Optimization and Evaluation of the Vegetation Photosynthesis Respiration Model (VPRM) parameters for coniferous forests in Northern Europe

Master Thesis

Author: Martha Domhoefer

NIU: 1586407

Director: Dr. Gara Villalba

Supervisor: Ricard Segura Barrero

30 August 2022

Abstract

Due to the anthropogenic climate change, multiple organisations are pursuing different strategies to limit the CO_2 emissions, especially in urban areas. By limiting the uncertainties in biogenic carbon fluxes, one can suppose an improvement in the estimation of anthropogenic fluxes in top down atmospheric CO_2 inversion processes. The Vegetation Photosynthesis and Respiration Model (VPRM) is a satellite-based assimilation scheme that estimates hourly values of the net carbon exchange between the ecosystem and the atmosphere, also known as the Net Ecosystem Exchange (NEE), from a local to continental scale. The VPRM combines meteorological and satellite-sensed data to derive the two components of the ecosystem fluxes. The first component is the light dependent part, the Gross Ecosystem exchange (GEE) and refers to the total amount of carbon fixed in the process of photosynthesis by plants in the ecosystem. The second component, the light independent part, is the Ecosystem Respiration (RECO) consisting of autotrophic (plant) respiration and heterotrophic (soil) respiration. For this study, Northern Europe is used as the study region, especially boreal coniferous forests, which represent almost 60% of the land in our study domain and it is one of the largest land biomes in the world. Observational data from 6 eddy covariance flux towers covering a period from 2017-2020 are used to calibrate the model. The vegetation indices (EVI, LSWI) for the model were extracted in a novel methodology from the Sentinel2 satellite data. Meteorological forcing of the model was obtained from in-situ meteorological measurements from the flux stations.

Two methodologies were conducted and compared for the model parameters optimization: the optimization of all the significant parameters in the model simultaneously, and the separated optimization of the parameters from the GEE and RECO. Before optimizing, a Morris sensitivity analysis was performed to detect the most critical VPRM parameters. Moreover, different optimization methods were compared: a simple linear Gradient Descent, a least squares method using the Levenberg-Marquardt algorithm and a Differential Evolution. The best optimization was with the Levenberg-Marquardt algorithm ($RMSE = 3.323 \mu mol \frac{CO_2}{m^2 * s}$) with an statistically significant difference, estimated with the Diebold-Mariano test, compared to the default parameters ($RMSE = 3.705 \mu mol \frac{CO_2}{m^2 * s}$). Overall, all optimization techniques yielded a better performance than the default parameters and can be used for a large scale implementation of the model over Northern Europe.

Acknowledgements

Contents

List of Figures and Tables

List of Figures

List of Tables

1   Introduction

One of the primary drivers of climate change is the impact of human activity on the global carbon cycle since industrialisation. Due to the excessive emission of greenhouse gases (GHG) like CO_2, more heat gets trapped inside the Earth's atmosphere when sunlight gets reflected back (IPCC, 2014). Particularly, urban and suburban areas are accounting for 70% of the fossil fuel CO_2 sources. Therefore, governments and different organisations worldwide are pursuing different ways to monitor, report and reduce GHG emission in metropolitan areas.

There are two types of approaches to estimate greenhouse gas emissions to the atmosphere, namely 'bottom-up' and 'top-down'. In 'bottom-up' approaches, the emissions are estimated based on reported consumption and production data across various activities, such as vehicle fuel consumption or energy produced. 'Top-down' approaches involve atmospheric measurements and usually include an atmospheric inversion of the CO2 concentrations, using atmospheric transport models to estimate carbon fluxes (i.e. posterior fluxes) by adjusting the fluxes (i.e. prior fluxes) to be consistent with observed CO2 concentrations (Feng et al., 2016). The uncertainties in this method are attributed to physical representation errors in the atmospheric transport model, the errors related to the observations and the uncertainties in the 'prior' emissions. Moreover, biogenic emissions estimates are necessary in 'top-down' approaches in order to constrain the anthropogenic/biogenic proportion of the fluxes (Mahadevan et al., 2008, Wu et al., 2018). There is therefore a great interest in improving the biogenic carbon estimations to time-invert the anthropogenic fluxes. With optimized a priori assumptions about anthropogenic emissions uncertainty is reduced. Additionally, biogenic emission fluxes applied at high-resolution can be used to study the GHG balance of different greening strategies at city-scale.

The Vegetation Photosynthesis and Respiration Model is a satellite-based light-use efficiency model that estimates the biogenic carbon emissions (Mahadevan et al., 2005). VPRM estimates the two components of the net ecosystem carbon exchange: the atmospheric carbon uptake by the vegetation due to the photosynthesis and the carbon release to the atmosphere from the vegetation and soil respirations. The former component is a satellite-based assimilation scheme that estimates the gross ecosystem exchange (GEE) from light-use efficiency assumptions. The other part is the loss of CO_2 from the vegetation and soil to the atmosphere (ecosystem respiration) related to air temperature. The main data sources of the model are meteorological data, satellite-derived vegetation indices and model parameters specifically obtained for each plant typology (Mahadevan et al., 2008).

VPRM parameters were optimized for different plant functional types (PFTs) by Mahadevan et al. (2008), and by Kountouris et al. (2018). The flux data used in this study stems from 6 eddy covariance flux towers covering a period from 2017-2020. The eddy covariance technique is used to measure and calculate vertical turbulent fluxes within the boundary layer of the atmosphere. The method analyses high-frequency data series of wind and atmospheric scalar values and generates the fluxes of these properties(Liang and Xi, 2012). It is a statistical method to determine emission rates from land or water surfaces and their momentum, heat, water vapour, carbon dioxide and methane fluxes (Liang and Xi, 2012).

Due to the eddy flux covariance techniques there is an increasing availability measurements over different PFTs at more than 400 sites. The data products include the net ecosystem exchange (NEE), GEE, RECO and meteorological drivers (air temperature, radiation, precipitation etc.) aggregated at different timescales (e.g. half-hourly, daily, annually) and consistent for data treatment (Migliavacca et al., 2011). Researchers analyse eddy covariance data, and then offer their data through regional network data portals. The international net-

work FLUXNET (`https://fluxnet.org/`) links these regional networks together, collecting their datasets to create processed and harmonized eddy covariance data products (Pastorello et al., 2020). The data product used was the Warm Winter 2020 (Team and Centre, 2022) dataset that extends the work from the FLUXNET 2015 (Pastorello et al., 2020).

Overall, this study is integrated in the by the Integrated System Analysis of Urban Vegetation and Agriculture (URBAG) project, which is an ERC funded project aimed to determine to what degree green infrastructures can be a source of sustainable food, reduce environmental impacts, and promote a more efficient use of resources in urban regions. This project is developing a new approach in which land and resource use are used to drive both life cycle analysis and atmospheric modelling. The project has two case study cities and their metropolitan regions: The Metropolitan Area of Barcelona and the Greater Oslo Region. In order to study the urban carbon balance in the Greater Oslo Region, the biogenic fluxes in the natural areas within and around the city must be optimized to represent better the daily and seasonal cycles.

According to Korhonen and Stahl (2020), 227 million ha of forests in Europe cover 35% of the total land area. 46% of European forests are predominantly coniferous, 37% are predominantly broadleaved, and the rest are mixed. Currently, in Finland forests cover 76.2% of the land area, in Sweden 74.5% and in Norway 47.1%. The most dominant plant functional type in Northern Europe are evergreen needle forests (De Rigo et al., 2016; FAO and UNEP, 2020). This is the PFT, in which most eddy flux stations of our region are situated. Therefore, this study focuses on evergreen needle forests. The main objective of this study is to identify and optimize the most important parameters in the VPRM model for the coniferous forest biomes in Northern Europe. The parameters are optimized using three different techniques:

- Gradient Descent (GD): The method is used in numerics to solve general optimisation problems. In this process, a path gets proceeded from a starting point along a descent direction until convergence or a certain number of steps was achieved.

- Levenberg-Marquardt (LM): The method is a combination of the Gradient Descent method and the Gauss-Newton principle.

- Differential Evolution (DE): The method does not require a differentiable function, like the two other methods, as it does not employ the gradient of the function to be optimized. In theory, the algorithm is used in evolutionary computation and seeks to iteratively enhance a candidate solution with respect to a specified quality metric.

All three algorithms find a local minimum.

2 Methodology

2.1 Description VPRM

The Vegetation Photosynthesis and Respiration Model (VPRM), is a satellite-based assimilation scheme that estimates hourly values of the Net Ecosystem Exchange (NEE) of CO_2. It consists of a light-dependent part, the Gross Ecosystem Exchange (GEE) and light-independent part, the Ecosystem Respiration (RECO).

The GEE refers to the total amount of carbon fixed in the process of photosynthesis by plants in an ecosystem. The RECO refers to the combination of autotrophic respiration i.e. the fixed carbon, that is lost by internal plant metabolism and the heterotrophic respiration i.e. the carbon, that is lost by organisms in ecosystems other than plants (Kirschbaum et al., 2001). As the uptake of CO_2 by plants is a negative flux (removal from the atmosphere), a minus sign gets put in front of GEE.

$$NEE = -GEE + RECO \tag{1}$$

The VPRM model consists of seven parameters that change for different plant functional types (PFT). An incorrect estimation of these parameters is leading to large uncertainties and systematic errors (Lin et al., 2011). By default, the VPRM uses 8 plant functional types and estimates the biogenic fluxes for a land region, integrating the fluxes derived for each plant functional type weighted by the vegetation cover fraction of each one.

The RECO is represented by:

$$RECO = \alpha * T + \beta \tag{2}$$

where α is the temperature sensitivity of respiration (Gourdji et al., 2022) and β can be interpreted biogeochemically as the flux-weighted mean size of the respiring pools of organic matter in the ecosystem. In Equation 2, T is the air temperature. T gets set to T_{low} if it decreases below T_{low}, representing the persistence of the soil respiration in winter, when air temperatures are cold but soil remains warm. The GEE is represented by:

$$GEE = \lambda * T_{scale} * P_{scale} * W_{scale} * EVI * \frac{1}{(1 + Rad/Rad_0)} * Rad \tag{3}$$

where λ is the maximum light use efficiency factor. The Enhanced vegetation index (EVI) (see Appendix A, Equation (21)) is used to estimate the fraction absorbed of photosynthetically active radiation. Also, the Photosynthetically Active Radiation (PAR in $\mu mol m^{-2} s^{-1}$) is needed. As PAR is related to the shortwave radiation (W/m^2) with $PAR \approx 0.505 Rad$, Rad gets used in the Equation 3, as it can be more easily measured. Rad_0 is the half-saturation value of the Photosynthetically Active Radiation for a plant functional type. Moreover, T_{scale} represents the temperature sensitivity of photosynthesis, using the Equation developed for the Terrestrial Ecosystem Model (Raich et al., 1991).

$$T_{scale} = \frac{(T - T_{min}) * (T - T_{max})}{(T - T_{min}) * (T - T_{max}) - (T - T_{opt})^2} \tag{4}$$

where T_{min} is a minimum, T_{max} a maximum and T_{opt} the optimal temperature in ºC for photosynthesis. T_{scale} gets set to zero, if the temperature falls below T_{min} or above T_{max}. (Xiao et al., 2004). P_{scale} accounts for effects of leaf age on the canopy photosynthesis. For evergreen classes, so the main vegetation class in this study, P_{scale} is assumed to be 1 for the whole year. Otherwise, the EVI (Enhanced Vegetation Index) and LSWI (Land Surface Water Index) (see Appendix A, Equation (22)), are used to determine the plant growing phases. The first phase is from bud burst till a full canopy, the second phase is the time of

full canopy, and the third phase is the phase of leaf senescence. During the second phase P_{scale} is set to one (full canopy). In the other phases of not evergreen plant functional types, P_{scale} is calculated as:

$$P_{scale} = \frac{1 + LSWI}{2} \tag{5}$$

In the default VPRM model the LSWI is obtained from the MODIS satellite derived surface reflections. W_{scale} expresses the effect of water stress in GEE.

$$W_{scale} = \frac{1 + LSWI}{1 + LSWI_{max}} \tag{6}$$

where $LSWI_{max}$ is the maximum LSWI within the plant growing season for each calibration site. The goal of this study is to optimize the key VPRM parameters λ, Rad_0 of the GEE function (see Equation 3) and α, β of the RECO function (see Equation 2). In Figure 1, the structure of the VPRM can be observed.
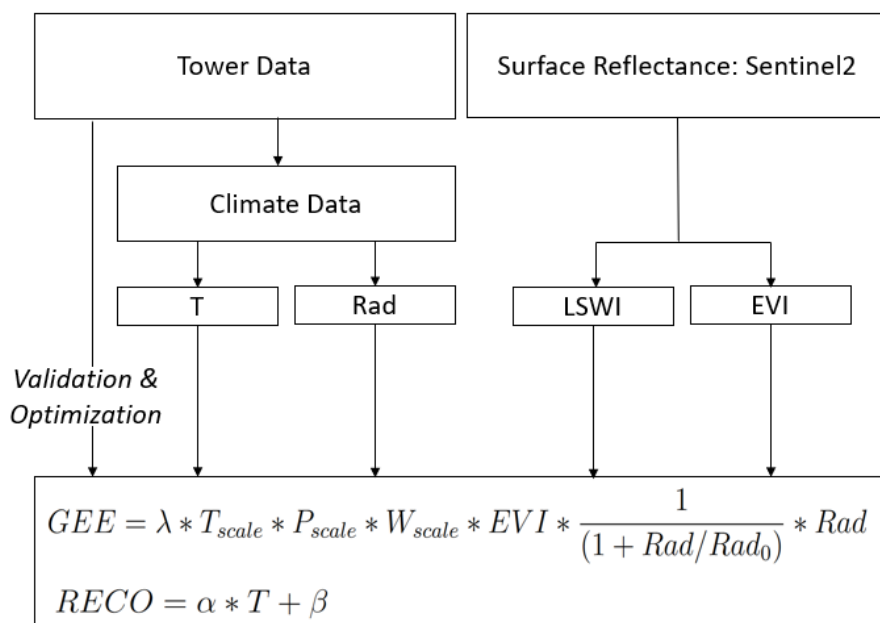
Figure 1: Schematic diagram of the Vegetation Photosynthesis Respiration Model (VPRM). Mahadevan et al. (2008).

2.2 Data Sources

All data downloaded was reduced to the spring, summer and autumn months. The winter months (December, January, February) were excluded from all datasets. This is due to the low temperatures and the lack of sunlight in our area of interest. Due to the strong conditions during the winter season in Northern Europe regions, the GEE is severely reduced and RECO dominates the carbon exchange. Moreover, there was less data available for the satellite-derived vegetation indices described in the winter months due to long nights, larger cloud presence and snow.

2.2.1 Station Data

Observational data from 6 eddy covariance flux towers covering a period from 2017-2020 and representing the evergreen needleleaf forest PFT was downloaded (see Figure 3) from the Warm Winter 2020 dataset (Team and Centre, 2022). The Warm Winter 2020 dataset (Team and Centre, 2022) is an extension of the FLUXNET 2015 dataset (Pastorello et al., 2020) for the period between 2015 and 2020 (see Appendix A). The FLUXNET2015 Dataset includes data collected at sites from multiple regional flux networks. The stations and years for which

data was downloaded can be observed in Table 2. The meteorological data contained and extracted from the dataset were: Temperature (TA_F), Shortwave Radiation (SW_IN_F), GEE (GPP_DT_VUT_REF), RECO (RECO_DT_VUT_REF), NEE (NEE_VUT_REF). In the abbreviations, DT stands for Daytime, a specific way to estimate the ecosystem components from observed NEE based on the daytime fluxes and fitting a light-response curve of the observations. VUT stands for Variable Ustar Threshold, a method to filter the data on low turbulent periods, which can affect adversely the Eddy covariance method. The area of interest was in long: 0°-29° and lat: 53°-66°. The station names their exact latitude, longitude and surrounding plant functional types can be found in Table 1. In Figure 2 we can observe the 1km x 1km PFT assignment done by the SYNMAP map (Jung et al., 2006, see Appendix A).

Table 1: Station abbreviation (SE = Sweden, FI = Finland), Station Name, Longitude and Latitude in degrees, Plant functional type

Station	Station Name	Longitude	Latitude	PFT	
FI-Hyy	Hyytiala	24.29477	61.84741	ENF	Evergreen needleleaf forest
FI-Let	Lettosuo	23.95952	60.64183	ENF	Evergreen needleleaf forest
SE-Ros	Rosinedal-3	19.738	64.1725	ENF	Evergreen needleleaf forest
SE-Svb	Svartberget	19.7745	64.25611	ENF	Evergreen needleleaf forest
SE-Htm	Hyltemossa	13.41897	56.09763	ENF	Evergreen needleleaf forest
SE-Nor	Norunda	17.47950278	60.08649722	ENF	Evergreen needleleaf forest

Table 2: Station abbreviations and years, 1 indicates that data was available for this year

Station/Year	SE-Nor	FI-Let	SE-Ros	FI-Hyy	SE-Svb	SE-Htm
2017	1	1	1	1		1
2018	1	1	1	1	1	1
2019	1	1	1	1	1	1
2020	1	1	1	1	1	1
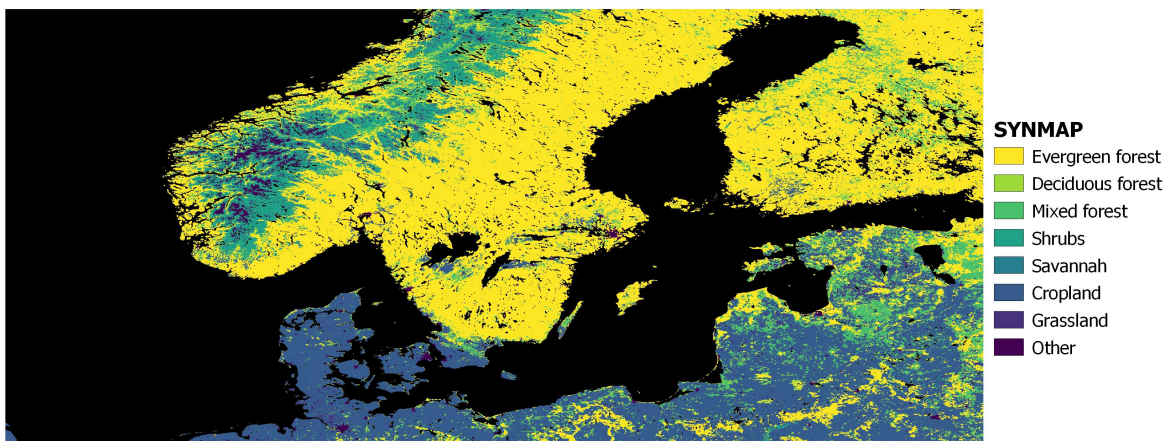
Figure 2: This Map displays the different PFTs (one classification per 1km x 1km) as used by the VPRM model, 8 classes, classified by SYNMAP (Jung et al., 2006).
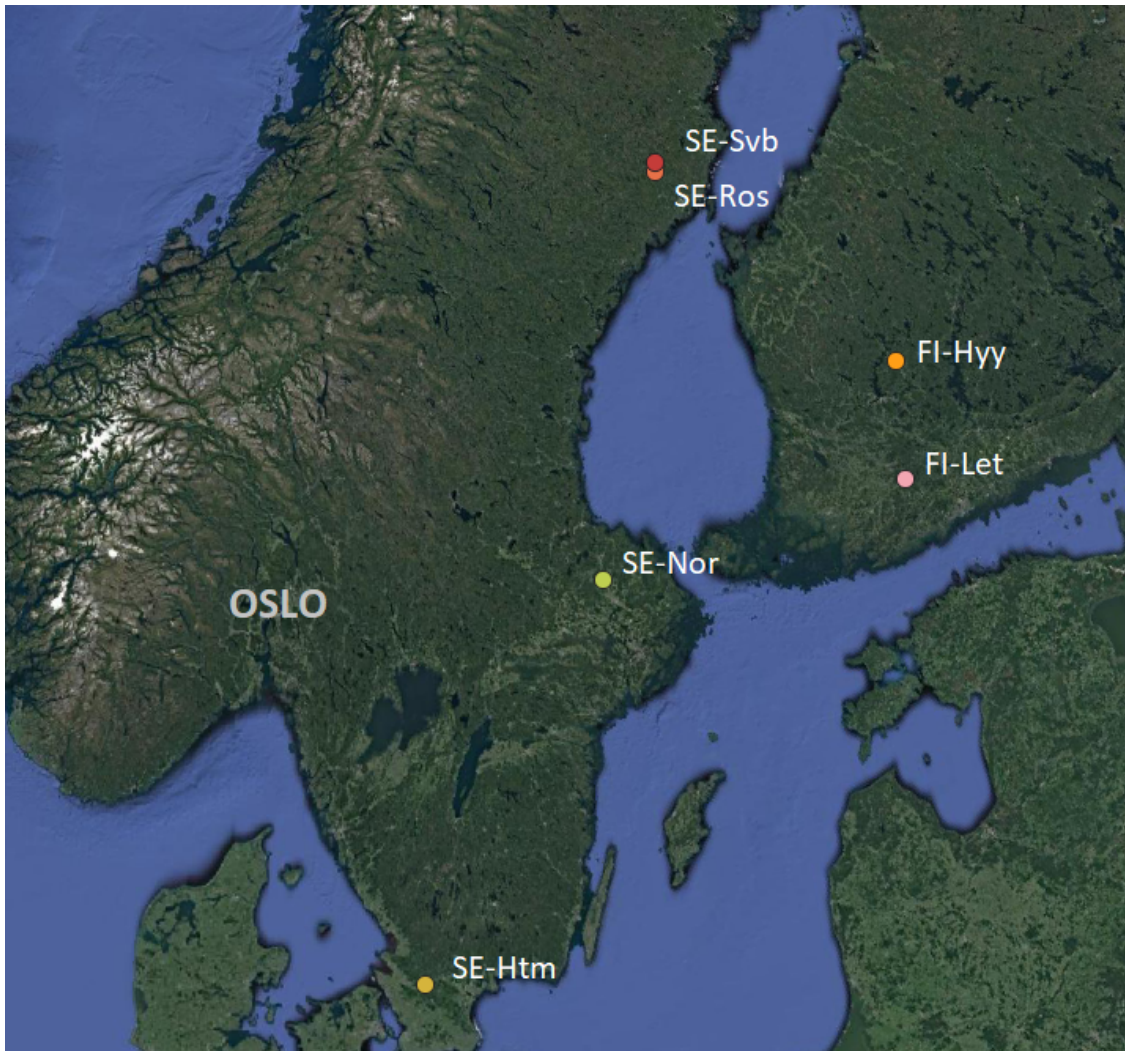
Figure 3: Map of the area of interest in Nothern Europe with the stations of PFT ENF

2.2.2 MODIS Satellite Data

For the vegetation indices (EVI and LSWI), we have made a prior comparison of two satellite-derived datasets. MODIS is a satellite-based sensor for measuring electromagnetic radiation. The translation of the term refers to a Moderate Resolution Imaging Spectroradiometer. MODIS was launched into orbit by NASA on the Terra (1999) and Aqua (2002) satellites (Brandon Maccherone, 1995). Further information can be found in the Appendix A. First, we worked with the MOD09A1 product, the Terra Surface Reflectance Version 006, that provides an estimate of the surface spectral reflectance of Terra MODIS Bands 1 through 7, corrected for atmospheric conditions such as gasses, aerosols, and Rayleigh scattering. It covers at a per pixel basis of 500m spatial resolution and an 8 day temporal resolution (Vermote, 2015). This means, the satellite product is the averages of the two images per day (Terra and Aqua) per 8 days. It is stored on an equal angle lat-long grid. The grid tiles are 1 degree by 1 degree. For MOD09A1, the relevant grid tiles ([18,2], [18,3], [19,2], [19,3]) for the years 2016-2018 were downloaded. Then, in the grid tiles, the station latitude and longitude had to be mapped onto the tile to extract the values in the correct region (Giglio et al., 2016). Unfortunately, the EVI and LSWI data was completely missing for two stations or missing in the winter month for another station.

Therefore, the grid cells of the MOD13A1, i.e. Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices 16-Day Version 6.1 product was downloaded for the year 2018. It provides Vegetation Index (VI) values at a per pixel basis at 500 meter (m) spatial resolution and 16 days temporal resolution. The low temporal resolution is a drawback for exact estimations. Moreover, as can be observed, for example in Figure 4 for the [18,2] tile of the year 2018, there is missing data, especially in the winter months. Fur-

thermore, due to a lack of the ρ_{nir} band we could not calculate the LSWI. Hence, we decided to use a different data source for the satellite-derived indices, such as the Sentinel-2 dataset.
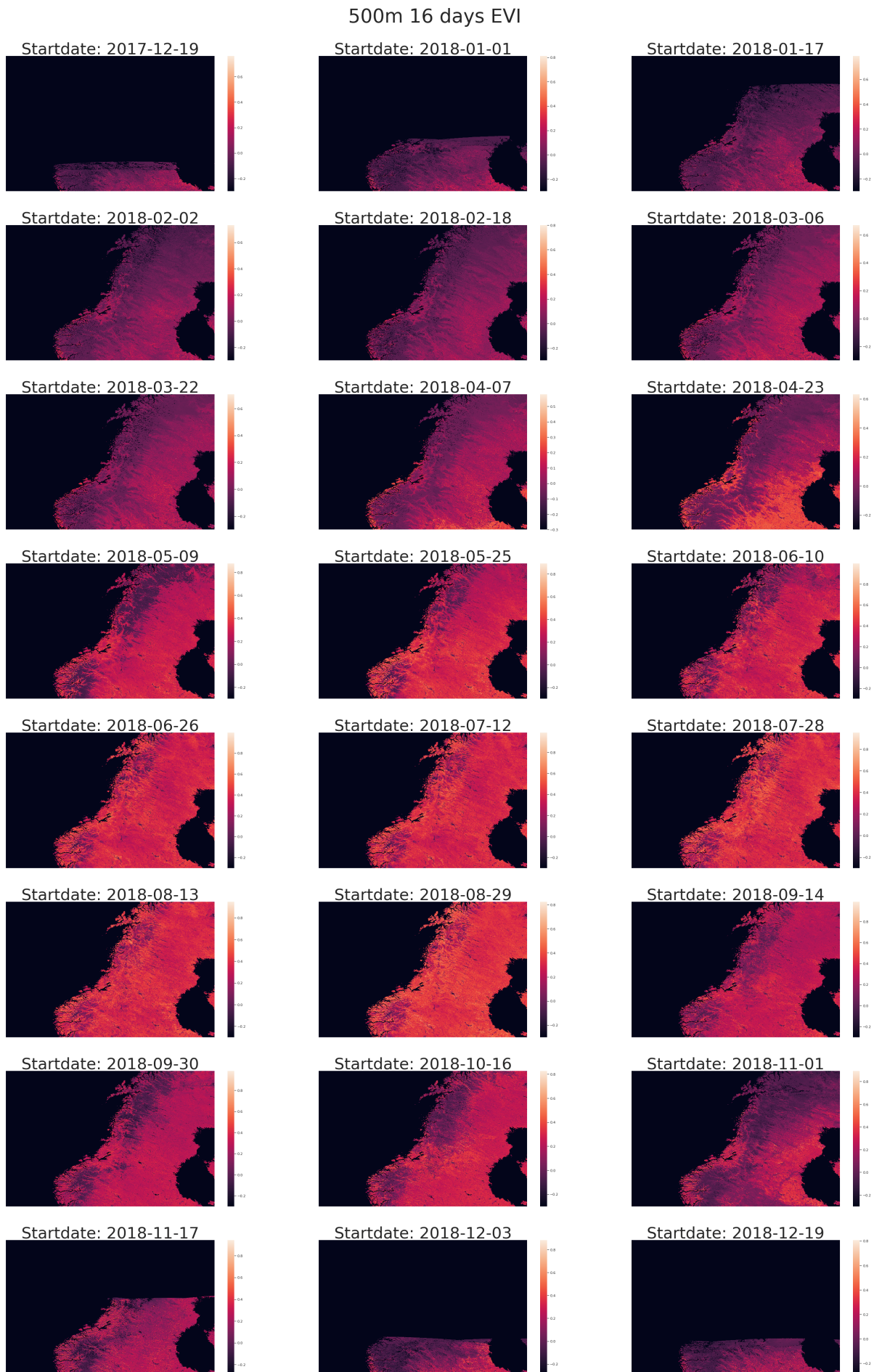
Figure 4: EVI plots of MOD13A1, (MODIS/Terra Vegetation Indices 16-Day L3 Global 500m resolution, tile [18,02], year 2018). The satellite product is the averages of the two images per day (Terra and Aqua) per 16 days. The Start date, indicated the first day of the 16days.

2.2.3 Sentinel-2 Satellite Data

The Sentinel-2 (S2) space mission, consisting of Sentinel-2A and Sentinel-2B, is a pair of optical Earth observation satellites in a sun-synchronous Earth orbit. Like the satellite pairs Sentinel-1 and Sentinel-3, they belong to the European Union's Copernicus program for Global Monitoring for Environment and Security and are part of the Sentinel satellite series. Sentinel-2 is intended to provide data for climate protection, land monitoring and disaster and crisis management on Earth. The images are recorded in spatial resolutions of 10 or 20 m per pixel, in ten spectral channels, in the wavelength range from 490 to 2200nm (Suhet and Hoersch, 2015). To calculate the LSWI (Equation (22)) the bands 8 (NIR) and 11 (wavlength B8 = 832.8nm, wavelength B11 = 1613.7nm) are needed. To calculate the EVI (Equation (21)) the bands 8, 4 (red), and 2 (blue) (wavalength B4 = 664.6nm, wavelength B2 = 492.4nm) are needed. Sentinel 2 data can be freely accessed and processed using google earth engine (`https://earthengine.google.com/`). To obtain the EVI and LSWI for the stations in the years 2017-2020, first, round shapefiles with a circular buffer zone of 500 m radius around each station were created using QGIS. The indices were derived inside the buffer zones using a google earth engine script. Every 5 days the Sentinel-2 mission covers the whole earth. Even though, there is not a common pipeline pre-processor to derive the Sentinel-2 data at longer spatial scales, for example covering the whole Northern Europe domain, as it is for the MODIS data (Ahmadov et al., 2007, Ahmadov et al., 2009).

2.3 Data Preparation

First, the NEE of all stations and years was plotted to examine missing years. Also, the count and percentage of each quality occurrence of data was plotted. This means, that the NEE, GEE and RECO measured at the flux towers consisted of measured but also interpolated data and each datapoint received a quality tag from 0-3, (0 = measured; 1 = good quality gap fill; 2 = medium; 3 = poor). All stations and years that had a quality tag ratio of 0 and 1 that was lower than 67% of the half-hours during the year were sorted out. Furthermore, the predicted NEE using the default parameters coming from Kountouris et al. (2018) was plotted against the correct NEE. The EVI and LSWI data downloaded from Sentinel-2 data had to be cleaned and interpolated to half hourly data. For the EVI, first all nan values and all values below 0 or higher than 1 were removed. In the case of the LSWI, also all nan values and all values below -1 and above 1 were removed. Then, the datapoints in between the 5 days were linearly interpolated temporally to the half-hour frequency. Finally, the interpolated data was smoothed out using the LOWESS filter. Plots of the data pre-processing can be observed in Section 3.2. The Morris Sensitivity Analysis and the parameter optimization was performed for all years and stations individually (small datasets) and for all years and stations together (big dataset). The results of the former are not presented in this study.

2.4 Morris Sensitivity Analysis

The Morris method (Morris, 1991) analyses one variable/factor/parameter/input value at a time (OAT = one at a time). OAT is a local method where the model output is computed by changing the value of only one parameter between the simulation runs.

First each parameter is assigned a range in which it is expected, that can be observed in Table 3. Then the ranges get divided regularly to provide an equal amount of levels of each parameter. Δ refers to the offset/ interval size between these possible values. Imagining

our parameter space as grid space the algorithm starts at a random starting point in the grid space. Then each parameter once gets set either one level up or down and a trajectory gets build in our grid space. After one trajectory a new random starting point gets chosen. In our case 1000 trajectories were build. While a trajectory gets build, for each parameter change by their individual $Delta$, their elementary effect got calculated (Saltelli et al., 2004). To calculate the Morris indices, all the elementary effects are calculated at different points in the grid space of the parameters. A requirement is that the number of trajectories must be large enough compared to the number of levels, so that all levels are sufficiently explored (Valade et al., 2014). For our study we chose 6 levels. The elementary effect of a function f in front of the parameters X_{N_p} is defined as follows:

$$EE_n = \frac{f(x_1, ...x_{n-1}, x_n + \Delta, x_{N_p}) - f(x_1, ...x_{n-1}, x_n, x_{N_p})}{\Delta} \tag{7}$$

The EE_n is calculated by perturbing one parameter of the parameter set by Δ. In our case the parameter set is $\alpha, \beta, \lambda, Rad_0, T_{min}, T_{max}, T_{opt}$. The mean $\mu*$ assesses the absolute overall influence of the parameter on the response.

$$\mu_n^* = \frac{1}{L} \sum_{l=1}^{L} |EE_n^{(l)}| \tag{8}$$

The mean μ does the same, just with non-absolute values. Therefore, the $\mu*$ are more meaningful, as the μ metric is less reliable as the effect from parameters within a group becomes averaged out. If the direction of effects is important, then Campolongo, Cariboni, and Saltelli (2007) suggest comparing mu with $mu*$. If mu is low and $mu*$ is high, then the effects are of different signs. A large mean indicates a large change in response when perturbing a parameter. Also, a 95% confidence interval around $\mu*$ got calculated. The standard deviation or variance σ_n^2 is a measure of non-linear and/or interaction effects.

$$\sigma_n^2 = \frac{1}{L-1} \sum_{l=1}^{L} (EE_n^{(l)} - \mu_n)^2 \tag{9}$$

A large value indicates that the elementary effects depend highly on the choice of the sample point at which it is computed. A small σ_n^2 indicates similar values in elementary effects for different sample points, implying that the effect of the parameter is independent of other parameter values (Scheidt, Li, and Caers, 2018, Saltelli et al., 2004). The results of the

Table 3: Parameters and their minimum and maximum boundaries in Morris Sensitivity Analysis

Parameter	min	max
α	0	0.45
β	-1.2	1.2
λ	0.05	0.4
Rad_0	150	350
T_{min}	-1	1
T_{max}	39	41
T_{opt}	19	21

$\mu*$ and σ_n^2 of all parameters served to decide which parameters were best to be optimized. Moreover, the parameter combination with the lowest error (RMSE Equation (10)) served as initial parameter combination to be further optimized.

$$RMSE = \sqrt{\frac{1}{n} \Sigma_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2} \tag{10}$$

where $\hat{Y}$ are the estimated and Y the calibration values.

2.5 Parameter Optimization

There are two initial sets of parameters to be optimized. One are the default parameters from Kountouris et al., 2018 and the other are the best result (lowest RMSE) obtained by the Morris Sensitivity Analysis (see 2.4). Three different techniques were used to optimize the parameters $\alpha, \beta, \lambda, Rad_0$. As explained before, the parameters were optimized for only one plant functional type, namely evergreen needle forests. Two approaches are employed for the optimization process: in the first approach, the two parameters of the GEE and the two of the RECO are optimized separately; while in the second approach the four parameters are optimized at the same time. All three optimization algorithms are iterative optimization algorithms.

2.5.1 Gradient Descent

The first optimisation technique is the simple Gradient Descent (GD). Gradient Descent is a first-order iterative optimization procedure for locating a local minimum of a differentiable function in mathematics. The goal is to repeatedly move in the opposite direction of the function's gradient (or approximate gradient) at the present location Du et al., 2019. In our case, the loss function we use is mean squared error function.

$$f(\theta) = \frac{1}{n} \sum_{i=0}^{n} (Y - \hat{Y})^2 \tag{11}$$

In the Equation (11) the error computed between the modelled variable $\hat{Y}$ using parameters θ ($\hat{Y}$ can be NEE, GEE or RECO) and the observed variable Y is calculated. The length of our dataset is n (half-hours*siteyears). First the estimation of either RECO, GEE or NEE, with the default or the Morris parameters as initial parameters, gets calculated. We then have to calculate the partial derivatives $f'(\theta)$ of the loss function with respect to our parameter of interest θ, and plug in the current values. The derivatives can be observed in the Appendix Ax, Table 13. To calculate the derivatives with respect to the loss of NEE, the partial derivatives of loss functions can be changed by adding either RECO or subtracting the GEE and use the NEE values as Y. Then the parameters θ get updated with the following Equation:

$$\theta_{t+1} = \theta_t - \gamma * \nabla f(\theta_t) \tag{12}$$

Where $\gamma = 0.0001$ is the learning rate, i.e. how much the value of our parameter changes with each step. $\nabla f(\theta_t)$ is our gradient, so the derivative of the loss function with respect to the parameter of interest in step t (Ruder, 2016). With the new parameters we again calculate the new RECO, GEE or NEE estimations for each point in the dataset and their partial derivatives. These steps are repeated until a maximum amount of repetitions was done (epochs = 2000) or if a convergence criterion is reached ($loss \leq 0.001$). More information can be found in Appendix A.

2.5.2 Levenberg-Marquardt Algorithm

The Levenberg-Marquardt (LM) algorithm, named after Kenneth Levenberg and Donald Marquardt, is a numerical optimisation algorithm for solving nonlinear balancing problems using the least squares method. The method combines the Gauss-Newton method (GN) (see Appendix A) and the method of Gradient Descent. The latter has been described in section 2.5.1. This is the Gauss- Newton iteration step:

$$\theta_{t+1} = \theta_t - \gamma * (\nabla f(\theta_t)^T * \nabla f(\theta_t))^{-1} * \nabla f(\theta_t) * f(\theta_t) \tag{13}$$

Where γ is the learning step and $\nabla f(\theta)$ the Jacobian matrix (matrix of first order partial derivatives). As mentioned before the Levenberg-Marquardt interpolates between the Newton/ Gauss–Newton algorithms and the Gradient Descent algorithm and balances the converge-speed of the former and the converge-robustness of the latter. The iteration step is defined as following:

$$\theta_{t+1} = \theta_t - (\nabla^2 f(\theta_t) + \gamma \mathbb{I})^{-1} * \nabla f(\theta_t) \tag{14}$$

where $\mathbb{I}$ is the identity matrix and γ the damping parameter. Levenberg-Marquardt is more robust than the Gauss- Newton. This means, that in many cases it finds a solution even if the initial parameters start very far off the final minimum. This is due to the damping parameter γ, which when γ is small results in a Gauss-Newton update and large values result in a Gradient Descent update. The damping parameter γ is initialized to be large so that first updates are small steps in the steepest-descent direction. If any iteration happens to result in a worse approximation $f(\theta_{t+1}) > f(\theta_t)$, then γ is increased. Otherwise, as the solution improves, γ is decreased, the Levenberg-Marquardt method approaches the Gauss-Newton method, and the solution typically accelerates to the local minimum. The error function to minimize was the sum of squared residuals, so the Levenberg-Marquardt method is a least squares method.

$$f(\theta) = \sum_{i=0}^{n} (Y - \hat{Y})^2 \tag{15}$$

(Levenberg, 1944, Marquardt, 1963, Gavin, 2019). With this procedure we reach a faster convergence than with Gradient Descent, but the Gauss-Newton update is more prone to not finding a local minimum, therefore the Levenberg-Marquardt method combines the converge-speed of the GN and the converge-robustness of the GD.

2.5.3 Differential Evolution

Differential Evolution (DE) is used for multidimensional real-valued functions. It is also applicable to optimization issues that are not even continuous, are noisy, change over time and therefore very suited for our optimisation problem. The technique is used in evolutionary computation and seeks to iteratively enhance a candidate solution with respect to a specified quality metric. Since they can search very wide spaces of potential solutions and make little or no assumptions about the problem being optimized, such techniques are frequently referred to as metaheuristics. Metaheuristics though, do not ensure that an ideal solution will ever be discovered. Differential Evolution does not require the problem to be differentiable (Storn and Price, 1997; Storn, 1996).

First, a fixed number of individuals in a population gets set, in our case 15. Each individual θ can be imagined as a parameter set. Therefore, each individual gets initiated with randomly set values for $\alpha, \beta, \lambda, Rad_0$ within these boundaries: $[0, \alpha + 0.5], [\beta - 1.0, \beta + 1.0], [\lambda - 0.1, \lambda + 0.1][Rad_0 - 25, Rad_0 + 25]$ (where $\alpha, \beta, \lambda, Rad_0$ are once the default parameters and once the best fit Morris parameters). For each individual the resulting RECO, GEE and NEE gets calculated and then the error to the calibration data: $f(\theta)$. Then one individual θ^1 gets chosen and three other, distinct individual that are not θ^1, let's call them $\theta^2, \theta^3, \theta^4$. Moreover, one of the four indices i of θ^1 i.e. one of the parameters θ_i^1 gets chosen randomly. It is here referred to as R. The other parameters that were not chosen, get assigned a random number in the range of $[0, 1]$. If this number exceeds the crossover probability of 0.9, the parameter stays the same. If the random number does not exceed the crossover probability or if it is the parameter R, a new value for the parameter gets calculated.

$$\theta_i^{new} = \theta^2 + W * (\theta^3 - \theta^4) \tag{16}$$

Here W refers to the differential weight, that gets also chosen randomly for each generation (each renewal of all individuals) by a uniform distribution between 0 and 2. Then with the new parameters θ^{new} we calculate the error $f(\theta^{new})$. If $f(\theta^{new}) < f(\theta^1)$, then, θ^1 gets set to θ^{new}. This procedure gets completed for the whole generation of individuals. Then, it gets repeated, either for a fixed amount of maximum iteration steps or till a convergence criterion is reached (Rocca, Oliveri, and Massa, 2011).

2.6   Diebold- Mariano Statistical Test

To compare if the best fit (lowest RMSE) optimized parameters yield statistically significant better estimations than the default parameters a Diebold- Mariano Statistical test is performed. For the test we first calculated the squared residuals of both predictions $\hat{y}$ to the real NEE values y.

$$r_i = (y_i - \hat{y}_i)^2 \tag{17}$$

Then the time series d_i, called the loss-differential, is calculated as $d_i = r0_i - r1_i$, where $r0_i$ refers to the residuals of the fluxes estimated the parameters based on the best fit optimization and $r1_i$ obtained based on the fluxes estimated by the model using the default parameters. The Diebold-Mariano statistic states as follows:

$$DM = \frac{\overline{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{n}}} \tag{18}$$

where $\overline{d}$ is the mean of the loss differential, n the length of our timeseries ($n = 13200$, half hours of one year without winter months) and $2\pi \hat{f}_d(0)$ is the consistent estimate of the loss differential at frequency 0.

$$2\pi \hat{f}_d(0) = \hat{\gamma}_d(0) + 2 * \sum_{\tau=1}^{k-1} \hat{\gamma}_d(\tau) \tag{19}$$

where $\hat{\gamma}_d(\tau)$ is the autocovariance at lag τ and k is the total time lag we want to observe. In our example a k of 0,6,12,24 and 48 was chosen which represents 0, 3, 6, 12 and 24 hours in our dataset.

$$\hat{\gamma}_d(\tau) = \frac{1}{T} \sum_{i=\tau}^{T} (d_t - \overline{d})(d_{t-\tau} - \overline{d}) \tag{20}$$

For each station and year, the Diebold- Mariano Statistic (DM) was calculated, as using the big dataset, would not have represented a time series. Under the assumption that $\mu = 0$ (the null hypothesis), DM follows a standard normal distribution. Thus, there is a significant difference between the forecasts if $|DM| > z_{crit}$ where z_{crit} is the two-tailed critical value for the standard normal distribution. In our case $z_{crit} = 2.576$ for an $\alpha = 0.01$. (Diebold and Mariano, 2002)

The code in which the methodology was implemented can be found in the following github repository: `https://github.com/mdomhoefer/VPRM_NothernEurope`

3   Results

3.1   Satellite Data Comparison

For all satellite data products there are certain drawbacks. To compare them, we created plots and tables of the different satellite data products of the year 2018. In the case of the MOD09A1 there was missing EVI and LSWI data for two of our stations, namely FI-Let and SE-Ros. This can be observed in the Figures 5 and 6. Also, there are less data points than for the Sentinel2, as can be observed in Tables 4 and 5 for the other stations. The MOD13A1 was supposed to be the alternative to the MOD09A1 and there was EVI data for all stations. But as the temporal resolution of the MOD09A1 is higher than of the MOD13A1, we of course got the least amount of data points (see Table 4). Moreover, the MOD13A1 has only a ρ_{mir} band and no ρ_{swir} band, which is needed to calculate the LSWI. Therefore, as can be observed in the Figure 6 and Table 5, there is no LSWI data available. MODIS works with grid tiles. When accessing those, they are also divided into small cells. Therefore, when searching in the grid tiles with a certain latitude and longitude you cannot find the exact spot of the station but rather the cell in which it is located. With the Sentinel2 data, we could exactly pin down the locations of our stations, so the data was more precise. Due to the higher spatial resolution of Sentinel2, there is no WRF-VPRM pipeline yet to automatically integrate Sentinel2 data, to make a larger scale NEE analysis (Ahmadov et al., 2007, Ahmadov et al., 2009). A future step is therefore, to upscale the Sentinel 2 indices to a regional scale. Moreover, as Sentinel2 was launched 2015, there is no data available before this year. From the Sentinel2 there was EVI and LSWI data available for all years and stations. As can be observed in the Figures 5 and 6, all satellite data products have less data points in the winter months, compared to the other seasons. This is one reason, why we did not include them in our analysis.

Table 4: Counts of EVI datapoints from different satellites (MOD09A1, MOD13A1, Sentinel2) in year 2018

	FI-Hyy	FI-Let	SE-Ros	SE-Svb	SE-Htm	SE-Nor
MOD13A1_EVI	24	24	24	24	24	24
MOD09A1_EVI	43	0	0	38	46	45
Sentinel2_EVI	61	62	86	84	69	62

Table 5: Counts of LSWI datapoints from different satellites (MOD09A1, MOD13A1, Sentinel2) in year 2018

	FI-Hyy	FI-Let	SE-Ros	SE-Svb	SE-Htm	SE-Nor
m13_LSWI	0	0	0	0	0	0
m09_LSWI	46	0	0	39	46	45
s02_LSWI	61	63	89	86	71	62
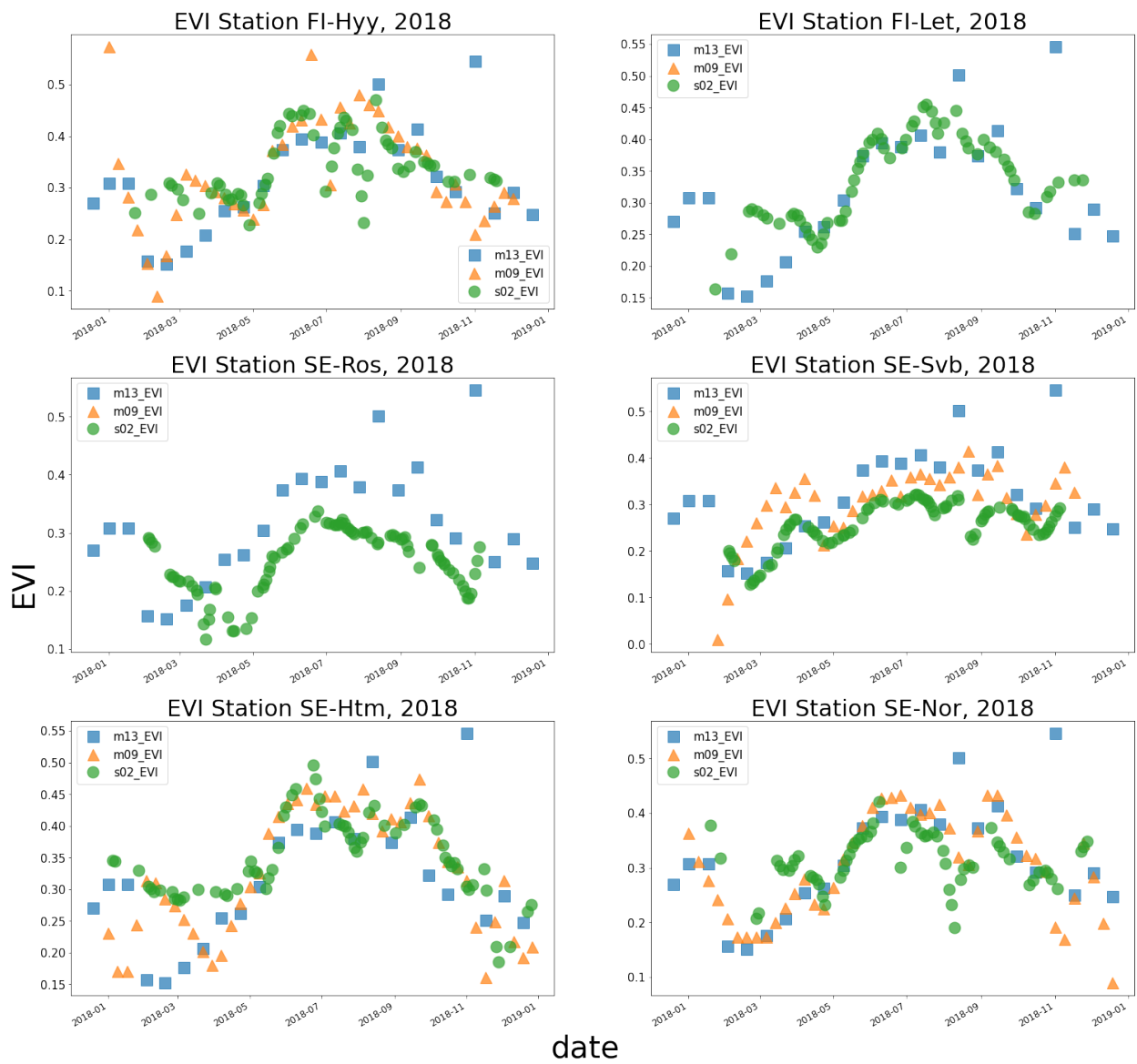
Figure 5: Comparison of the EVI values in year 2018 from different satellites (MOD09A1 = mod09, MOD13A1 = mod13, Sentinel2 = s02) at stations
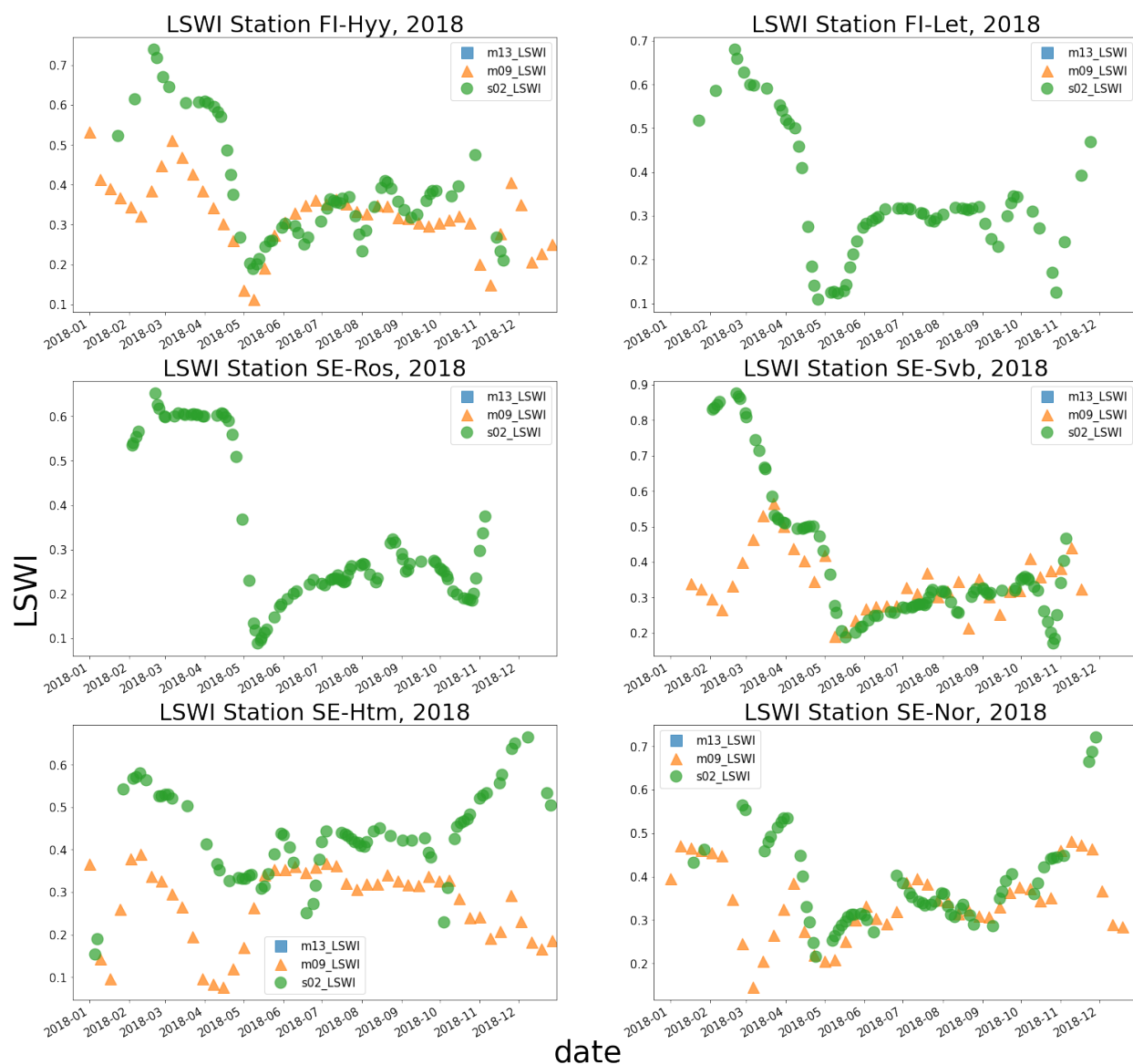
Figure 6: Comparison of the LSWI values in year 2018 from different satellites (MOD09A1 = mod09, MOD13A1 = mod13, Sentinel2 = s02) at stations

3.2   Data Visualisation

In Figure 7, we can observe the NEE of the calibration stations. In this Figure, we can detect the difference in the magnitude of fluxes over the four seasons.

As mentioned before, the quality of data for all stations was also visualized. In Figure 8, we can observe the quality tag count and percentage frequency of the NEE measures. For the station "SE-Svb", the year 2017 is missing. For the rest of the siteyears, the percentage of medium and poor quality gap filled half-hourly NEE does not exceed the 33%. Therefore, all other stations and years could be used.
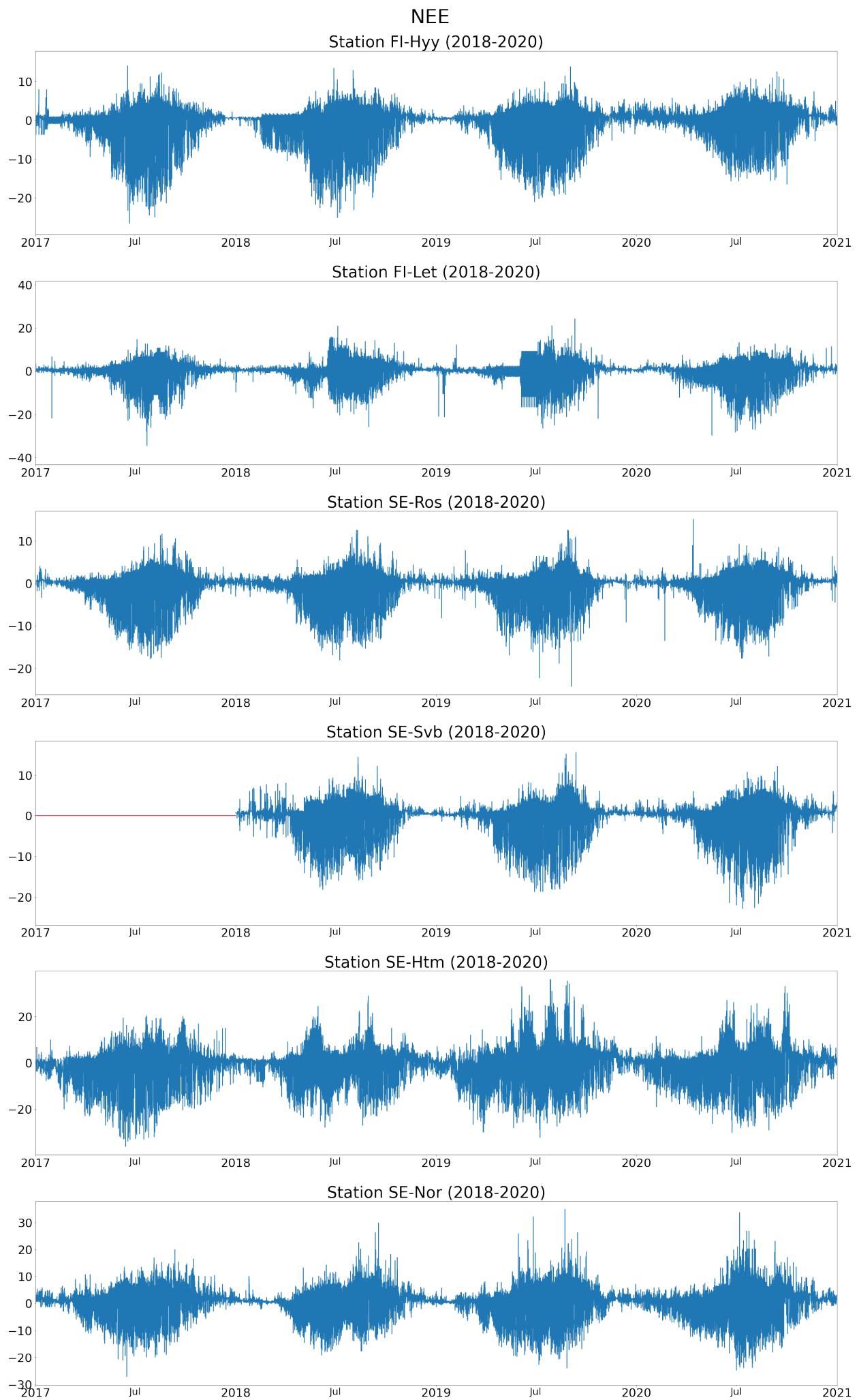
Figure 7: NEE in $\mu mol \frac{CO_2}{m^2*s}$ measured at the 6 eddy flux towers with the PFT Evergreen needle forest over the years 2017-2020
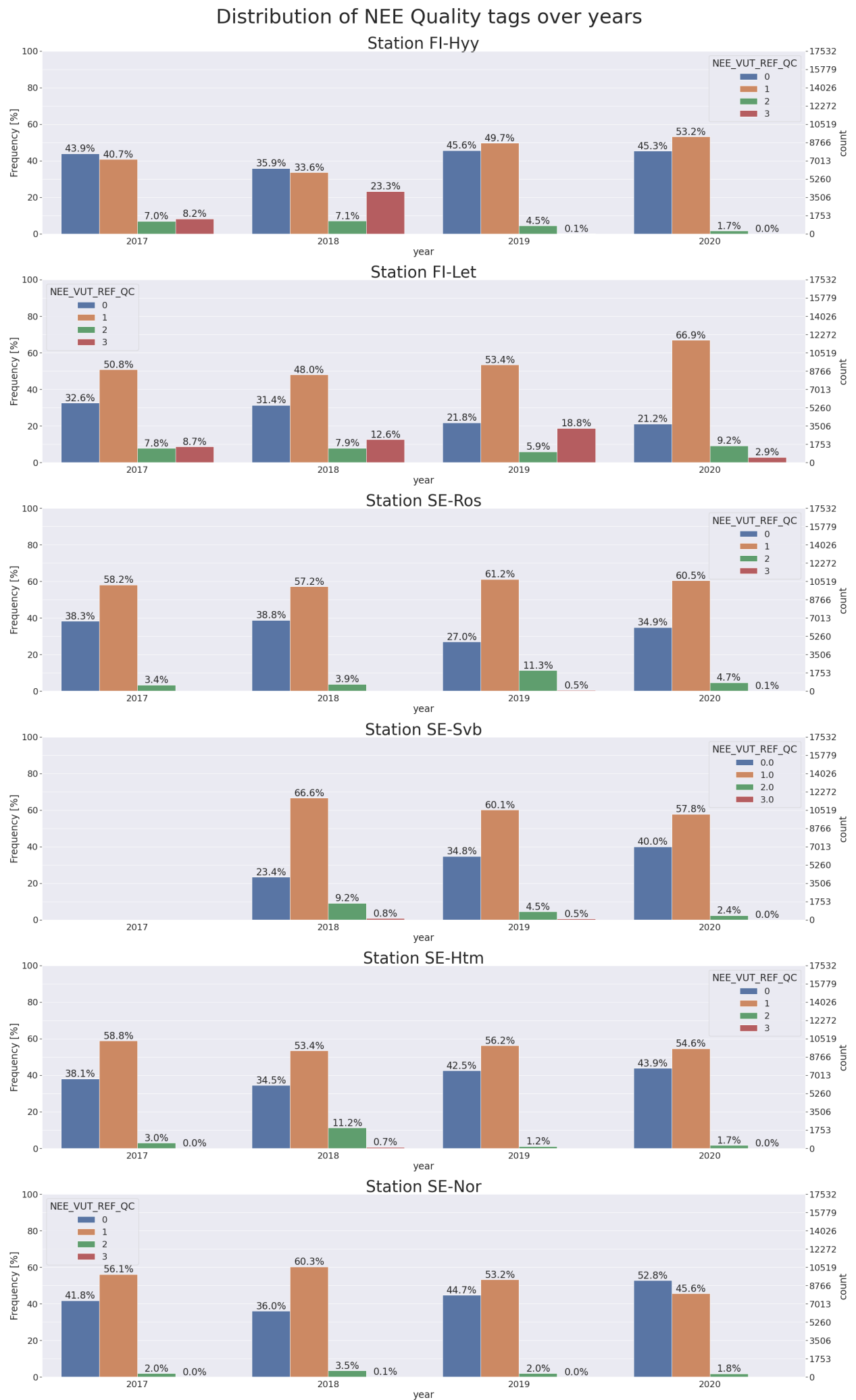
Figure 8: Percentage per year (2017-2020) of Quality tags of NEE measured at the 6 eddy flux towers with the PFT Evergreen needle forest

3.3   Morris Sensitivity Analysis

In Table 6 we inspect the results of the Morris Sensitivity Analysis on the whole dataset. λ has the highest $\mu*$ value of 15152.160 and T_{max} the lowest with 93.825. The second and third highest value had α and β and the fourth highest Rad_0 with 4638.220. Followed by T_{opt} with 459.895. The $mu*$ of Rad_0 is almost 10 times bigger than the one of T_{opt}. Due to this result, it was decided to only optimize the parameters $\alpha, \beta, \lambda, Rad_0$, as their μ was very high, showing they have a larger effect on the model outcome. What is interesting to observe, is that α, β have a very small σ^2. On the other hand, λ, Rad_0 have a high sigma, which means these parameters have non-linearities with other parameters, which can be noticeable in the GEE equation. The parameter combination with the lowest $RMSE$, our best fit result was: $\alpha = 0.3, \beta = 0.24, \lambda = 0.33, Rad_0 = 230.0$ (units are: $\alpha : \mu molCO_2 m^{-2}s^{-1}({}^0C){-}1; \beta : \mu molCO_2 m^{-2}s^{-1}; \lambda : \mu molCO_2 m^{-2}s^{-1}(Wm^{-2})^{-1}; Rad_0 : Wm^{-2}$).

Table 6: Results of Morris Sensitivity Analysis on all Parameters

|	$\mu*$	σ^2
α	10766.323837	9.440396e-11
β	8640.000000	3.448698e-12
λ	15152.159908	2.422744e+03
Rad_0	4638.219363	2.508989e+03
T_{min}	412.446756	2.298029e+02
T_{max}	93.824930	5.284808e+01
T_{opt}	459.894736	2.577275e+02

3.4   Individual Parameter Calibration

First, the parameters for the RECO and the GEE individually got optimized for the complete dataset. In Table 7, we can see the RECO parameter optimization results. In Table 8, we can observe the GEE parameter optimization results. Then, the RMSE of the estimated results of RECO and GEE to the calibration data (variables in the FLUXNET dataset: GPP_DT_VUT_REF, RECO_DT_VUT_REF) was calculated. Moreover, using the separately obtained parameters, the RMSE of the estimated NEE to the calibration data was calculated. The results can be observed in Table 9. First of all, the RMSE is lowest in all cases when the Morris parameters were used as initial parameters. Furthermore, we can observe that the RMSE of the RECO estimation with α and β calculated with all optimization methods (Morris init) is equally good (RMSE = 2.096). For the GEE estimation the best results (RMSE = 2.987) were obtained with the differential evolution optimization (Morris init). Interestingly, the best NEE estimation (-GEE + RECO) was with the parameters of the Morris Sensitivity Analysis. Apart from that, Gradient Descent and Differential Evolution algorithms with a RMSE of 3.374, very closely followed by the Levenberg- Marquardt algorithm (RMSE = 3.380), with the Morris parameters as initial parameters had the second and third lowest RMSE for the NEE estimation.

Table 7: α, β optimized with different initial parameters, once the Default and once the parameters obtained from Morris Sensitivity Analysis, units are: α : $\mu mol CO_2 m^{-2} s^{-1} (^0 C) - 1; \beta : \mu mol CO_2 m^{-2} s^{-1}$

	Default	Morris Sensitivity	Gradient Descent	Differential Evolution	Levenberg-Marquardt
α (init Default)	0.288	0.300	0.415	0.351	0.330
β (init Default)	-1.093	0.240	-0.958	-0.093	0.197
α (init Morris)	0.288	0.300	0.334	0.342	0.342
β (init Morris)	-1.093	0.240	0.231	0.122	0.122

Table 8: λ, Rad_0 optimized with different initial parameters, once the Default and once the parameters obtained from Morris Sensitivity Analysis, units are: λ : $\mu mol CO_2 m^{-2} s^{-1} (Wm^{-2})^{-1}; Rad_0 : Wm^{-2}$

	Default	Morris Sensitivity	Gradient Descent	Differential Evolution	Levenberg-Marquardt
λ (init Default)	0.226	0.330	0.287	0.326	0.380
Rad_0 (init Default)	275.460	230.0	275.460	225.460	178.867
λ (init Morris)	0.226	0.330	0.322	0.347	0.380
Rad_0 (init Morris)	275.460	230.0	229.999	205.0	178.867

Table 9: RMSE of estimated RECO, GEE and NEE (-GEE+RECO) values with individually calculated parameters

	Default	Morris Sensitivity	Gradient Descent	Differential Evolution	Levenberg-Marquardt
RECO (init Default)	2.752	2.131	2.169	2.100	2.097
RECO (init Morris)	2.752	2.131	2.096	2.096	2.096
GEE (init Default)	3.309	2.998	3.008	2.992	3.175
GEE (init Morris)	3.309	2.998	2.993	2.987	3.050
NEE (init Default)	3.705	3.368	3.416	3.369	3.409
NEE (init Morris)	3.705	3.368	3.374	3.374	3.380

3.5   Combined Parameter Calibration

In this section, the results of optimizing all parameters at once with the different strategies will be presented (see Table 10). The RMSEs of the NEE calculated with different parameters, obtained with different strategies to the calibration data, can be observed in Table 11. First of all, the RMSE is lowest in all cases when the Morris parameters were used as initial parameters.

The parameters obtained with the Levenberg- Marquardt algorithm yield the best result. We even got the same RMSE (3.323) no matter if the starting parameters were the Default parameters or the Morris parameters. That might be due to the behaviour of the damping parameter γ in the optimisation step of the LM algorithm. The result is better than the cases in which the parameters for RECO and GEE were optimized separately. There the best RMSE (3.368) was yielded by the Morris Analysis. In Figure 9 we can see the NEE that was measured at the stations in comparison to the estimated NEE one time with the

default and one time with the best fit parameters (LM). In all cases, the estimated NEE using the parameters obtained with the LM method seems to be more aligned to the real values, than by VPRM with the default parameters. In Figure 10, the diurnal profile of the VPRM modelled NEE and observed NEE for each month and siteyear is shown. First, these plots show how the NEE gets lower every day, as the GEE becomes higher. Moreover, the NEE presents smaller values in summer due to the stronger photosynthesis during this season. Furthermore, we cannot really observe a strong difference between the two optimization techniques. This is why a statistical analysis had to be done.

Table 10: $\alpha, \beta, \lambda, Rad_0$ optimized at the same time with different initial parameters, once the Default and once the parameters obtained from Morris Sensitivity Analysis, units are: $\alpha : \mu mol CO_2 m^{-2} s^{-1} (^0C) - 1; \beta : \mu mol CO_2 m^{-2} s^{-1}; \lambda : \mu mol CO_2 m^{-2} s^{-1} (Wm^{-2})^{-1}; Rad_0 : Wm^{-2}$

	Default	Morris Sensitivity	Gradient Descent	Differential Evolution	Levenberg-Marquardt
α (init Default)	0.288	0.300	0.472	0.384	0.401
β (init Default)	-1.093	0.240	-1.042	-0.093	-0.057
λ (init Default)	0.226	0.330	0.329	0.326	0.396
Rad_0 (init Default)	275.460	230.0	275.395	263.844	203.423
α (init Morris)	0.288	0.300	0.369	0.396	0.401
β (init Morris)	-1.093	0.240	0.246	-0.098	-0.057
λ (init Morris)	0.226	0.330	0.360	0.346	0.396
Rad_0 (init Morris)	275.460	230.0	230.014	247.002	203.419

Table 11: RMSE of NEE estimation, calculated with different parameters obtained with different optimization methods, to calibration data

	Default	Morris Sensitivity	Gradient Descent	Differential Evolution	Levenberg-Marquardt
NEE (init Default)	3.705	3.368	3.36643677	3.333	3.323
NEE (Morris Default)	3.705	3.368	3.329	3.328	3.323

NEE observation to default and LM estimation

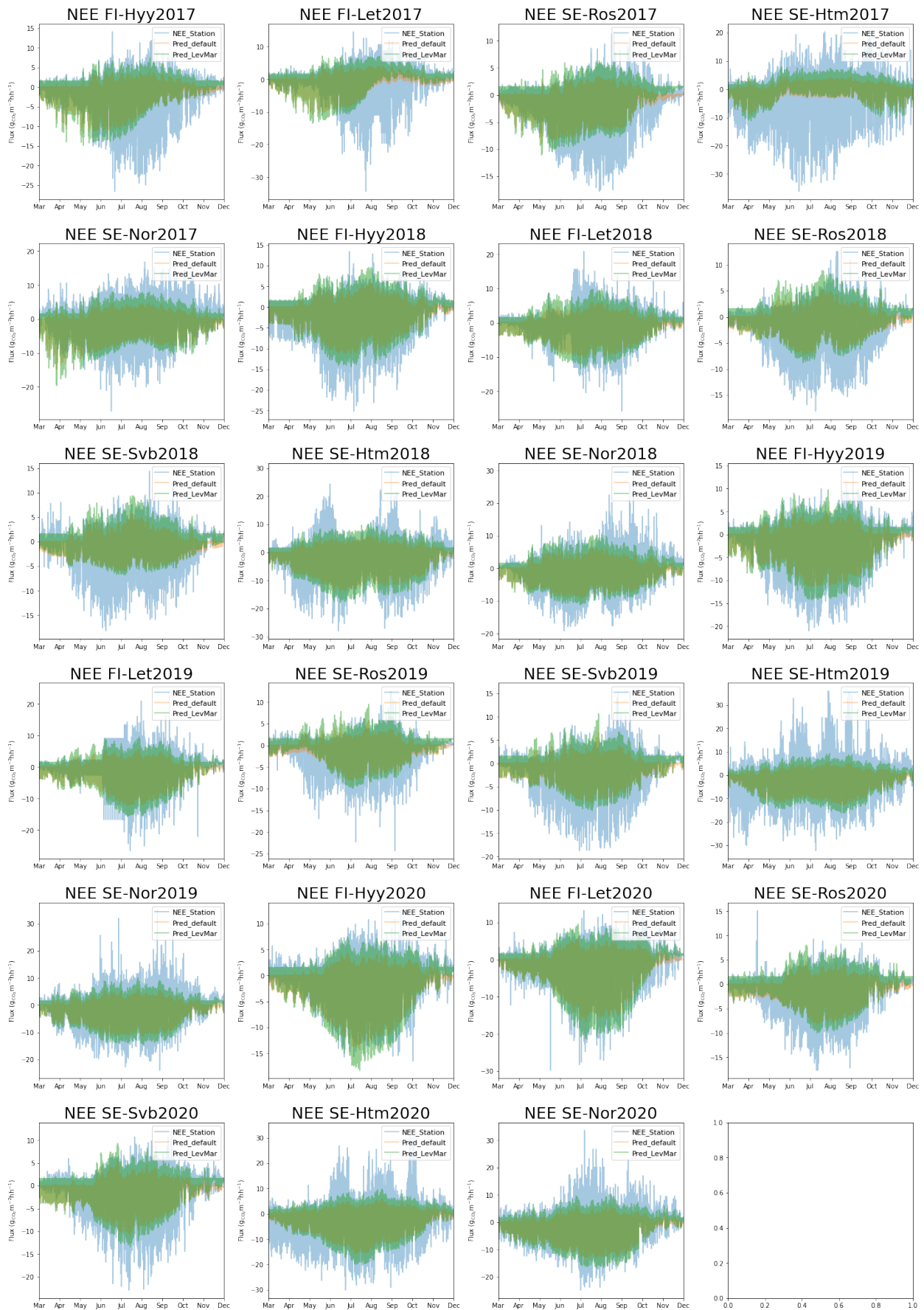*Figure 9: NEE (half-hourly) observation and prediction with two sets of parameters (default, Levenberg- Marquardt) of each station and year*
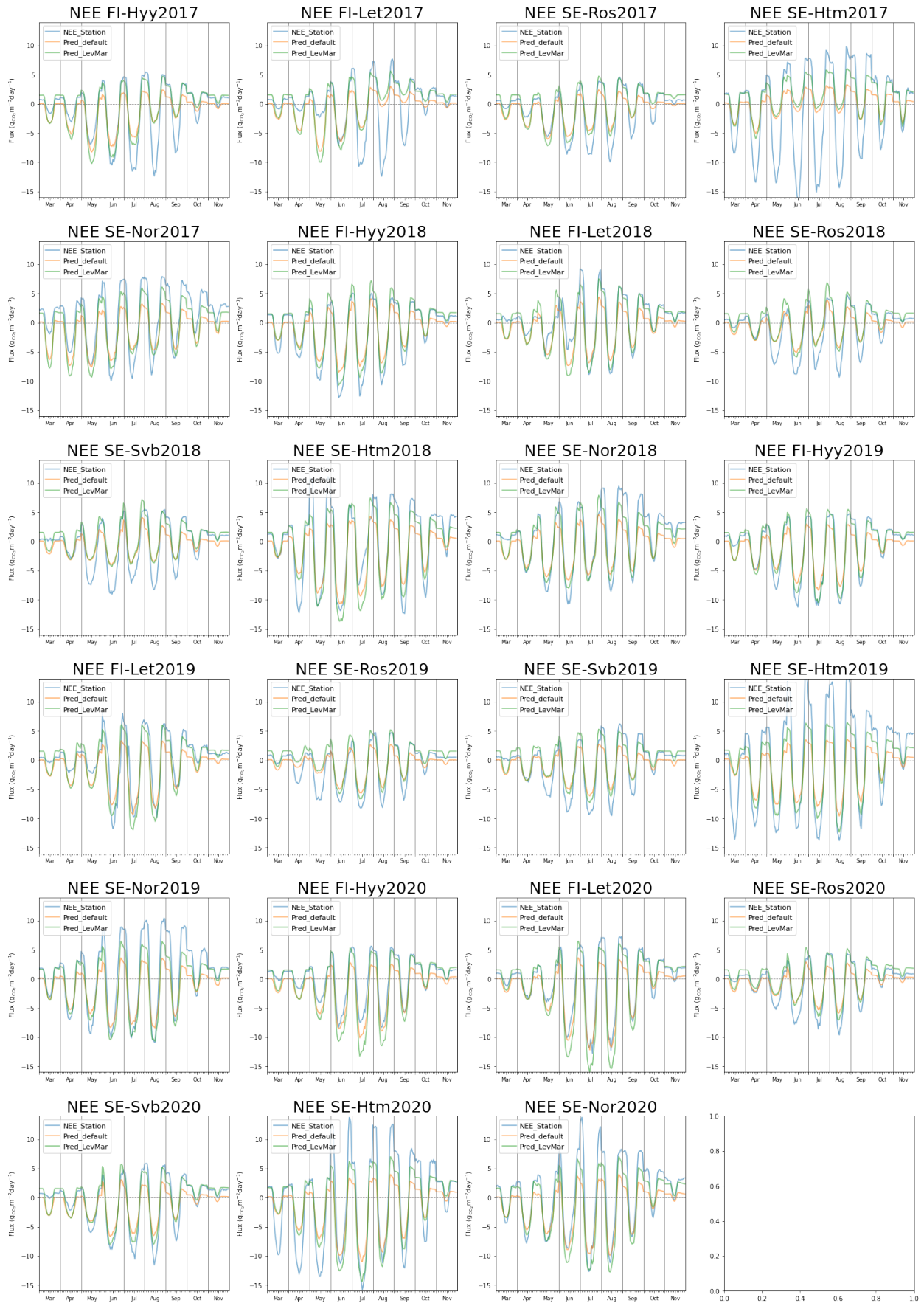
Figure 10: NEE (diurnal) observation and prediction with two sets of parameters (default, Levenberg- Marquardt) of each station and year

3.6 Statistical Evaluation

For each station and year, a Diebold-Mariano Test, to determine whether the two forecasts are significantly different, was performed. It was tested, if the NEE estimations made with the Levenberg- Marquardt parameters (lowest RMSE) was statistically better than the forecast made with the Default parameters. The results of that test can be observed in Table 12. As can be observed in Table 12, for $k = 0$ only for one station the $H0$ cannot be rejected. For all other k, for only three siteyears, the null hypothesis could not be rejected. This means that for all other examples there is a 99% chance for the other stations that the parameters estimated with the Levenberg- Marquardt method are significantly different to the Default parameters.

Table 12: In the first column are the abbreviations of the station names and the year of which the data stems. The following columns show the Diebold-Mariano Statistic DM for different time lags $k = [0, 6, 12, 24, 48]$. If $|DM| > z_{crit}$ the null hypothesis gets rejected (rej. $H0 = T$). Otherwise, the null hypothesis cannot be rejected (rej. $H0 = \textbf{F}$)

Station, Year	$k = 0$ DM	rej. $H0$	$k = 6$ DM	rej. $H0$	$k = 12$ DM	rej. $H0$	$k = 24$ DM	rej. $H0$	$k = 48$ DM	rej. $H0$
FI-Hyy2017	42.403	T	15.610	T	12.812	T	12.201	T	9.320	T
FI-Let2017	-0.251	**F**	-0.084	**F**	-0.068	**F**	-0.088	**F**	-0.069	**F**
SE-Ros2017	32.312	T	12.498	T	10.155	T	9.599	T	7.173	T
SE-Htm2017	15.679	T	5.327	T	4.227	T	4.802	T	3.691	T
SE-Nor2017	61.827	T	21.674	T	17.918	T	20.323	T	16.313	T
FI-Hyy2018	45.461	T	17.778	T	14.687	T	13.895	T	10.649	T
FI-Let2018	30.519	T	10.946	T	8.788	T	7.607	T	5.534	T
SE-Ros2018	4.925	T	1.884	**F**	1.499	**F**	1.321	**F**	0.997	**F**
SE-Svb2018	19.440	T	6.815	T	5.444	T	5.396	T	4.066	T
SE-Htm2018	45.951	T	16.032	T	12.846	T	11.821	T	8.765	T
SE-Nor2018	66.197	T	23.414	T	18.802	T	17.879	T	13.364	T
FI-Hyy2019	65.883	T	26.498	T	21.928	T	19.349	T	14.699	T
FI-Let2019	21.181	T	8.731	T	8.255	T	8.824	T	7.174	T
SE-Ros2019	-4.614	T	-1.611	**F**	-1.234	**F**	-1.030	**F**	-0.751	**F**
SE-Svb2019	31.187	T	10.750	T	8.331	T	7.346	T	5.398	T
SE-Htm2019	66.454	T	23.478	T	19.217	T	19.525	T	14.625	T
SE-Nor2019	69.636	T	25.195	T	21.175	T	23.370	T	17.895	T
FI-Hyy2020	14.380	T	5.227	T	4.307	T	4.891	T	3.827	T
FI-Let2020	17.152	T	6.867	T	5.794	T	6.597	T	5.517	T
SE-Ros2020	27.300	T	10.005	T	7.812	T	6.489	T	4.780	T
SE-Svb2020	56.079	T	21.421	T	17.607	T	15.722	T	11.663	T
SE-Htm2020	48.906	T	17.507	T	14.273	T	13.635	T	10.015	T
SE-Nor2020	55.534	T	20.115	T	17.546	T	20.704	T	16.532	T

4   Discussion

The goal of this study was to optimize the most important parameters of the VPRM model in order to estimate the carbon exchange in the Northern Europe region, in specific the evergreen needle forests. Generally, one of the most crucial constraints in the optimization of biogenic terrestrial models is the lack of measurement stations that provide the data necessary in order to optimize the model (Team and Centre, 2022). For this area only 6 eddy flux towers provided usable data, the rest did either not provide the relevant variables or did not fit the quality demands. As for other PFTs are not represented at all, or very few and as the Nothern Europe area consists of mainly coniferous forests (see Figure 2; Reinhart et al., 2021) the parameters of only this PFT were optimized. In other studies, like Kountouris et al. (2018) and Mahadevan et al. (2008) there was a parameter optimization for all plant functional types. It is open to discuss, if optimizing with less data for other PFTs, or enlarging the search space would lead to more insight, as the prediction of NEE for a bigger scale scenario would be dependent on the ratio of the PFT Lin et al. (2011). Moreover, there is no suitable station data available for the region very close to Oslo, nor any suitable station in Norway.

We have decided to omit the winter season of our analysis due to the lack of proper satellite measurements during this period to derive the vegetation indices. Moreover, the vegetation is practically dormant during this season due to the strong climatic conditions.

The most important part for discussion is that there were no scaling-up fluxes from site-specific measurements to regional levels. The main idea of the VPRM is to predict the NEE for a bigger region, than only at the flux towers. In other studies like Mahadevan et al. (2005) and Mahadevan et al. (2008), early VPRM studies, satellite images and larger scale climate data were integrated to predict the NEE for the region of interest. For our case, we did not use MODIS, but Sentinel2 data as explained in section 3.1. The main problem with the Sentinel2 data is that there is no pipeline yet to process the big amount of EVI and LSWI data that would come from the google earth engine. Therefore, it is not possible to perform a larger scale analysis yet, but only a point analysis at the flux tower stations.

In the paper by Mahadevan et al. (2008), the parameters were optimized with nonlinear least squares (Newton-Raphson, tangent linear approximation). Newton-Raphson is also known as Newton algorithm, which is explained in the Appendix A.8. The Newton method has the disadvantage to the Newton- Gauss method that the latter only works with Jacobian matrices (matrices of first order partial derivatives) and therefore avoids the calculation of the second order partial derivatives. The Levenberg- Marquardt method integrates the gradient descent and the Newton- Gauss method and has therefore the advantage of not using the second order derivatives, the faster convergence of Newton- Gauss over Gradient Descent and the higher convergence robustness of Gradient Descent (Puiseux, 2019).

Nevertheless, there are some problems with Gradient Descent. If the learning rate is chosen too high, it is possible to overshoot a local minima and never reach convergence due to "bouncing" around the minima. Choosing the learning rate too low on the other hand can result in a very long convergence time, or when working with a maximum number of epochs lead to an interruption of the algorithm before convergence. Another problem is preventing being stuck in multiple sub optimal local minima by minimizing extremely non-convex error functions. In a further study the usage of a momentum term could be another approach that can assist the algorithm to get rid of local minima and improve convergence speed (Haji and Abdulazeez, 2021).

As mentioned before, LM is more robust than the Newton- Gauss (Nelles, 2001). It's faster

to converge than either the Newton- Gauss or Gradient Descent on its own. Moreover, if your initial guess is far from the mark, the algorithm can still find an optimal solution. One problem is, that for flat functions (i.e. smooth functions with derivatives that vanish at a certain point), the algorithm can get "…lost in parameter space" (Transtrum and Sethna, 2012). Additionally, LM can be slow to converge, if the model has more than ten parameters (Waterfall et al., 2006, Gutenkunst et al., 2007), which is not the case for the VPRM model. Lastly, the Differential Evolution optimization could be used when you can formulate the optimization's goal using a fitness function where the individuals in the population are vectors of real numbers, which is the case in our study (Storn and Price, 1997). For problems like the travelling salesman or other combinatorial optimization problems you can choose a better suited genetic algorithm (Larranaga et al., 1999). Like in genetic algorithms, DE employs mutation and crossover. As explained in 2.5.3 the mutating is what happens to our R, the randomly chosen parameter of our individual θ^1 and the crossover is determined by our crossover probability. The transformation formula is the same, but the selection procedure different (Eftekhari et al., 2008). DE is rather simple and has the advantage of not needing a differentiable function as input. Moreover, it can be improved easily like in Yan, Ling, and Sun (2006) with Simulated Annealing for example. Another advantage is, that it is not a blackbox algorithm and the processes could be monitored pretty easy and help with further fine-tuning.

When the RECO and GEE equation got optimized separately the best fit NEE prediction came from the Morris sensitivity analysis. This is interesting, as we used those parameters to optimize them further with our three different techniques. Furthermore, when looking at Table 9, at the RMSEs of the RECO and GEE only, the results of them are better in all cases then the default parameters and in most cases than the Morris parameters. In conclusion, the newly optimized parameters work good for the Equations of RECO And GEE individually, but might cancel each other out when used to calculate the NEE. Therefore, it is better, to directly optimize the complete NEE function.

The Diebold-Mariano test compares the forecast accuracy of two forecast methods. Therefore, it analyses the autocovariance of the residual differences with different time lags k. Hyndman and Athanasopoulos (2018) suggest for seasonal data to use a k = 2m, where m is the period of seasonality i.e our season could be day and night (12h, approx. light and no light) so k = 48 (2*24half hours). As a season could also be 3 or 6 hours, as the temperature and light conditions are fairly the same in those time periods (k = 2*6hh, k = 2*12hh). This is why those lags have been tested. They all give the same result in terms of $H0$ rejection. The only difference comes when testing with a time lag of k = 0, then only one siteyear is not significantly different. All in all, it is very hard to choose the lag parameter. The original paper only suggests to choose a lag dependant on how much time into the future we want to make a forecast with our model (Diebold and Mariano, 2002). Another paper by Mohammed and Mousa (2019) is not stating a value at all and just presents results and another paper by Rathod et al. (2021) sets k to $k = n^{\frac{1}{3}} + 1$ without further explanations (in our case this would be $k \approx 24$).

5 Conclusion

In conclusion, the Morris Sensitivity Analysis was important to detect the most important parameters. Furthermore, it served to detect good initial parameters to further be optimised. All three optimization techniques worked well. The best result in this study was achieved by the Levenberg- Marquardt algorithm. The resulting estimated parameters were statistically different to the default parameters and yielded a better result. Unfortunately, the satellite-derived indices obtained from the Sentinel-2 data cannot be automatically integrated yet for a larger scale NEE analysis. Future work could centre on building a pre-processor pipeline for the Sentinel 2 data, in order to apply the model with our new parameters at a larger scale. All in all, this study can be used to represent better the daily and seasonal cycles of biogenic fluxes in natural areas within and around the city to study, in the future, the urban carbon balance in the Greater Oslo Region.

References

Ahmadov, R et al. (2007). "Mesoscale covariance of transport and CO2 fluxes: Evidence from observations and simulations using the WRF-VPRM coupled atmosphere-biosphere model". In: *Journal of Geophysical Research: Atmospheres* 112.D22.

Ahmadov, R et al. (2009). "Comparing high resolution WRF-VPRM simulations and two global CO 2 transport models with coastal tower measurements of CO 2". In: *Biogeosciences* 6.5, pp. 807–817.

Brandon Maccherone, Shannell Frazier (1995). *Moderate Resolution Imaging Spectroradiometer*. URL: https://modis.gsfc.nasa.gov/about/.

Campolongo, Francesca, Jessica Cariboni, and Andrea Saltelli (2007). "An effective screening design for sensitivity analysis of large models". In: *Environmental modelling & software* 22.10, pp. 1509–1518.

De Rigo, D et al. (2016). "European forests: an ecological overview". In: *European Atlas of Forest Tree Species. Publications Office of the European Union, Luxembourg, pp. e01e873+. https://w3id. org/mtv/FISEComm/v01/e01e873*.

Diebold, Francis X and Robert S Mariano (2002). "Comparing predictive accuracy". In: *Journal of Business & economic statistics* 20.1, pp. 134–144.

Du, Simon et al. (2019). "Gradient descent finds global minima of deep neural networks". In: *International conference on machine learning.* PMLR, pp. 1675–1685.

Eftekhari, Mehdi et al. (2008). "Eliciting transparent fuzzy model using differential evolution". In: *Applied soft computing* 8.1, pp. 466–476.

FAO and UNEP (2020). "Forests, biodiversity and people". In: *The State of the World's Forests 2020*. DOI: https://doi.org/10.4060/ca8642en.

Feng, Sha et al. (2016). "Los Angeles megacity: a high-resolution land–atmosphere modelling system for urban CO 2 emissions". In: *Atmospheric Chemistry and Physics* 16.14, pp. 9019–9045.

Gao, Bo-Cai (1996). "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space". In: *Remote sensing of environment* 58.3, pp. 257–266.

Gavin, Henri P (2019). "The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems". In: *Department of Civil and Environmental Engineering, Duke University* 19.

Giglio, Louis et al. (2016). "Collection 6 modis burned area product user's guide version 1.0". In: *NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA*.

Gourdji, Sharon M et al. (2022). "A modified Vegetation Photosynthesis and Respiration Model (VPRM) for the eastern USA and Canada, evaluated with comparison to atmospheric observations and other biospheric models". In: *Journal of Geophysical Research: Biogeosciences* 127.1, e2021JG006290.

Guillory, Anabelle (2017). *ERA5*. URL: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

Gutenkunst, Ryan N et al. (2007). "Universally sloppy parameter sensitivities in systems biology models". In: *PLoS computational biology* 3.10, e189.

Haji, Saad Hikmat and Adnan Mohsin Abdulazeez (2021). "Comparison of optimization techniques based on gradient descent algorithm: A review". In: *PalArch's Journal of Archaeology of Egypt/Egyptology* 18.4, pp. 2715–2743.

Hersbach, H et al. (2018). *ERA5 hourly data on single levels from 1959 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS)[data set]*.

Huete, Alfredo et al. (2002). "Overview of the radiometric and biophysical performance of the MODIS vegetation indices". In: *Remote sensing of environment* 83.1-2, pp. 195–213.

Hyndman, Rob J and George Athanasopoulos (2018). *Forecasting: principles and practice.* OTexts.

Jung, Martin et al. (2006). "Exploiting synergies of global land cover products for carbon cycle modeling". In: *Remote Sensing of Environment* 101.4, pp. 534–553.

Kirschbaum, MUF et al. (2001). "Definitions of some ecological terms commonly used in carbon accounting". In: *Cooperative Research Centre for Carbon Accounting, Canberra*, pp. 2–5.

Korhonen, Kari and Göran Stahl (2020). "Criterion 1: Maintenance and Appropriate Enhancement of Forest Resources and their Contribution to Global Carbon Cycles". In: *FOREST EUROPE, 2020: State of Europe's Forests 2020.*

Kountouris, Panagiotis et al. (2018). "Atmospheric CO 2 inversions on the mesoscale using data-driven prior uncertainties: quantification of the European terrestrial CO 2 fluxes". In: *Atmospheric Chemistry and Physics* 18.4, pp. 3047–3064.

Larranaga, Pedro et al. (1999). "Genetic algorithms for the travelling salesman problem: A review of representations and operators". In: *Artificial intelligence review* 13.2, pp. 129–170.

Levenberg, Kenneth (1944). "A method for the solution of certain non-linear problems in least squares". In: *Quarterly of applied mathematics* 2.2, pp. 164–168.

Liang, S and L Xi (2012). *Chapter 16-Vegetation Production in Terrestrial Ecosystems. In" Advanced Remote Sensing"(S. Liang, X. Li and J. Wang, eds.)*

Lin, JC et al. (2011). "Attributing uncertainties in simulated biospheric carbon fluxes to different error sources". In: *Global Biogeochemical Cycles* 25.2.

Mahadevan, P et al. (2005). "Modeling Large-Scale Biosphere NEE by Integrating Satellite Images and Climate Data-Vegetation Photosynthesis and Respiration Model (VPRM)". In: *AGU Fall Meeting Abstracts.* Vol. 2005, A41C–0055.

Mahadevan, P et al. (2008). "A satellite-based biosphere parameterization for net ecosystem CO2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM)". In: *Global Biogeochemical Cycles* 22.2.

Marquardt, Donald W (1963). "An algorithm for least-squares estimation of nonlinear parameters". In: *Journal of the society for Industrial and Applied Mathematics* 11.2, pp. 431–441.

Migliavacca, Mirco et al. (2011). "Semiempirical modeling of abiotic and biotic factors controlling ecosystem respiration across eddy covariance sites". In: *Global Change Biology* 17.1, pp. 390–409.

Mohammed, Firas Ahmmed and Moamen Abbas Mousa (2019). "Applying Diebold–Mariano Test for Performance Evaluation Between Individual and Hybrid Time-Series Models for Modeling Bivariate Time-Series Data and Forecasting the Unemployment Rate in the USA". In: *International Conference on Time Series and Forecasting.* Springer, pp. 443–458.

Morris, Max D (1991). "Factorial sampling plans for preliminary computational experiments". In: *Technometrics* 33.2, pp. 161–174.

Nelles, Oliver (2001). "Nonlinear dynamic system identification". In: *Nonlinear System Identification.* Springer, pp. 547–577.

Pastorello, Gilberto et al. (July 2020). "The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data". In: *Scientific Data* 7.1, p. 225. ISSN: 2052-

4463. DOI: `10.1038/s41597-020-0534-3`. URL: `https://doi.org/10.1038/s41597-020-0534-3`.

Puiseux, Thomas (Nov. 2019). "Numerical simulations for phase-contrast Magnetic Resonance Imaging". PhD thesis. DOI: `10.13140/RG.2.2.31338.95680`.

Raich, JW et al. (1991). "Potential net primary productivity in South America: application of a global model". In: *Ecological applications* 1.4, pp. 399–429.

Rathod, Santosha et al. (2021). "Two-Stage Spatiotemporal Time Series Modelling Approach for Rice Yield Prediction & Advanced Agroecosystem Management". In: *Agronomy* 11.12, p. 2502.

Reinhart, Vanessa et al. (2021). "High-resolution land-use land-cover change data for regional climate simulations over Europe-Part I: The plant functional type basemap for 2015". In: *Earth System Science Data Discussions*, pp. 1–43.

Rocca, Paolo, Giacomo Oliveri, and Andrea Massa (2011). "Differential evolution as applied to electromagnetics". In: *IEEE Antennas and Propagation Magazine* 53.1, pp. 38–49.

Ruder, Sebastian (2016). "An overview of gradient descent optimization algorithms". In: *arXiv preprint arXiv:1609.04747*.

Saltelli, Andrea et al. (2004). "Sensitivity analysis in practice: a guide to assessing scientific models". In: *Chichester, England*.

Scheidt, Céline, Lewis Li, and Jef Caers (2018). *Quantifying uncertainty in subsurface systems*. Vol. 236. John Wiley & Sons, pp. 109–111.

Storn, Rainer (1996). "On the usage of differential evolution for function optimization". In: *Proceedings of north american fuzzy information processing*. Ieee, pp. 519–523.

Storn, Rainer and Kenneth Price (1997). "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces". In: *Journal of global optimization* 11.4, pp. 341–359.

Suhet and Bianca Hoersch (2015). *Sentinel -2 User Handbook, ESA Standard Document Issue 1 Rev 2*. URL: `https://sentinel.esa.int/web/sentinel/missions/sentinel-2`.

Team, Warm Winter and ICOS Ecosystem Thematic Centre (2022). *Warm Winter 2020 ecosystem eddy covariance flux product for 73 stations in FLUXNET-Archive format—release 2022-1 (Version 1.0), ICOS Carbon Portal*. DOI: `https://doi.org/10.18160/2G60-ZHAK`.

Transtrum, Mark K and James P Sethna (2012). "Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization". In: *arXiv preprint arXiv:1201.5885*.

Tucker, Compton J (1979). "Red and photographic infrared linear combinations for monitoring vegetation". In: *Remote sensing of Environment* 8.2, pp. 127–150.

Valade, Aude et al. (2014). "ORCHIDEE-STICS, a process-based model of sugarcane biomass production: calibration of model parameters governing phenology". In: *Gcb Bioenergy* 6.5, pp. 606–620.

Vermote, E (2015). *MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006 [Data set] NASA EOSDIS Land Processes DAAC*. URL: `https://doi.org/10.5067/MODIS/MOD09A1.006`.

Vienna, Liaison Unit (2003). "Background information for improved Pan-European indicators for sustainable forest management". In: *Ministerial Conference on the Protection of Forests in Europe. Liaison Unit Vienna, Austria*.

Waterfall, Joshua J et al. (2006). "Sloppy-model universality class and the Vandermonde matrix". In: *Physical review letters* 97.15, p. 150601.

Wu, Kai et al. (2018). "Joint inverse estimation of fossil fuel and biogenic CO2 fluxes in an urban environment: An observing system simulation experiment to assess the impact of multiple uncertainties". In: *Elementa: Science of the Anthropocene* 6.

Xiao, Xiangming et al. (2004). "Satellite-based modeling of gross primary production in an evergreen needleleaf forest". In: *Remote sensing of environment* 89.4, pp. 519–534.

Yan, Jing-Yu, Qing Ling, and De-min Sun (2006). "A differential evolution with simulated annealing updating method". In: *2006 International Conference on Machine Learning and Cybernetics.* IEEE, pp. 2103–2106.

A   Appendix: Additional Information

A.1   EVI

The Enhanced Vegetation index EVI is monitoring the vegetation and is responsive to canopy structural variations, like leaf area index (LAI), canopy type, plant physiognomy, and canopy architecture. It is designed to isolate the canopy background signal and to reduce atmospheric influences. The vegetation signal gets enhanced with an improved sensitivity in regions of high biomass (Huete et al., 2002). It is calculated as:

$$EVI = G * \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + (C_1 * \rho_{red} - C_2 * \rho_{blue})} \tag{21}$$

The coefficients adopted in the MODIS-EVI algorithm and according to Mahadevan et al., 2008 are G(grain factor) $= 2.5, C_1 = 6, C_2 = 7.5, and L = 1$. The blue band ($wavelength \rho_{blue}$: $459 - 479nm$) helps to account for atmospheric contamination, and L helps to compensate for soil background reflectance. $C_1 and C_2$ are coefficients of the aerosol resistance term to correct for aerosol influences in the red band.

A.2   LSWI

The LSWI is the Land Surface Water index. It is calculated as:

$$LSWI = \frac{\rho_{nir} - \rho_{swir}}{\rho_{nir} + \rho_{swir}} \tag{22}$$

In the LSWI the variables ρ_{nir} and ρ_{swir} are the satellite-derived reflectance in the near infrared (wavelength ρ_{nir} : $841 - 876nm$) and the shortwave infrared (wavelength ρ_{swir} : $1628 - 1652nm$) bands (Tucker, 1979). As there is a strong light absorption by liquid water in the ρ_{swir} electromagnetic spectrum, the LSWI is known to be sensitive to the total amount of liquid water in vegetation and its soil background (Gao, 1996).

A.3   ERA5

The ERA5 dataset provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. The data covers the Earth on a 30km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80km. It also includes information about uncertainties for all variables at reduced spatial and temporal resolutions (Hersbach et al., 2018, Guillory, 2017). Before getting the meteorological data from the stations, we also obtained it from the ERA5 dataset. The ERA5 variables used are: 2m temperature and the Surface solar radiation downwards. A preprocessing script was used to divide each year into separate months and store the respective .nc data files. The ERA5 data was not used as the Temperature and Photosynthetically Active Radiation data was also measured

by the flux stations. When compared for the locations of the stations this data was not very different but the station data was more precise. When later a large scale analysis is executed the ERA5 data can be used.

A.4   MODIS

The MODIS was launched into orbit by NASA on the Terra (1999) and Aqua (2002) satellites. Further information can be found in the Appendix. The instruments are designed to detect large-scale geodynamic processes, including changes in cloud cover and radiation budget, but also to provide information about the deeper atmosphere or oceans. The MODIS instrument provides high radiometric sensitivity (12 bits) in 36 spectral bands ranging from 0.4 to 14.4 µm. The response has been specially adapted to the needs of the user and responds very slightly to signals from other frequency bands. Two bands acquire with a nominal resolution of 250 m nadir, 5 frequency bands acquire with 500 m resolution and the remaining 29 bands operate with one kilometre resolution. A ±55-degree acquisition pattern at the EOS orbital altitude of 705 km covers a 2330-kilometre-wide strip, allowing the Earth's surface to be completely imaged every two days (Brandon Maccherone, 1995).

A.5   SYNMAP

The SYNMAP is a new joint 1-km global land cover product with improved characteristics for land cover parameterization of the carbon cycle models that reduces land cover uncertainties in carbon budget calculations. The overall advantage of the SYNMAP legend is that all classes are properly defined in terms of plant functional type mixtures, which can be remotely sensed and include the definitions of leaf type and longevity for each class with a tree component (Jung et al., 2006). For our area of interest, the most represented PFT, as can be observed in the map (Figure 2) is Evergreen Needle Forest (Reinhart et al., 2021; Vienna, 2003). The SYNMAP has been compared to other vegetation classification maps like the Corine land cover map (https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=mapview) in QGIS.

A.6   Warm Winter 2020, ICOS

The Warm Winter 2020 dataset is the observational data product for eddy covariance fluxes at 73 stations in the ecosystem domain, part of them outside the Integrated Carbon Oberservation System (ICOS) network, covering the period 1989-2020. The ICOS network is a European-wide greenhouse gas research infrastructure. ICOS produces standardised data on greenhouse gas concentrations in the atmosphere, as well as on carbon fluxes between the atmosphere, the earth and oceans. The processing of the tower data has been done using the ONEFlux processing pipeline (https://github.com/icos-etc/ONEFlux).

A.7   Gradient Descent

A disadvantage of this procedure is the slow convergence, compared to other algorithms like the Gauss-Newton (GN) principle. Convergence is also dependant on the step size. The Gradient Descent is related to the hill climbing algorithm. It works with any finite number of dimensions. It is one of the most common optimisation techniques and served as base for other approaches (Ruder, 2016). Also, in the first but also current neural networks, the

optimisation of neurons in the backtracking step was based on Gradient Descent Du et al., 2019.

Table 13: *Derivatives $f'(\theta)$ of the loss function of RECO and GEE with respect to the parameter θ to be optimized. n is the length of the dataset.*

Parameter	Derivative
α (RECO)	$D_\alpha = \frac{-2}{n} \sum_{i=0}^{n} (T_i * (RECO_i - R\hat{E}CO_i))$
β (RECO)	$D_\beta = \frac{-2}{n} \sum_{i=0}^{n} ((RECO_i - R\hat{E}CO_i))$
λ (GEE)	$D_\lambda = \frac{-2}{n} \sum_{i=0}^{n} \frac{T_{scale} * W_{scale} * P_{scale} * EVI * Rad}{1 + (Rad/Rad_0)} * (-1) * (GEE_i - G\hat{E}E_i)$
Rad_0 (GEE)	$D_{Rad_0} = \frac{-2}{n} \sum_{i=0}^{n} \frac{\lambda * T_{scale} * W_{scale} * P_{scale} * EVI * (Rad^2)}{(Rad + Rad_0)^2} * (-1) * (GEE_i - G\hat{E}E_i)$

A.8 Gauss- Newton Method

In the Newton method, the learning rate/step size γ gets replaced by $\frac{1}{f''(\theta_t)}$, so 1 over the second order derivative of our function. This can be rewritten as the inverse of the hessian matrix: $\nabla^2 f(\theta_t)^{-1}$. The hessian matrix is the matrix of second order partial derivatives of $f(\theta)$ evaluated at θ.

$$\mathbb{H}_{i,j} = \nabla^2 f(\theta) = \frac{\partial^2 f}{\partial\theta_i \partial\theta_j} \tag{23}$$

Where $i, j = 1...n$ and $n = 4$ as it is the amount of our parameters. The Equation for updating all the parameters θ in timestep t to θ_{t+1} with the Newton method is:

$$\theta_{t+1} = \theta_t - \nabla^2 f(\theta_t)^{-1} * \nabla f(\theta_t) \tag{24}$$

The Newton method has quadratic convergence and is sensitive to the initial point. Moreover, it has a high scalability, which in our case is probably not a problem as we only have four parameters. The Gauss- Newton method, is similar with the advantages that it only works with Jacobian matrices (matrices of first order partial derivatives) and therefore avoids the calculation of the second order partial derivatives.

$$\mathbb{J}_{i,j} = \nabla f(\theta) = \frac{\partial f}{\partial\theta_i} \tag{25}$$

Where $i, j = 1...n$ and $n = 4$. The Hessian matrix of a function $f(\theta)$ is the Jacobian matrix of the gradient of the function $\mathbf{H}(f(\theta)) = \mathbf{J}(\nabla f(\theta))$. After further small transformations this leaves us with the Gauss- Newton method iteration step.