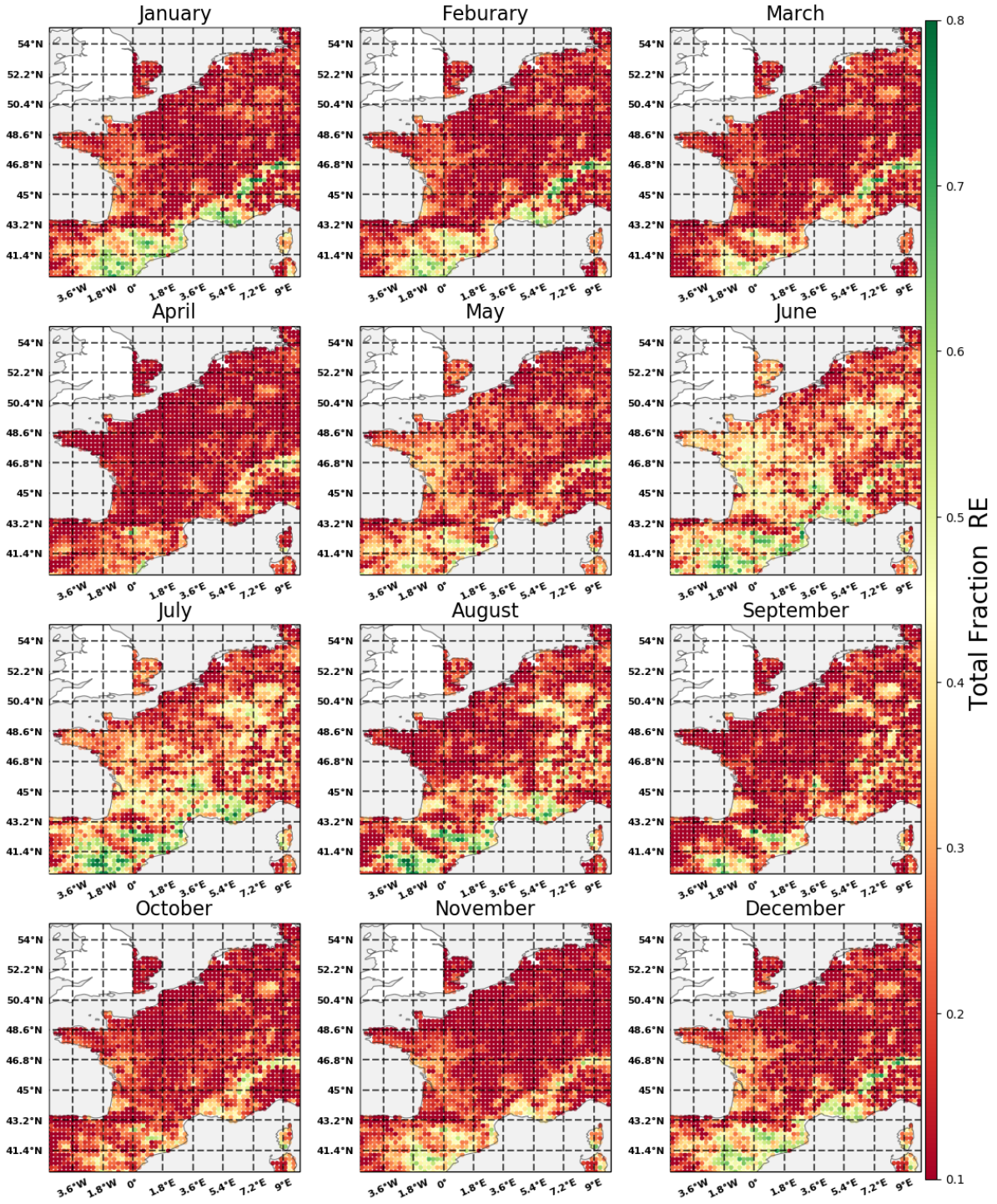

Figure A.31: Backscatter: RE feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

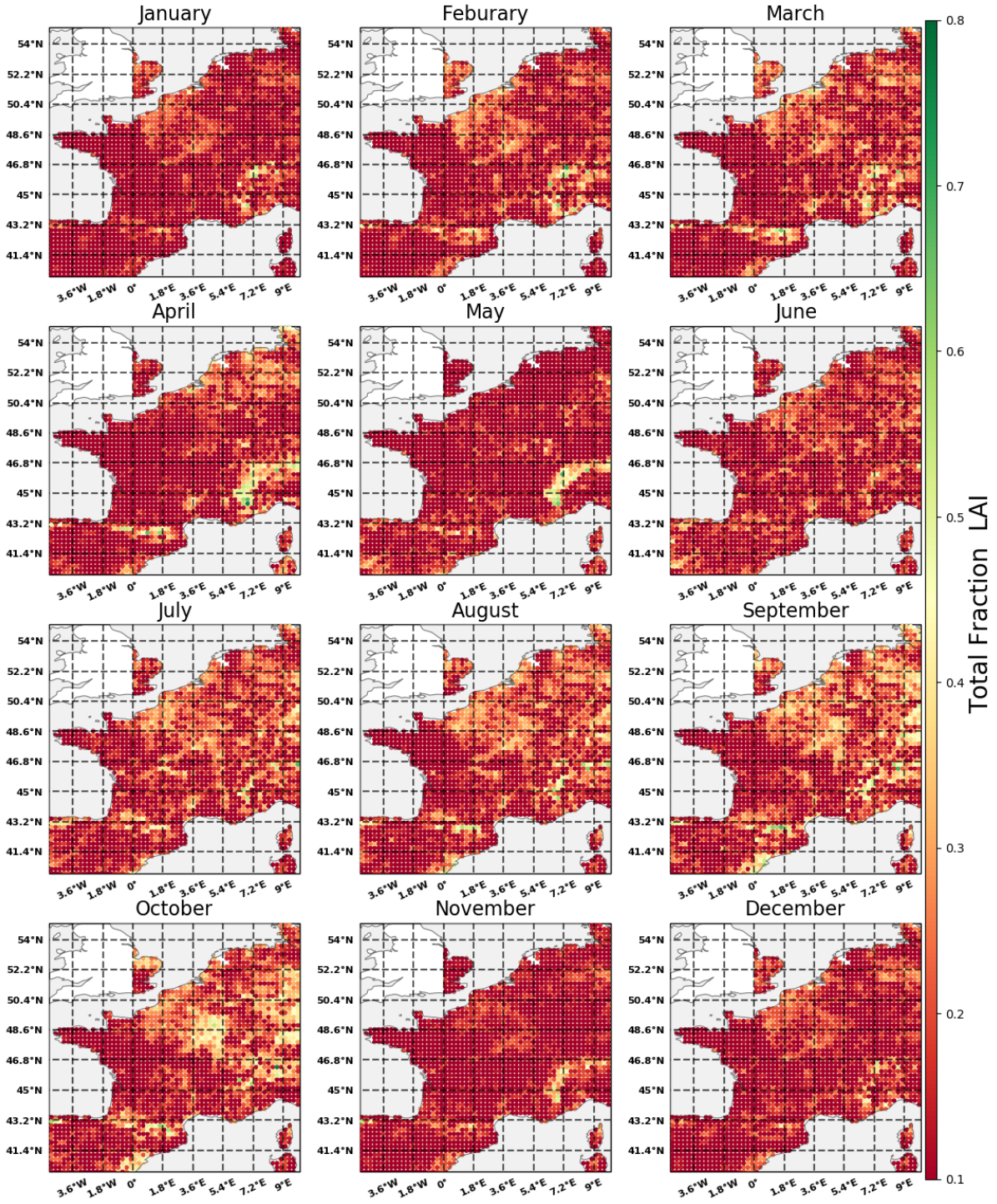

Figure A.32: Backscatter: LAI feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

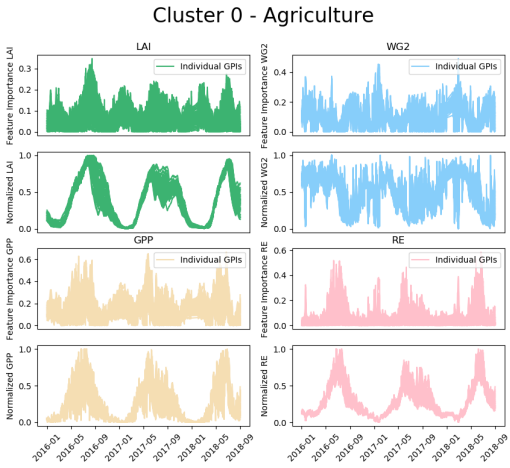

Figure A.33: Backscatter: Feature importance and land surface parameter time series for the named cluster

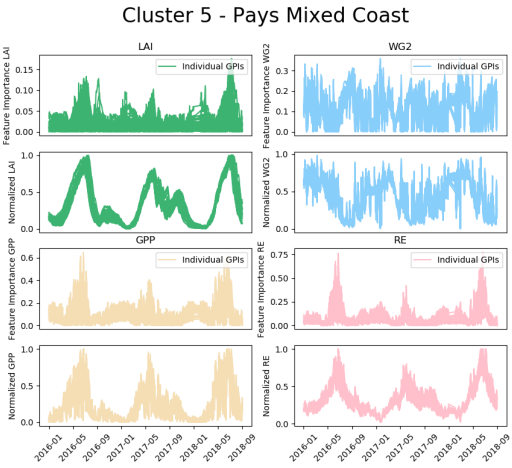

Figure A.34: Backscatter: Feature importance and land surface parameter time series for the named cluster

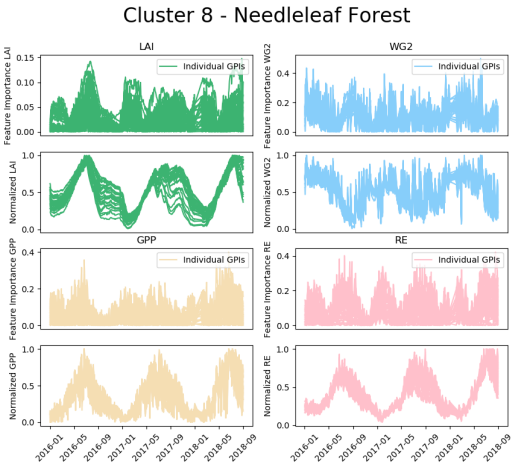

Figure A.35: Backscatter: Feature importance and land surface parameter time series for the named cluster

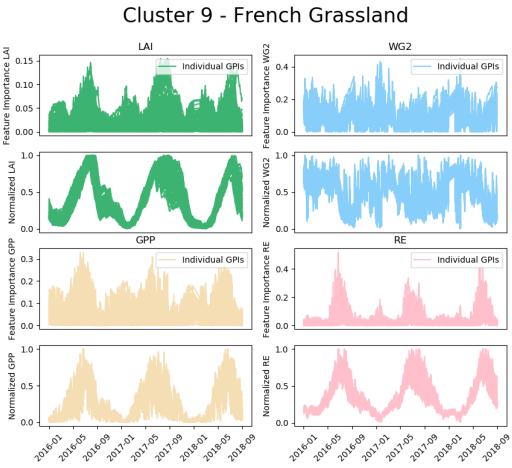

Figure A.36: Backscatter: Feature importance and land surface parameter time series for the named cluster

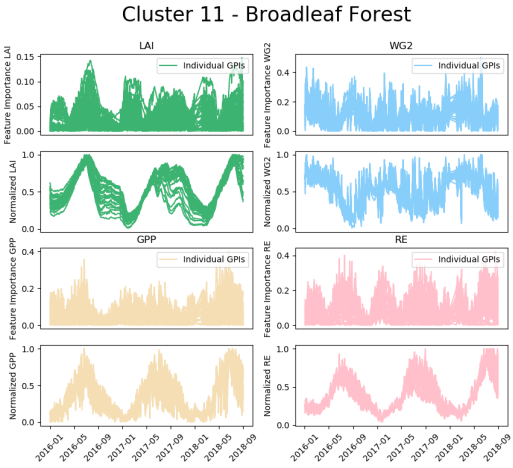

Figure A.37: Backscatter: Feature importance and land surface parameter time series for the named cluster

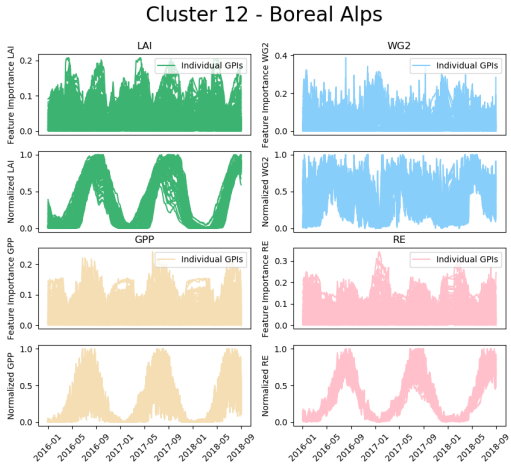

Figure A.38: Backscatter: Feature importance and land surface parameter time series for the named cluster

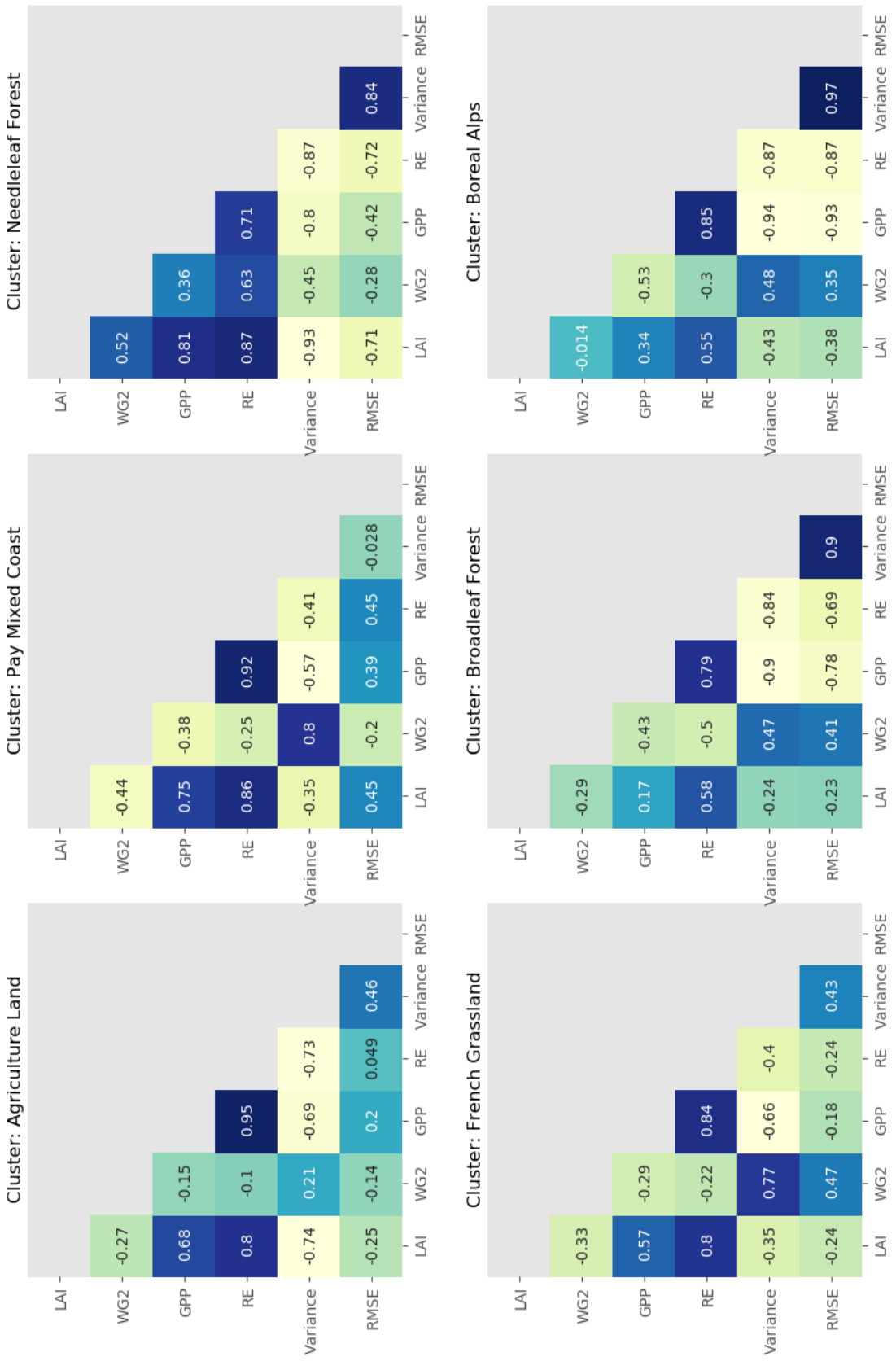

Figure A.39: Backscatter: Spearman Correlation coefficient matrix between the inter seasonal variance of the input and output data as well as the RMSE values. The data used for the variance calculation is based on monthly averages.

A.4. Slope Analysis

In the following section, all extra figures are provided to support the results and discussion section to the slope analysis.

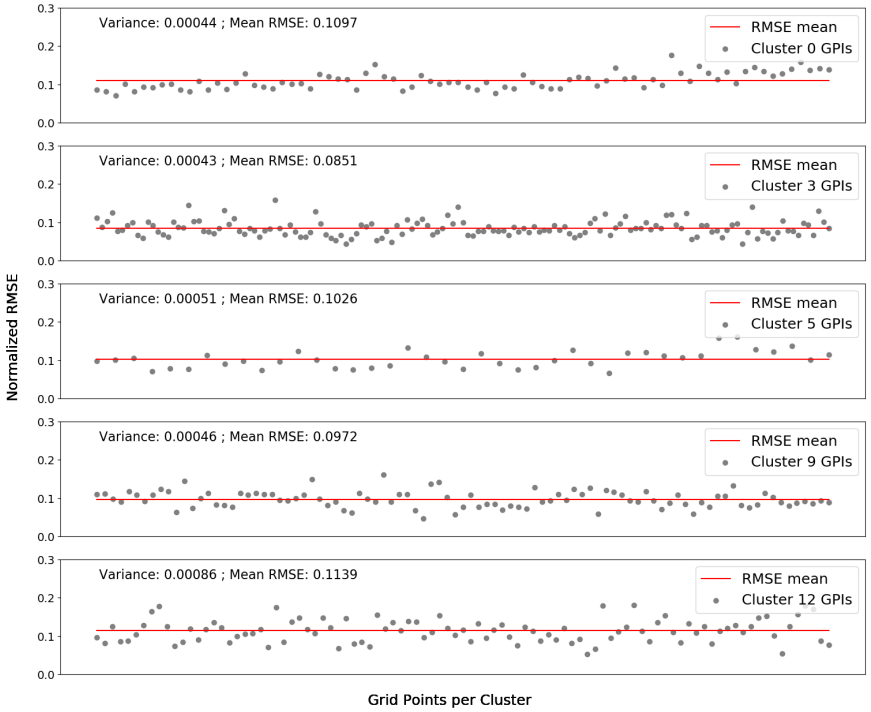

Figure A.40: Slope: Normalized RMSE estimations for 5 cluster types (Types: 0,3,5,9,12). The grey dots represent one grid point of the according cluster.

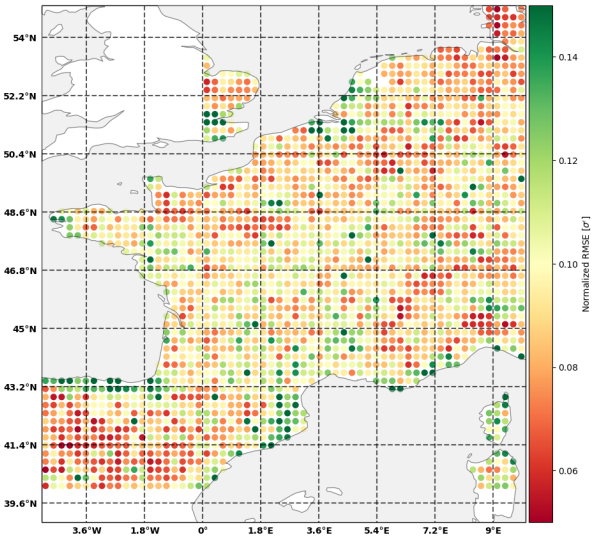

Figure A.41: Slope: Cross Validation Slope: 2012 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

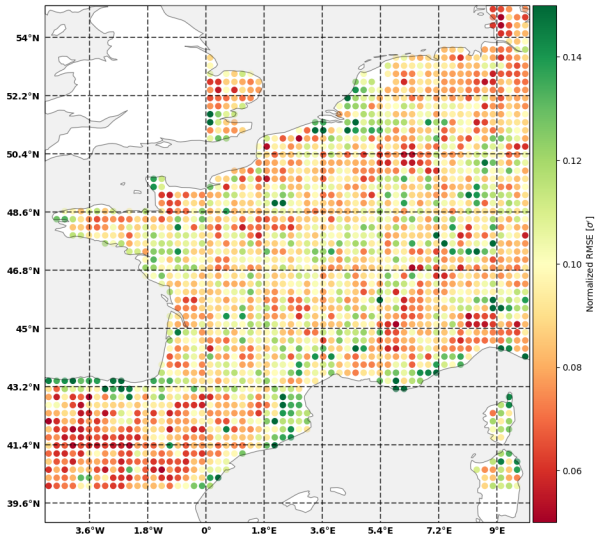

Figure A.42: Slope: Cross Validation Slope: 2014 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

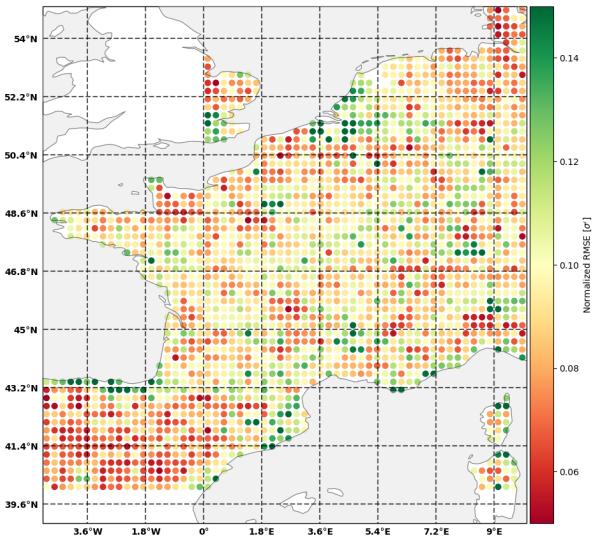

Figure A.43: Slope: Cross Validation Slope: 2017 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

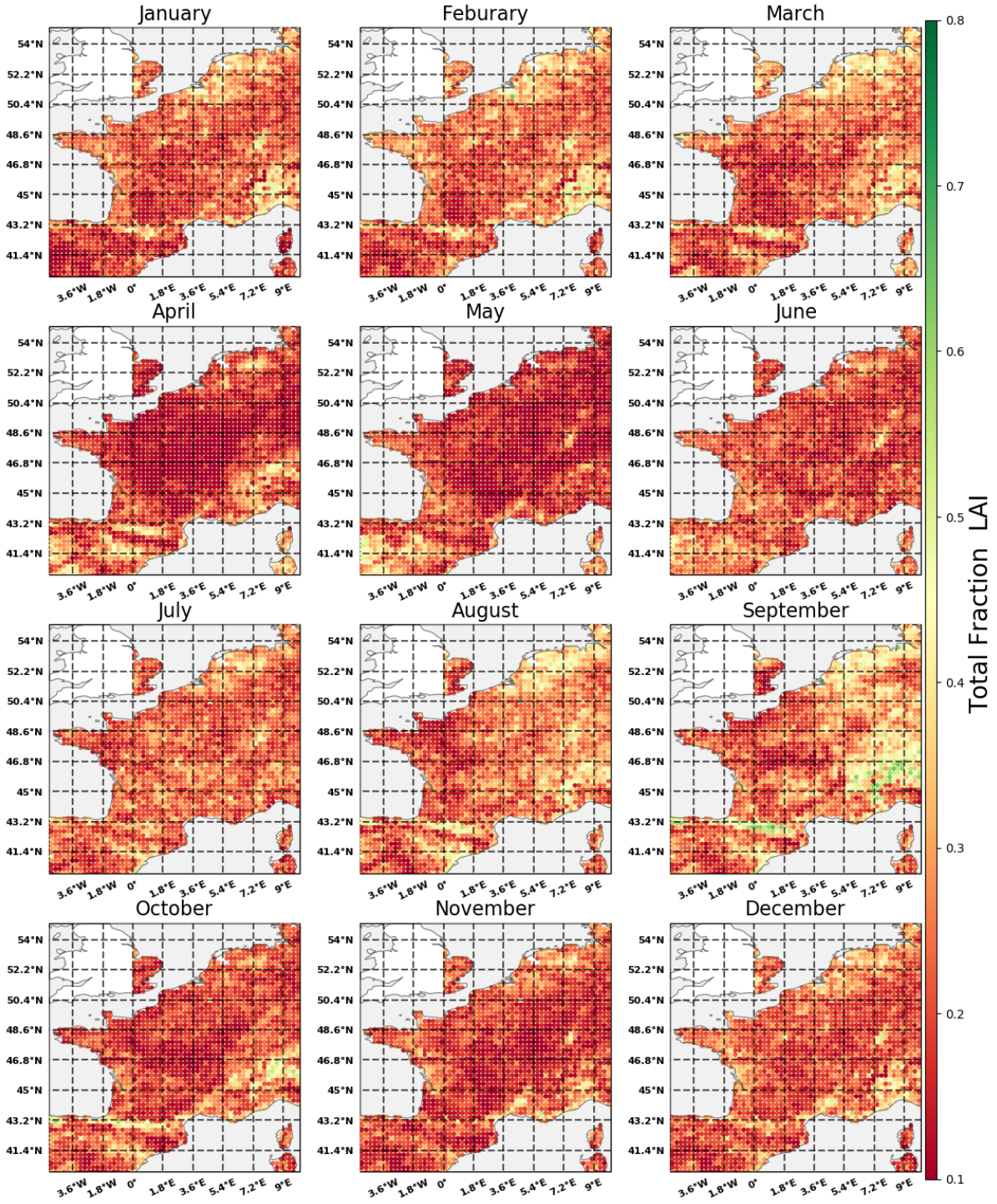

Figure A.44: Slope: LAI feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

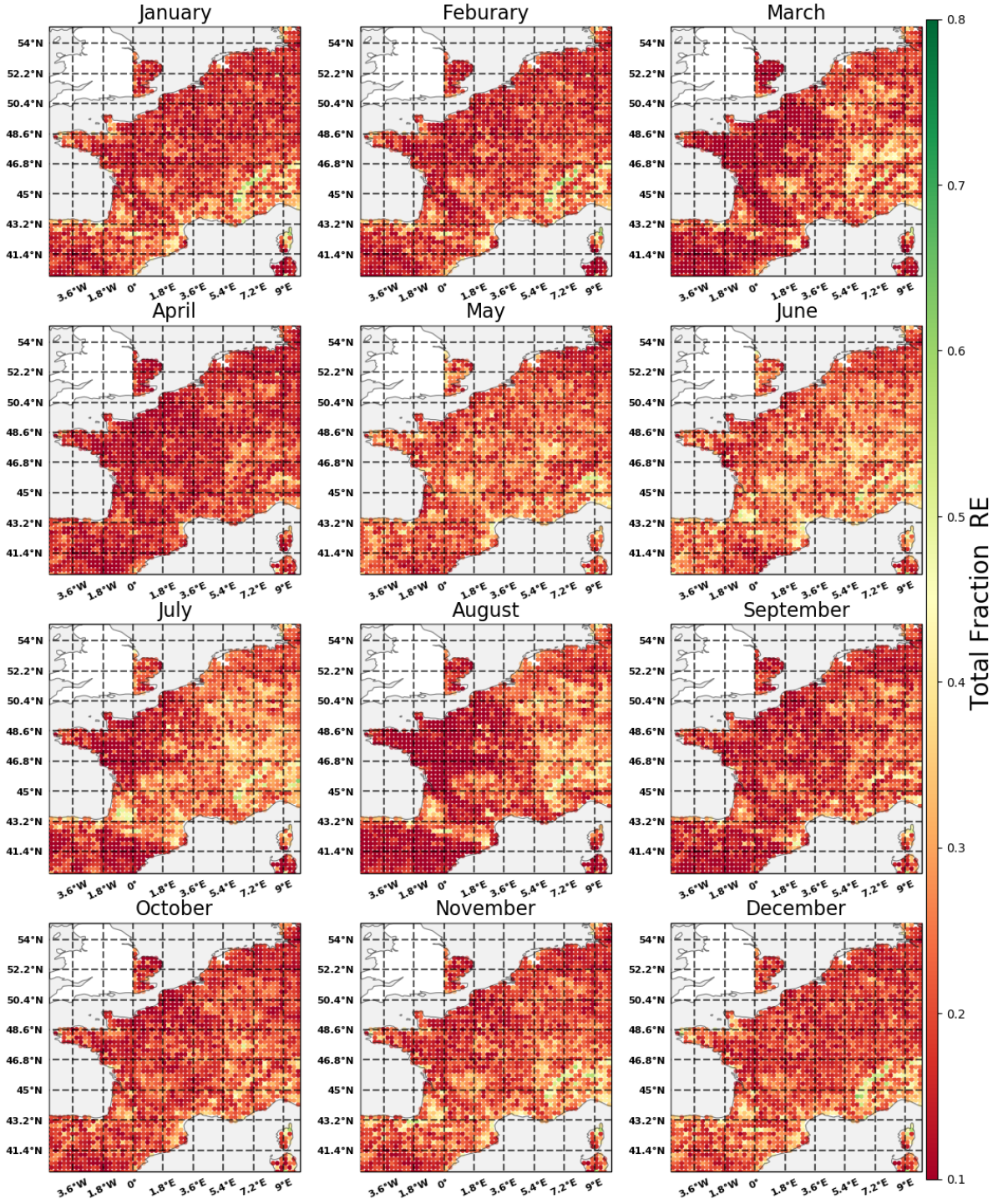

Figure A.45: Slope: RE feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

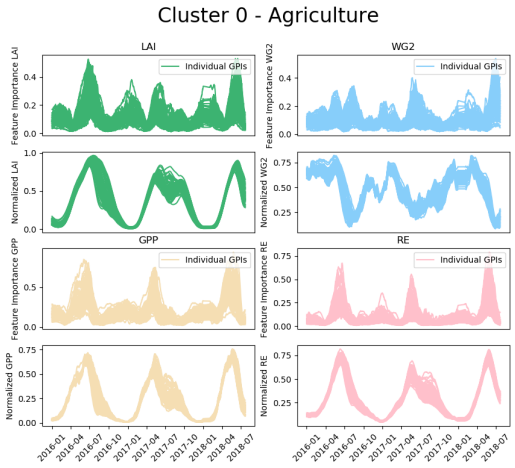

Figure A.46: Slope: Feature importance and land surface parameter time series for the named cluster

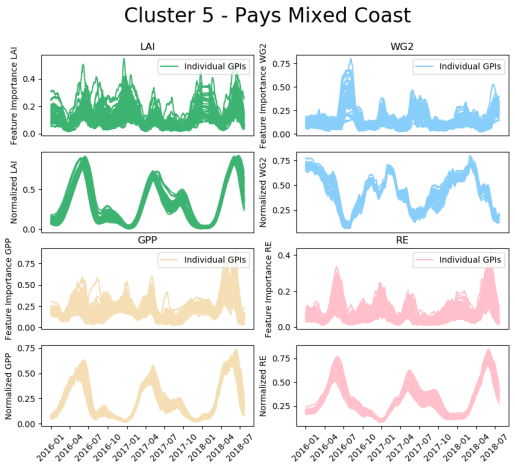

Figure A.47: Slope: Feature importance and land surface parameter time series for the named cluster

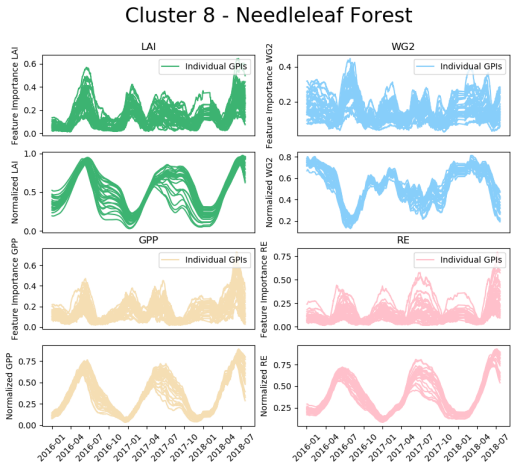

Figure A.48: Slope: Feature importance and land surface parameter time series for the named cluster

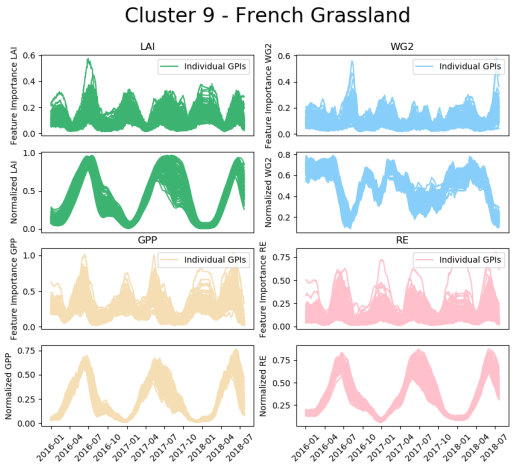

Figure A.49: Slope: Feature importance and land surface parameter time series for the named cluster

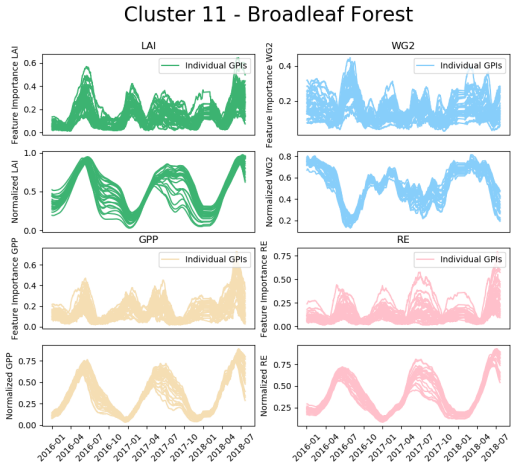

Figure A.50: Slope: Feature importance and land surface parameter time series for the named cluster

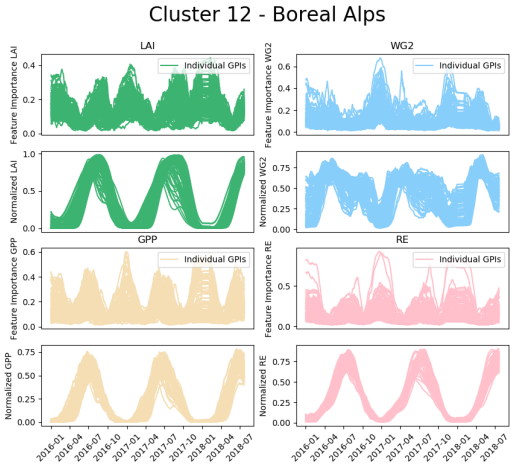

Figure A.51: Slope: Feature importance and land surface parameter time series for the named cluster

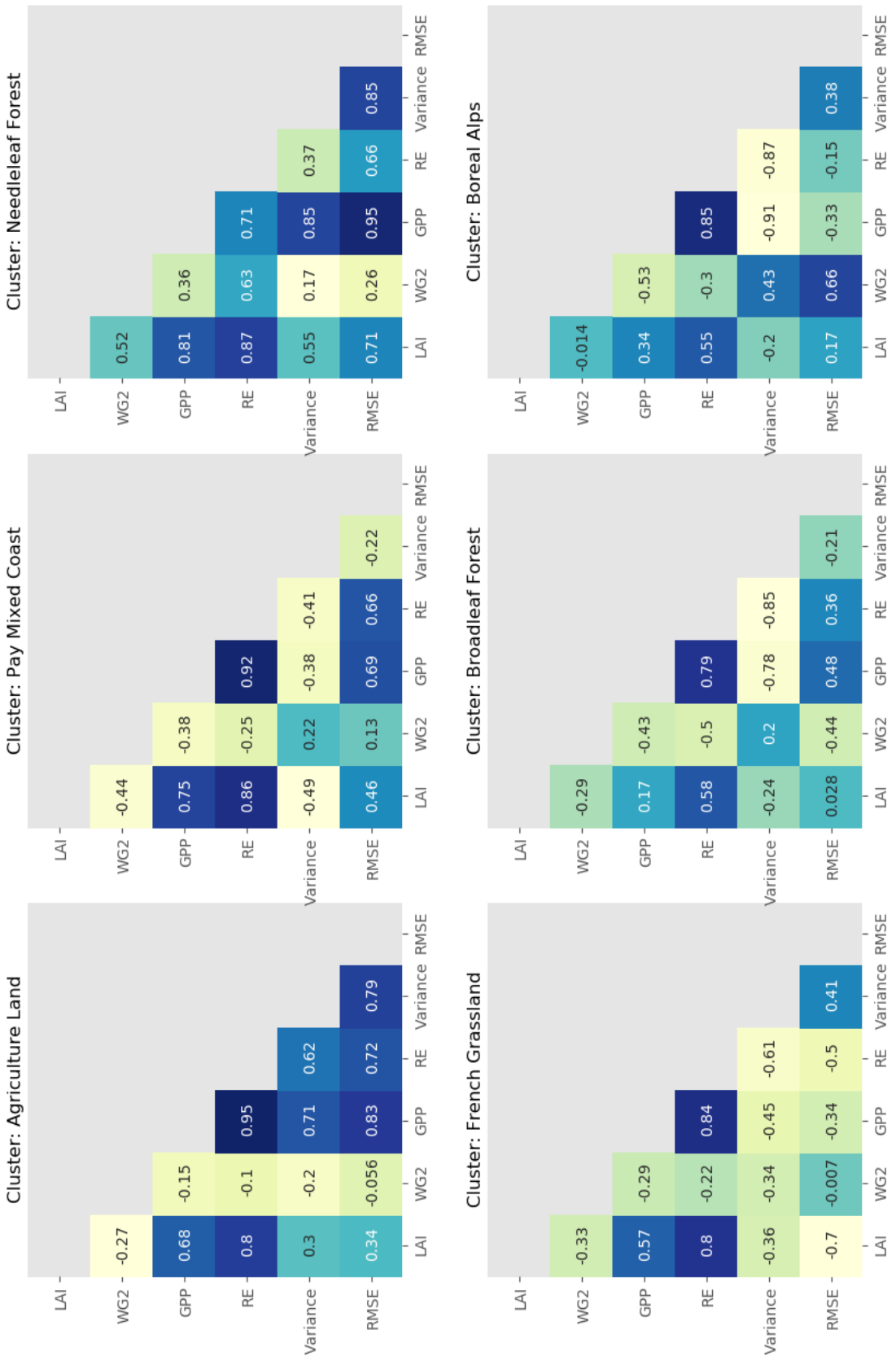

Figure A.52: Slope: Spearman Correlation coefficient matrix between the inter seasonal variance of the input and output data as well as the RMSE values. The data used for the variance calculation is based on monthly averages.

A.5. Curvature Analysis

In the following section, all extra figures are provided to support the results and discussion section to the curvature analysis.

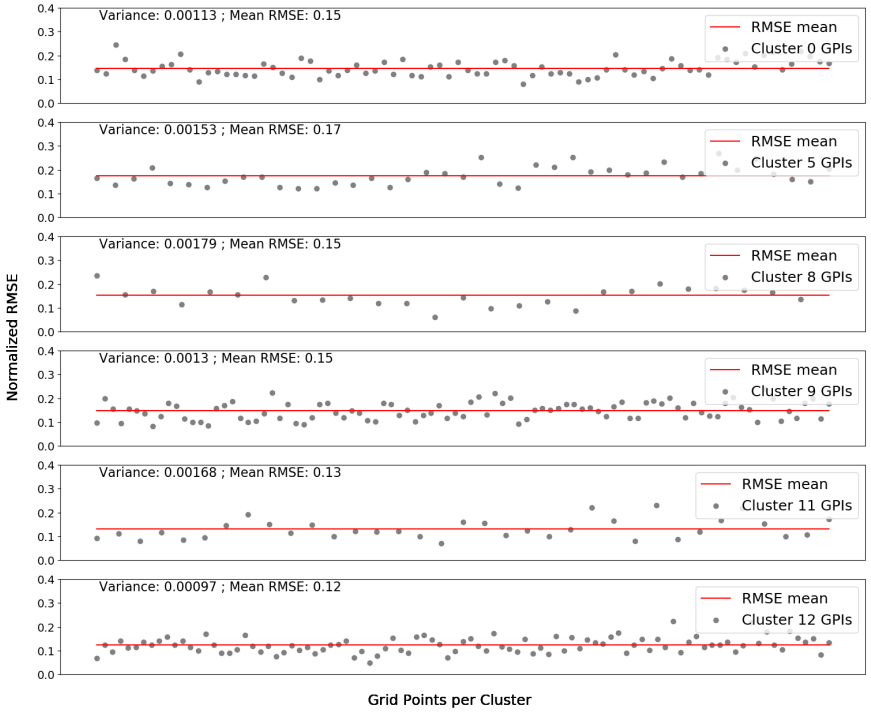

Figure A.53: Curvature: Normalized RMSE estimations for 5 cluster types (Types: 0,3,5,9,12). The grey dots represent one grid point of the according cluster.

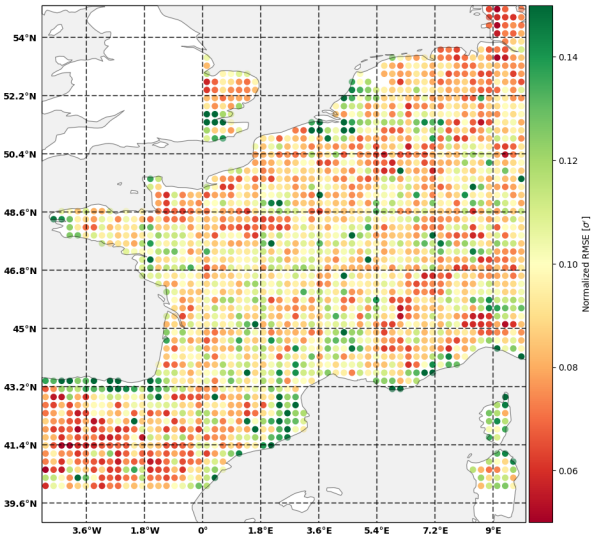

Figure A.54: Curvature: Cross Validation Curvature: 2012 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

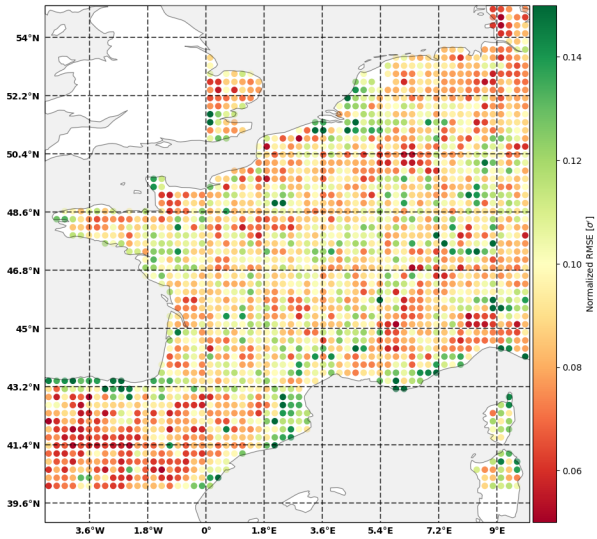

Figure A.55: Curvature: Cross Validation Curvature:2014 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

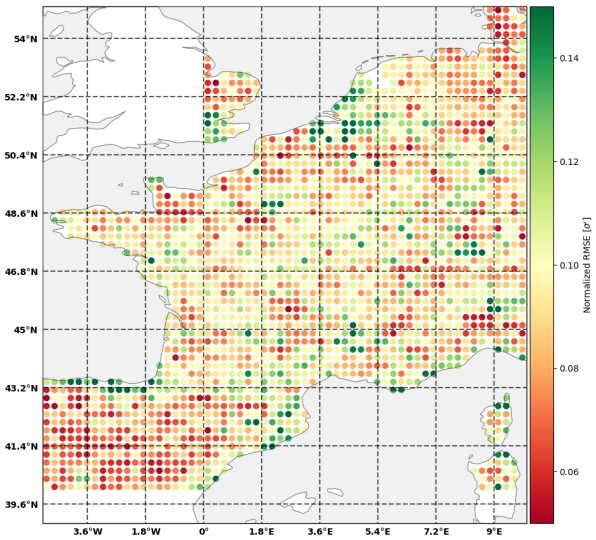

Figure A.56: Curvature: Cross Validation Curvature: 2017 Simulation - RMSE calculated over the whole time period from 2010 to 2018 for each GPI.

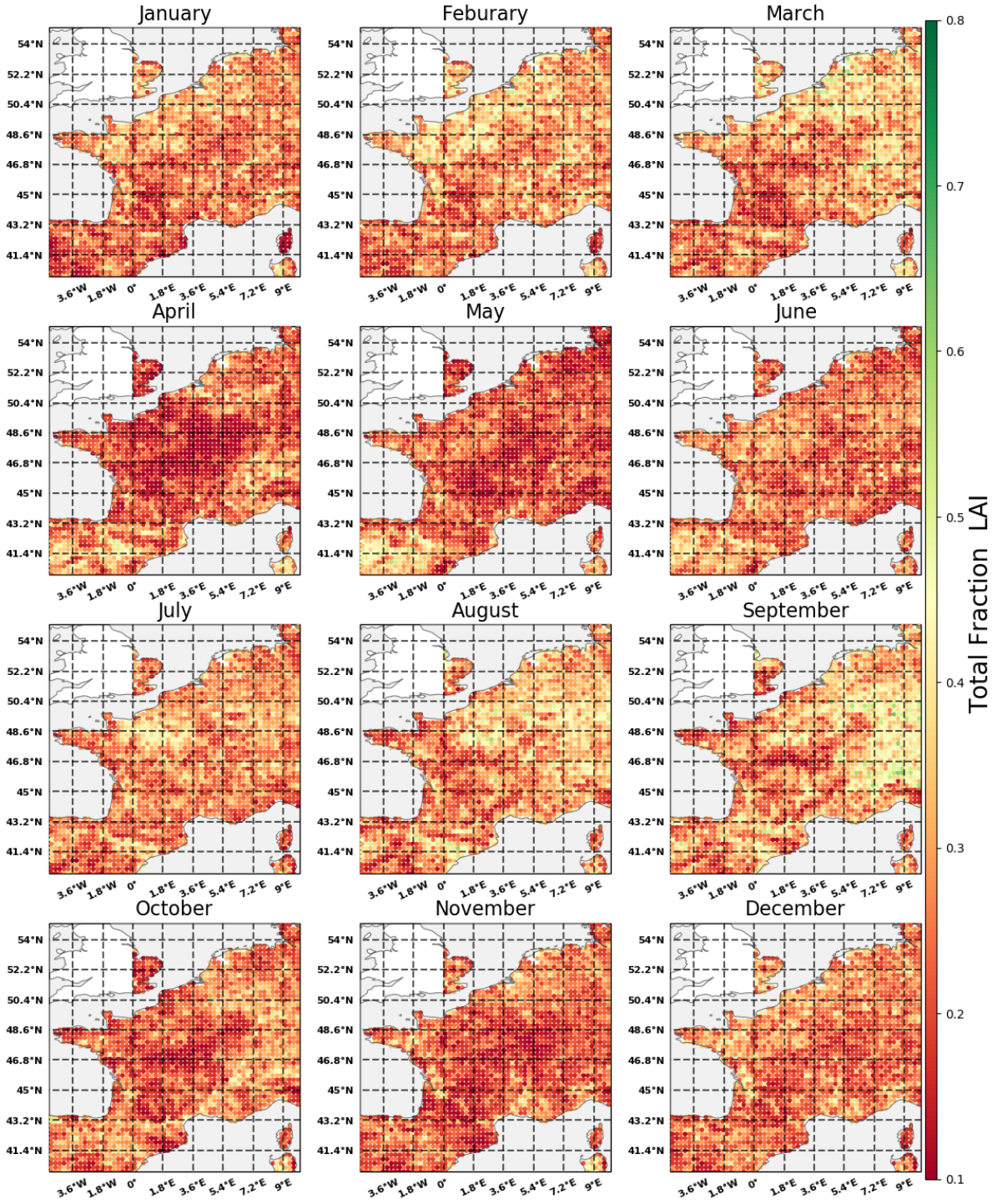

Figure A.57: Curvature: LAI feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

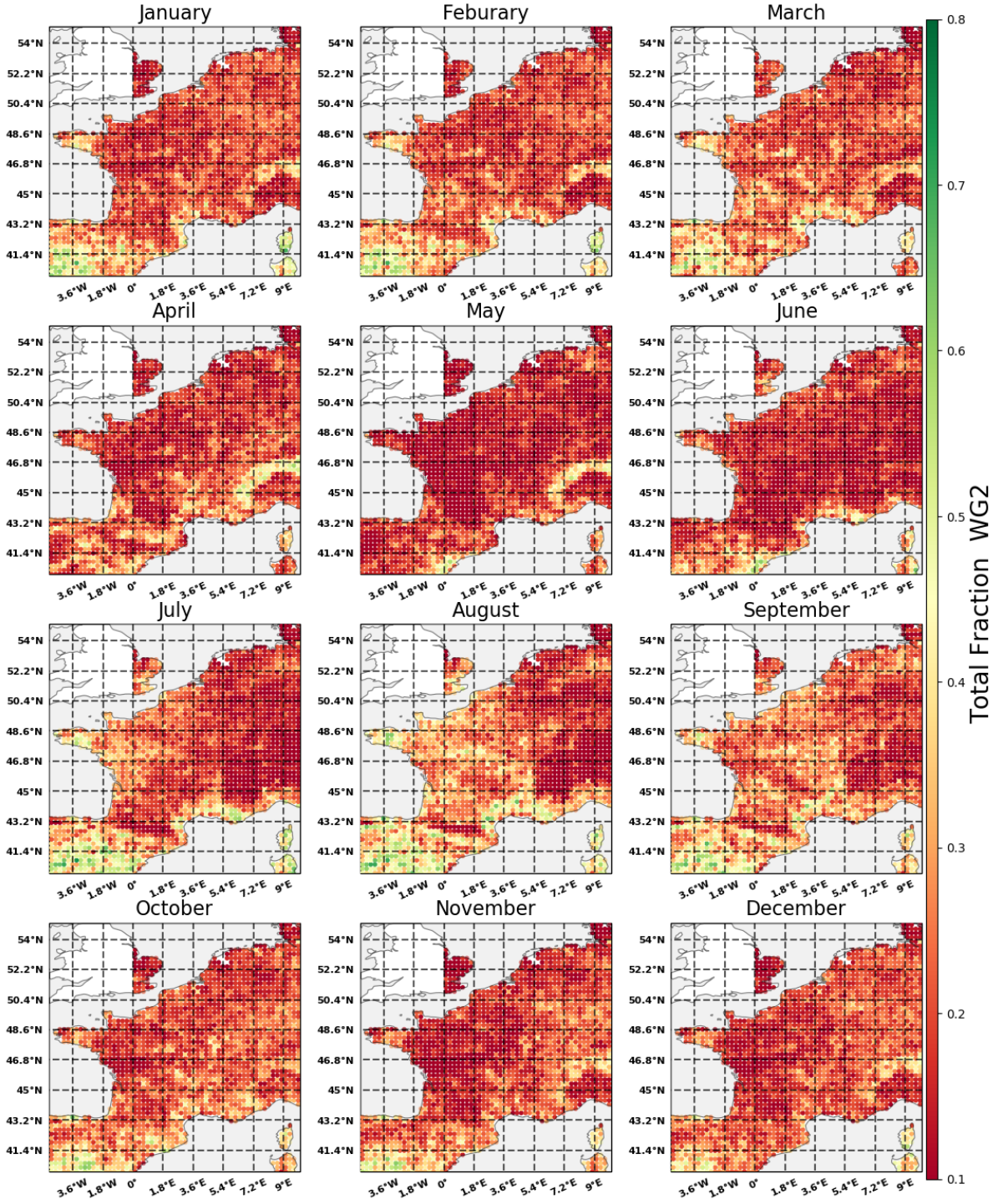

Figure A.58: Curvature: WG2 feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

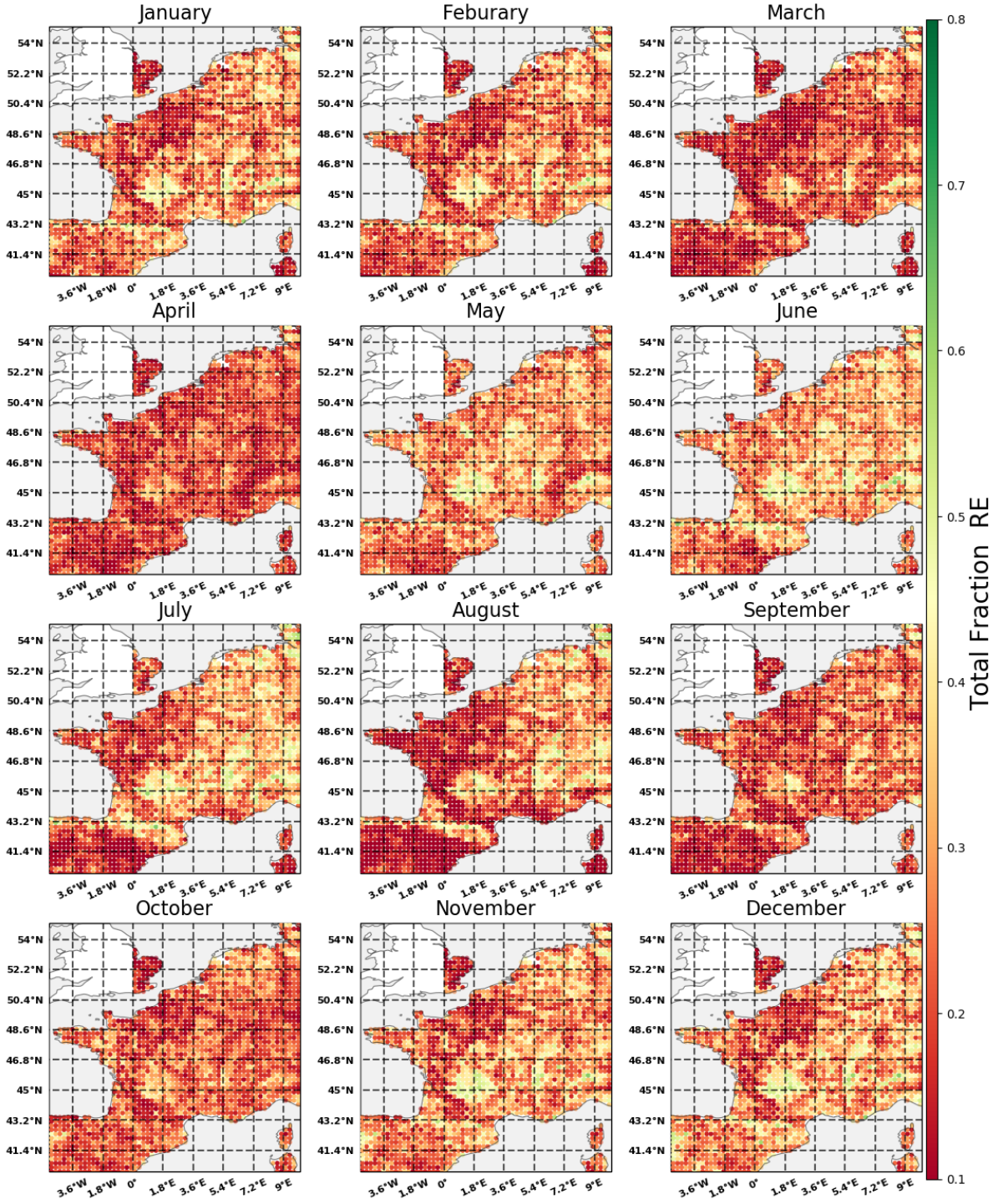

Figure A.59: Curvature: RE feature importance for each month and the whole study domain. The values are expressed in fractions of the total feature importance

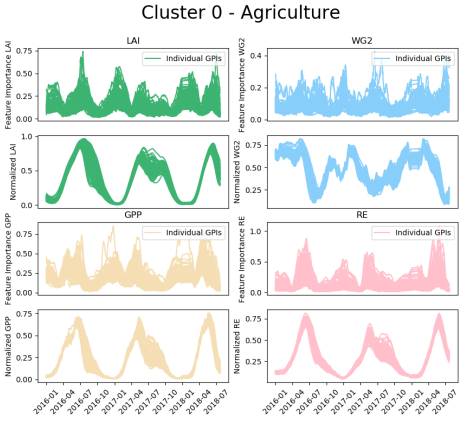

Figure A.60: Curvature: Feature importance and land surface parameter time series for the named cluster

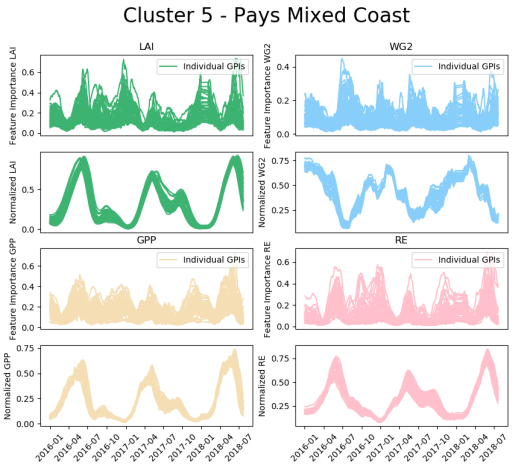

Figure A.61: Curvature: Feature importance and land surface parameter time series for the named cluster

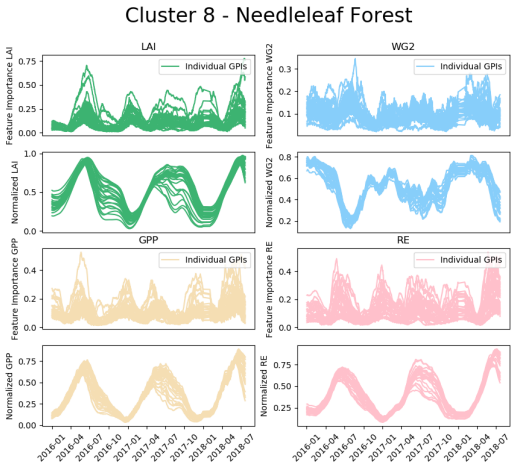

Figure A.62: Curvature: Feature importance and land surface parameter time series for the named cluster

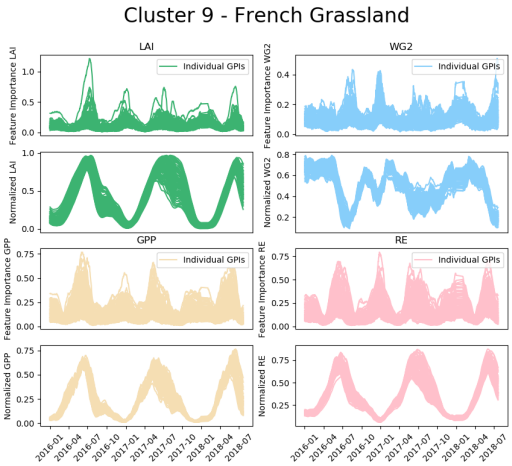

Figure A.63: Curvature: Feature importance and land surface parameter time series for the named cluster

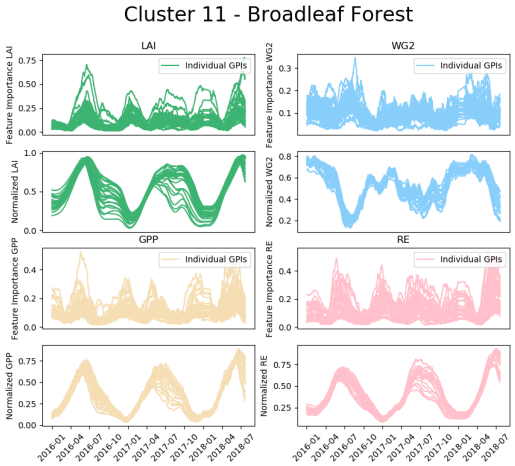

Figure A.64: Curvature: Feature importance and land surface parameter time series for the named cluster

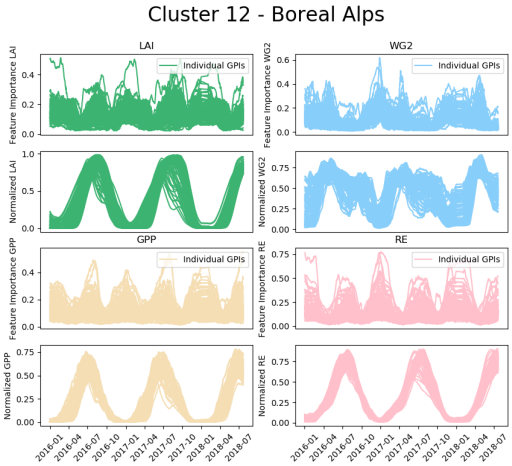

Figure A.65: Curvature: Feature importance and land surface parameter time series for the named cluster

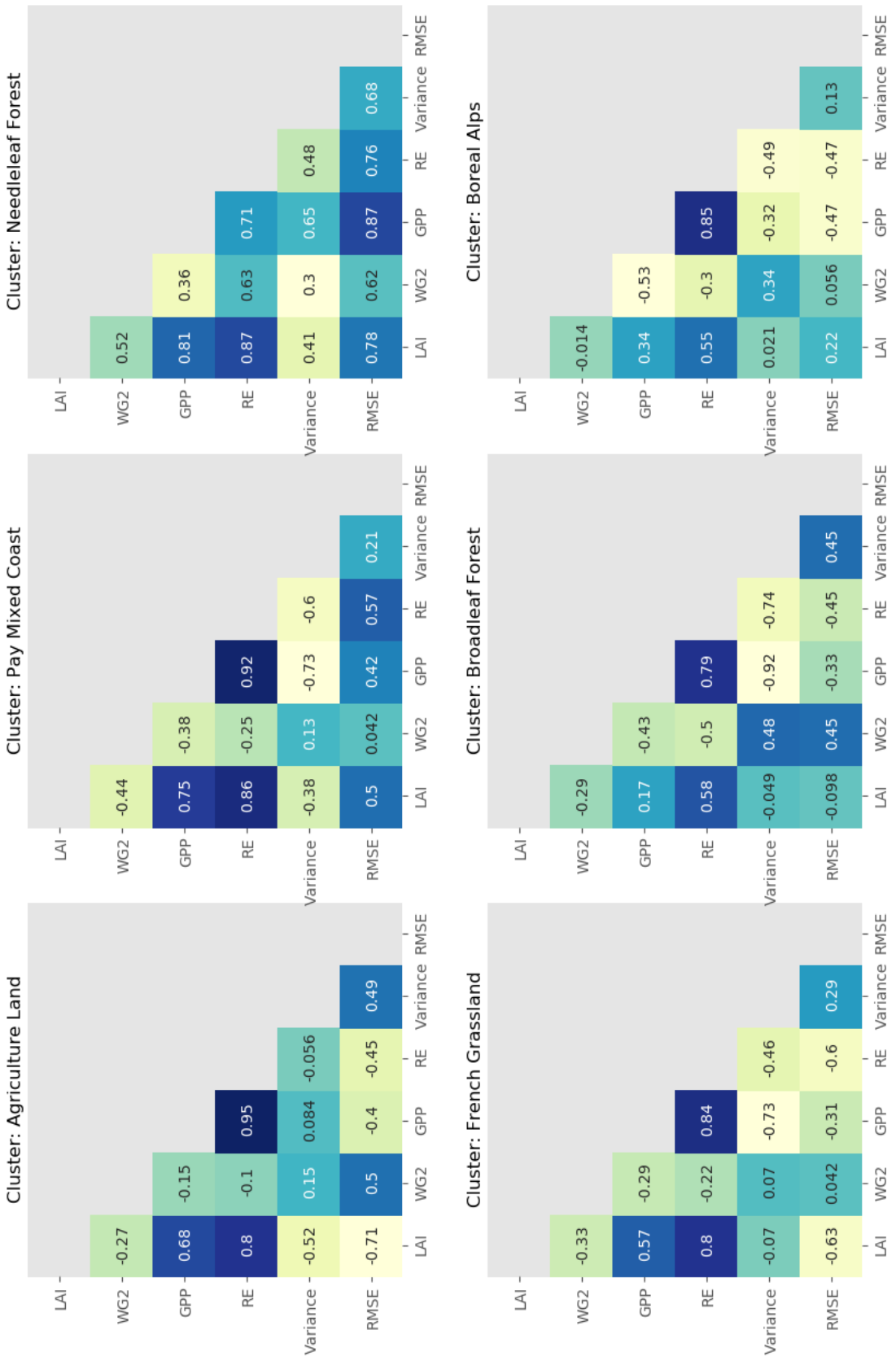

Figure A.66: Curvature: Spearman Correlation coefficient matrix between the inter seasonal variance of the input and output data as well as the RMSE values. The data used for the variance calculation is based on monthly averages.