CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

Department of Water Resources and Environmental Modelling

Assimilation of Chosen Satellite Data into the Hydrological Balance Modelling

Diploma Thesis

Supervisor: doc. Ing. Petr Máca, Ph.D.

Author: Fortune Sam Okon

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

DIPLOMA THESIS ASSIGNMENT

B.Sc. Fortune Sam Okon, BSc

Landscape Engineering
Environmental Modelling

Thesis title

Assimilation of chosen satelite data into the hydrological balance modeling

---

Objectives of thesis

The aim of the thesis is to assimilate chosen satellite products within the lumped hydrological modeling. For this purpose student will calibrate lumped hydrological model on selected set of catchments. Student will derive the NDVI data form Modis dataset, and will compare the calibration results based on input datasets with and without NDVI information.

Methodology

The solution of thesis will be based on following steps:

1. The preparation of inputs data for hydrological modeling for selected set of catchments

2. The preparation of satellite information on NDVI data

3. The calibration of lumped hydrological model without the satellite information

4. the calibration of lumped hydrological model with the satellite information

5. The comparison of model outputs

The proposed extent of the thesis

standard

Keywords

runoff, NDVI,ET,entropy

Recommended information sources

C. Gomez et al. / ISPRS Journal of Photogrammetry and Remote Sensing 116 (2016) 55–72 (Optical remotely sensed time series data for land cover classification: A review).

C. Pohl & J. L. Van Genderen (1998) Review article Multisensor image fusion in remote sensing: Concepts, methods and applications, International Journal of Remote Sensing, 19:5, 823-854, DOI: 10.1080/014311698215748 .

H.E. Beck et al. / Remote Sensing of Environment 115 (2011) 2547–2563 (Global evaluation of four AVHRR–NDVI data sets: Intercomparison and assessment against Landsat imagery).

Izquierdo, J., Pérez, R., & Iglesias, P. L. (2004). Mathematical models and methods in the water industry. Mathematical and Computer Modelling, 39(11-12), 1353-1374. ISSN 0895-7177. https://doi.org/10.1016/j.mcm.2004.06.012

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software, 21(5),

Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modelling, and doing hydrology backward, Water Resour. Res., 45, W02429, doi:10.1029/2008WR006912.

Rey, S. J. (2015). Mathematical Models in Geography (J. D. Wright, Ed.). International Encyclopedia of the Social & Behavioral Sciences (Second Edition), 00(00), 785-790. ISBN 9780080970875. https://doi.org/10.1016/B978-0-08-097086-8.72033-2

Singh, V. P. (2013). Entropy theory and its application in environmental and water engineering. John Wiley & Sons

W. Song et al. / International Journal of Applied Earth Observation and Geoinformation 58 (2017) 168–176 (Estimating fractional vegetation cover and the vegetation index of bare soil and highly dense vegetation with a physically based method).

X. Zhou et al. / Journal of Hydrology 470–471 (2012) 269–279 (Benchmarking global land surface models against the observed mean annual runoff from 150 large basins) .

Expected date of thesis defence

2020/21 SS – FES

The Diploma Thesis Supervisor

doc. Ing. Petr Máca, Ph.D.

Supervising department

Department of Water Resources and Environmental Modeling

Electronic approval: 26. 3. 2021	Electronic approval: 26. 3. 2021
prof. Ing. Martin Hanel, Ph.D.	**prof. RNDr. Vladimír Bejček, CSc.**
Head of department	Dean

Prague on 29. 03. 2021

Declaration of Ownership

I, Fortune Sam Okon, hereby declare that I have independently elaborated the diploma/final thesis with the topic of: Assimilation of Chosen Satellite Data into the Hydrological Balance Modelling and that I have cited all the information sources that I used in the thesis and that are also listed at the end of the thesis in the list of used information sources. I am aware that my diploma/final thesis is subject to Act No. 121/2000 Coll., on copyright, on rights related to copyright and on amendment of some acts, as amended by later regulations, particularly the provisions of Section 35(3) of the act on the use of the thesis. I am aware that by submitting the diploma/final thesis I agree with its publication under Act No. 111/1998 Coll., on universities and on the change and amendments of some acts, as amended, regardless of the result of its defence.

With my own signature, I also declare that the electronic version is identical to the printed version and the data stated in the thesis has been processed in relation to the GDPR.

In Prague, 30.03.21

Acknowledgement

I thank God for the success and perseverance needed to finish this research project. I also want to appreciate my parents and sisters for their love, care and moral support through this time. Most especially, I want to appreciate my supervisor for being patient with me, and guiding me through this phase. I believe this is not the last of my victories and triumphs, they are still many more to come.

Abstract

English

This project entails the assimilation of chosen satellite products within the lumped hydrological modeling. For this purpose the student will calibrate a lumped hydrological model on a selected set of catchments. The student will derive the NDVI data from Modis dataset, and will compare the calibration on results based on input datasets with and without NDVI information. The area of study will be fifteen chosen catchments in the czech republic.

keywords:
runoff,NDVI,ET,entropy

Česke

Tento projekt zahrnuje asimilaci vybraných satelitních produktů v rámci koncentrovaného hydrologického modelování. Za tímto účelem student provede kalibraci koncentrovaného hydrologického modelu na vybrané sadě povodí. Student odvodí data NDVI z datové sady Modis a porovná kalibraci s výsledky na základě vstupních datových sad s informacemi NDVI a bez nich. Studijní obor bude patnáct vybraných povodí v České republice.

klíčová slova:
odtok, NDVI, ET, entropie

Table of Contents

1. Introduction

So many definitions have been postulated about what a model is but the one I stuck with the most is:

“A model is a simplified description of reality”

As simple as it sounds this may be the best definition, I understand that acutely describes what a model is meant to represent.

Rain is a natural occurrence as we all know, still it is but one of the state waters takes as it transverses various states as shown in the water cycle.
As scientists and curious minds, we seek to understand the process of rainfall to surface runoff of designated catchments for the creation of water balance models using observed time series data while seeking to implement remote sensing data gotten from satellite imagery.
We will be using a lumped conceptual model to simulate this process. A lumped conceptual model is a model that assumes that there is no infiltration and so we have just the surface runoff output to model without considering the infiltration.
As the age of Big Data is upon us, we as well need to make the effort to adapt to this change and make the processing and assimilation of data for research and practical purposes as efficient and clean as possible.
In terms of models we take some words of guidance from the great German scientist;

“*Every model must be made as simple as possible, but not simpler.*”

– Albert Einstein

Emerging Technologies has made it possible for us to be more efficient with our processing of data with terms like Remote Sensing being the lead in terms of technologies used in environmental science projects and research topics. Thus, models built based on data gotten from remotely sensed locations using satellites or drones has been proven to provide more accurate results after calibration and validation has been done.
The Water Balance Package in CRAN called BILAN that was designed by a team at T.G Masaryk Water Research Institute in Prague, Czech Republic was used to build this model.

The idea is not to describe this aspect of reality completely or perfectly for that would be impossible, but to provide a simplified description of its processes to aid in educational research purposes or practical purposes if applicable.

In the last phase of this project, we will be applying the principle of Entropy to our final dataset to further optimise our results.
Entropy is basically associated with randomness, and it's a method of working with a random variable in a probability distribution while associating this variable with information and how much uncertainty is present as a result of this information. In simpler terms, entropy links information with uncertainty in random variables before observation (Singh, 2013, 21). It's also a term that has been in the Machine Learning industry for a pretty long time and it was made even popular by Shannon's Information Theory definition.

According to (Singh, 2013, 1) , some of the definitions of information include ‘‘information’’ is variously defined. In Webster’s International Dictionary, definitions of ‘‘information’’ encompass a broad spectrum from semantic to technical, including ‘‘the communication or reception of knowledge and intelligence,’’ ‘‘knowledge communicated by others and/or obtained from investigation, study, or instruction,’’ ‘‘facts and figures ready for communication or use as distinguished from those incorporated in a formally organized branch of knowledge, data,’’ ‘‘the process by which the form of an object of knowledge is impressed upon the apprehending mind so as to bring about the state of knowing,’’ and ‘‘a numerical quantity that measures the uncertainty in outcome of an experiment to be performed.’’

Along the lines of my project I came across a situation where the correlation of the major variables of NDVI and ET produced a non-linear value. More accurately, the final dataset from the daily time series data produced a non-linear correlation. From the definition of entropy, we know that this is the most ideal situation for its application because the non-linear correlation indicates a state where the data exhibits randomness according to observation.

1.1 Literature Review

1.1.1 Modelling

Models have become a mainstay in our society today as we see it traces in countless fields. With the most identifiable being in the economics and environmental science fields respectively. According to (Jakeman et al., 2006, 2), models are increasingly being relied upon to inform and support natural resources management, they are incorporating an even broader range of disciplines and now often confront people without strong quantitative or strong modelling building backgrounds. This just shows how important it is for people to have an idea of what models are and what modelling can do hence the sub-topic of this part in this literature review.

In simpler terms, a model is an abstract of reality; it is a mechanism that converts input to output by means of a set of relationships in the form of algebraic equations, differential equations, ordinary differential equations, partial differential equations and integral equations when looking from a mathematical point of view, (Izquierdo et al., 2004,2).
To understand models, it is important to know that there exists a classification for models as follows:

- Deterministic.
- Stochastic.
- Fuzzy.

According to (Rey, 2015,2), deterministic models are a class of models that are strict and inflexible, it is a model in which the values for the dependent variables of the system are completely determined by the parameters of the model while stochastic models are the direct contrast because they introduce a randomness in such a way that the outcomes of the model can be viewed as probability distributions rather than unique values.

Fuzzy models are a bit unique as they have something close to randomness but not quite, they are not also very precise and in a sense are said to be vague. It uses principles gotten from the fuzzy logic and fuzzy set theory. Fuzzy sets are a generalisation of conventional set theory that were introduced by Zadeh in 1965 as a mathematical way to represent vagueness in everyday life. Its interpretations of data structures are a very natural and intuitively plausible way to formulate and solve various problems, (Bezdek, 1993, 2).
As with any professional practice, it is easy for the inadequately skilled to make errors or mistakes in their determination of results or outputs from a model. Another insight from (Jakeman et al., 2006, 2) on this is that the uses of models by managers and interest groups, as well as by modellers, bring dangers. It is easy for a poorly in- formed non-modeller to remain unaware of limitations, uncertainties, omissions and subjective choices in models. The risk is then that too much is read into the outputs and/or predictions of the model. There is also a danger that a model is used for purposes different from those intended, making invalid conclusions very likely.
There should be a clearly defined goal for each model so as to keep the modeller in focus and not deviate from the final results.

1.1.2 Hydrology and Modelling

Hydrology as a field is important for humans because this is a field that studies one of the most necessary resources for our survival on the earth. Water as a source of life and thirst quencher can also pose some problems on its own if left unchecked hence the need for a science that studies the process through which this water is created, recycled and recreated. This process looks like a loop, so it is termed a water cycle with water as a resource having varying states such as gaseous, liquid or solid depending on the temperature at that particular state.

According to (Ojha et al., 2008, 9), Hydrology is associated with the circulation of water in nature and the human influence on this system. The water transport can be conceptualized as a combination of the natural circulation in an exterior system and an inner man- made system where humans tap water from the outer system and return it back after shorter or longer use, unfortunately quite polluted.

1.1.3 Remote Sensing, Satellite Imagery & Formats.

With the advent of the satellite came the various possibilities associated with it. The presence of the various orbiting satellites made it possible for things like analysis of land topography, vegetation cover, atmospheric conditions and classification of cloud types. These advantages open up a door of possibilities when it comes to monitoring, planning and prevention of natural disasters occuring on the surface of planet Earth.

For this thesis, the focus is going to be on data provided by orbiting satellites from covering different portions of the electromagnetic spectrum at different spatial, temporal and spectral resolutions. This data comes in the form of fused images which may provide increased interpretation capabilities and more reliable results since data with different characteristics are combined. The images vary in temporal, spectral and spatial resolutions which give a more complete view of the observed objects. (Pohl & Genderen, 1998).

Along this lines according to (Pohl & Genderen, 1998), the concept of image fusion used in the provision of this satellite data comes with some key considerations to take note of including:

- What is the objective of the user?
- What types of data are most useful for meeting these needs?
- Which is the best technique for fusing these data types for that application?
- What are the necessary pre-processing steps involved?
- Which combination of the data is most successful?

These questions above are behind the production of most if not all satellite imagery used by professionals in the field who apply these images into their various needs. For us to understand how and what images would best fit our needs, we have to understand the components of this data and how they are produced or processed. At a glance, the questions asked might seem plain but the parameters associated with them encompass a large number. The first question to be answered and the most important is what is the objective of the user that needs this data?
Knowing our project needs and the parameters that are associated with it we would be right to consider the spatial and spectral resolutions which ultimately have an impact on the choice of our remote sensing data (Pohl & Genderen, 1998). The next question asks which types of data are suitable to satisfy our needs. Which should take into consideration that specific data types are mostly needed for respective projects or analysis and this also depends on things like satellite coverage, operational aspects of the agency running the satellite, atmospheric constraints, etc (Pohl & Genderen, 1998). Next we have the best technique for fusing the data types into a presentable imagery. This is also an important aspect because the subsequent steps also depend on this. For this step, a lot of things have to be taken into consideration. Things like the season and weather condition, observed area and selection of appropriate interpretation methods. For most of these parameters, the data has to be critically evaluated using ground truth for verification (Pohl & Genderen, 1998).

Refer to the picture below showing a description of the processing involved in the image fusion aspect.

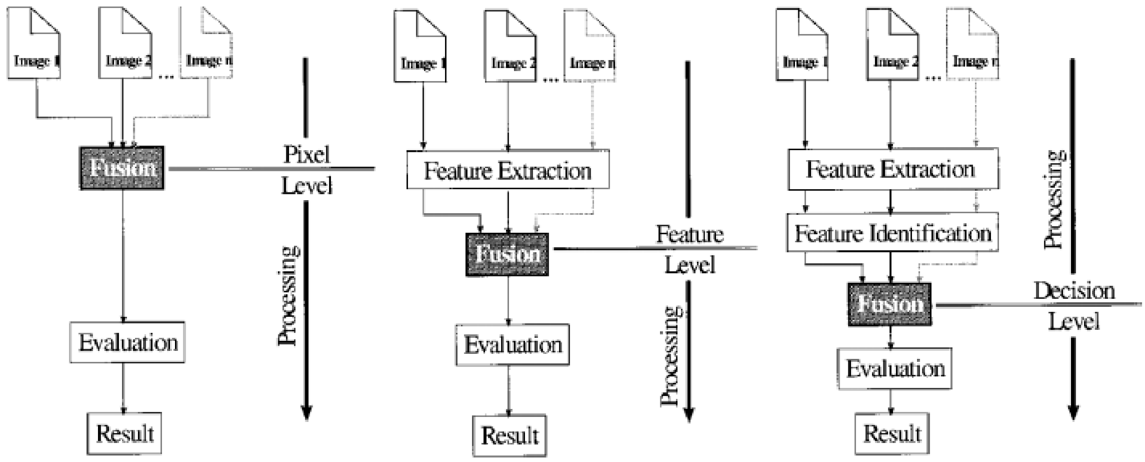


Figure 1. Processing Levels Associated with image fusion. (Pohl & Genderen, 1998, 825)

From the picture above, we can see the process involved in the fusion of satellite images. From the pixel level, feature level and decision level, we see a systematic approach to image fusion and this provides a new perspective while also opening our eyes to the steps used in producing satellite data which we so readily use for various projects.

For the purpose of this thesis, the format of satellite data we are mostly interested in is called the Normalized Difference Vegetation Index (NDVI). And the formula for calculating NDVI is shown below as follows:

$$NDVI = (\rho_{nir} - \rho_r) \div (\rho_{nir} + \rho_r) , \qquad (1)$$

Where ρ_r and ρ_{nir} are the spectral reflectance in the red and infrared channels respectively (Beck et al., 2011). In some studies, NDVI has been shown to be related to other biophysical variables like the Leaf Area Index (LAI) (Beck et al., 2011, 2547).

1.1.4 BILAN Water Balance Model.

The BILAN water balance model is a hydrological model that can be used in water balance modelling, hydrological analysis and forecasting of hydrological parameters directly related to the weather and climate change.

This novel hydrological modelling tool was designed and developed by the research team at TG Masaryk Water Management research institute, Prague Czech Republic. The aim and purpose of the modelling package is the same as almost any other modelling tool around these days, the only difference being that it is a water balance modelling tool otherwise known as a hydrological modelling tool. The use of such tools is mainly to preserve or better equip hydrological experts with the knowledge of how to preserve nature's gift to us in the form of water and to manage its resources properly. This also involves preventing the damage or loss of life and property that normally accompanies most natural disasters that relate to water (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).

According to (T. G. Masaryk Water Management Research Institute, vvi et al., 2015), the model has been designed to simulate the various components that contribute to water balance in a catchment. Mainly these components are contributors to what we also call a water cycle and the model has been designed such that it describes and functions in a similar manner only more simple in complexity. As we know, man has yet to be able to simulate environmental variables perfectly. This has remained a cause to question when we will be able to perfectly simulate environmental variables, but that is a quest for the future. (Grutzner, 1996).

For the input, the data used can either be a daily or monthly series of precipitation, air temperature and relative humidity which can be optional while for the calibration, the parameters which are gotten by applying the optimization algorithm uses the daily or monthly series of simulated or observed runoff at the outlet of the basin (T. G. Masaryk Water Management Research Institute, vvi et al., 2015). The model also simulates as a result a monthly or daily time series of actual evapotranspiration, potential evapotranspiration, infiltration to the soil and recharge from the soil to the aquifer. (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).
Because this model is a lumped conceptual model, total runoff for the simulation consists of 3 components namely direct runoff, interflow which only applies to the monthly time step and finally the baseflow. (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).

There is a conscious effort put in by the research team who developed this model to describe the internal structure of the model and these structures are aided by the appropriate mathematical and statistical formulas or functions that are used in the field of modelling not just in hydrological modelling but in a much more broader scope. Specific components refer to formulas that relate to hydrology especially since this is a hydrological model. For the components that make up the internal structure of this tool first we have Oudin's method for PET estimation (Oudin et al., 2010), (T. G. Masaryk Water Management Research Institute, vvi et al., 2015). PET simply means Potential Evapotranspiration and it is one of the components of the bilan model that is estimated using variables like air temperature from Oudin's method. Mathematically, we have;

$$\mathrm{PET(i)} = \left\{\frac{0.408Re(T(i)+5)}{100}\right\} \qquad \text{For } \mathrm{T(i)} + 5 > 0$$
$$\text{For } \mathrm{T(i)} + 5 \leq 0 \qquad (2)$$

Where, i= days which requires a single inlet air temperature, (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

Re = extraterrestrial radiation [MJ, m -2 d -1], (T. G. Masaryk Water Management Research Institute, vvi et al., 2015), according to (Allen et al., 1998)

$$\mathrm{Re\ (i)} = \frac{24.60}{\pi} G_{SC}\, d_r [\omega_s \sin\varphi \sin\delta + \cos\varphi \cos\delta \sin\omega_s] \qquad (3)$$

G_{SC} = solar constant 0.082 MJ, m -2 min -1, (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

d_r = inverse relative distance of the Earth and the Sun, (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

$$d_r = 1 + 0.033 cos\left(\frac{2\pi}{365} J\right) \quad (4)$$

J = serial number of the day of the year, (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

δ = declination from the Sun or angular distance from the equator [rad], (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

$$\delta = 0.409 sin\left(\frac{2\pi}{365} J - 1.39\right) \quad (5)$$

ω_s = hour angle of sunset [rad], (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

$$\omega_s = arccos[- tan\varphi \; tan\delta] \quad (6)$$

Next among the components is the description of the daily/monthly model type. For the daily model type we have;

Depending on the conditions of the given day, the model identifies which conditions represent summer and winter respectively according to the average daily air temperature. (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

Summer conditions are chosen when temperature is:

$$T(i) \geq 0 \quad (7)$$

The model also simulates the total runoff RM(i) of the daily type as follows:

$$RM(i) = DR(i) + BF(i) \quad (8)$$

Where, DR(i) = direct outflow, (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

BF(i) = basic outflow. (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

The direct outflow (DR(i)) represents the rapid response of the river basin, while the basic outflow (BF(i)) represents a situation where the residence time in the river basin is longer than the direct outflow and consists basically of an outflow from the groundwater supply. (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).

Now we have conditions that affect territorial evaporation and hydrological balance in the soil under summer conditions. Under these, we have the conditions that apply if infiltration which is equal to precipitation is greater than or equal to the potential evapotranspiration. Mathematically this is represented as;

$$\mathrm{INF(i)} \geq PET(i) \tag{9}$$

In this case then territorial evaporation is equal to potential evapotranspiration, (T. G. Masaryk Water Management Research Institute, vvi et al., 2015)

$$\mathrm{ET(i) = PET(i)} \tag{10}$$

Some other conditions include, excess water can make up soil moisture,

$$\mathrm{SW(i) = SW(i\text{-}1) + INF(i) - PET(i)} \tag{11}$$

Incase the soil moisture capacity exceeds its limits,

$$\mathrm{SW(i)} > spa \tag{12}$$

Then the left over water percolates down,

$$\mathrm{PERC(i) = SW(i)} - spa \tag{13}$$

Also the soil moisture supply SW(i) is then equal to the *spa* (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).
Bilan also takes note if the potential evapotranspiration exceeds the precipitation (infiltration), then the territorial evapotranspiration is subsidized from the soil moisture supply which is emptied (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).

$$\mathrm{SW(i) = SW(i\text{-}1)}\ .\ e^{\frac{INF(i) - PET(i)}{spa}} \tag{14}$$

Where, *e* = natural logarithm.

For the conditions of territorial evapotranspiration and hydrological balance in winter using bilan, the most peculiar condition would most likely be that infiltration is equal to zero (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).
Under certain conditions like was experienced in this research project the following apply to scenarios that involve winter condition internal calculations.

According to the bilan user guide, if the sum of precipitation and water supply exceeds the potential evapotranspiration on a given day, it is assumed that the territorial evapotranspiration is equal to the potential evapotranspiration (T. G. Masaryk Water

Management Research Institute, vvi et al., 2015). Mathematically, the following formulas give a better description of this condition.

$$ET(i) = PET(i) \text{ , for } SS(i\text{-}1) + P(i) \geq PET(i) \quad (15)$$

Adversely, the territorial evapotranspiration is calculated as the sum of the precipitation and water supply in the snow (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).

$$ET(i) = SS(i\text{-}1) + P(i), \text{ for } SS(i\text{-}1) + P(i) < PET(i) \quad (16)$$

Then the amount of leftover water in the snow becomes,

$$SS(i) = \max(SS(i\text{-}1) + P(i) - PET(i), 0) \quad (17)$$

Where SS(i-1) is the water supply in the snow cover on the day i-1,
Infiltration therefore is zero,

$$INF(i) = 0 \quad (18)$$

Previously, it was stated that bilan is a package that was used to develop a lumped conceptual model for this thesis. Now we will briefly describe the internal calculations related to groundwater and basic runoff while also providing mathematical descriptions of that process.
Groundwater supply is represented as GS, so the groundwater supplyon day i is calculated as the sum of the supply on the previous day and the groundwater subsidy represented as RC(i) (T. G. Masaryk Water Management Research Institute, vvi et al., 2015). Then the basic runoff is represented by the runoff from groundwater, which is directly proportional to the stock at the beginning of the day and is controlled by the parameter Grd (T. G. Masaryk Water Management Research Institute, vvi et al., 2015). Mathematically we have,

$$BF(i) = Grd.GS(i\text{-}1) \quad (19)$$

The groundwater at the end of the day therefore becomes,

$$GS(i) = RC(i) + (1 - Grd).GS(i\text{-}1) \quad (20)$$

It is important to also note that if water use variables are included, the groundwater supply is reduced by the abstraction from groundwater POD(i). Also note that the value of the stock cannot be negative (T. G. Masaryk Water Management Research Institute, vvi et al., 2015).

$$GS(i) = \max(RC(i) + (1 - Grd).GS(i - 1) - POD(i), 0) \quad (21)$$

1.1.5 Actual Evapotranspiration (ET).

The field of agriculture island water management is a very broad field, but the concept of actual evapotranspiration is not one that is uncommon among professionals and the like in this area of expertise. Actual evapotranspiration which is mostly represented as ETa has major significance in agriculture and water management as it is one of the best ways to monitor and better assess water management and improve decision making in the agricultural sector (Kibria et al., 2021). For a better understanding of actual evapotranspiration, we will first review what evapotranspiration as a standalone term means in environmental science.

Evapotranspiration is basically water released to the atmosphere from the soil surface through the evaporation process and from crops through the transpiration process (Kibria et al., 2021). According to (NC State University, n.d.), evapotranspiration accounts for 15% of the planet's water vapor and this makes it an important process in the water cycle.

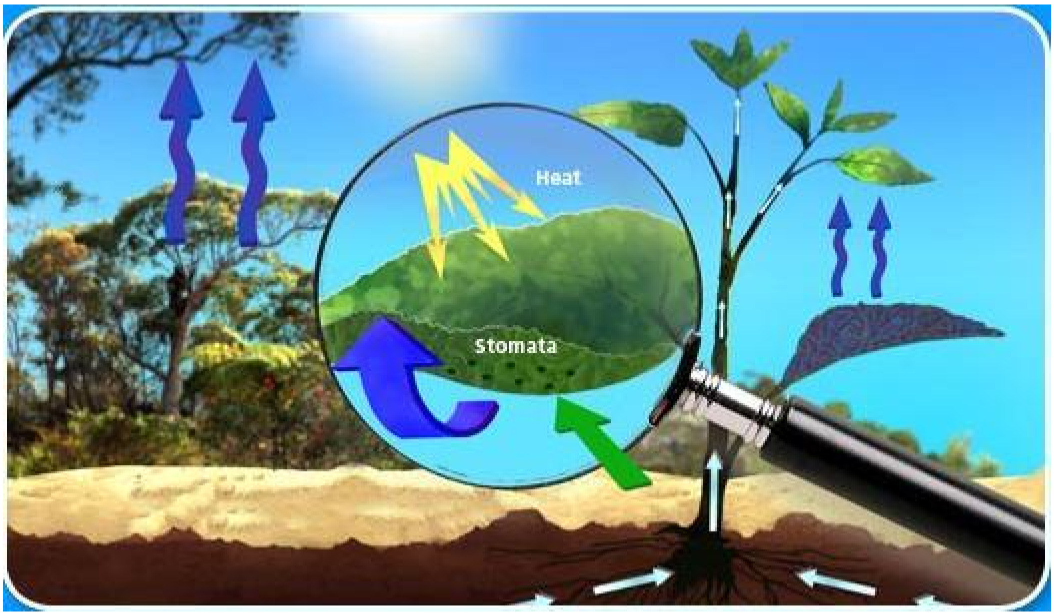


Figure 2. Image showing evapotranspiration process (NC State University, n.d.)

It is so important because it is a main component in crop maintenance when we are talking about agriculture and also in terms of hydrology it is equally important due to its level of importance among the processes that make up the water cycle (NC State University, n.d.).

It is also important to note that apart from its importance to earth's water cycle, ET also helps in surface energy balance and therefore is a crucial role player in climate modulation and the hydrological cycle (FAO United Nations, n.d.). Humidity and precipitation among other things are also influenced by ET (Wang et al., 2015). With the increase in world population we also see an increase in agriculture or more specifically irrigated agriculture; adversely this indicates an increase in the demand for water (Kibria et al., 2021). The effect

on ET can also be evident by the activities of man through land use and water management (Gordon et al., 2008). Coming along with rise in water demand, the availability of water is increasingly threatened (Rockström et al., 2009). All this just goes to show how important ET is to the environment, it is a non-negotiable component of hydrology on earth.
To be able to accurately estimate the ET, it is important to have the right instruments for measurements. Measuring ET from crops under actual growing conditions require special instruments (Allen et al., 1998). And what we know as actual evapotranspiration is relevant to the agricultural industry, it is also relevant to the engineering and modelling industry. Water balance encompompasses components like Evapotranspiration so it is inevitable that an important component such as ET is needed to get a more accurate result from the model.

The pressing question among hydrologists/environmental scientists and engineers around the world right now would be how to mitigate the hydrologic effects of climate change even as they also require better understanding of how rural vegetation management affects water budgets and stream flows (Younger et al., 2020). Because of higher leaf area indices, interception and bit more deeper rooting, forest cover increases ET and reduces the average streamflows relative to lawns, pastures and croplands within the same hydroclimatic region (Teuling, 2018). Nevertheless, like all data, measurements of water-budget are subject to a level of uncertainty while accurate and precise measurements of ET is normally difficult and expensive (Younger et al., 2020).

1.1.6 Mutual Information and Entropy

Mutual information according to (Li et al., 2021), is a metric based on the mutual independence between two random variables and it can quantify the amount of information obtained from one variable through observing another variable.
This way makes it objectively the best way to show the correlation between two variables. (Li et al., 2021).

Given A Pair Of Random Variables, (X,Y),over the space X×Y.Their joint density function is p(x,y), and the marginal prob-ability density functions are p(x) and p(y) respectively. (Li et al., 2021)
Mutual Information can be defined as follows.

$$I(X;Y) = H(X)+H(Y)-H(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(X)*p(Y)} \quad (22)$$

Many challenges plague the performance of real world applications and improving the classification of performances is one of the main challenges. Undesirable features of data unfortunately seem to be drawn to many classification models, features such as redundancy and irrelevance (M et al., 2021).

For calculating entropy of a probability distribution, here a number of steps that may prove useful putting in mind the formula above;

- Define probability
- Create probability distribution
- Calculate entropy for each distribution
- Plot results.

Relative to the step above, the solution it provides provides us the ability to deal with multiple probability events. The next steps we give insight into situations where we have a probability distribution for single events;

- Specify the number of events.
- Get the probability of one event
- Calculate the entropy using the given formula.

Entropy works with a principle where a probability distribution which is skewed is less surprising than a balanced probability, whereas a balanced probability distribution has or is more surprising hence it has a higher entropy. We should note that entropy quantifies the amount of information in a random variable in bits. (Singh, 2013)

For a dataset with classifications, (Singh, 2013)

$$H(S) = -(P(X) * Log (p(X)) + P(Y)* Log(P(Y)) \quad (23)$$

Where, S= dataset,
H= entropy,
X=sample class 1,
Y=sample class 2

Also note that binary classifications is the same as splitting a dataset into two classes in layman terms. In this way entropy can be used to calculate the purity of a dataset meaning how balanced the distribution of the classes happens to be. A smaller entropy suggests more purity or less surprise. To evaluate the impact on purity by splitting a dataset S by a random variable *a* with a range of values we use the following;

$$IG(S,a) = H(S) - H(\frac{S}{a}) \quad (24)$$

Where, H(S) = entropy for the dataset before any change,
$H(\frac{S}{a})$ = conditional entropy for the dataset given variable *a,*
IG(S,*a*) = Information for dataset S for the variable *a* for a random variable.

The above formula describes the gain in the dataset S for the variable *a.* It is the number of bits saved when transforming the dataset.

In practical terms, mutual information may be used to substitute certains statistical terms for data analysis especially because the complexity and uncertainty in data increases when the distribution is non-linear.

2. Objectives

The aim of the thesis is to assimilate chosen satellite products within the lumped hydrological modeling. For this purpose, the student will calibrate a lumped hydrological model on selected sets of catchments. The student will derive the NDVI data from Modis dataset, and will compare the calibration results based on input datasets with and without NDVI information.

3. Methodology

3.1 The preparation of input data for hydrological modeling for a selected set of catchments.

The data used for this project accommodates 15 catchments and was cleaned and prepared using R and Rstudio.

First let us have a look at the study area we're going to be dealing with. The area of concentration for this project is going to be in Czech Republic also called Czechia, a country in Central Europe with historical provinces namely- Bohemia and Morovia respectively and total land area in square kilometers of about 78, 865km^2 (Hauner et al., 2021). Czechia has a population of about 10.7 million people as of the year 2020 (Hauner et al., 2021). The basin shown in the map below are the basins we are going to be working with.

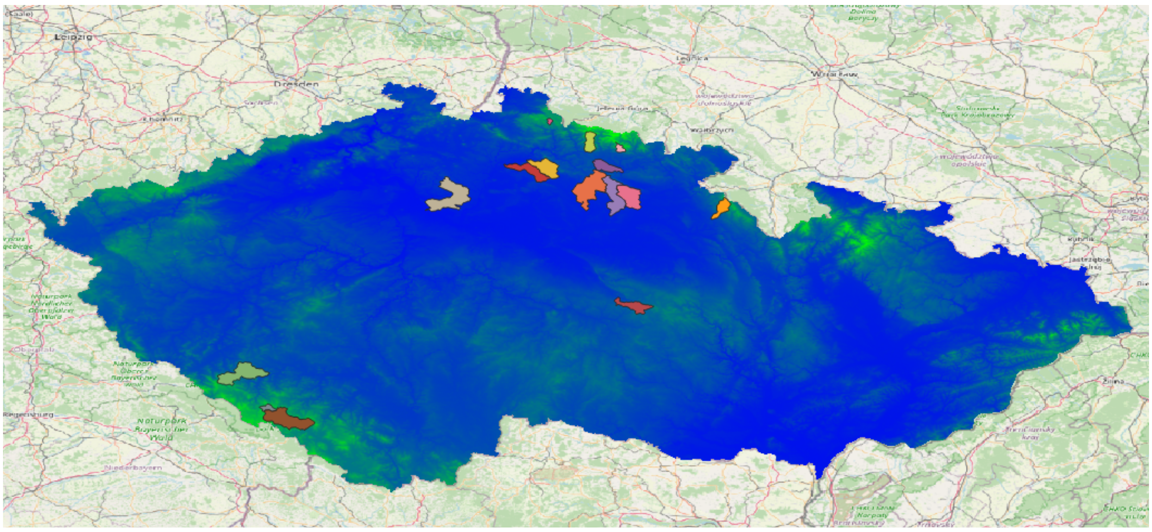
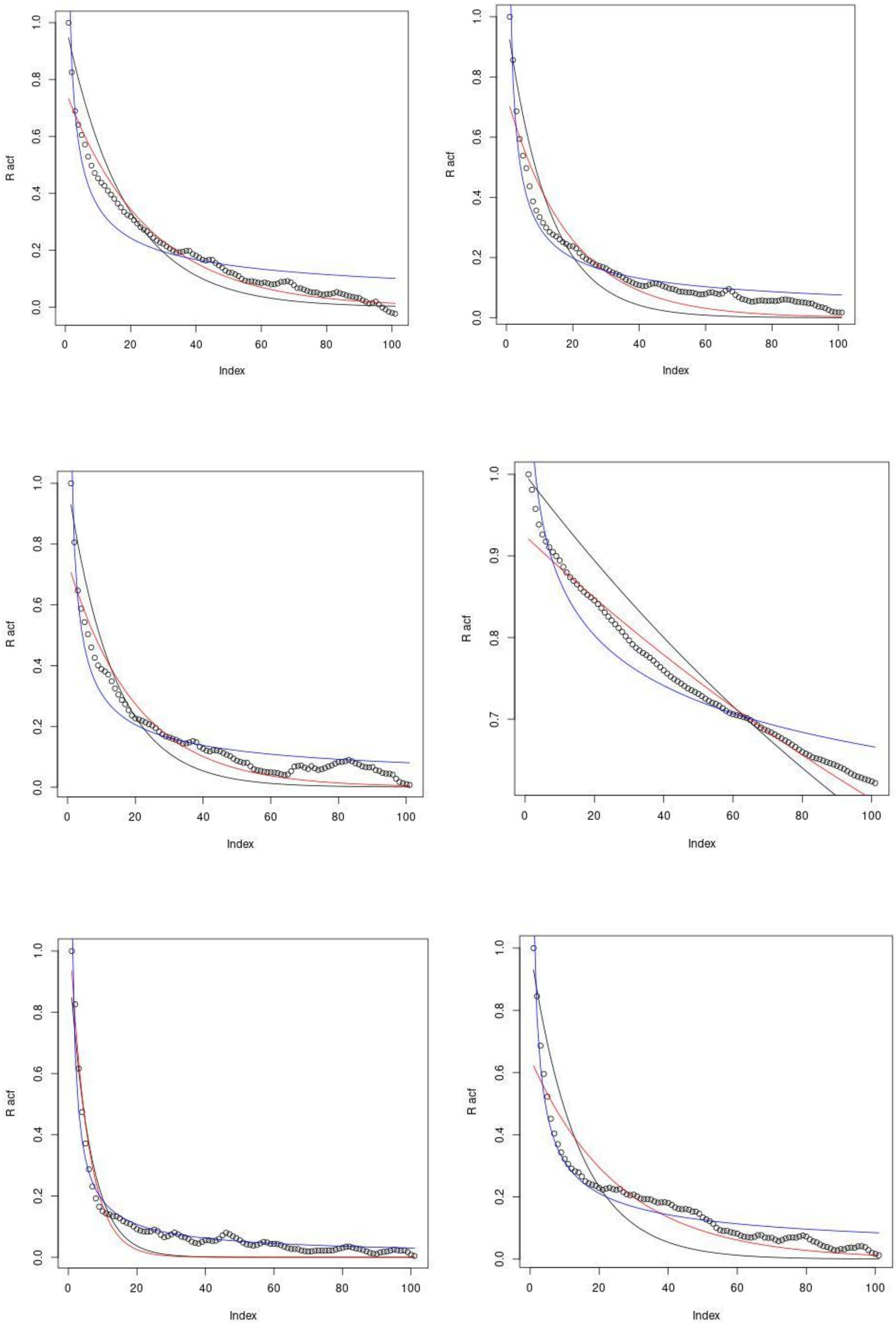


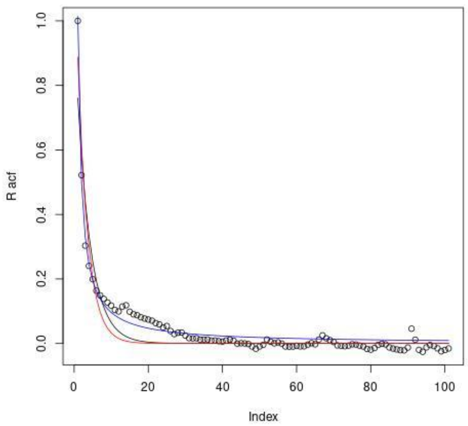
Figure 3. Digital Elevation Model QGIS image showing marked catchments.

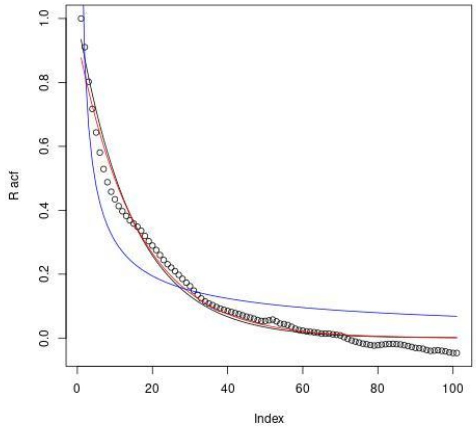
Key:

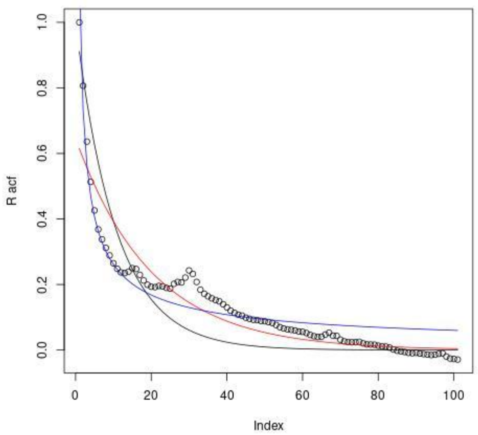
The input data was estimated from an aggregation of meteorological data provided by the Department of Water Resources and Environmental Modelling to estimate and calibrate the model. The time duration of the data used ranges between 1990 till 2010 for the meteorological data.

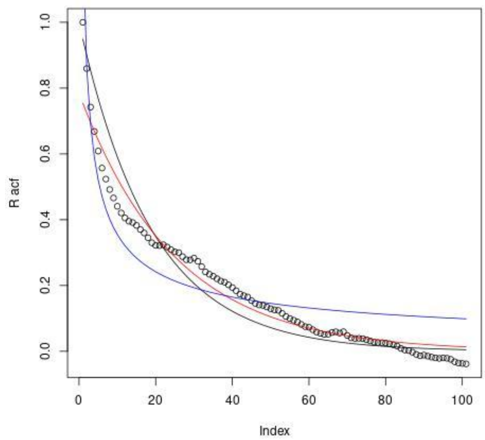
The following plots were estimated using the aggregated data to calculate the autocorrelation of the runoff of the selected catchments before modelling.

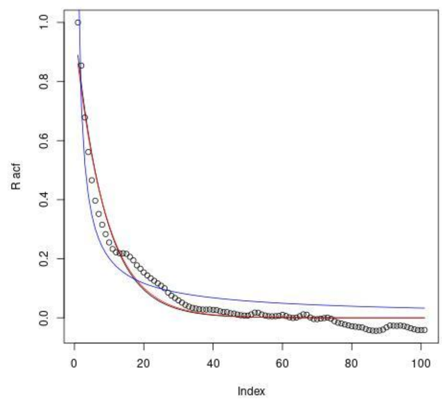

R acf
Index

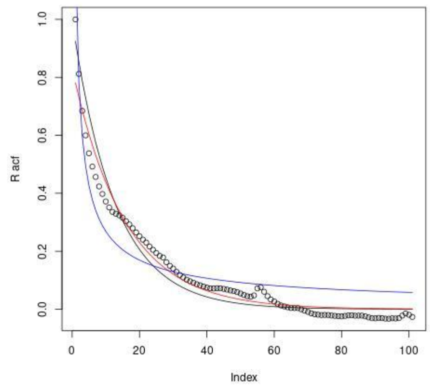

R acf
Index

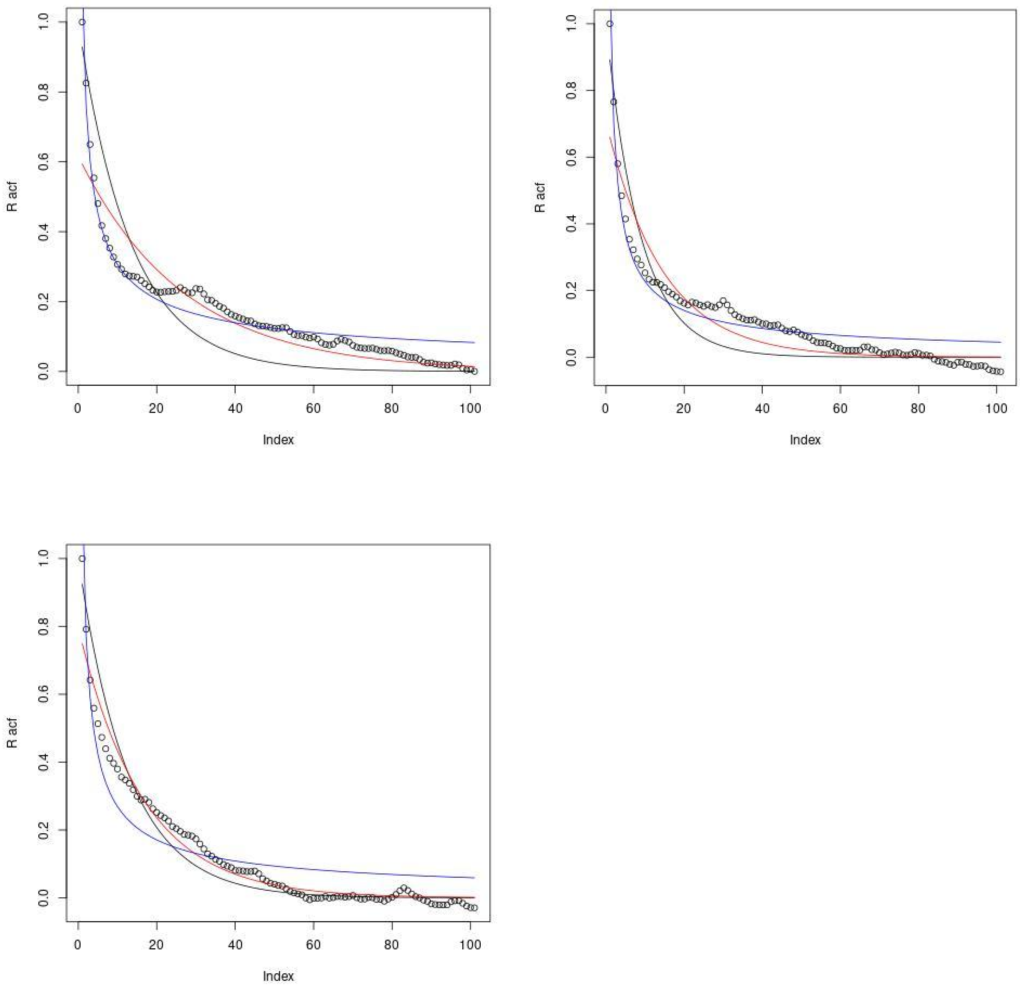

R acf
Index


R acf
Index


R acf
Index

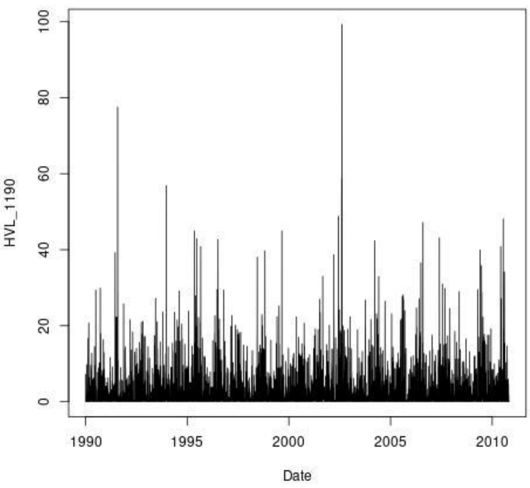

R acf
Index

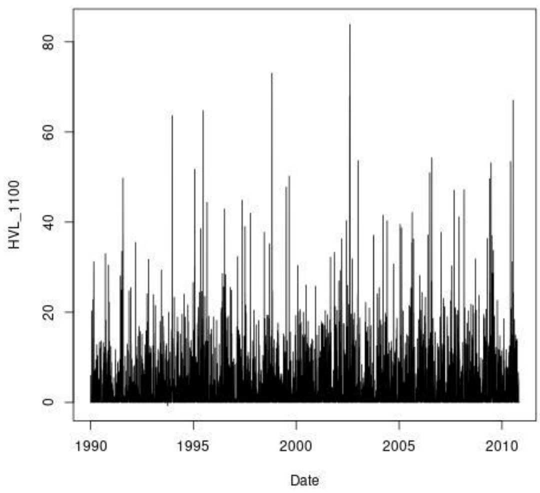

R acf
Index

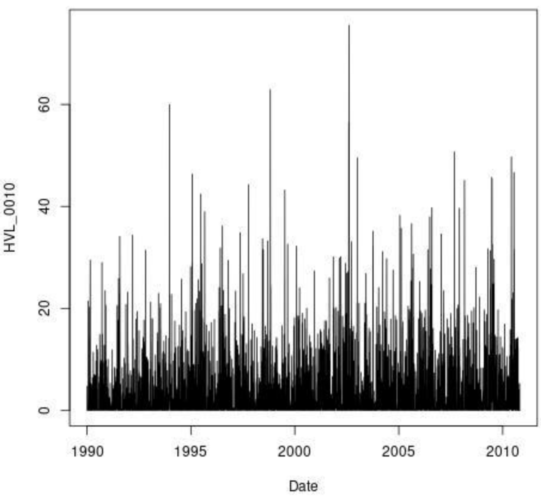


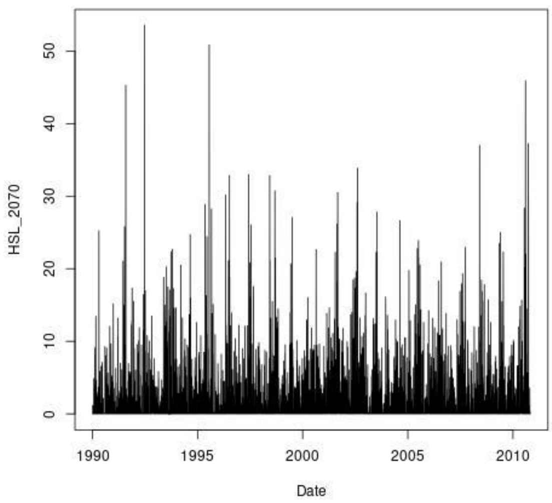
Figures 4-19. Autocorrelation plot for 15 catchments derived from aggregated data.

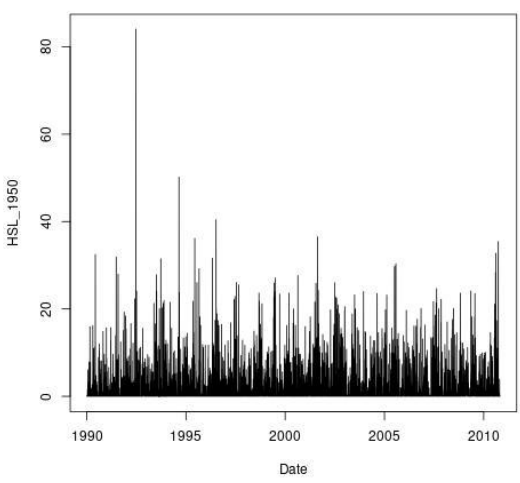
The following plots were estimated using the aggregated data to visualise the estimated rainfall intensity of the selected catchments before modelling.

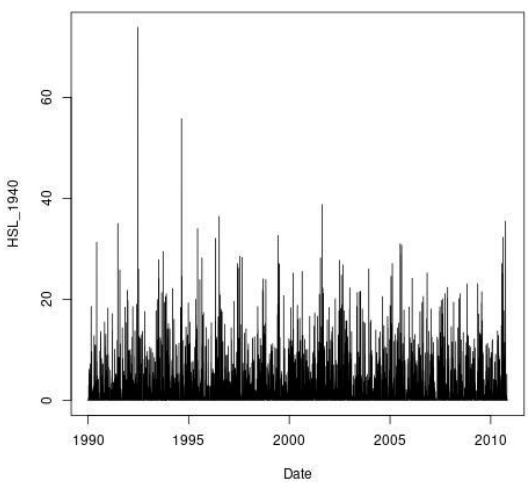

HVL_1190
0
20
40
60
80
100
1990
1995
2000
2005
2010
Date

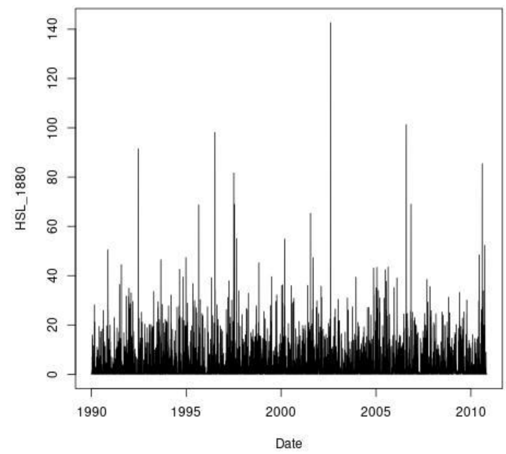

HVL_1100
0
20
40
60
80
1990
1995
2000
2005
2010
Date

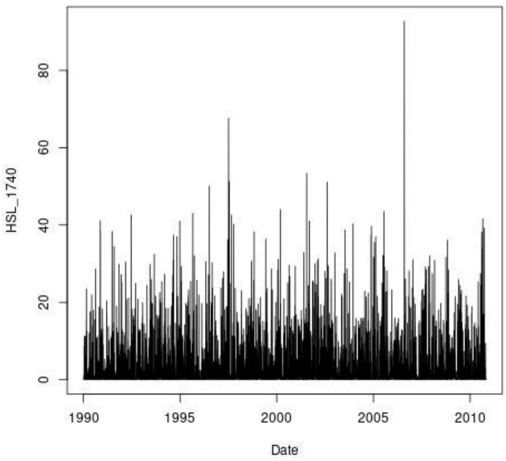

HVL_0010
0
20
40
60
1990
1995
2000
2005
2010
Date

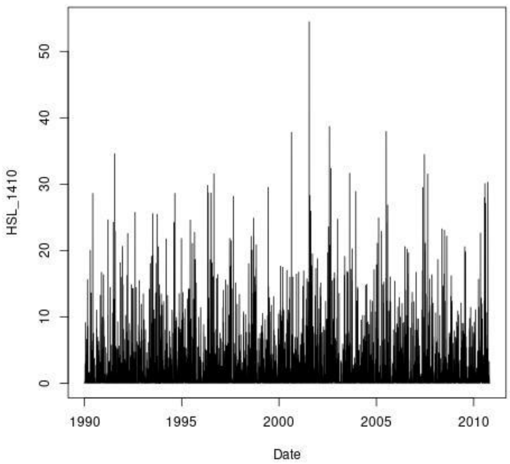

HSL_2070
0
10
20
30
40
50
1990
1995
2000
2005
2010
Date

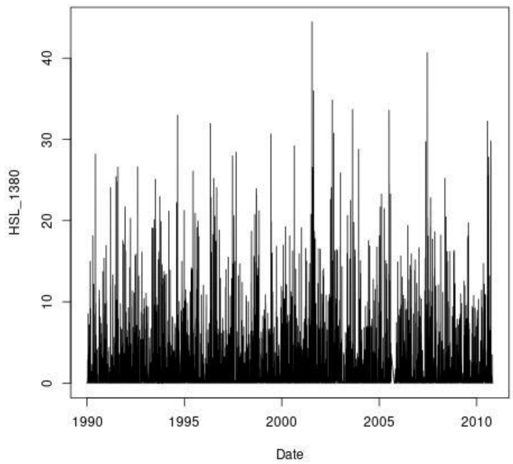

HSL_1950
0
20
40
60
80
1990
1995
2000
2005
2010
Date

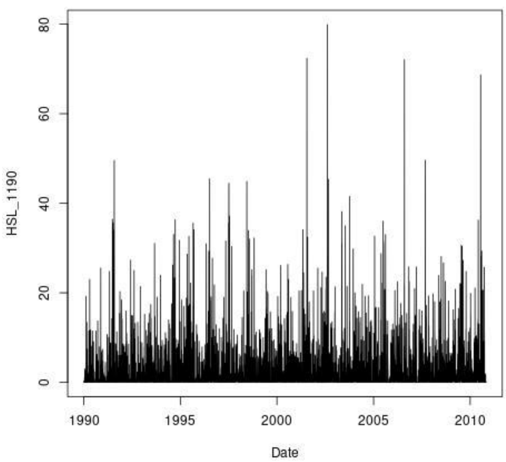

HSL_1940
0
20
40
60
1990
1995
2000
2005
2010
Date

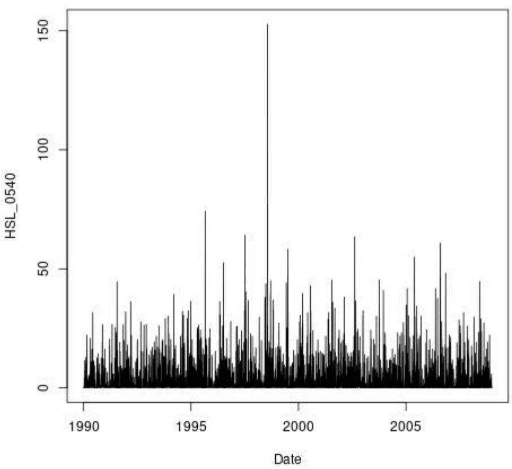

HSL_1880
0
20
40
60
80
100
120
140
1990
1995
2000
2005
2010
Date

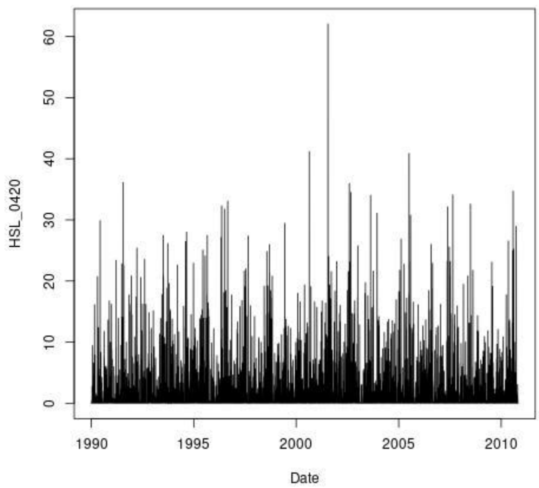

HSL_1740
0
20
40
60
80
1990
1995
2000
2005
2010
Date

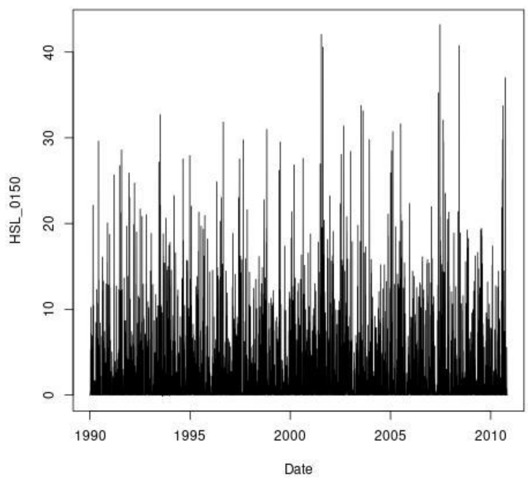

HSL_1410
0
10
20
30
40
50
1990
1995
2000
2005
2010
Date


HSL_1380
0
10
20
30
40
1990
1995
2000
2005
2010
Date


HSL_1190
0
20
40
60
80
1990
1995
2000
2005
2010
Date


HSL_0540
0
50
100
150
1990
1995
2000
2005
Date





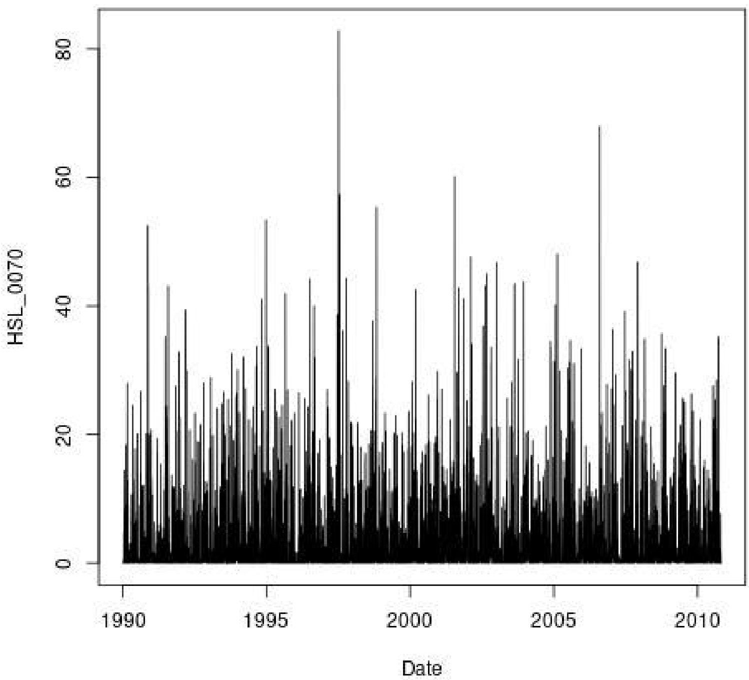


Figure 20-35. Plots showing rainfall intensity for 15 catchments.

3.2 The preparation of satellite information on NDVI data.

The product used for this project is called VIIRS/S-NPP Vegetation Indices (NDVI/EVI) 16-Day L3 Global 500m SIN Grid and is part of the products available using the MODISTools package on R. This remote sensing data provides vegetation indices by a process of selecting the best available pixel over a 16-day acquisition period at 500 meter (m) resolution.

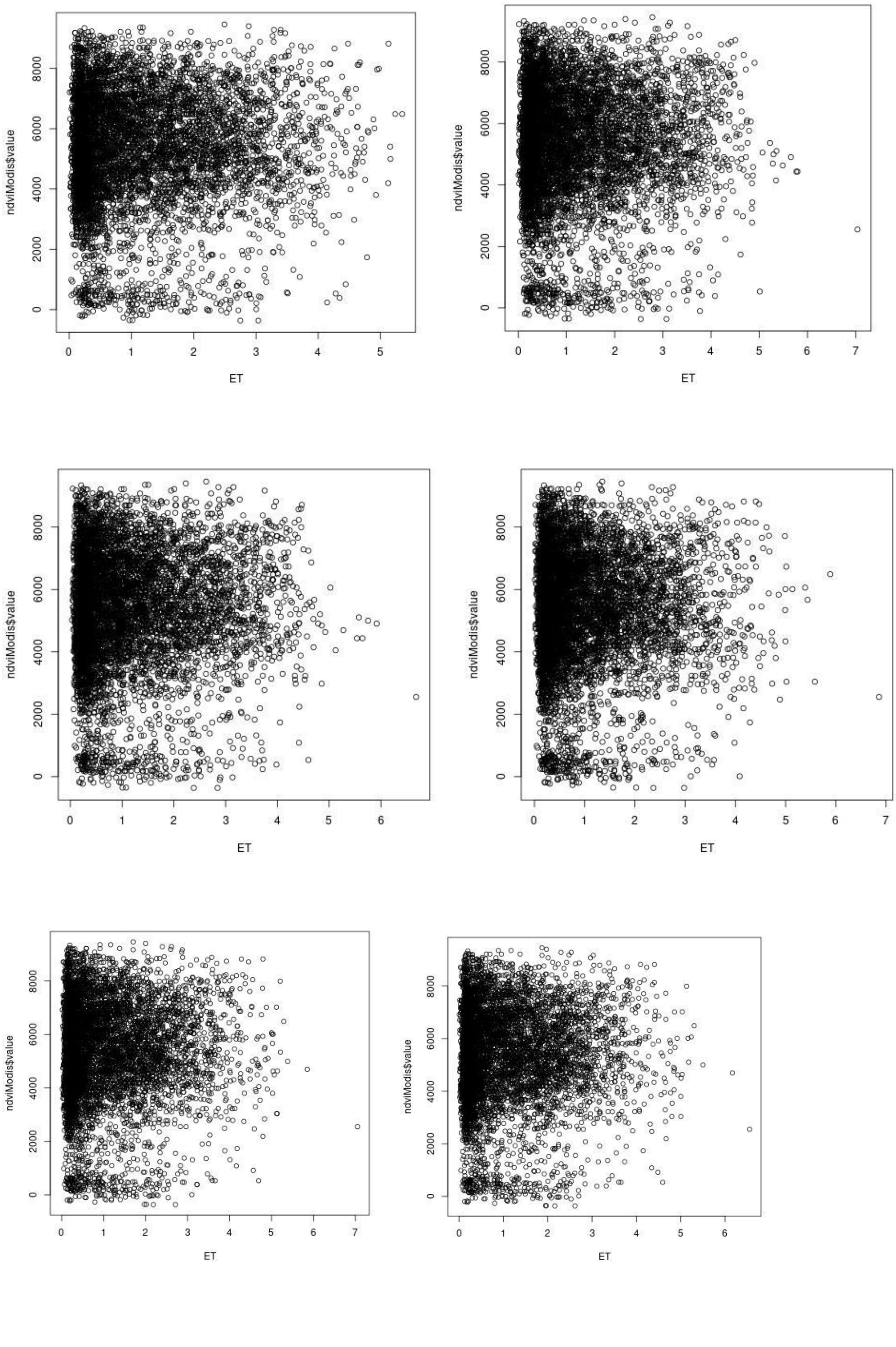
The VNP13 data products are designed after the Moderate Resolution Imaging Spectroradiometer (MODIS) Terra and Aqua Vegetation Indices product suite to promote the continuity of the Earth Observation System (EOS) mission. (NASA/NOAA NPP Project., n.d.).
This is one of the more recent and advanced projects from NASA/NOAA and the package is available for installation on Rstudio using the syntax `install.packages("MODISTools")`.

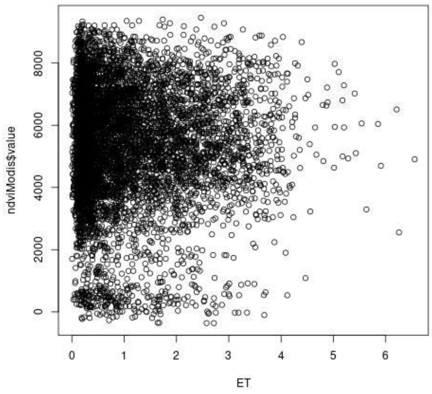
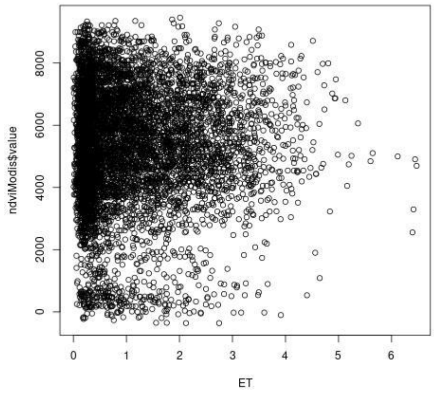
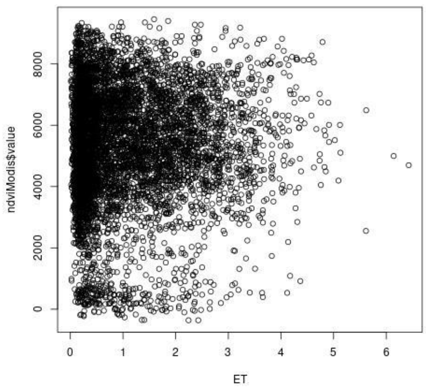
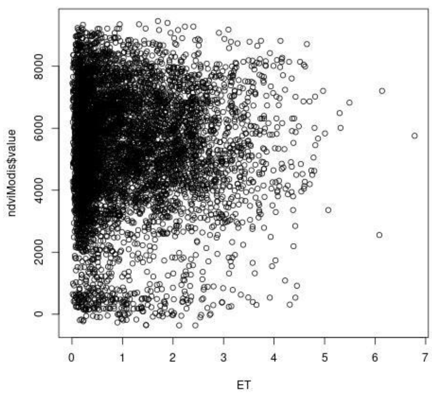
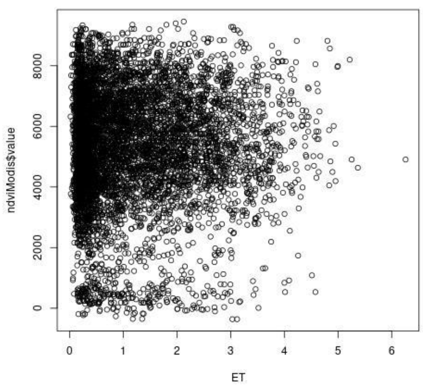
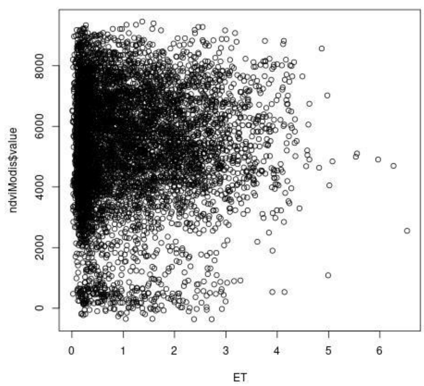
For a bit more understanding, the VNP13 algorithm process produces three vegetation indices: Normalized Difference Vegetation Index (NDVI),the Enhanced Vegetation Index (EVI), and Enhanced Vegetation Index-2 (EVI2). NDVI is one of the longest continual remotely sensed time series observations, using both the red and near-infrared (NIR) bands (NASA/NOAA NPP Project., n.d.). EVI is a slightly different vegetation index that is more sensitive to canopy cover, while NDVI is more sensitive to chlorophyll. EVI2 is a reformation of the standard 3-band EVI, using the red band and NIR band. This reformation addresses arising issues when comparing VIIRS EVI to other EVI models that do not include a blue band. EVI2 will eventually become the standard EVI (NASA/NOAA NPP Project., n.d.).

The next step after getting the required data from the modis package will be to make sure the number of observations for the NDVI data matches the number of observations of the aggregated meteorological input data. For the purpose of this research project that will be about 7609 observations or 6940 observations in cases where there are missing or not available data. For R it is possible to limit the number of observations to the desired number using the code `myData %>% slice(1:6940)` or `myData %>% slice(1:7609)` for both number of observations. Note that to use the methods described above, the package called ***dplyr*** has to be installed first.

The subsequent step will be to create a data table from the data obtained so far and the data table will contain variables date, NDVI value and pixel value alone. This data table now becomes the main input data table for the model when it gets to the calibration phase.

Refer to the plots below to visualise the correlation between aggregated meteorological data and the NDVI data table.


ndviModis$value
ET
ndviModis$value
ET
ndviModis$value
ET
ndviModis$value
ET
ndviModis$value
ET
ndviModis$value
ET


ndviModis$value
0
2000
4000
6000
8000
ET
0 1 2 3 4 5 6

ndviModis$value
0
2000
4000
6000
8000
ET
0 1 2 3 4 5 6

ndviModis$value
0
2000
4000
6000
8000
ET
0 1 2 3 4 5 6

ndviModis$value
0
2000
4000
6000
8000
ET
0 1 2 3 4 5 6 7

ndviModis$value
0
2000
4000
6000
8000
ET
0 1 2 3 4 5 6

ndviModis$value
0
2000
4000
6000
8000
ET
0 1 2 3 4 5 6

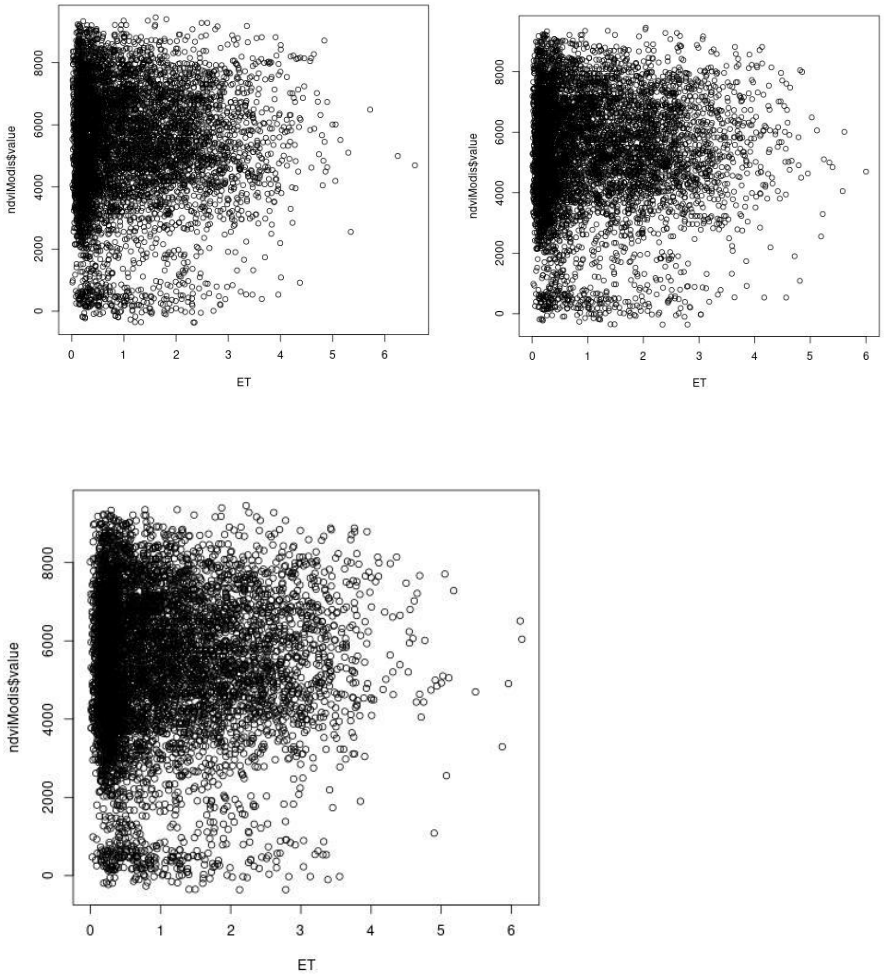


Figure 36-51. Plots showing the correlation between ETa and NDVI for each catchment

3.3 The calibration of a lumped hydrological model without the satellite information.

The following will be about the process of calibrating the bilan model using the aggregated meteorological input data.

Previously we showed the visualisation of the data for each catchment using the autocorrelation function plots. It is important to note that the fitting of the model during calibration so that the simulated data will be as accurate as possible largely depended on the autocorrelation function applied in the form of an objective function as a criteria in bilan.

Objective functions have been used in modelling to calibrate models and all objective functions have adjustable parameters. These parameters are typically tuned using expert knowledge or heuristic rules (Jafrasteh & Suárez, 2021). For this calibration, we involve a number of objective functions developed based on three different indices/criteria of performance. These indices were embedded into the objective functions and adjusted accordingly until a satisfactory result was obtained. The three criteria for performance used in the objective function to calibrate the model are a modified KGE (Kling-Gupta Efficiency), an autocorrelation function and the MAE (Mean Absolute Error). The respective performance indices focused on calibration had parameters like ETa (Actual Evapotranspiration), RM (Total Runoff), R(Direct Runoff) and SW(Soil water content). Mathematically, KGE is described as;

$$KGE = 1 - \sqrt{(r - 1)2 + (\alpha - 1)2 + (\beta - 1)2} \quad (25)$$

Where r = linear correlation coefficient between observed and simulated dataset,

$\alpha = \frac{\sigma s}{\sigma o}$,

$\beta = \frac{\mu_s}{\mu_o}$

Based on the above, KGE can be explained as the euclidean distance to an ideal value in a three dimensional space defined by three components of the modelling error (Liu, 2020).

As for the autocorrelation function, it was used at the beginning of this chapter to visualize the data for this section. Like the KGE, the coefficients for the function are also adjustable based on expert knowledge or heuristic values. For the MAE, mathematically we have,

$$MAE = \left[n^{-1} \sum_{i=1}^{n} \left| e_i \right| \right] \quad (26)$$

Calculation of MAE is relatively simple, it involves summing the magnitudes of the errors to get the total error and then dividing the error by n. This is taking into consideration that $\gamma = 1$ (Willmott & Matsuura, 2005).

These are the performance criteria involved in the calibration of the model for this part of the project.

3.4 The calibration of a lumped hydrological model with the satellite information.

Like the above section, we still use the KGE as one of the performance criteria and also the autocorrelation function as part of the objective functions for the calibration. The only new objective function introduced was the function containing the value from the correlation between the ETa and NDVI.

In reference to a similar topic where the visualisation of correlation between these values was presented above, we see that the correlation shows a non-linear and possibly negative correlation. The parameters used as input for this section still remain the same as the parameters used previously.

The difference between the calibration process used for both scenarios was the use of the objective function related to the correlation of ET and NDVI.

3.5 The comparison of model outputs.

For a comparison of both outputs, see the first set of visuals below for each catchment respectively;

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	595
Number of observations with all compared variables equal	6345
Number of values unequal	595

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	825
Number of observations with all compared variables equal	6115
Number of values unequal	825

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	6940
Number of observations with all compared variables equal	0
Number of values unequal	6940

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	102
Number of observations with all compared variables equal	6838
Number of values unequal	102

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	6940
Number of observations with all compared variables equal	0
Number of values unequal	6940

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	3624
Number of observations with all compared variables equal	3316
Number of values unequal	3624

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	0
Number of variables compared with all values equal	1
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	0
Number of observations with all compared variables equal	6940
Number of values unequal	0

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	6686
Number of observations with all compared variables equal	254
Number of values unequal	6686

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	4394
Number of observations with all compared variables equal	2546
Number of values unequal	4394

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	0
Number of variables compared with all values equal	1
Number of observations in common	6940
Number of observations in x but not y	616
Number of observations in y but not x	0
Number of observations with some compared variables unequal	0
Number of observations with all compared variables equal	6940
Number of values unequal	0

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	669
Number of observations in y but not x	0
Number of observations with some compared variables unequal	6940
Number of observations with all compared variables equal	0
Number of values unequal	6940

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	1
Number of variables compared with all values equal	0
Number of observations in common	6940
Number of observations in x but not y	0
Number of observations in y but not x	0
Number of observations with some compared variables unequal	6871
Number of observations with all compared variables equal	69
Number of values unequal	6871

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	1
Number of variables compared	1
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	0
Number of variables compared with all values equal	1
Number of observations in common	7609
Number of observations in x but not y	0
Number of observations in y but not x	0
Number of observations with some compared variables unequal	0
Number of observations with all compared variables equal	7609
Number of values unequal	0

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	19
Number of variables compared	19
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	11
Number of variables compared with all values equal	8
Number of observations in common	7609
Number of observations in x but not y	0
Number of observations in y but not x	0
Number of observations with some compared variables unequal	7609
Number of observations with all compared variables equal	0
Number of values unequal	47351

Table: Summary of overall comparison

statistic	value
Number of by-variables	0
Number of non-by variables in common	19
Number of variables compared	19
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	11
Number of variables compared with all values equal	8
Number of observations in common	7609
Number of observations in x but not y	0
Number of observations in y but not x	0
Number of observations with some compared variables unequal	7609
Number of observations with all compared variables equal	0
Number of values unequal	55356

Figure 52-67. Images showing summary of overall comparison for model outputs

The following images show the variable differences between the two outputs for each catchment respectively;

Table: Differences detected by variable

var.x	var.y	n	NAs
PERC	PERC	595	0

Table: Differences detected by variable

var.x	var.y	n	NAs
INF	INF	825	0

Table: Differences detected by variable

var.x	var.y	n	NAs
GS	GS	6940	0

Table: Differences detected by variable

var.x	var.y	n	NAs
SS	SS	102	0

Table: Differences detected by variable

var.x	var.y	n	NAs
SW	SW	6940	0

Table: Differences detected by variable

var.x	var.y	n	NAs
ET	ET	3624	0

Table: Differences detected by variable

var.x	var.y	n	NAs
PET	PET	0	0

Table: Differences detected by variable

var.x	var.y	n	NAs
DR	DR	6686	0

Table: Differences detected by variable

var.x	var.y	n	NAs
DS	DS	4394	0

Table: Differences detected by variable

var.x	var.y	n	NAs
B	B	0	0

Table: Differences detected by variable

var.x	var.y	n	NAs
BF	BF	6940	0

Table: Differences detected by variable

var.x	var.y	n	NAs
RM	RM	6871	0

Table: Differences detected by variable

var.x	var.y	n	NAs
R	R	0	0

Table: Differences detected by variable

var.x	var.y	n	NAs
DTM	DTM	0	0
P	P	0	0
R	R	0	0
RM	RM	7609	0
BF	BF	7609	0
B	B	0	0
DS	DS	4893	0
DR	DR	4755	0
PET	PET	0	0
ET	ET	3804	0
SW	SW	7609	0
SS	SS	1199	0
GS	GS	7609	0
INF	INF	545	0
PERC	PERC	686	0
RC	RC	1033	0
T	T	0	0
H	H	0	0
WEI	WEI	0	0

Table: Differences detected by variable

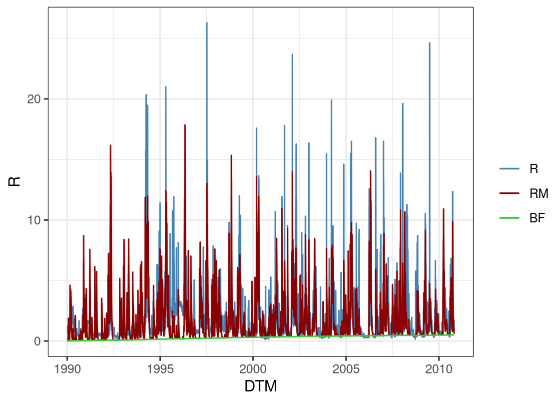
var.x	var.y	n	NAs
DTM	DTM	0	0
P	P	0	0
R	R	0	0
RM	RM	7609	0
BF	BF	7609	0
B	B	0	0
DS	DS	7560	0
DR	DR	7530	0
PET	PET	0	0
ET	ET	2670	0
SW	SW	7609	0
SS	SS	2725	0
GS	GS	7609	0
INF	INF	1120	0
PERC	PERC	1463	0
RC	RC	1852	0
T	T	0	0
H	H	0	0
WEI	WEI	0	0

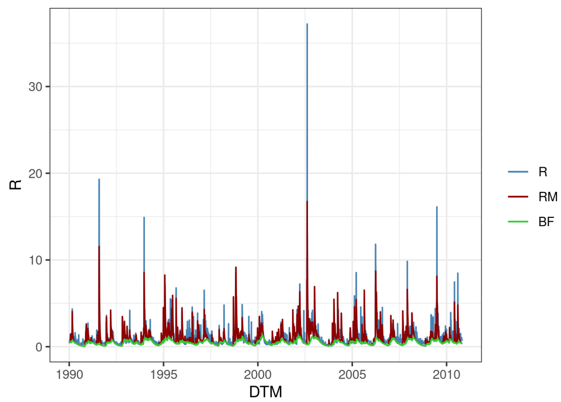
Figure 68-93. Images showing the variable differences per catchment.

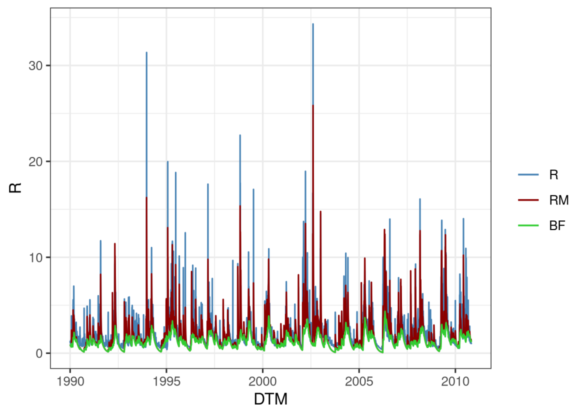
4. Results

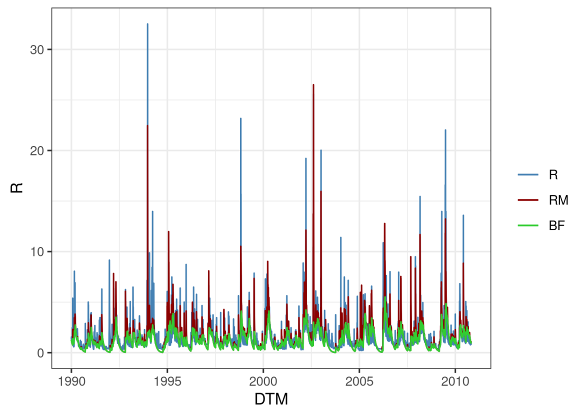
The next phase of this project will be presenting the results from the model outputs.

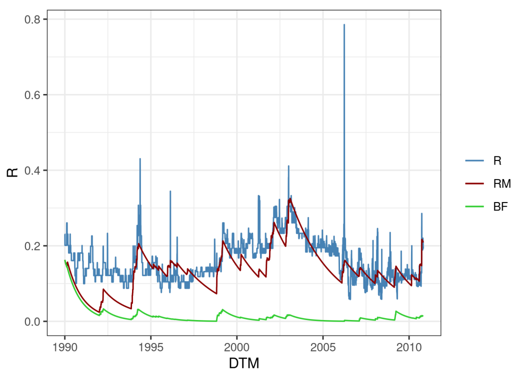
The plots below contain results from the simulated bilan model and represent the final part of this modelling process and the first set of plots will be the result from the model output before the assimilation satellite data into the model.

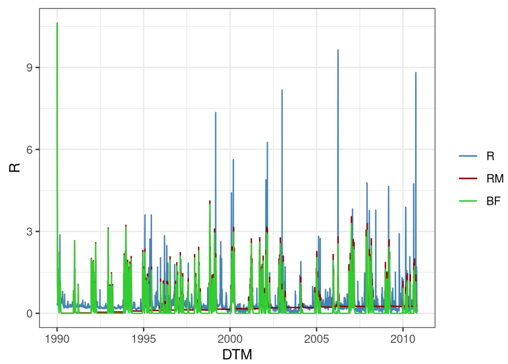


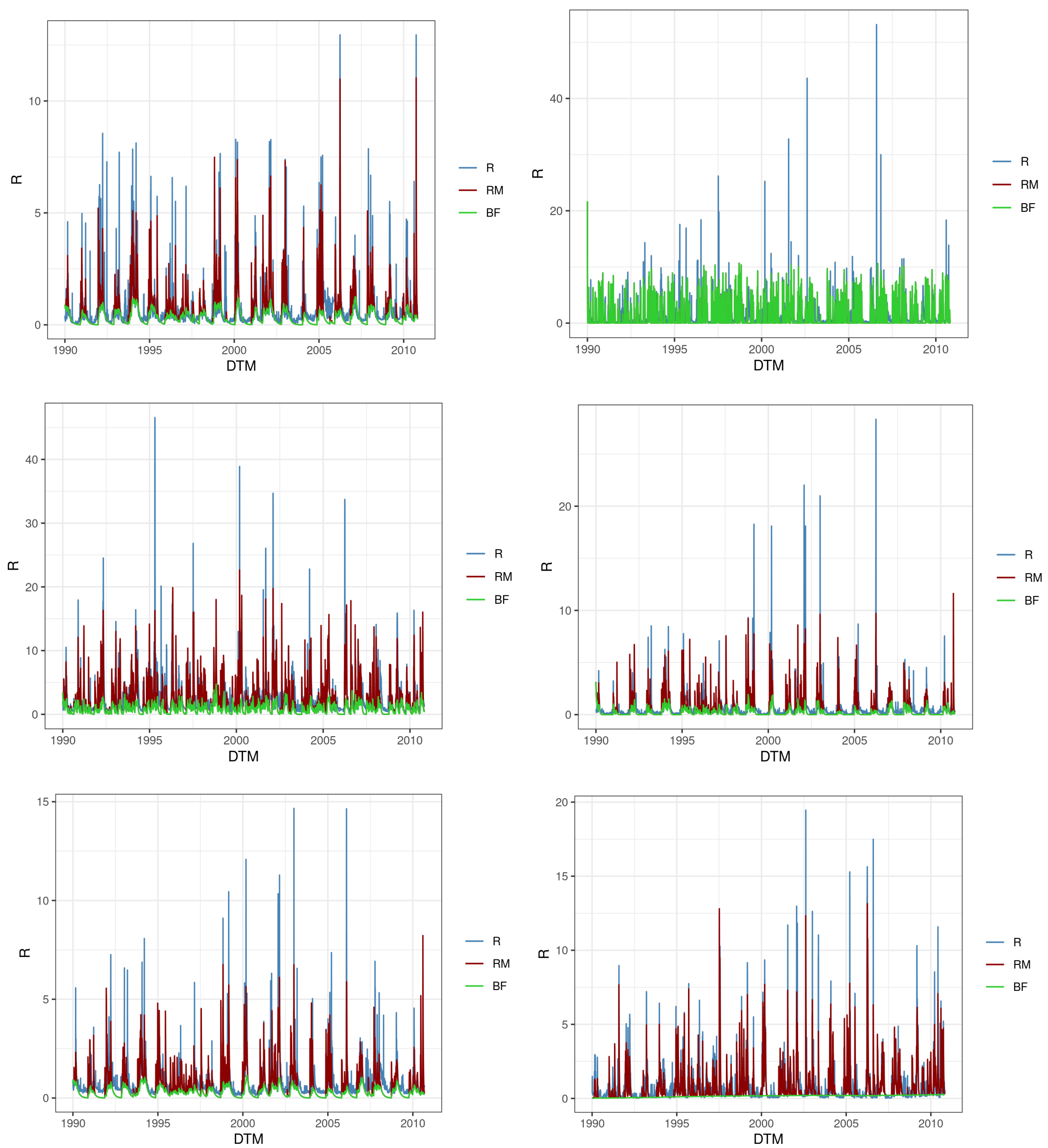











R
RM
BF
DTM
R
1990
1995
2000
2005
2010

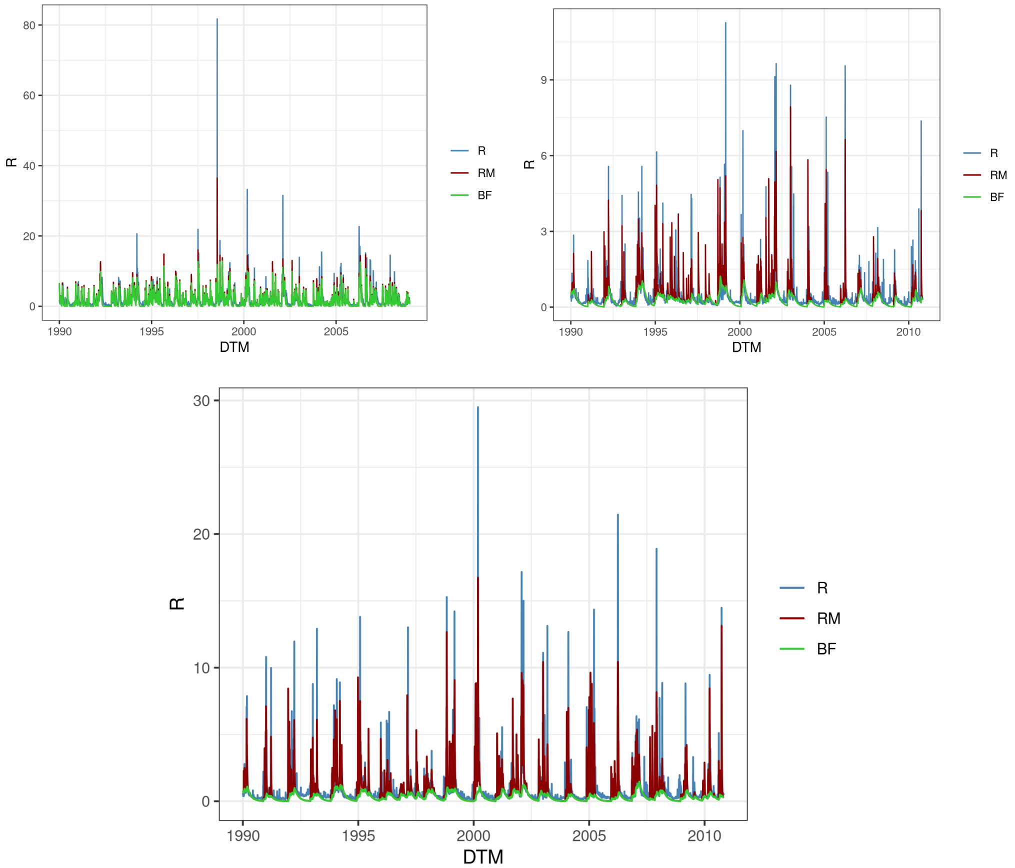
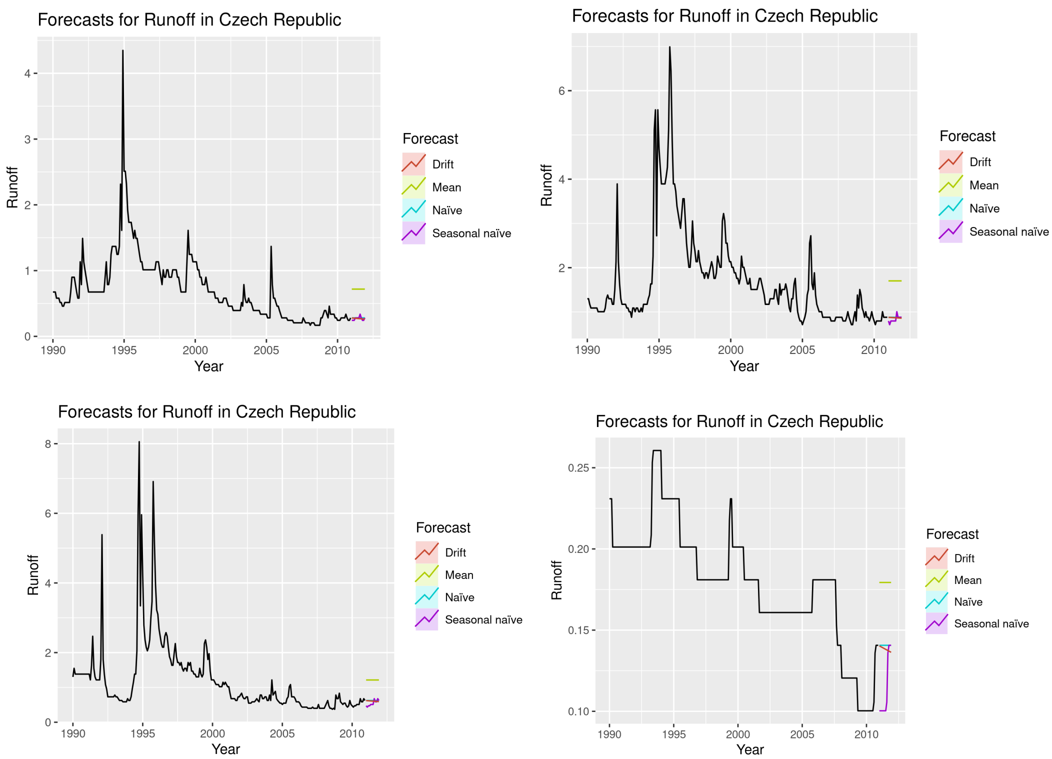
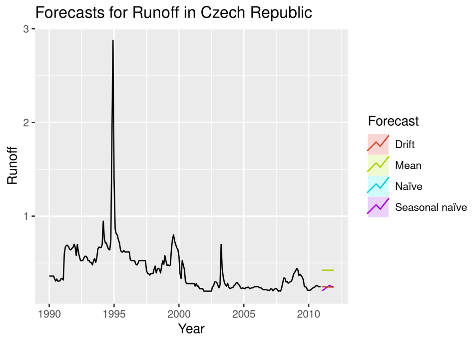
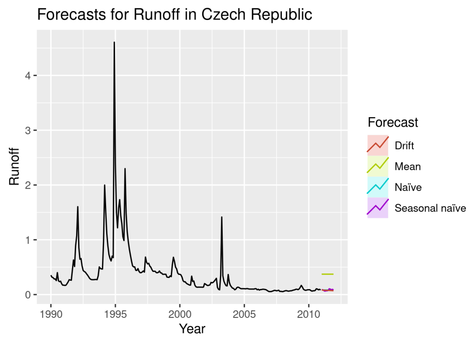


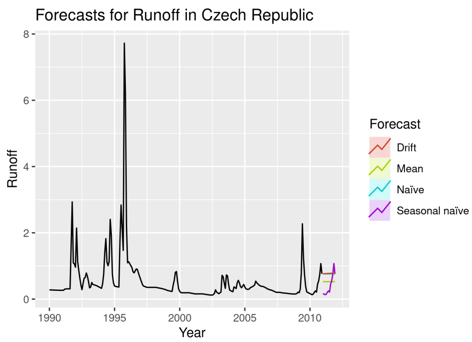
Figure 94-109. Plots for R,RM and BF before satellite data assimilation

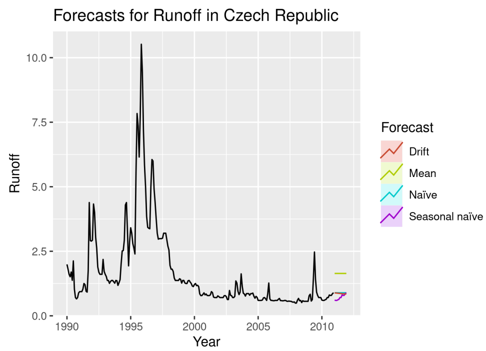
Subsequently we also have these from the first phase of the model calibration;

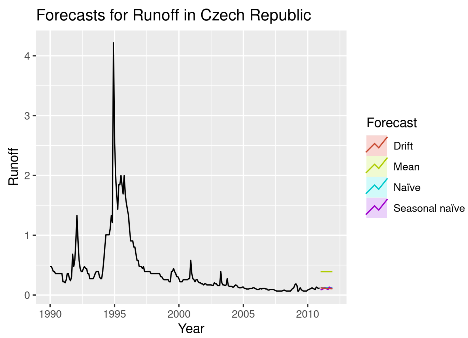


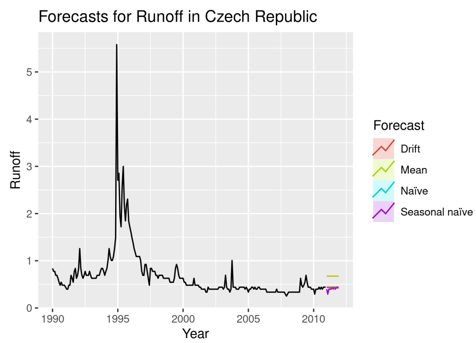

Forecasts for Runoff in Czech Republic
Runoff
Year
Forecast
Drift
Mean
Naïve
Seasonal naïve

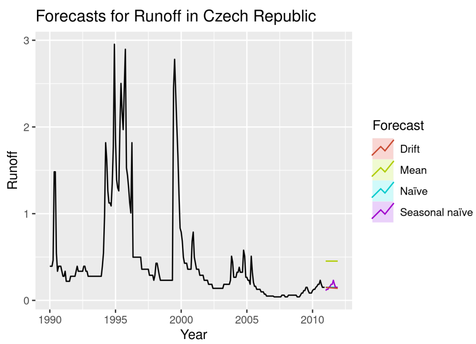

Forecasts for Runoff in Czech Republic
Runoff
Year
Forecast
Drift
Mean
Naïve
Seasonal naïve

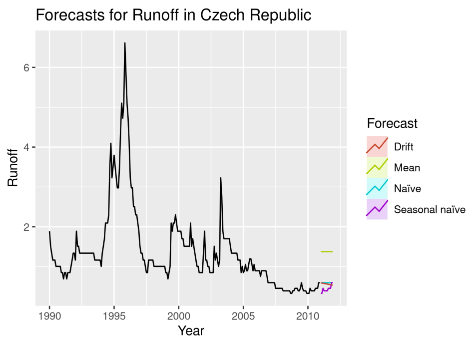

Forecasts for Runoff in Czech Republic
Runoff
Year
Forecast
Drift
Mean
Naïve
Seasonal naïve


Forecasts for Runoff in Czech Republic
Runoff
Year
Forecast
Drift
Mean
Naïve
Seasonal naïve


Forecasts for Runoff in Czech Republic
Runoff
Year
Forecast
Drift
Mean
Naïve
Seasonal naïve


Forecasts for Runoff in Czech Republic
Runoff
Year
Forecast
Drift
Mean
Naïve
Seasonal naïve


Forecasts for Runoff in Czech Republic
Runoff
Year
Forecast
Drift
Mean
Naïve
Seasonal naïve


Forecasts for Runoff in Czech Republic
Runoff
Year
Forecast
Drift
Mean
Naïve
Seasonal naïve

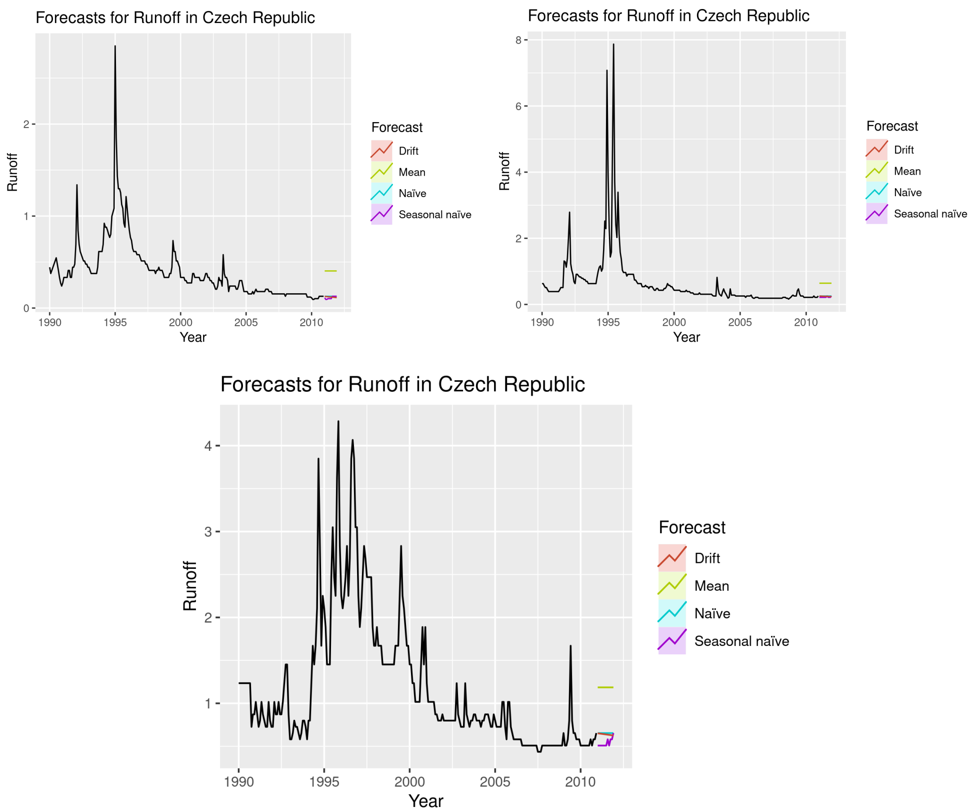
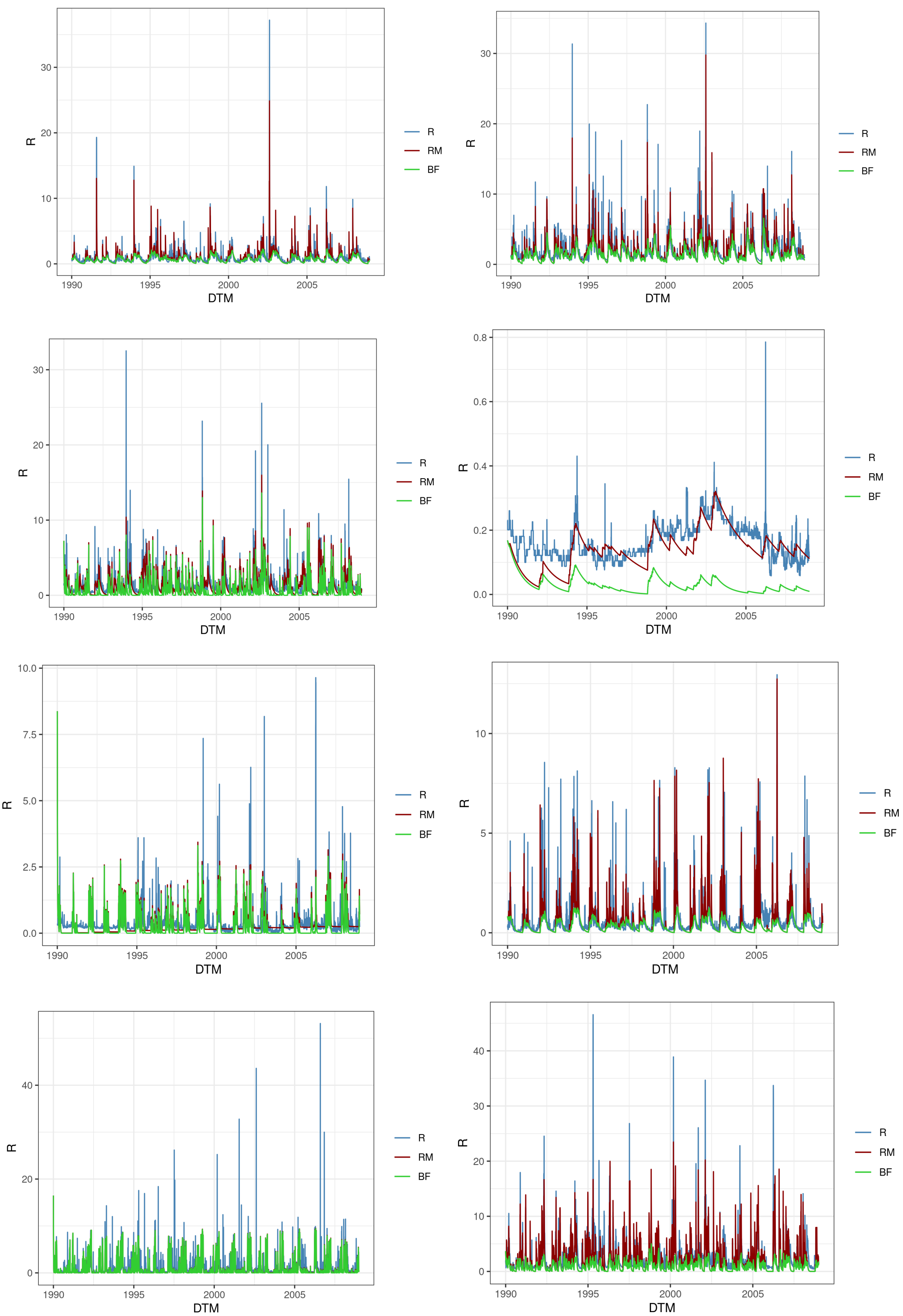


Figure 110-125. Plots for the forecasts of runoff time series from model outputs

Now we will see the results derived from the model when we assimilate satellite data using the various calibration methods already elaborated on in previous chapters.

First, like the above first set of plots for this chapter, we have the visualization of R,RM and BF from the simulated outputs with the assimilation of NDVI satellite data.


R
RM
BF
R
DTM
1990
1995
2000
2005

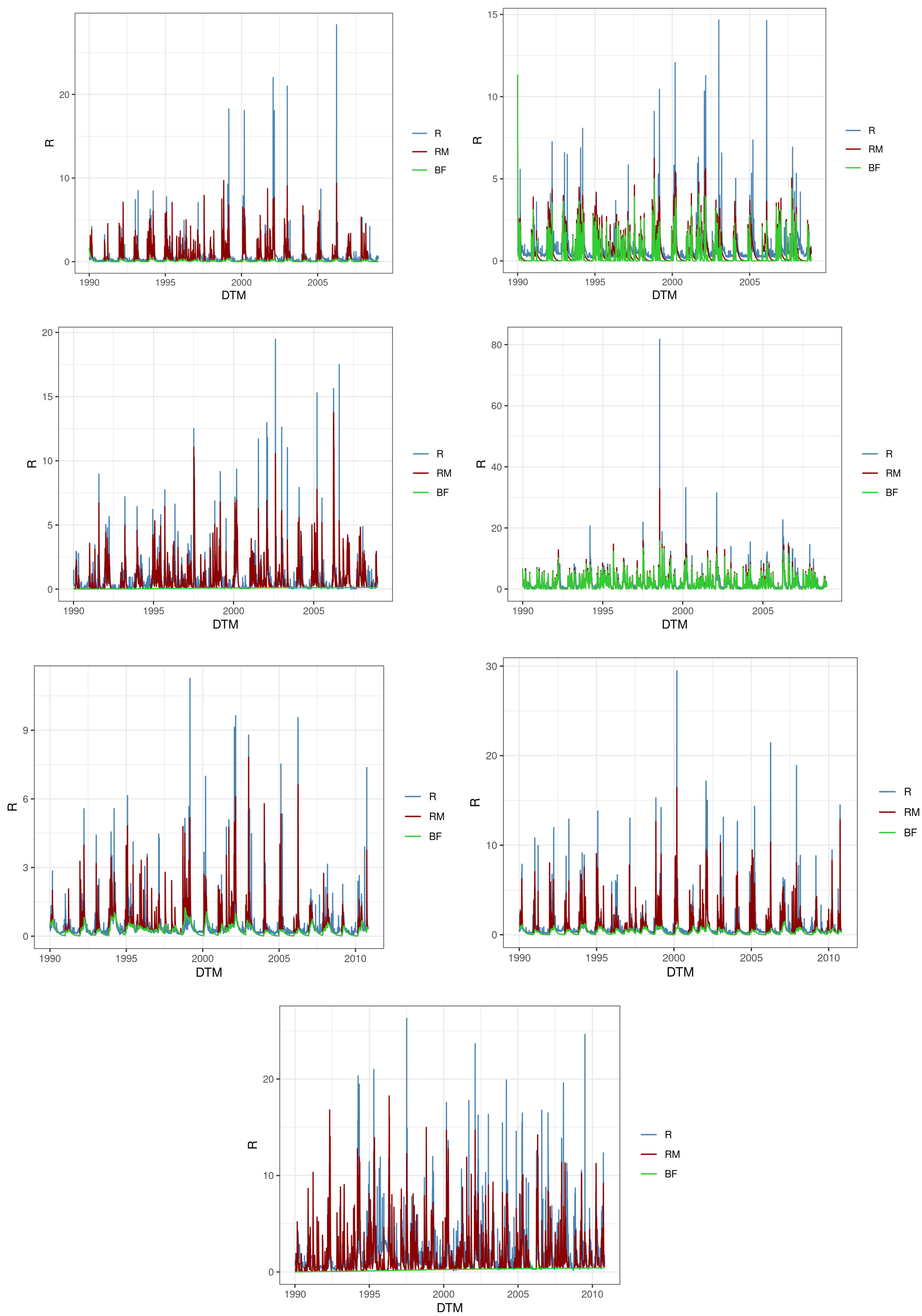


Figure 126-141. Plots showing R, RM and BF after assimilation of NDVI data

The tables below show the difference in MAE values for selected variables of each catchment:

ET	P	R	RM
0.004845099406908	0	0	0.161696940627811

ET	P	R	RM
0.003912992360949	0	0	0.021559824339986

ET	P	R	RM
0.006601552985938	0	0	0.011800125085206

ET	P	R	RM
0.001835340186319	0	0	0.034141145037617

ET	P	R	RM
0.004399260579902	0	0	0.124992941462793

ET	P	R	RM
0.017427021898062	0	0	0.217673179516507

ET	P	R	RM
0.019402188584842	0	0	0.092291906260266

ET	P	R	RM
0.000204507046693	0	0	0.03896668322299

ET	P	R	RM
0.095865121937983	0	0	1.3059834139141

ET	P	R	RM
0.000244932514022	0	0	0.060703223085872

ET	P	R	RM
0.001156004453503	0	0	0.045497834379377

ET	P	R	RM
0.015324063954339	0	0	0.007223888388801

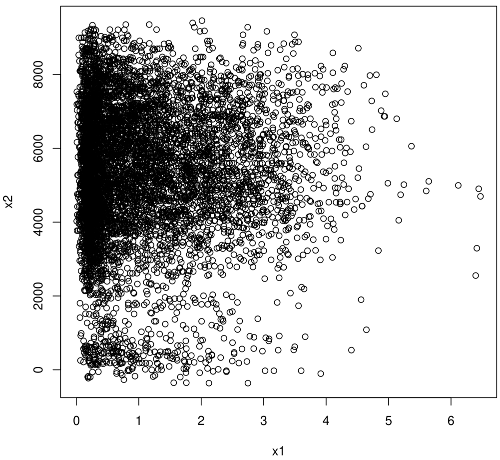
ET	P	R	RM
0.057524239020634	0	0	0.558208483590741

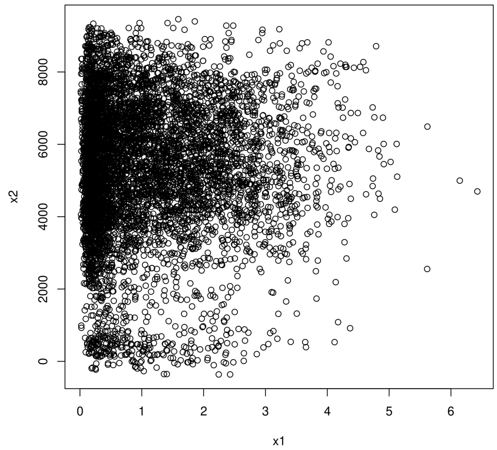
ET	P	R	RM
0.050252176099331	0	0	0.195527288167834

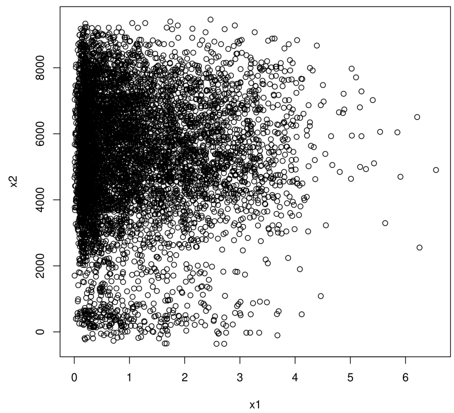
ET	P	R	RM
0.014497454149542	0	0	0.170230142000295

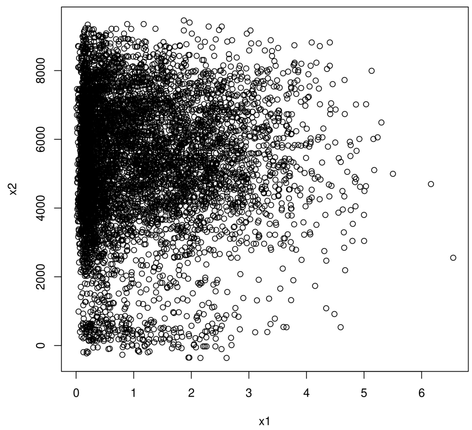
Table 1-15. Tables showing the difference in MAE (Mean Absolute Error) values between both model outputs.

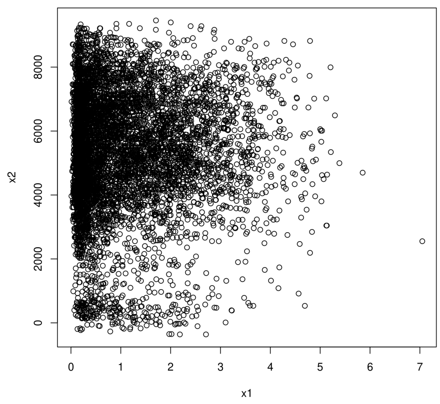
Next we have the mutual information results from the model before and after the calibration of the model for the part of the project where we have the assimilation of the NDVI data. First we have the plots before the calibration.

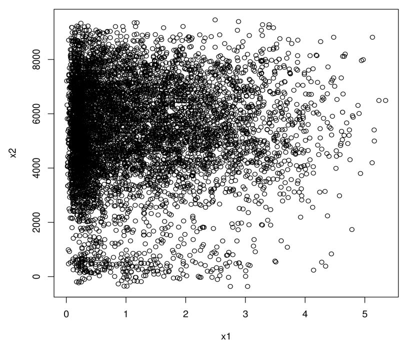

x1
x2


x1
x2


x1
x2


x1
x2


x1
x2


x1
x2

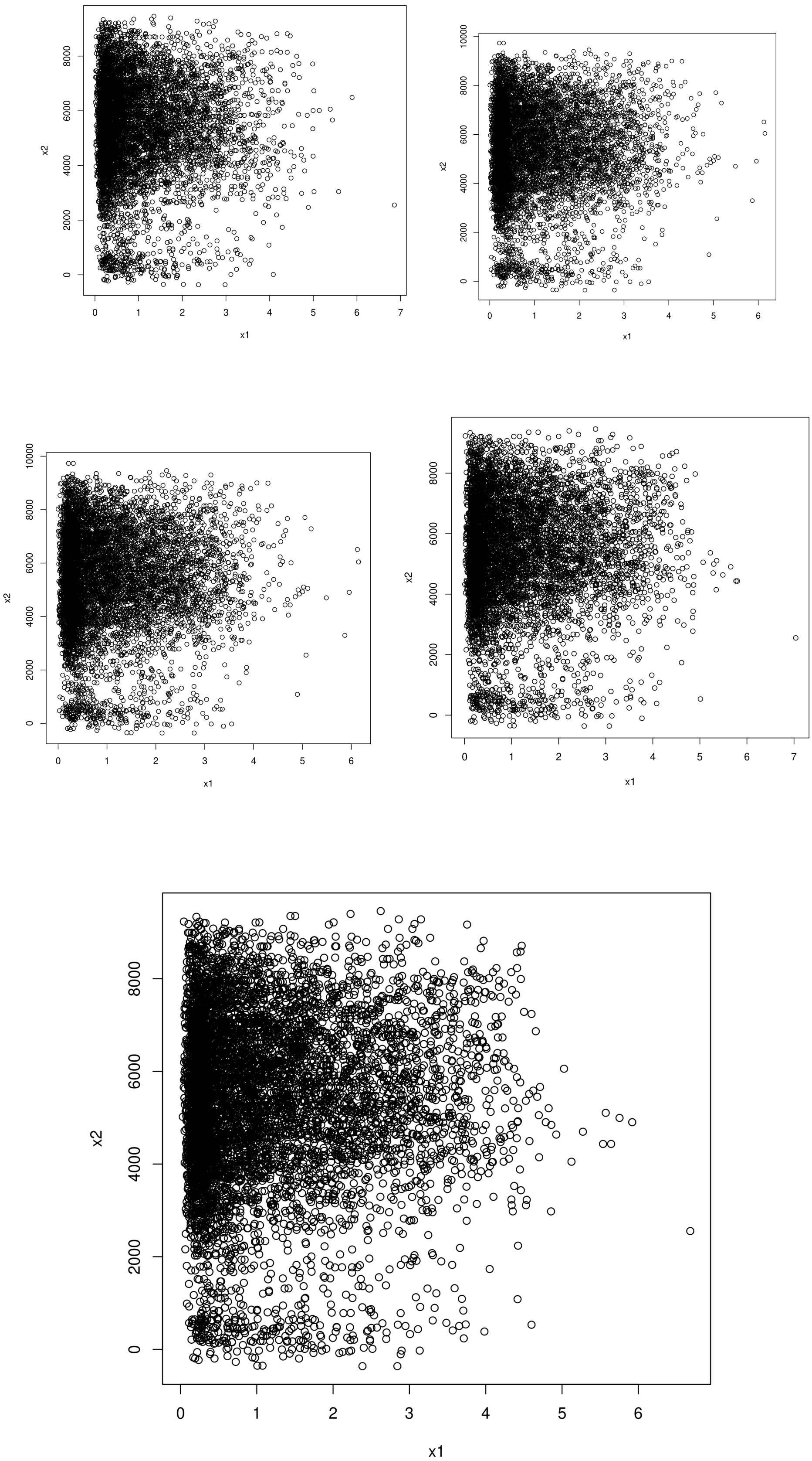
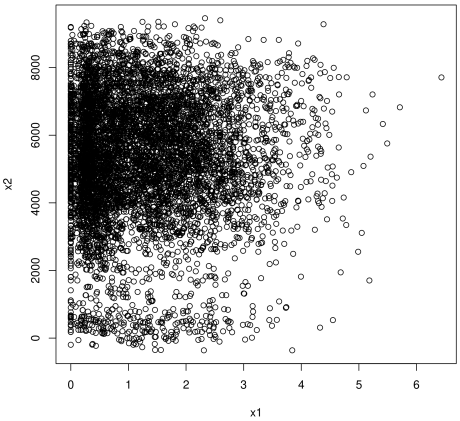
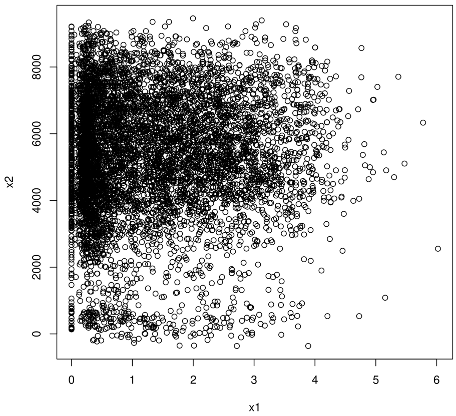
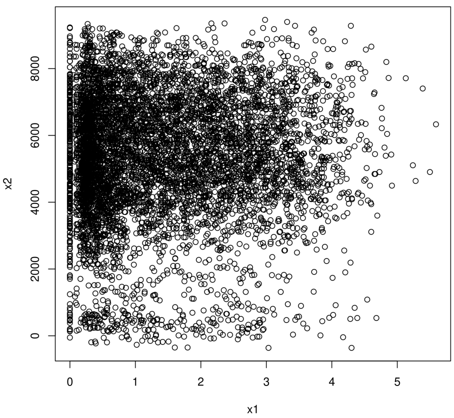


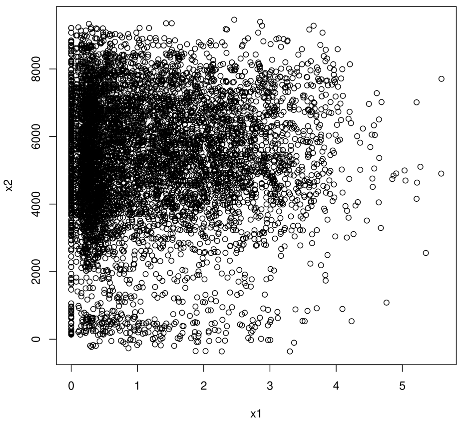
Figure 142-156. Mutual Information of observed random variables before simulated data.

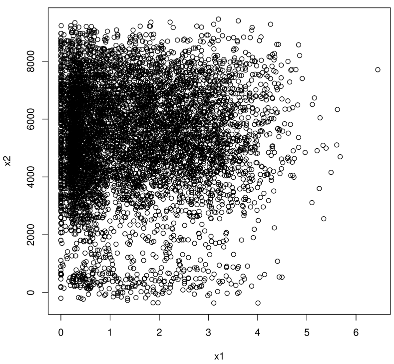
The next plots will show the Mutual information of the observed random variables after the simulation of the assimilated satellite data model.
It is important to note that the observed random variables we are working with to calculate the mutual information is the actual evapotranspiration (ETa) and the NDVI values from Modis.

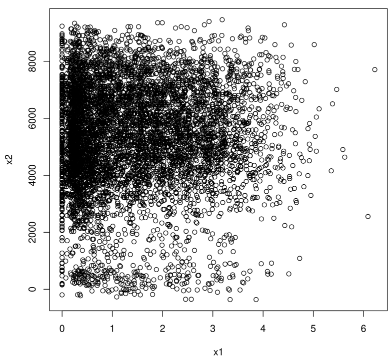


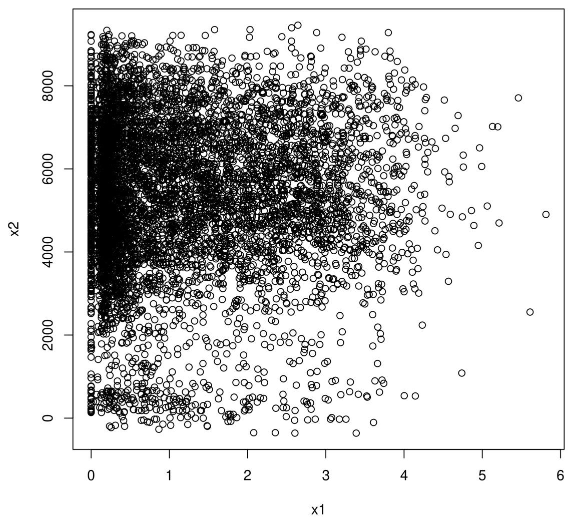


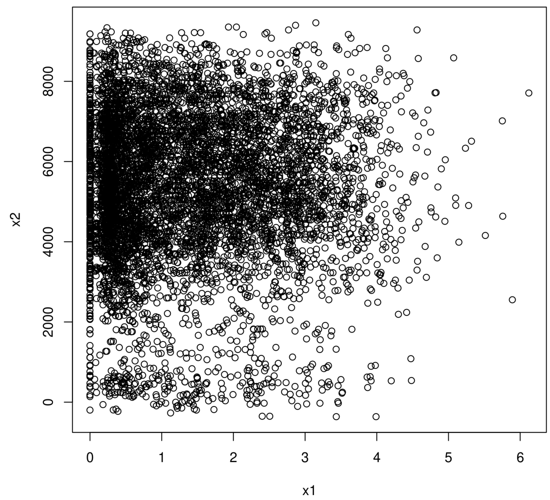


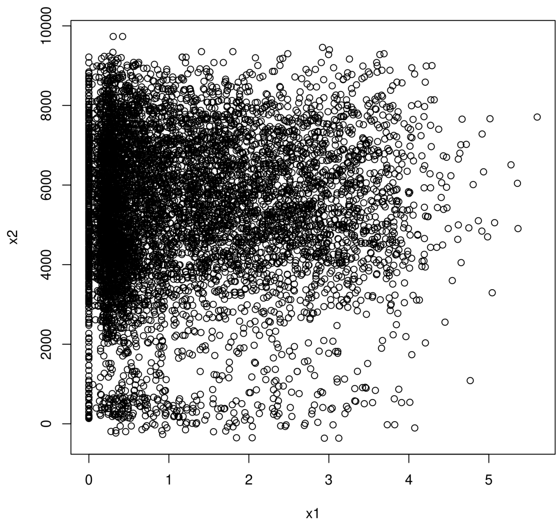


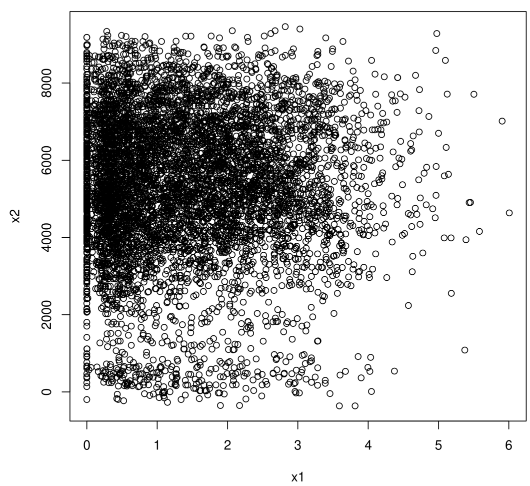





x2
x1


x2
x1


x2
x1


x2
x1

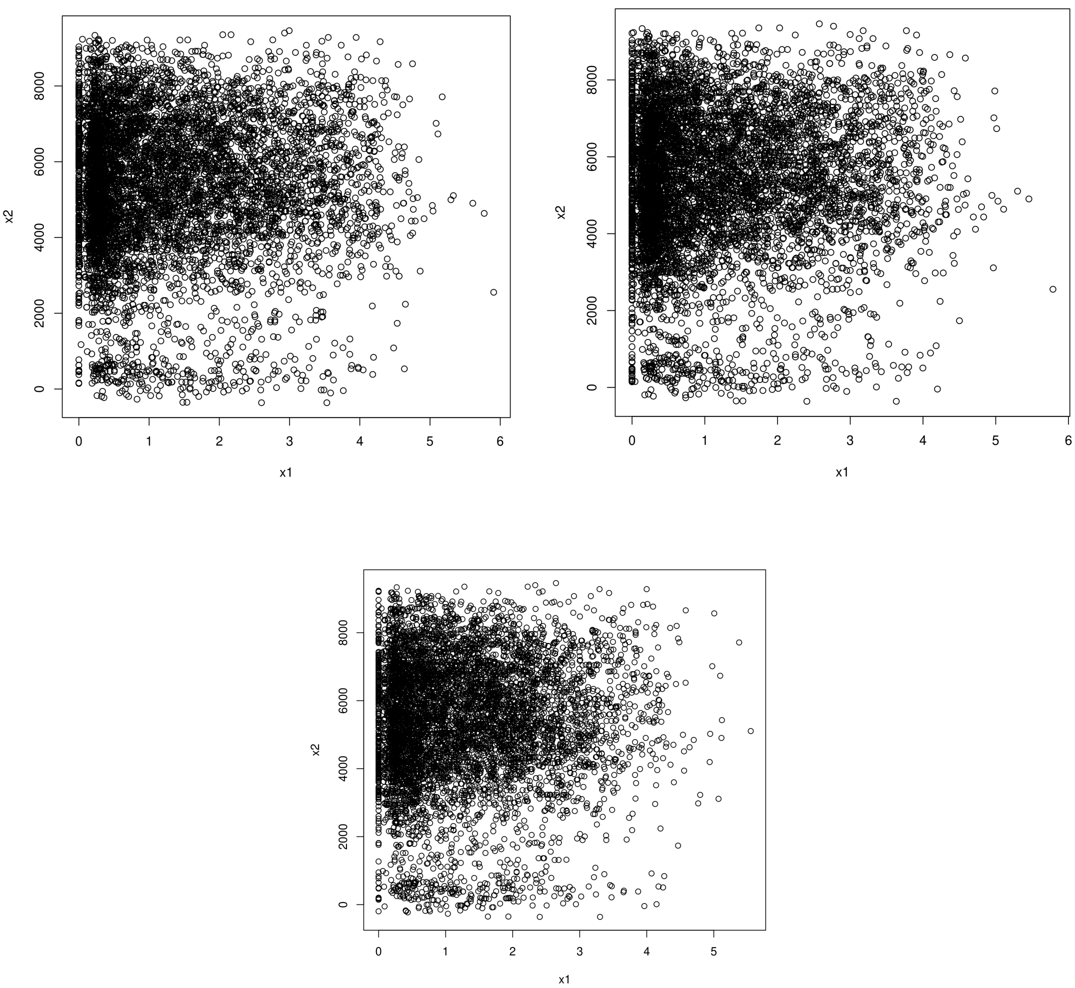


Figure 157-182. Mutual Information visualised after model calibration

5. Discussions

This project has had precedence with some studies done on this major as a way to gradually bring the world of remote sensing and modelling together. I would say they belong in the same field for they both on a foundational level do the same thing.

According to (Kouiti Hasegawa et al., 2010), at a time when NDVI was mainly used for Vegetation indices there was the advent of new indices that proved more superior to NDVI. We have come a long way in our development and we not only are introduced to new indices but also have remote sensing and hydrological modelling finally mentioned in the same field.

Another thing worthy of note is the addition of Mutual Information to modelling practices and how it will come to change our perspective when dealing with data. We see methods like the studies by (M et al., 2021) where they introduced the method based on fuzzy joint mutual information coming to light more frequently. There's also the MIM (Mutual

Information Maximization) according to a study by (Lewis, 1992) even though the study was done more than 20 years ago not many people readily applied these techniques until now. Its principle is basically to select the best relevant feature subset based on the feature-class relation. Although the computational efficiency of MIM is high, it may return features with no additional information because it does not consider feature-feature relation (M et al., 2021). In this project we used Mutual information principles and entropy to visualise our data as a way of testing the validity of the output. Similarly , mutual information has become a tool used for test software application even beyond the scope of environmental modelling. According to (Ibias et al., 2021), their study focuses on reducing the number test suites used in testing software applications because as we know the process of software development is a tedious process that goes through so many rigorous phases before it is certified ok to be released. Such processes would greatly benefit from a method whereby they can achieve maximum efficiency with half the effort. The study by (Ibias et al., 2021), hopes to shine the light on this idea so as to motivate researchers into looking at testing in software development from a different perspective. In this study we see (Miura et al., 2020), merging the concept of entropy and what it means across different areas of concentration in physics. From fluid dynamics, to newtonian particle theory to statistical mechanics and then back to information theory. From their point of view we see that entropy is both a macro and micro agent constantly on the move in newtonian particles at both states. It can be defined from both perspectives hence its broad range of application (Miura et al., 2020).

Another interesting study we see is the study by (Allahverdyan et al., 2021), where they postulate an interesting concept of MAXENT or maximum entropy whereby the method addresses the problem of recovering unknown probabilities of a random variable via the maximization of entropy (Allahverdyan et al., 2021). Nominally when entropy is said to be maximum we say there is no information to be gained but the study sheds new light on concepts which probably aren’t new but not just open to a lot of researchers.

6. Conclusions And Contributions

I would conclude this research work by making a hypothesis based on the results I got that, for the assimilation of satellite data; the correlation between the values of the chosen satellite products and the any arbitrary variable from the input to the model might be negative or non-linear but using the autocorrelation function as part of the performance criteria for the objective function used to calibrate the model will yield positive results.

From my results, the difference of the MAE between the results gotten from the model without the assimilated satellite data and the one with the assimilated data, we see that the difference for each catchment did not exceed 0.1. This tells us a lot about the accuracy of the results and also supports the assimilation of chosen satellite data into the model. It is also important to note the Mutual information and entropy results show a difference in correlation between the two entropies of the observed random variables. Noting that the value for entropy was a little, we can conclude that the dataset gotten from the modelling exercise had a high amount of purity and less uncertainty.

This research work would not have been made possible without relentless support and guidance of my supervisor and I appreciate his efforts. I enjoyed the sessions, they were eye opening and for my aspirations in this field of research I would be obliged to further the topic on entropy and mutual information. I see myself eventually setting my sights on machine learning because the world is more open to data than it was some years ago and it will be even more advanced in years to come.

7. Bibliography And Sources

Bibliography

Allahverdyan, A. E., Khalafyan, E. A., & Martirosyan, N. H. (2021). Validity limits of the maximum entropy method. *Chinese Journal of Physics*, *71*, 95-111. ISSN 0577-9073. https://doi.org/10.1016/j.cjph.2021.02.009

Allen, Richard G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration - Guidelines for computing crop water requirements. *FAO Irrigation and drainage paper: Rome: FAO - Food; Agriculture Organization of the United Nations.*, *56*.

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). FAO Irrigation and Drainage Paper No. 56. *Food and Agriculture Organization of the United Nations, Rome*, *56*(97), e156.

Beck, H. E., McVicar, T. R., Van Dijk, A. I.J.M., Schellekens, J., De Jeu, R. A.M., & Bruijnzeel, L. A. (2011). Global evaluation of four AVHRR–NDVI data sets: Intercomparison and assessment against Landsat imagery. *Remote Sensing of Environment*, *115*(10), 2547-2563. ISSN 0034-4257. https://doi.org/10.1016/j.rse.2011.05.012

Bezdek, J. C. (1993). Fuzzy models—What are they, and why? [Editorial]. *EEE Transactions on Fuzzy Systems*, *1*(1), 1-6. 10.1109/TFUZZ.1993.6027269. 10.1109/TFUZZ.1993.6027269.

FAO United Nations. (n.d.). *Food and Agricultural Organisation of United Nations*. FAO United Nations. http://www.fao.org/home/en/

Gordon, L. J., Peterson, G. D., & Bennett, E. M. (2008). Agricultural modifications of hydrological flows create ecological surprises. *Trends Ecol. Evol*, *23*(4), 211-219. http://dx.doi.org/10.1016/j.tree.2007.11.011

Grutzner, R. (1996). Environmental modeling and simulation - applications and future requirements. *Environmental Software Systems*, --(--), 114. --. 10.1007/978-0-387-34951-0_9

Hauner, Horsley, R., Osborne, M., Auty, Robert, Carter, William, F., Zeman, B, Z. A., & Blazek. (2021). *Miroslav. "Czech Republic".* Britannica. Retrieved 2021, from https://www.britannica.com/place/Czech-Republic

Ibias, A., Núñez, M., & Hierons, R. M. (2021). Using mutual information to test from Finite State Machines: Test suite selection. *Information and Software Technology*, *132*, 106498. ISSN 0950-5849. https://doi.org/10.1016/j.infsof.2020.106498

Izquierdo, J., Pérez, R., & Iglesias, P. L. (2004). Mathematical models and methods in the water industry. *Mathematical and Computer Modelling*, *39*(11-12), 1353-1374. ISSN 0895-7177. https://doi.org/10.1016/j.mcm.2004.06.012

Jafrasteh, B., & Suárez, A. (2021). Objective functions from Bayesian optimization to locate additional drillholes. *Computers & Geosciences*, *147*, 104674. ISSN 0098-3004. https://doi.org/10.1016/j.cageo.2020.104674

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*(5), 602-614. ISSN 1364-8152. https://doi.org/10.1016/j.envsoft.2006.01.004

Kibria, S., Masia, S., Sušnik, J., & Hessels, T. M. (2021). Critical comparison of actual evapotranspiration estimates using ground based, remotely sensed, and simulated data in the USA. *Agricultural Water Management*, *248*. ISSN 0378-3774. https://doi.org/10.1016/j.agwat.2021.106753

Kouiti Hasegawa, Matsuyama, H., Tsuzuki, H., & Sweda, T. (2010). Improving the estimation of leaf area index by using remotely sensed NDVI with BRDF signatures. *Remote Sensing of environment*, *114*. http://dx.doi.org/10.1016/j.rse.2009.10.005

Lewis, D. D. (1992). Feature selection and feature extraction for text categorization. *Proceedings of the Workshop on Speech and Natural Language. Association for Computational Linguistics*, 212-217.

Li, W., Liang, Z., Neuman, J., Chen, J., & Cui, X. (2021). Multi-generator GAN learning disconnected manifolds with mutual information. *Knowledge-Based Systems*, *212*, 106513. ISSN 0950-7051. https://doi.org/10.1016/j.knosys.2020.106513

Liu, D. (2020). A rational performance criterion for hydrological model. *Journal of Hydrology*, *590*. https://doi.org/10.1016/j.jhydrol.2020.125488

M, O. A., Salem, Liu, F., Chen, Y.-P. P., & Chen, X. (2021). Feature selection and threshold method based on fuzzy joint mutual information. *International Journal of Approximate Reasoning*, *132*, 107-126. ISSN 0888-613X. https://doi.org/10.1016/j.ijar.2021.01.003

Miura, A., Tomoeda, A., & Nishinari, K. (2020). Formularization of entropy and anticipation of metastable states using mutual information in one-dimensional traffic flow. *Physica A: Statistical Mechanics and its Applications*, *560*, 125152. ISSN 0378-4371. https://doi.org/10.1016/j.physa.2020.125152

NASA/NOAA NPP Project. (n.d.). *VIIRS/NPP Vegetation Indices 16-Day L3 Global 500m SIN Grid V001*. VNP13A1. https://cmr.earthdata.nasa.gov/search/concepts/C1392010616-LPDAAC_ECS.html

NC State University. (n.d.). *Evapotranspiration*. climate.ncsu.edu/edu/Evap. https://climate.ncsu.edu/edu/Evap

Ojha, C., Berndtsson, R., & Bhunya, P. K. (2008). *Engineering hydrology*. Oxford University Press. 00

Oudin, Ludovic, Moulin, L., Bendjoudi, H., & Ribstein, P. (2010). Estimating potential evapotranspiration without continuous daily data: possible errors and impact on water balance simulations. *Hydrological Sciences Journal*, *55*, 209. 10.1080 / 02626660903546118.

Pohl, C., & Genderen, J. V. (1998). Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications. *International Journal of Remote Sensing*, *19*(00), 823-854. --. 10.1080/014311698215748

Rey, S. J. (2015). Mathematical Models in Geography (J. D. Wright, Ed.). *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, *00*(00), 785-790. ISBN 9780080970875. https://doi.org/10.1016/B978-0-08-097086-8.72033-2

Rockström, J., Falkenmark, M., Karlberg, L., Hoff, H., Rost, S., & Gerten, D. (2009). Future water availability for global food production: the potential of green water for increasing resilience to global change. *Water resources. res*, *45*(7).

Singh, V. P. (2013). *Entropy theory and its application in environmental and water engineering*. John Wiley & Sons.

Teuling, A. J. (2018). A Forest Evapotranspiration Paradox Investigated Using Lysimeter Data. *Vadose Zone Journal*, *17*(1). https://doi.org/10.2136/vzj2017.01.0031

T. G. Masaryk Water Management Research Institute, vvi, The Elbe River Basin, sp, Vltava River Basin, sp, & The Ohře River Basin, sp. (2015, September 18). *Sustainable use of water resources in conditions of climate change*. http://bilan.vuv.cz/. http://bilan.vuv.cz/

Wang, S., Pan, M., Mu, Q., Shu, X., Mao, J., Brümmer, C., & Black, T. A. (2015). Comparing evapotranspiration from eddy covariance measurements, water budgets, remote sensing, and land surface models over Canada. *Journal of Hydrometeorol*, *16*(4), 1540-1560. https://doi.org/10.1175/jhm-d-14-0189.1

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *CLIMATE RESEARCH*, *30*, 70-82.

Younger, S. E., Jackson, C. R., & Rasmussen, T. C. (2020). Relationships among forest type, watershed characteristics, and watershedET in rural basins of the Southeastern US. *Journal of Hydrology*, *591*. https://doi.org/10.1016/j.jhydrol.2020.125316

8. Appendix