Student thesis series INES nr 320

Assimilation of satellite data and in-situ data for the improvement of global radiation maps in the Netherlands.

Jurgen van Tiggelen

2014
Department of
Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Jurgen van Tiggelen. Assimilation of satellite data and in-situ data for the improvement of global radiation maps in the Netherlands
Master degree thesis, 30 credits in *Geomatics*
Department of Physical Geography and Ecosystems Science, Lund University

Lund University

Thesis

KNMI

Assimilation of satellite data and in-situ data for the improvement of global radiation maps in the Netherlands.

Author:
Jurgen van Tiggelen

June 25, 2014

Contents

Preface

This report is written as a thesis project from the Lund University. The research itself is done at, the Royal Netherlands Meteorological Institute (KNMI) in the Bilt, in the Netherlands.

The focus in this report will be on mapping global radiation. However, this research is part of a larger research, namely: "High resolution climatology based on integrated in-situ observations, satellite observations and model data.", within the KNMI which covers several climatic topics. The goal of the bigger research at the KNMI is to improve the existing climatological and meteorological products. This is done by increasing the resolution through the addition of external data sources or different modeling methods.

In this report, satellite data is used as an external data source in combination with the existing measurement network.

Acknowledgement

Hereby I would like to send out a thanks for all the people that helped me during my master and especially for the writing of this thesis.

First I would like to thank the people working at the KNMI for giving me the opportunity to write my thesis within this organization. A thanks goes out to the I-RD department where I was assigned to and got a lot of feedback and advice. I would specially like to thank my supervisor Raymond Sluiter for all the advice and tips on writing my thesis and motivating me to keep looking for solutions or different angles to look at problems. Another thanks goes out to Jan Fokke Meirink (KNMI) and Richard Müller (CM-SAF) for providing me with excellent satellite data and feedback on the models and algorithms used. Finally I would like to thank all my friends and family for supporting me through my entire study! Especially Patrizia Vollmar whom I worked with a lot and was always ready to help me during the entire master!

Thank you!

Abbreviations

AOD	Aerosol Optical Depth	**MAPE**	Mean Absolute Percentile Error
AOT	Aerosol Optical Thickness	**MB**	Mean Bias
AU	Astronomical Unit	**ME**	Mean Error
AVHRR	Advanced Very High Resolution Radiometer	**MEmean**	Mean Error mean
BSRN	Baseline Surface Radiation Network	**MetOp**	Meteorological Operational Satellite
CM-SAF	Climate Monitoring Satellite Application Facility	**MSG**	Meteosat Second Generation
COT	Cloud Optical Thickness	**NIR**	Near InfraRed
CPP	Cloud Physical Properties	**NN**	Nearest Neighbour
DEM	Digital Elevation Model	**NOAA**	National Oceanic and Atmospheric Administration
DU	Dotson Unit	**NRMSE**	Normalized Root Mean Squared Error
ER	Electromagnetic Radiation	**OK**	Ordinary Kriging
ESA	European Space Agency	**RMSE**	Root Mean Square Error
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites	**RMSEsd**	Root Mean Square Error standard deviation
GERB	Geostationary Earth Radiation Budget	**RSS**	Rapid Scan Service
GG	Asymmetric Parameters	**RTM**	Radiative Transfer Model
GOES	Geostationary Operational Environmental Satellite	**SAF**	Satellite Application Facility
GSIE	Geospatial Interpolation Environment	**SEVIRI**	Spinning Enhanced Visible and InfraRed Imager
HRV	High Resolution Visible	**SICCS**	Solar Insolation under Clear and Cloudy Skies
IB	Interpolated Bias	**SIS**	Shortwave Incoming Solar radiation
IDW	Inverse Distance Weighting	**SPOT**	
IWV	Integrated Water Vapour	**SSA**	Single Scattering Albedo
KED	Kriging with External Drift	**SZA**	Solar Zenith Angle
KED-EXP	Kriging with External Drift, Exponential	**TOA**	Top Of Atmosphere
KED-SPH	Kriging with External Drift, Spherical	**TPS**	Thin Plate Splines
KNMI	Royal Netherlands Meteorological Institute	**VIS**	Visible
LSA-SAF	Land Surface Analysis Satellite Application Facility	**WMO**	World Meteorological Organisation
LUT	Look Up Table		

Summary

For this research, two satellite products were used to see if it was possible to improve the resolution and quality of the global radiation interpolation in the Netherlands. The first data source was from the Climate Monitoring Satellite Application Facility (CM-SAF). The second data source was the Surface Insolation under Clear and Cloudy Skies (SICCS) from the KNMI. Both products were available for the period of January 2006 to December 2011 and came in the form of images with monthly and daily averages.

To combine the satellite images with the input provided by the KNMI's 32 measurement stations, these interpolation/merging methods were used:

1. Thin Plate Splines (TPS)
2. Mean Bias interpolation (MB)
3. Interpolated Bias interpolation (IB)
4. Kriging with External Drift, Exponential model (KED-EXP)
5. Kriging with External Drift, Spherical model (KED-SPH)

All these methods made use of the in-situ measurements as main input for the interpolation and all methods except TPS used the satellite products as auxiliary data.

Interpolations were made for the average of the six year period and on monthly measurements, for each month, in each year. Daily interpolations were made for April 2010 until July 2010.

Different validation methods were used to analyze the output. The results showed that; for the six year average both products and all interpolation methods did a good job on predicting global radiation. The R^2 was lowest for the IB on the CM-SAF product with a value of 0.19. However, the MAPE (mean absolute percentage error) did not exceed 1.39% on the CM-SAF product and 1.42% on the SICCS product. These values corresponded with an absolute bias of 1.77 W/m^2 and 1.8 W/m^2.

The monthly results showed similar results. The R^2 values tended to differ more, especially in the IB and MB interpolation. In most cases this could be explained by the quality of the satellite images. The MAPE was low in all cases.

A maximum MAPE of 8.38% was found (when using proper satellite images), in November, which corresponded with an absolute bias of $\pm$ 4 W/m^2. Datasplitting returned similar results. MAPE's did increase up to 9.27% when leaving out 1/4th of the measurement stations but this value corresponds with an absolute bias of $2.71 W/m^2$. These low absolute errors showed that all interpolation methods return an accurate interpolation. However, because the interpolation methods rely on the quality of the satellite images, the SICCS product would be a better product. These images were complete in all months while the CM-SAF product lacked data in December.

Since it turned out that all interpolations perfored well, daily data was analyzed for the period of April until July 2010.

It turned out that for the daily data KED and the IB interpolations performed significantly better than the TPS or MB interpolation. The biggest average MAPE was found for the TPS method (10.7% in May). The smallest average error of 0 % was found for the IB method. However this method was paired with very low R^2 values which made the model unpredictable. The average KED R^2 and MAPE ranged from 0.57 to 0.75 and from 0.08% to 0.95 %. This made the method a stable and accurate interpolation method. The satellite images on their own would not be good enough to use directly as a global radiation map, for this time interval. The over- and underestimated bias of the satellite images ranged from -89.63 to 64.49 W/m^2. This showed that, a combination of station data and satellite data would improve the quality and resolution of daily global radiation maps.

1 Introduction

Global radiation is the main driver of nearly every dynamic process on Earth. It drives both air and ocean circulations, thereby influencing weather and climate. It has a direct climatic impact and it is the main energy source for nearly all life on Earth. Therefore it is fundamental to understand and keep a good record of global radiation measurements. With proper knowledge and a long term well preserved database with global radiation measurements, it is possible to get a better understanding of the climate system and possibly find solutions to mitigate global warming. At the same time, global radiation measurements can be used to satisfy request from agricultural-, medical-, biological-, industrial- and energy-sectors (Greuell et al., 2013; Kipp & Zonen; WMO, 2008).

Global radiation data has been recorded in the Netherlands since 1957 by means of a network of meteorological stations. The measurements started with only one meteorological station, located in De Bilt, the Netherlands, in 1957. Since that time, the amount of meteorological stations has increased to 32.

At the same time, there is an increasingly growing demand of high resolution global radiation maps, both internal within the Royal Netherlands Meteorological Institute (KNMI), as external. An increasing number of applications make use of high resolution data. Therefore current low resolution maps should be improved. The current resolution of global radiation maps in the Netherlands is determined by the density and distribution of the existing stations. Using station data and interpolating this data has some drawbacks. The density of the ground stations is often lower then the pixel density of satellite images. Also, due to the limited number of stations and the distribution of these stations, spatial patterns cannot always be described. New possibilities should be explored to increase the spatial resolution (Greuell et al., 2013).

Currently, in-situ measurements, satellite observations and model outputs are all treated separately. The integration of these data sources could lead to an improvement in the resolution of the radiation maps and should therefore be explored.

Several studies have already shown that both polar orbiting as well as geostationary satellites have the possibility to improve the potential of global radiation mapping. Polar orbiting satellites have the advantage of being able to provide data with a spatial resolution of several meters as source data compared to geostationary satellites which provide data with a spatial resolution of 1 to several kilometers. On the downside, temporal resolution of these satellites can be rela-

tively low. Especially when one compares the temporal resolution of a polar orbiting satellite with that of a geostationary satellite, like Meteosat second generation (MSG). MSG provides the user with data on a 15 minute temporal resolution with a spectral resolution of 12 channels ranging from 0.635 μm to 14.4 μm (Schmetz et al., 2002). Perez et al. (1994, 1997) have already proven that satellite data from a geostationary meteorological satellites becomes more accurate than local ground measurements if the distance to the station exceeds 40 to 50 km and in some cases even 34 km. With the stations in the Netherlands lying 6 km to 60 km apart, this could lead to data improvements. Greuell et al. (2013) have also created a model able to successfully capture global irradiance at the ground with a 15 minute interval using the Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument on the Meteosat second generation satellite. Noia. M., et al., has evaluated several techniques to get global radiation data from satellite images. Both using physical as well as statistical models. (Noia. M., et al. 1993). Statistical as well as physical models have the possibility to improve the resolution of global radiation mapping compared to the mapping of radiation by using meteorological stations only. Journée. et al. have proven that combining satellite data with ground measurements by using Kriging or an interpolated bias leads to superior maps compared to using either one of those sets on its own. (Journée et al. 2010, Journée et al. 2012).

Aim of the research:

"The aim of this report is to see if it is possible to improving the existing global radiation maps in the Netherlands by assimilating in-situ observations with MSG satellite measurements."

Objectives of the research:

Objectives are the following research questions:

1. Is Kriging with external drift the best interpolation method, as expected?
2. Is the physically detailed Surface Insolation under Clear and Cloudy Skies (SICCS) product a better auxiliary data source than the Climate Monitoring - Satellite Application Facility (CM-SAF) product?

To test and answer these question, three data sets were used. The main data source were the measurements made by the KNMI's meteorological stations. This data was used as the main input for interpolation. It was also be used to validate the output by for instance cross-referring with a "leave one out cross validation" technique. The second data set that was used was irradiance data obtained by MSG. This data set was created by the climate-monitoring satellite application facility (CM-SAF). The final data source that was used was another irradiance data set. Data from this Surface Insolation under Clear and Cloudy skies (SICCS) model was also obtained by the MSG satellite. The difference with this data set however is that the SICCS used a more detailed physical model; taking more cloud physical properties and atmospheric properties into account. This model is used by the KNMI and is designed and created by Greuell, W., Meirink, J.F. and Wang, P. (Greuell et al., 2013).

The organization of this report is as follows. Chapter 1 will continue with an introduction of a related study, explaining what other researchers have done. Afterwards, a notification is given for fields that could benefit from better global radiation data. In chapter 2, the background information and data which was used is described. In chapter 3, the methods are described, starting with a description of which data was used, followed by an overview of how this data could be combined. After this, the interpolation and validation methods are described. In chapter 4, the results will be represented. In chapter 5, a summarizing discussion will be given, followed by a conclusion. Finally, chapter 6 will give an outlook on what further possibilities could be explored in relation to this research.

Due to the fact that different interpolation methods have to be compared with each other, both statistically and visually, metadata is not provided in the example maps due to the limited space available on the paper. Therefore, the research area within Europe and the metadata for all the maps of the Netherlands is represented in appendix A.

1.1 Related study

Improving the spatio-temporal distribution of surface solar radiation data by merging ground and satellite measurements (Journée et al., 2012).

In this research, Journée and Bertrand (2010), from the royal meteorological institute of Belgium, combined in-situ data with data from Meteosat. They made use of different algorithms and functions provided by two Satellite Application Facility's (SAF). First they created a cloud mask by using the MSG/SEVIRI cloud detection algorithm. The surface albedo came from the LSA-SAF, which provided a near real-time surface albedo product. Other parameters like transmittance, water vapour, ozone and aerosols were taken from other models or measurements. Once all parameters were known and the transmittance for a specific atmospheric state was determined, the solar surface irradiation could be calculated. This was done by multiplying the transmittance with the extraterrestrial incoming solar flux density. Another approach used in this research was to extract the transmittance of a given atmospheric state from a look up table and multiplying this value with the extraterrestrial incoming solar flux density. Other parameters like ozone and water vapour concentrations were still used from external sources. Several methods were used by Journée and Bertrand to combine the data. They came to the conclusion that out of these methods, Kriging with external drift gave the best results. They also pointed out that the process of combining satellite data with in-situ data for an area with a relative dense network of measurement stations and a relative high cloud frequency could improve previously obtained data.

1.2 Fields of interest

Solar energy

The energy provided by the Sun is so abundant that the Earth receives energy at a rate that is 10.000 greater then mankind consumes it (Arvizu et al., 2011). It is an energy source which is available in every country in the world and besides the emissions present at the production of solar energy collectors, CO_2 and other greenhouse gas emissions are very low for the use of direct solar energy. With new solar technologies and an increase in the use of photovoltaic solar energy, thermal power plants, passive solar heating/cooling and daylight systems, proper solar data is required (Journée et al., 2012). Therefore to make efficient use of this energy source, it is important to know how much energy the Sun provides at each location by measuring the total irradiation (Arvizu et al., 2011). A map providing information about the amount of potential solar energy is the basic essential information for solar power designing (Otani et al., 1994). That is a reason why it is important for (solar)energy companies to get a good and relatively high resolution map of solar irradiation. With this information it becomes easier to pinpoint locations that would have the optimum benefit from the Suns energy.

Other sectors

There are many fields that could benefit from global radiation data. Here, a few of these fields will be shortly mentioned to get an impression where the datasets from this research can be used for.

One of the fields that could defenitly benefit from high resolution global radiation maps are the nature and agricultural sectors. "Solar radiation in the visible region of the spectrum affects the growth rates of crops, and it is used in numerical models to estimate soil moisture, potential evapotranspiration and photosynthesis." (Tarpley, 1979). Besides crop growth rates other properties of landscapes and soils are affected by global radiation.

One of the operational products that the KNMI delivers to one of its associates is the evaporation product of the Netherlands. For the KNMI evaporation product, Makkink is used (Hiemstra et al., 2011). Makkink (a calculation method used in the Netherlands for evaporation) evaporation uses the shortwave incoming solar

radiation and the mean daily temperature as input parameters. From these two the radiation parameter is the most important one and has the largest influence on the evaporation (Hiemstra et al., 2011). Therefore high resolution radiaton data could also lead to higher resolution evaporation data.

Another product that makes use of global radiation data as input is the Sunshine duration product (Greuel et al., 2013). Radiation is used in an algorithm to calculate the hours of Sunshine at a location within a specific time interval. There are quite some models that also work the other way around and calculate global radiation by the use of Sunshine duration as input. The Angström Prescott equation is an example where Sunshine duration is used to calculate global radiation (Yorukoglu et al., 2005).

2 Radiation

2.1 General description

This paper only deals with electromagnetic radiation (ER). ER is a form of radiant energy and the most important mode of transportation of energy in the Earths system. The source of the Earths energy is the Sun. The Sun emits ER to the Earth true a vacuum with the speed of light. In general the amount of energy received at the top of the atmosphere (TOA) by a surface perpendicular to the Sun is set to be 1.367 W/m^2, which is the value recommended to use by the world radiometric center (Huashan et al., 2010) . The temperature of the Sun is about 5.800 Kelvin, leading to a solar spectrum with wavelengths between 0.25 μm to 3 μm (Arvizu et al., 2011).

2.1.1 Global radiation

The maps that were created for this research deal with global radiation. Radiation received by the Earth's surface can be divided in two different types: direct radiation (also referred to as beam radiation) and diffuse radiation. Direct radiation, is radiation that is directly received by the Sun. Diffuse radiation, is radiation that is indirectly received. For example, radiation from the Sun can be reflected and scattered by clouds and other molecules or particles in the air before reaching the sensor. Global radiation is the sum of both direct and diffuse radiation received on a horizontal surface under a solid angle of 2π steradian. The unit used for global

radiation is the amount of energy per time unit per area (W/m^2 or $Js{-}1/m{-}2$).

Extraterrestrial solar radiation reaches the TOA. The fraction that infiltrates the atmosphere interacts with molecular gases, aerosols and cloud droplets. The electromagnetic radiation can be reflected upon the interacting and be redirected back to space. It can be absorbed or it continues towards the Earths surface. Once it reaches the Earth's surface it can either be absorbed by it or it can yet again be reflected back towards space. The fraction of the amount of radiation that is reflected back into space is referred to as the Albedo. This leads to the fact that the radiation leaving the atmosphere, which is captured by the satellite, consists of the back-scatter from the particles and molecules in the atmosphere and the fraction of radiation reflected by the Earths surface (Noia et al., 1993). An overview of these interactions is given in figure 1.

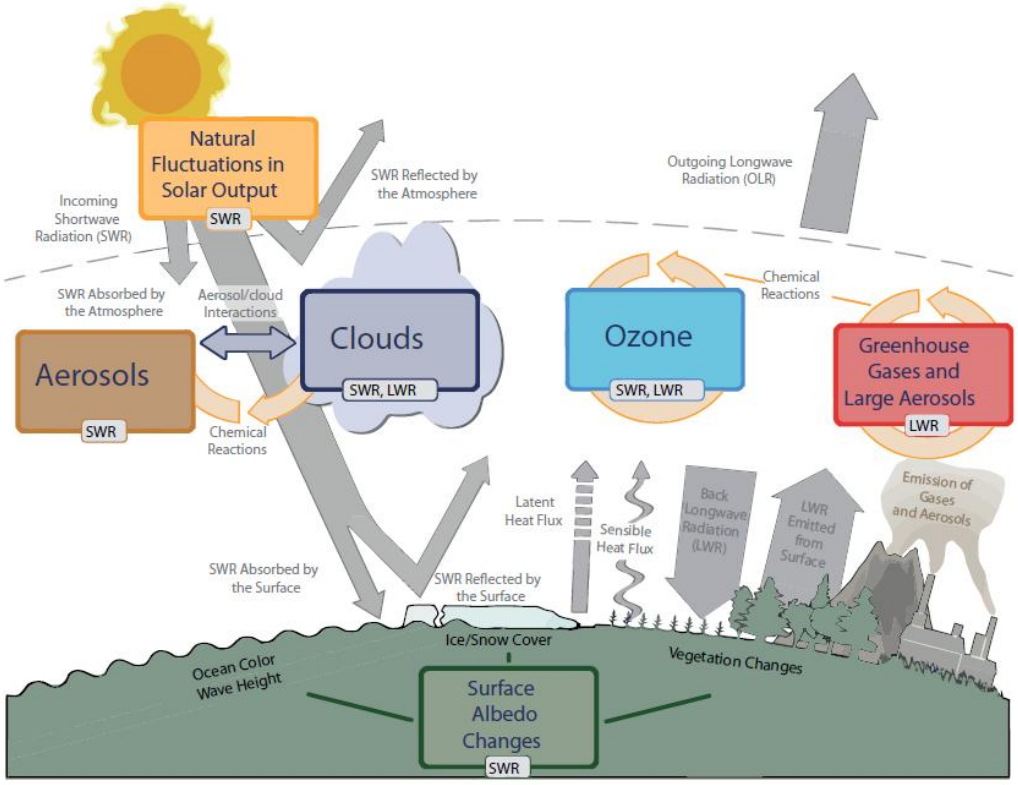


Figure 1: Overview of what happens to radiation when interacting with the atmosphere and Earth (Cubasch et al., 2013).

These interactions lead to the energy balance in equation 1. This can be calculated for each pixel individually.

$$\mathrm{IE}_{\downarrow}(\mathrm{i,j}) = \mathrm{IE}_{\uparrow}(\mathrm{i,j}) + \mathrm{E_A}(\mathrm{i,j}) + \mathrm{E_G}(\mathrm{i,j}) \tag{1}$$

Where:

- $IE_{\downarrow}(i,j)$ = The flux density incident on TOA at location (i,j) in W/m^2.
- $IE_{\uparrow}(i,j)$ = The flux density of radiation leaving the atmosphere measured by the satellite at location (i,j) in W/m^2.
- $E_A(i,j)$ = The fraction of $IE_{\downarrow}$ absorbed by the atmosphere at location (i,j) in W/m^2.
- $E_G(i,j)$ = The fraction of $IE_{\downarrow}$ absorbed by the surface at location (i,j) in W/m^2.

The flux density incident on the atmosphere is dependent on the distance between the Earth and the Sun and the zenith angle of the Sun. The incident flux density can be calculated with equation 2, for each individual pixel.

$$\mathrm{IE}_{\downarrow}(\mathrm{i,j}) = \mathrm{F_{es}} \left(\frac{\mathrm{r_o}}{\mathrm{r}}\right)^2 \cos\theta(\mathrm{i,j}) \tag{2}$$

Where:

- $IE_{\downarrow}(i,j)$ = The flux density incident on TOA at location (i,j) in W/m^2.
- F_{es} = The solar constant $\simeq 1367\ W/m^2$.
- r_o = The mean distance between the Earth and the Sun in AU.
- r = The real distance between the Earth and the Sun in AU.
- θ = The Sun's zenith angle at location (i,j).

The mean distance between the Sun and the Earth is set to 1 AU (astronomical unit) which is equal to 149.597.870.700 meter. The real distance between the Earth and the Sun can be calculated with a relative easy equation, 3. There are more complicated and precise equations to calculate the distance between the Earth and the Sun, however this equation is adequate and accurate enough for most engineering calculations (Duffie et al., 1991).

$$\mathrm{r} = 1 + 0.033\cos\left(\frac{360\mathrm{n}}{365}\right) \tag{3}$$

Where:

- r = The real distance between the Earth and the Sun in AU.
- n = The day of the year (1-365) with January 1^{st} being 1.

When radiation reaches the surface, it can either be reflected or be absorbed. To calculate the amount of radiation that is absorbed by the surface (E_g) equation 4 can be used. This equation takes the incoming radiation at the surface and multiplies it with the fraction that is not reflected due to the surface albedo.

$$\mathrm{E_g(i,j)} = \mathrm{IG(i,j)}(1 - \mathrm{A(i,j)}) \tag{4}$$

Where:

- E_g(i,j) = The fraction of $IE_{\downarrow}$ absorbed by the surface.
- IG = Solar radiance at ground level.
- A(i,j) = The Albedo of the ground at location (i,j), varying between 0 and 1.

The solar radiance at ground level can either be calculated with the help of equation 5 or it can be measured by a Pyranometer.

$$\mathrm{IG} = \frac{1}{1 - \mathrm{A}}[\mathrm{IE_{\downarrow}} - \mathrm{IE_{\uparrow}} - \mathrm{E_A}] \tag{5}$$

With knowledge about the incoming radiation at TOA and incoming radiation at the ground it is possible to determine the transmittance of the atmosphere with equation 6.

$$\mathrm{T} = \frac{\mathrm{IG}}{\mathrm{IE_{\downarrow}}} \tag{6}$$

Where:

- T = The fraction of transmittance of the atmosphere.

Now it is possible to calculate most of the incoming and outgoing radiation fluxes. However it is important to know that all these fluxes get influenced by several factors. The most important factor influencing these fluxes is the zenith angle of the Sun. This directly influences the air mass the radiation has to interact with and thereby influences the absorption and scattering of radiation in the atmosphere. Since it is possible to calculate the solar zenith angle the first factor influencing the flux measured by the satellite can be corrected. Another factor influencing the flux measured by the satellite and the incident flux are clouds. Water droplets and ice particles in clouds have a big influence on the absorption and scattering of solar radiation. Cano et al. (1985), Otani et al. (1994), Diabaté et al. (1989) and several other researchers assume that the cloud cover over a certain area statistically determine the amount of incoming radiation at that location. If clouds can be detected by the satellite it is possible to determine their impact by comparing insolation under clear sky conditions with those of cloudy conditions. It is possible to distinguish clouds on satellite images, since clouds tend to have a high fraction of reflectivity in the visible solar spectrum (much higher then the Earth's surface if not covered with snow or ice)(Cano et al., 1986; Diabaté et al., 1989; Noia et al., 1993; Otani et al., 1994).

2.1.2 physical patterns

Several researches have concluded that clouds are one of the main factors that determinte the amount of incoming global radiation (Cano et al., 1986; Diabaté et al., 1989; Tovar et al., 2001). This means that the formation of clouds over the Netherlands will results in lower radiation at certain locations and higher radiation in other places. One factor influencing cloud formation are big lakes and other waterbodies. In general, during the winter period more clouds are expected above waterbodies due to the relative warmer temperatures compared to the surrounding land. In the summer period this pattern is the other was around and more clouds are expected to form over land (Ackerman et al., 2013). This would result in a relative cloud free shore in the Netherlands during the summer.
Another factor influencing cloud formation is land cover and vegetation. Vegetation processes such as respiration release water vapour which is needed for the formation of clouds (Nc-climate). This means that in the Netherlands more clouds are expected above vegetated areas. This is true especially in the summer months due to higher respiration and vegetation activities. The Veluwe area (appendix G) for example will therefore experience more clouds and less incoming global

radiation.

2.2 Methods to capture global radiation from satellites

A method for the determination of the global solar radiation from meteorological satellite data. (Cano et al.,1986)

Cano et al. have developed a statistical approach using satellite data and training data from the ground to determine the parameters of the regression model that is used to predict the global radiation. Satellite data is taken from the geostationary satellite Meteosat. They made use of both the visual as well as the thermal infrared spectrum. In a first step albedo map of a cloudless sky is created over the research area. This is done by taking the pixels with the lowest value out of time series of satellite data. This can be done because in general clouds will reflect more radiation then other surfaces (except snow and some desert soils). In a second step the cloud cover index is computed. This index is computed by taking the original satellite image and extracting the ground albedo. What is then left is the reflectance from the clouds which can be standardized to become a value from 0 to 1 indicating the percentage of clouds covering a pixel. With this knowledge and data measured by stations the atmospheric transmittance has been calculated for each pixel. These values range from 0.2 to 0.8. With these known parameters Cano et al. created a model to predict global radiation at any pixel of the satellite image with success.

Description of an operational tool for determining global solar radiation at ground using geostationary satellite images. (Diabaté et al., 1989)

Diabaté et al. have delivered the Helion station. This is a relative cheap package of software and hardware that will calculate global radiation at ground level by using geostationary satellite images. They used the same method as Cano et al. (1986) which has been proven to be efficient. They took a satellite image and used clear sky conditions to create a reference albedo map from the surface. In a second step the seasonal variation in albedo was taken into account. When this map was complete they use it as a reference map to be able to create a cloud index. To determine the transmittance of the atmosphere Diabaté et al. refer to the linear relation between the cloud cover index and the transmittance, proven by Cano et

al. (1986). With these parameters known, Diabaté et al. were able to estimate global solar radiation at the surface by taking the transmittance at a specific pixel and multiplying it with the incoming radiation at the top of the atmosphere.

Solar radiation mapping from NOAA AVHRR data in Catalonia, Spain. (Tovar et al., 2001)

Tovar et al. used a statistical approach to determine the global solar radiation in Catalonia, a location in the north-eastern corner of Spain. In this research a similar approach was used as by Diabaté et al. (1989). They assumed that the amount of cloud cover for each location (pixel) determines the global solar radiation received by that area. Therefore the first step in this research consists of detecting clouds by using a multispectral cloud detection procedure. With the results of this procedure a cloud cover index could be determined for each point (pixel) of the area. This cloud cover index was then used in a final step to statistically determine the global solar radiation model. For this global solar radiation model the transmittance from the atmosphere was used as a factor. This was determined before by looking at the relationship between total incoming radiation and radiation measured by stations at the ground. Ground measurements were also used to determine the regression coefficients of the model. The results of this research showed an excellent correlation between the estimated global solar radiation and the measurements from the stations. It resulted in a coefficient of determination which was greater then 0.98 for every case and it had a RMSE ranging from 9.6% to 15.8%. The bias varied from 1.3% to 9.5%. The research also showed that the estimated global solar radiation tended to be better if the measurement stations were more sparse (RMSE of 7%) than those who were obtained by interpolation of station data which had a RMSE of 11% to 16%.

Mapping a topographic global solar radiation model implemented in a GIS and refined with ground data. (Pons et al., 2008)

Pons et al. have computed a physically-based model to predict potential solar radiation in Catalonia, Spain and refined the data by using meteorological data. The main challenge in this research was to include the elevation of the area and to be able to obtain a potential solar radiation map with only a Digital Elevation Model (DEM) and data from meteorological measurement stations. The model took several parameters into account when predicting potential solar radiation.

These parameters were: The solar constant, the distance between the Earth and the Sun, the solar geometry, the angles of incident Sunbeams in each cell, shadows and the relation between direct and diffuse solar radiation. The methodology used in this research has shown to be valid for computing solar radiation both on a monthly and annual time scale. It has provided new maps with RMSE of 10kJ $m^{-2}day^{-1}$. The result showed an error of 7.3% in March, 6.1% in June, 6.4% in September and 13.1% in December. These results are better then those obtained by classic interpolation techniques, especially in rugged terrain like in this research. The model is still very much influenced by the quality of the DEM and the winter months need to be accurately adjusted.

Estimating incident solar radiation at the surface from geostationary satellite data. (Tarpley et al., 1979)

Tarpley et al. used a statistical approach to determine solar radiation at the surface. However compared to most other methods, Tarpley et al. used three different equations to calculate the irradiance. One equation for clear sky pixels, one equation for partially clouded pixels and finally, one for cloudy pixels. To be able to apply the right equation to the pixels, first a clear sky condition had to be computed. This was done with data captured by the satellite before the start of the time serie of this research. To compute the cloud detection process all data sets with a solar zenith angle of 85° and higher were discarded. Data with a standard deviation which was to large was also discarded. The remaining data was fit to a regression model to get a set of coefficients. If the brightness values of a picture were still greater than the predicted value, they were discarded as well. The last 2 steps were performed three times. The procedure then left a cloud free data set and reliable regression coefficients. These results were then manually checked for special conditions like mountains and lakes etc. The next step in the research was to create a cloud index to determine the amount of clouds in a pixel. As a final step the transmittance of the atmosphere was calculated. The results show that the model has an error of less than 10% of the mean. Knowledge of cloud type and thickness could further improve this model.

Retrieval and validation of global, direct, and diffuse irradiance derived from SEVIRI satellite observations. (Greuell et al., 2013)

In this research Greuell et al. created a new model to obtain solar irradiance at

the surface derived from SEVIRI imagery on board of the MSG satellite (Schmetz et al., 2002). The main input data used for this research were; a cloud mask and cloud physical properties, which were both obtained by SEVIRI observations. The Surface Insolation under Clear and Cloudy skies (SICCS) model works with a physical based algorithm. This means that it does not use any ground control points for the determination of the surface insolation. Instead, surface insolation was obtained by using a detailed radiative transfer model. SICCSs algorithm consists of two parts. The first part is the input and describes the state of the atmosphere. The second part is the algorithm itself which calculates radiative transfer of the atmosphere based on the input parameters. When comparing the hourly results obtained by the SICCS with eight stations that measured radiation on the ground the following results were received. The median values of the station biases was 6W/m2 (5%) for direct irradiance and 1W/m2 (1%) for diffuse irradiance. The global irradiance had an bias of 7W/m2 (2%).

2.2.1 Global radiation data challanges

In general there are several problems when comparing ground data with satellite data. The first problem is that you have to find the Pyranometer location on the satellite image. This can be complicated due to the relatively high latitude location of the Netherlands compared to the centre of the area observed by MSG. The high latitude changes the resolution of the pixels and the viewing angle making it harder to pinpoint the exact location of the Pyranometers. The second problem is the difference of what is measured. The satellite measures radiation over a small solid viewing angle. The Pyranometer measures radiation over a solid angle of 2 π. Another small possible problem is the time scale at which the data is captured. However with MSG capturing data every 15 minutes this problem is close to resolved. The biggest challenge is coming from clouds and atmospheric distortion. As stated before water droplets and ice crystals have an influence on the scattering and absorption of radiation in the atmosphere. Aerosols are another type of particles in the atmosphere that have a similar effect on radiation. All of these factors have to be accounted for, making equation 1 to 6 more complicated. The principle of these equations still stands however, explaining the differences between each location will be harder to explain.

Another problem with clouds is the so called cloud parallax effect. The Sun does not stand directly above the area of interest. This produces difficulties with clouds in respect to the viewing angle and the location on the satellite image. A

10km high cloud at 50 degrees North is displaced by more than 10km northwards in the satellite image (Journée et al, 2012).

2.3 Current situation

Before trying to improve data, it is important to take a look at the current data and how this data is obtained. In 2011 The KNMI has published its newest edition of the climate atlas. It is both available as a book (Sluijter et al., 2010) and a selection is visible on a website (www.klimaatatlas.nl). One part of this atlas consists of the global radiation maps that are open for improvements. These maps were created by taking data from official measurement stations and interpolating the missing data to create a complete map. The data that was used in the klimaatatlas is the average of the 30 year period from 1981 to 2010. An example of a global radiation map is given in figure 2. The global radiation data is currently being used as an input to compute Makkink evaporation. This dataset is operational and is being deliverd to and used by Rijkswaterstaat. Rijkswaterstaat is a part of the Dutch Ministry of Infrastructure and the Environment and responsible for the design, construction, management and maintenance of the main infrastructure facilities in the Netherlands.

2.3.1 Radiation measurements

Current radiation data is captured at 32 different meteorological measurement stations in the Netherlands. The locations of these stations can be seen in figure 3. The meteorological stations located at the North Sea do not capture global radiation data.

The stations that capture global radiation, use a Pyranometer. These Pyranometers are of type CM 11 manufactured by Kipp & Zonen, Delft, The Netherlands . The measurements at the stations are performed automatically every 12 seconds. Mean, minimum maximum and standard deviation levels of irradiance are computed from these 12 second interval measurements for time series of 1 and 10 minutes. In general the 10 minute interval measurement data is used. This data is the average of 50 measurements performed by the Pyranometer. The measurement resolution for average global radiation for the stations is 1 W/m^2. The thermo-electric Pyranometer can measure global or diffuse irradiation. For the latter, the pyrometer's measurement tool can be blocked from direct Sun beams by means of a shadow disk (Kipp & Zonen, 1992; KNMI, 2005).

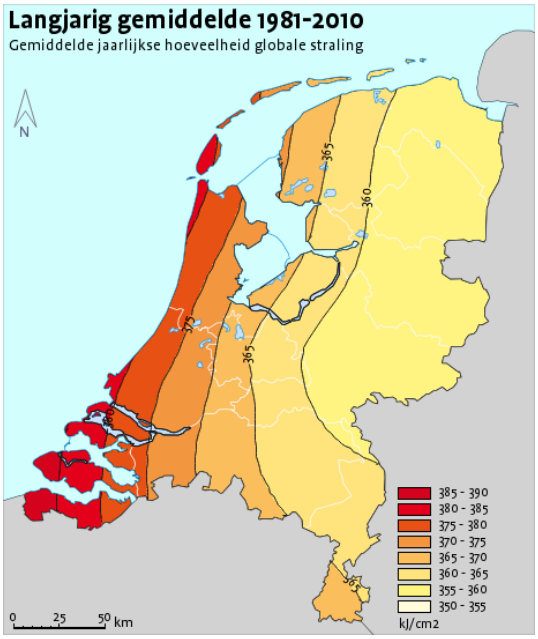


Figure 2: Longterm average 1981-2010. Average yearly global radiation. (www.klimaatatlas.nl).

.

The CM 11 works by using two heat absorbing detectors. One of these detectors is exposed to short-wave radiation the other one acts like a reference detector and its thermal state is not altered by radiation. The temperature difference between these two detectors is used to determine the amount of irradiation by converting it to a voltage. The irradiance is modelled by a linear equation, equation 7 (Kipp & Zonen, 1992).

$$\mathrm{E}_{\downarrow \mathrm{Solar}} = \frac{\mathrm{U}_{\mathrm{emf}}}{\mathrm{S}_{\mathrm{ensitivity}}} \tag{7}$$

Where:

- $E_{\downarrow Solar}$ = Global Radiation $[W/m^2]$.
- U_{emf} = Output of the Pyranometer $[\mu V]$.
- $S_{ensitivity}$ = Sensitivity of the Pyranometer $[\mu V/W/m^2]$.

Natural factors like precipitation, temperature changes, winds etc. can influence the accuracy of the measurements. Another possible source of offset is that,

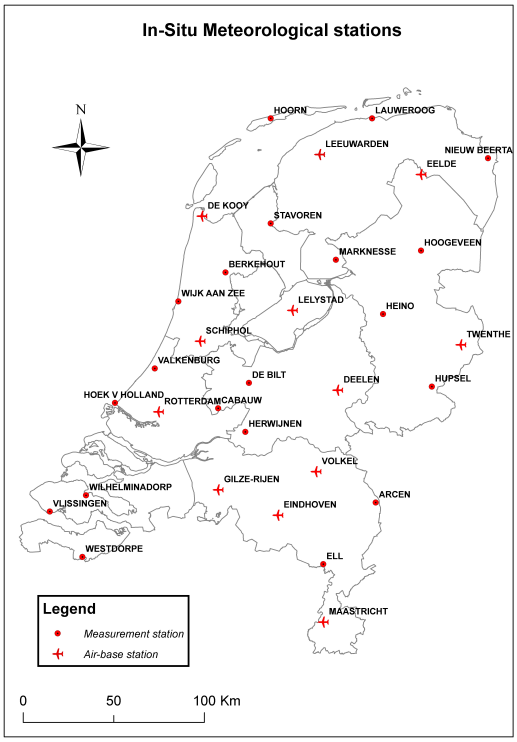


Figure 3: The 32 meteorological stations owned by the KNMI that are used in the interpolations for global radiation.

the sensors are non-selective, they absorb all radiation, both short-wave as well as long-wave radiation. Since the CM 11 is designed to accurately measure solar irradiance, long-wave radiation and environmental factors have to be blocked out. The CM 11 is therefore protected by two glass domes. These glass domes are designed in such a way that they do not interfere with direct solar irradiance and serve as a filter for shortwave radiation (Figure 4). This makes the CM 11 Pyrometer capture radiation data within a spectral interval of 285 nm to 2800 nm (Kipp & Zonen, 1992).

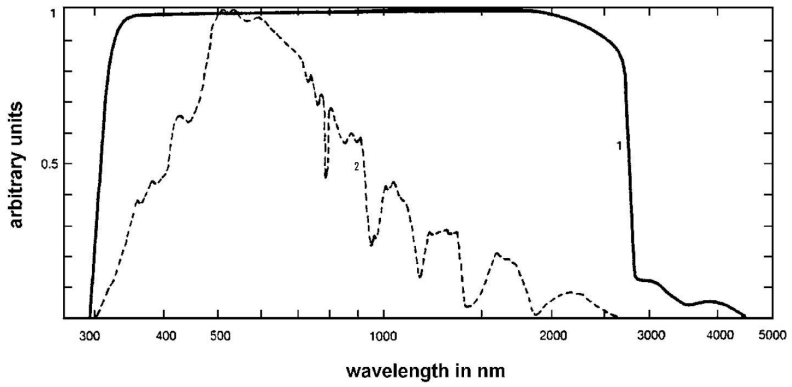


Figure 4: The thick line (1) represents the relative spectral transmittance of the two pyranometer domes of the CM11. The dotted line (2) represents the spectral distribution of solar radiation at sea level. Sun at zenith (Airmass 1). (Kipp & Zonen, 1992).

.

The World Meteorological Organization (WMO) has classified the CM 11 Pyranometer as a secondary standard device (Kipp & Zonen, 1992). This means that the device is placed in the best category according to the ISO 9060 standard. To be classified as secondary standard the WMO expects maximum errors in hourly radiation to not exceed 2 to 3% (WMO, 2008). In order to perform proper measurements and to prevent errors, maintenance of the Pyranometers is of utter importance. Therefore the Pyranometers are cleaned every half year and on airfields this happens more frequently with an average of every two months. If errors or abnormalities occur in the measurements, local maintainers can clean the Pyranometers on request. To ensure the consistency and accuracy of the Pyranometers a routine calibration is performed every 26 months at the KNMI (KNMI 2005).

2.3.2 Data processing

The 30 year averaged data obtained by stations was used to interpolate the current maps in the climate atlas. For average yearly insolation, 20 stations were used and for monthly insolation 16 stations were used. The data was taken for the period of 1981 to 2010 which is the normalized period for climatic data. The interpolations were performed in the program R, which is a language and environment for statistical computing and graphics (R-Project). Both for the monthly and for the yearly average the same R script was used. The only difference is the input data

that was used. R used the auto-map function and thin plate splines for the interpolation and mapping of the data. In the interpolation with thin plate splines, λ was set to be constant at 0.004 for all interpolations. The data was validated by cross referring the interpolated data with the measured station data by the "leave one out cross validation". The R^2 was given as an output alongside the map and information about the cross-validation. The R^2 for yearly averages was 0.51 and for monthly averages it was 0.61. The resulting maps are also evaluated by expert judgement, done by R. Sluijter, within the KNMI itself (Hiemstra et al., 2011; Sluiter, 2012).

2.4 Satellites

Satellites can be classified into two groups. Polar orbiting satellites and geostationary satellites.

Polar orbiting satellites move in a path around the world at an average height of 800 km (EUMETSAT), passing over both poles, hence the name. These polar orbiting satellites move from North to South and cross the equator at an angle of approximately 90 degrees. However each rotation around the Earth they pass the equator at a different longitude. Therefore it can take several days to months before these satellites visit the exact same location again. The biggest advantage of these satellites is the high resolution they allow data to be captured in.

Geostationary orbiting satellites are located at a specific point above the Earth (often the equator at 0 degrees latitude) at an average height of 36.000 km (EUMETSAT). These satellites move at the same speed as the rotation of the Earth. This allows the satellite's instruments to constantly capture data above a certain area. Weather satellites are geostationary in most cases. The drawback of this type of satellite is the relative low resolution the instrument captures data in compared to polar orbiting satellites (NASA).

2.4.1 Meteosat

Meteosat Second Generation (MSG) is the current generation of European meteorological geostationary satellites. They are established by the cooperation of the European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). MSG has replaced the older Meteosat first generation. The first MSG satellite was Meteosat-8 and was launched in 2002. This satellite was followed by 2 more similar satellites, Meteosat-9, launched in 2005 and the current main meteorological satellite, Meteosat-10 in

2012. Meteosat-10 is located at a position of 0° at a height of 36.000km, Meteosat-9 is located at a position of 9.5° at a height of 36.000km and Meteosat-8 serves as a backup service at a position of 3.5° and a height of 36.000km (ESA). The Spinning Enhanced Visible and Infrared Imager (SEVIRI) instruments on the MSG satellites have a temporal resolution of 15 minutes (5 minutes for Meteosat 8 and 9 which are currently set on Rapid Scan Service (RSS)). The images produced by SEVIRI have a radiometric resolution of 10bit per pixel and a spatial resolution of 3km at nadir (directly under the satellite) for all channels. SEVIRI is also equipped with a high-resolution visible (HRV) channel which has a spatial resolution of 1km (Schmetz et al., 2002).

On the ground side, the MSG program exists of several components. The central facilities are located at the EUMETSAT headquarters in Darmstadt, Germany. Here data captured by the satellite is pre-processed up to level 1.5. This means that satellite data is: corrected for differences in detector responses, compensation for non-linearity and a geometric correction is performed to put the data in a standard reference system. The rest of the ground segment consists of stations for data acquisition, data control and back-ups. Another service provided by the EUMETSAT is the distribution of data by Satellite Application Facilities (SAFs). These facilities provide end-users with data and services that are fully operational and ready to use. At this point there are 7 SAFs operational, each covering different "themes" that are related to climate monitoring (Schmetz et al., 2002). To give some examples, the OSI-SAF (ocean and sea ice-SAF), provides information about the ocean and atmosphere interaction. The O3M-SAF (ozone and atmospheric chemistry monitoring-SAF) processes data related to ozone, aerosols, ultraviolet data and other trace gases (EUMETSAT SAF).

2.4.2 Other satellites

Due to the fact that satellites measure radiance, any satellite could be used to obtain global radiation data. However not all satellites would be equally useful when it comes down to their performance. All satellites have their plus and down sides. The SEVIRI instrument on MSG has a relatively low resolution but is able to capture data every 15 minutes. This makes the satellite able to create data sets that are usable for time series. Satellites like SPOT (Satellite Pour l'Observation de la Terre) and Landsat could provide a better resolution and therefore more detailed maps (NASA, SPOT). Besides the operational Landsat and SPOT satellites, the Sentinels from the EUMENTSAT sentinel program can also provide the data which

is required. The first Sentinal was launched at the 3th of April 2014. Sentinel-2 to Sentinal-6 will be launched in the near future. Though as a downside, the temporal resolution of these satellites could be too low to create good time series and due to the lower temporal resolution it becomes harder to correct cloud influences.

MSG will in time be replaced by Meteosat Third Generation (MTG). The MTG program will consists of 6 satellites and the first one is planned to be ready for launch in 2018. The MTG will have a higher spatial, temporal and radiometric resolution compared to MSG. MTG will have 16 channels with a spatial resolution ranging from 0.5 to 2 km. The temporal resolution of MTG will be 10 minutes for a full disk scan and 2.5 minutes when set to a European regional rapid scan (EUMETSAT MTG).

The choice of satellite is therefore dependent on the wishes of the user. If the user wants an end product with high resolution on a specific time, data from SPOT, Landsat or the Sentinels might be better than MSG. However, if the user wants a result based on time series, a higher temporal resolution is needed and MSG or GOES (Geostationary Satellite system) are better data sources. MTG could become a balance between spatial and temporal resolution when operational.

3 Methods

The methods to obtain global radiation from satellites described in chapter 2 can be devided into two approaches. The first approach is a statistical approach. The statistical approach is based on the relationship between the satellite and ground station data. The second approach is a physical approach. This approach uses a radiative transfer model to determine how the satellite data and the ground station data are linked together (Tovar et al., 2001; Noia et al., 1993). In general, statistical methods work better on small areas (Cano et al., 1985). Both approaches work with a similar physical basis. In this research data was taken from two sources, both using MSG and a physical approach. The SICCS (Greuell et al. 2013) is a pure physical model not taking ground measurements into account for its calculations (except for its validation). This is a very detailed model, taking cloud physical properties into account among other factors like aerosols and water vapor etc. The second set came from the CM-SAF (EUMETSAT CM-SAF 2013). The physical model used for the CM-SAF product is less complex than the SICCS model. It accounts for less physical properties in the atmosphere.

3.0.3 Input data used

For this research three input sources were used:

1. In-situ radiation data, measured at meteorological stations by means of Pyranometers.

2. Climate Monitoring Satellite Application Facility (CM-SAF) radiation data, modelled by the climate modelling satellite application facility using MSG images.

3. Surface Insolation under Clear and Cloudy Skies (SICCS) radiation data, modelled by the KNMI using a detailed physical model using MSG images.

3.0.4 In-situ data

Monthly, yearly and daily averages of global incoming radiation were used. These values were extracted from the KNMI database and contained values of J/cm^2. These values were converted into W/m^2 to match the satellite images. First J/sm^2 was converted to J/m^2 by multiplying by 10.000. After this Joules were converted to W. This was done by using equation 8.

$$\mathrm{P(W)} = \frac{\mathrm{E(J)}}{\mathrm{t(s)}} \tag{8}$$

Where:

- P(W) = Power in Watt [W].
- E(J) = Energy in Joules [J].
- t(s) = Time in seconds [s].

In total, 32 in-situ measurement stations were used. They were spread out over the country as shown in figure 3. Radiation data at these locations was recorded according to the method explained in section 2.3.1.

3.0.5 Meteosat data

Both the CM-SAF product as well as the SICCS product used MSG images as input data for their models both products used auxiliary data from external sources.

3.0.5.1 CM-SAF: From the CM-SAF daily and monthly averages were used (montly averages are also used to compute the yearly averages).This data was obtained from the CM-SAF website (CM-PRODUCT). The start point of the time series for montly averages was January 2006 and the end data was December 2011. Daily data was obtained for the period of April 1st 2010 untill July 31st 2010.

The CM-SAF product used a Radiative Transfer Model (RTM) to compute the radiation. The instruments providing the input data for the algorithms used are: SEVIRI and the Geostationary Earth Radiation Budget (GERB) on MSG. The Advanced Very High Resolution Radiometer (AVHRR) on the National Oceanic and Atmospheric Administration (NOAA) and the Meteorological Operational (MetOp) satellites for the northern latitudes (Mueller et al. 2009).

The CM-SAF algorithm used for calculating shortwave incoming solar radiation (SIS) is based on the following underlying fundamental assumption:

$$\mathrm{SIS} = \mathrm{E_o cos}(\theta_0)\mathrm{T} \tag{9}$$

where E_0 is the incoming solar flux at TOA. θ_0 is the solar zenith angle and T is the transmissivity of the atmosphere.

This algorithm to obtain SIS is based on several look up tables (LUT). This was done to decrease the computational time of the algorithm. LUT's were constructed after running calculations in radiative transfer models (RTM) for different atmospheric compositions and states. Once the tables were computed Transmittance of the atmosphere could be extracted from these tables by simple interpolation. This transmittance was then used in the fundamental assumption as shown in the equation above (equation 9).

When computing the SIS with the CM-SAF algorithm pixels are first classified as cloud free or cloudy. Depending on the classification different approaches were used.

If a pixel was classified as cloud free, a LUT which considers: Aerosol Optical Depth (AOD), single scattering albedo (ssa), asymmetry parameters (gg), water vapour, ozone and surface albedo were used to obtain the transmittance of the atmosphere. Water vapour, Ozone and surface albedo use fixed values in the model, these values correspond with: $15 \mathrm{kg}/m^2$ for water vapour, 345 DU (Dotson Unit) ozone and a surface albedo of 0.2 (seasonal changes are not considered). The satellite image was not used any further in the process to determine SIS. The full schema used for clear sky conditions is displayed in figure 5 (Mueller et al., 2009; CM-SAF, 2013).

When a pixel was classified as cloudy a different approach had to be used since cloud albedo's are considered to determine the incoming surface irradiance. All the other input data was identical to that of the clear sky model. Surface albedo, water vapour, ozone, aerosol properties etc. were the same. However in this case the satellite image was used to determine the radiation under cloudy conditions. It did so by deriving the top of the atmosphere albedo.

The full schema used for cloudy sky conditions is displayed in figure 6 (CM-SAF, 2013).

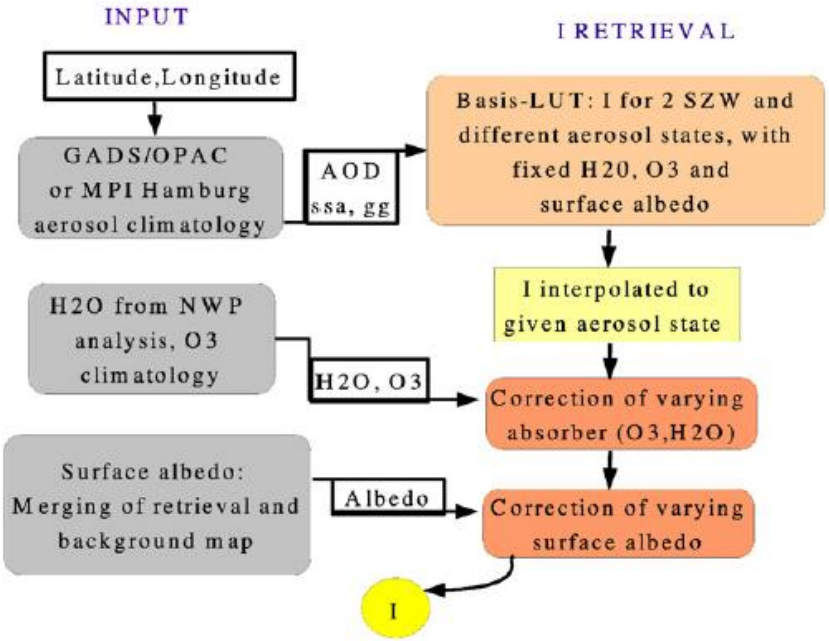


Figure 5: Schema of the clear sky conditions. I is the Solar surface irradiance) (Mueller et al., 2009).

.

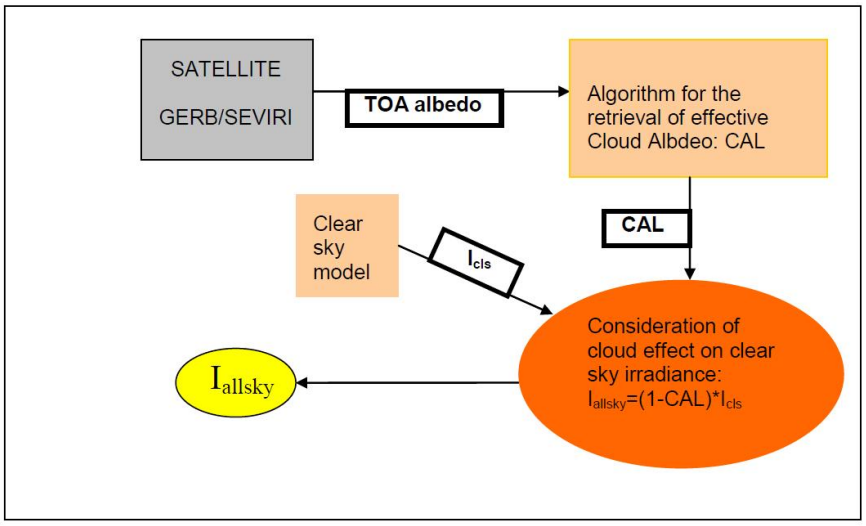


Figure 6: Schema of the cloudy sky conditions (CM-SAF, 2013).

.

To validate the results of the CM-SAF surface incoming solar radiation method, results from the model were compared with measurements of a baseline surface radiation network (BSRN). To calculate the accuracy of the monthly mean maps 12 stations were used. From these 12 stations 578 measurements were taken between 2006 and 2011 for the validation. The optimal accuracy was set to 8 W/m^2 both for the bias and the absolute bias. The bias of the results compared with the BSRN is $1.6W/m^2$ and the absolute bias is $7.2W/m^2$. Both were well below the optimal accuracy, indicating that the obtained measurements were of high quality. When looking at the quality of the daily data. The target accuracy was set to $20W/m^2$. The bias obtained by this method was $1.6W/m^2$ and has an absolute bias of $14.8W/m^2$. This indicated that the hourly obtained radiances were also well below the target, again indicating high quality measurements (EUMETSAT 2013).

Table 1 shows the statistics that were obtained by the validation with the Cabauw station in the Netherlands, which is part of the BSRN. These statistics were provided by R. Müller, one of the creators of this data set.

	Year	Month	MonthlyMeanSatellite	MonthlyMeanMeasurement	Bias.W.m².	Bias...	RMSE	RMSE.1	correlation	absBias.W.m2	absPercBias
1	2006	1	30.7	37.3	-6.57	-17.61	28.14	75.42	0.93	6.57	17.61
2	2006	2	40.6	42.9	-2.28	-5.32	39.94	93.10	0.90	2.28	5.32
3	2006	3	101.2	111.3	-10.17	-9.14	75.52	67.83	0.91	10.17	9.14
4	2006	4	166.2	170.0	-3.77	-2.22	98.43	57.91	0.91	3.77	2.22
5	2006	5	203.5	196.7	6.74	3.43	98.02	49.82	0.93	6.74	3.43
6	2006	6	256.6	236.9	19.71	8.32	94.34	39.82	0.95	19.71	8.32
7	2006	7	278.3	266.8	11.53	4.32	85.88	32.19	0.96	11.53	4.32
8	2006	8	146.4	143.2	3.17	2.22	104.86	73.24	0.87	3.17	2.22
9	2006	9	160.5	155.1	5.49	3.54	53.43	34.46	0.97	5.49	3.54
10	2006	10	55.9	61.5	-5.59	-9.10	42.62	69.30	0.94	5.59	9.10
11	2006	11	34.2	38.0	-3.86	-10.14	34.81	91.52	0.90	3.86	10.14
12	2006	12	11.4	17.7	-6.30	-35.66	23.32	132.09	0.86	6.30	35.66
13	2007	1	22.0	25.3	-3.29	-13.02	29.64	117.12	0.87	3.29	13.02
14	2007	2	40.1	44.5	-4.35	-9.78	35.51	79.86	0.91	4.35	9.78
15	2007	3	112.7	112.3	0.39	0.35	59.33	52.84	0.95	0.39	0.35
16	2007	4	222.7	219.2	3.51	1.60	65.75	30.00	0.97	3.51	1.60
17	2007	5	208.0	199.9	8.15	4.08	105.15	52.61	0.92	8.15	4.08
18	2007	6	198.2	173.5	24.70	14.24	128.06	73.83	0.88	24.70	14.24
19	2007	7	180.2	167.7	12.49	7.45	135.00	80.51	0.84	12.49	7.45
20	2007	8	192.1	187.3	4.75	2.54	95.67	51.07	0.93	4.75	2.54
21	2007	9	118.0	116.6	1.33	1.14	77.15	66.14	0.91	1.33	1.14
22	2007	10	63.5	68.3	-4.76	-6.97	45.78	67.06	0.94	4.76	6.97
23	2007	11	32.1	35.1	-3.05	-8.67	37.21	105.92	0.87	3.05	8.67
24	2007	12	18.9	25.0	-6.17	-24.62	24.09	96.20	0.90	6.17	24.62
25	2008	1	21.0	23.9	-2.87	-12.01	22.61	94.61	0.91	2.87	12.01
26	2008	2	63.6	64.9	-1.26	-1.94	37.45	57.73	0.95	1.26	1.94
27	2008	3	85.9	88.3	-2.40	-2.71	67.36	76.26	0.91	2.40	2.71
28	2008	4	181.9	169.9	12.01	7.07	100.97	59.42	0.91	12.01	7.07
29	2008	5	249.4	241.2	8.21	3.40	92.55	38.37	0.95	8.21	3.40
30	2008	6	250.6	239.2	11.47	4.80	130.24	54.46	0.90	11.47	4.80
31	2008	7	179.7	175.5	4.23	2.41	101.99	58.12	0.92	4.23	2.41
32	2008	8	165.8	164.8	1.00	0.61	96.80	58.75	0.91	1.00	0.61
33	2008	9	118.9	121.1	-2.25	-1.86	57.51	47.48	0.95	2.25	1.86
34	2008	10	74.6	78.1	-3.53	-4.51	49.27	63.09	0.93	3.53	4.51
35	2008	11	28.4	32.3	-3.95	-12.23	32.54	100.72	0.90	3.95	12.23
36	2008	12	21.6	25.6	-4.05	-15.80	27.25	106.25	0.89	4.05	15.80
37	2009	1	30.9	38.3	-7.36	-19.22	32.33	84.39	0.92	7.36	19.22
38	2009	2	45.8	41.8	3.96	9.46	36.92	88.26	0.91	3.96	9.46
39	2009	3	103.1	107.3	-4.13	-3.85	65.17	60.77	0.93	4.13	3.85
40	2009	4	174.8	168.5	6.23	3.70	66.71	39.58	0.96	6.23	3.70
41	2009	5	221.4	223.6	-2.18	-0.98	97.82	43.76	0.94	2.18	0.98
42	2009	6	251.5	244.2	7.25	2.97	121.08	49.58	0.91	7.25	2.97
43	2009	7	230.8	223.8	6.99	3.12	121.58	54.32	0.90	6.99	3.12
44	2009	8	207.0	200.1	6.88	3.44	99.83	49.89	0.93	6.88	3.44
45	2009	9	131.8	134.5	-2.72	-2.02	71.31	53.03	0.94	2.72	2.02
46	2009	10	67.6	72.7	-5.12	-7.04	46.35	63.74	0.94	5.12	7.04
47	2009	11	26.3	30.5	-4.22	-13.83	27.62	90.53	0.91	4.22	13.83
48	2009	12	15.1	24.1	-8.98	-37.27	35.90	148.99	0.76	8.98	37.27
49	2010	1	22.6	34.0	-11.39	-33.54	39.35	115.83	0.84	11.39	33.54
50	2010	2	42.5	45.6	-3.04	-6.67	48.81	107.07	0.84	3.04	6.67
51	2010	3	107.3	108.4	-1.12	-1.03	59.02	54.45	0.94	1.12	1.03
52	2010	4	193.5	190.4	3.04	1.59	81.23	42.65	0.95	3.04	1.59
53	2010	5	184.8	190.2	-5.35	-2.82	101.93	53.60	0.92	5.35	2.82
54	2010	6	272.2	261.5	10.73	4.10	94.01	35.96	0.95	10.73	4.10
55	2010	7	244.7	236.5	8.13	3.44	116.19	49.12	0.92	8.13	3.44
56	2010	8	156.1	154.2	1.87	1.21	100.89	65.42	0.90	1.87	1.21
57	2010	9	117.8	121.2	-3.33	-2.74	78.93	65.15	0.91	3.33	2.74
58	2010	10	70.5	77.4	-6.92	-8.93	51.85	66.96	0.93	6.92	8.93
59	2010	11	24.9	30.1	-5.21	-17.30	29.79	98.83	0.88	5.21	17.30
60	2010	12	12.0	23.5	-11.47	-48.86	40.24	171.43	0.65	11.47	48.86
61	2011	1	26.2	28.4	-2.19	-7.71	26.73	94.00	0.92	2.19	7.71
62	2011	2	39.6	41.1	-1.56	-3.80	27.32	66.42	0.95	1.56	3.80
63	2011	3	115.3	114.2	1.12	0.98	40.37	35.36	0.98	1.12	0.98
64	2011	4	206.1	201.3	4.80	2.38	70.13	34.83	0.97	4.80	2.38
65	2011	5	236.6	233.1	3.50	1.50	97.85	41.99	0.94	3.50	1.50
66	2011	6	234.0	221.7	12.32	5.56	126.70	57.15	0.89	12.32	5.56
67	2011	7	196.6	193.3	3.27	1.69	102.75	53.16	0.92	3.27	1.69
68	2011	8	161.2	157.4	3.81	2.42	97.35	61.85	0.91	3.81	2.42
69	2011	9	128.7	129.5	-0.86	-0.66	78.32	60.48	0.92	0.86	0.66
70	2011	10	58.4	62.7	-4.36	-6.95	40.89	65.18	0.95	4.36	6.95
71	2011	11	38.8	40.3	-1.49	-3.71	27.22	67.58	0.95	1.49	3.71
72	2011	12	16.1	22.1	-5.96	-26.99	30.53	138.23	0.81	5.96	26.99

Table 1: Validation of the montly CM-SAF product in Cabauw, the Netherlands.

3.0.5.2 SICCS: From the SICCS daily and monthly averages were used (monthly averages were also used to compute the yearly averages). This data was supplied by J.F. Meirink (KNMI, climate department, the Netherlands), one of the creators of the dataset. The start point for montly data was January 2006 and the end data was December 2011. For daily data the time series started at April 1st 2010 and ended at the 31st of July 2010.

The SICCS product used a detailed RTM to compute radiation. The SICCS's algorithm consists of two parts. The first part is the input and describes the state of the atmosphere. The second part is the algorithm itself which calculates radiative transfer based on the input. The following parameters were considered when the irradiance and the atmospheric transmissivity were calculated:

1. Solar zenith angle (SZA).
2. Cloud optical thickness (COT).
3. Cloud particle radius.
4. Cloud phase.
5. Aerosol optical thickness (AOT) at 500 nm.
6. The *Ångström* exponent.
7. The aerosol single scattering albedo (SSA).
8. surface elevation.
9. Visible and near-infrared surface albedo.
10. Integrated water vapour (IWV).

Before the algorithm starts, 8 look up tables (LUT) were computed by radiative transfer calculations. These tables contained information about the transmissivity of the atmosphere depending on variables between the atmosphere and the surface. 4 LUT's were computed for clear sky conditions. Two LUT's contained transmissivity about global irradiance for VIS (240-704nm) and NIR(704-4606nm) wavelengths. The other 2 LUT's contained transmissivity about direct irradiance for VIS and NIR wavelengths. The other 4 LUT's were for cloudy skies. 2 LUT's contained transmissivity about global irradiance for VIS and NIR wavelengths for

water clouds and the other 2 contained transmissivity in the same wavelengths but for ice clouds.

When these LUT's were computed the algorithm can be run for each pixel in the SEVIRI image. This process goes as followed:

1) A cloud mask was computed. This was done with information about SEVIRI reflectances and brightness temperatures. Each pixel was classified in one of three categories in this step. A pixel is either "cloud free", "cloud contaminated" or "cloudy".

2) If a pixel was in the "cloud free" category. Direct and global transmissivity for both wavelengths is computed by using the LUT's. Since the LUT's work with discrete values, interpolation of the transmissivity might be used.

3) If a pixel was classified as "cloud contaminated" or "cloudy" information about the physical properties of the cloud had to be computed. Cloud phase, COT and cloud particle effective radius were determined with the Cloud physical properties (CPP) algorithm. A description of this algorithm can be found in Greuell et al. (2013).

4) With the CPP known, the global transmissivity could be found in the LUT's for clouds. The same interpolation was used as for clear sky pixels if the values did not perfectly match the LUT values. If it turned out that the CPP found a COT lower then a certain threshold. The pixel was treated like a cloud free pixel and the LUT's for clear sky conditions were used.

5) In this step the global transmissivity for all other wavelengths and pixels were computed and the direct transmissivity for clear sky pixels. This is done by taking the transmissivity of the VIS and NIR wavelengths and weighing them according to the fraction they contribute to the total incoming radiation on top of the atmosphere.

6) A correction was then performed to account for a bias found in ice clouds.

7) COT and clear sky direct transmissivity were used in an equation to calculate the cloudy direct transmissivity (Greuell et al., 2013).

8) Gaps in the data and missing data were filled. This missing data was computed by a correction algorithm (Greuell et al., 2013) or by taking the mean of the retrieved transmissivity on the same day.

9) The diffuse transmissivity was calculated. This was done by taking the global transmissivity and subtracting the direct transmissivity. The difference between these two is equal to the amount of diffuse transmissivity.

10) The global, diffuse and direct transmissivity(T_i) were recalculated into global, diffuse and direct irradiance(F_i). This was done by the following equation:

$$\mathrm{F_i} = \mathrm{S}\frac{\mathrm{u_o}}{\mathrm{d^2}}\mathrm{T_i} \tag{10}$$

In this equation, S is the reduced solar constant which is set on $1358.1 W/m^2$ and d is the distance from the Sun to the Earth in astronomical units.

When comparing the hourly results obtained by the SICCS with eight stations that measured radiation on the ground the following results were received. The median values of the station biases was $6 W/m^2$ (5%) for direct irradiance and $1 W/m^2$ (1%) for diffuse irradiance. The global irradiance had an bias of $7 W/m^2$ (2%) (Greuell et al., 2013).

3.1 Data processing

The data was processed in the KNMI's Geospatial Interpolation Environment (GSIE). GSIE is a virtual environment that allows the user to perform large amount of interpolations by running a pre-defined script which consists of 3 main input files. A recipe, a query and a, interpolation script written in R.

The recipe tells GSIE where to find all the required files and what parameters should be set for the input and output. All recipies that are used in GSIE have the same structure. First the time span of the data to be processed is given. The user has to define the start and stop time and has to set the resolution within this time serie (day, month or year for instance). After this, the query is defined. The query calls up the in-situ measurements from the KNMI database. The query contains information on which variable should be loaded in, which stations and the locations from these stations. The recipe describes where the results of the query are located and what kind of file and units it contains. Next, the folders where the R files are located are defined and the R scriped that has to be used is set.

The R script contains the interpolation and validation code that will be performed. The R script loads in all the data and works trough it line by line as written in the script. The R script will load in the in-situ and satellite data and will convert all radiation units to W/m^2 using the method described in 3.0.4. When this is done the different interpolation methods can be used on the input data. This is done by using codes that define the variable to be interpolated and parameters required for each interpolation. When the initial interpolation is done,

a cross-validation script is run and statistics on the interpolation are computed according to the code written in the R script. These statistics are exported as .txt files and can be analysed to get a better understaning of the interpolation output. The interpolated results are exported as maps in the form of .ASCII files. The R scripts used for global radiation can be found in appendix B to F.

Ones the interpolation has been performed and the maps have been created, GSIE uses a web mapping service to show the created maps to the user and evaluate them (figure 7). Within this mapping service maps can easily be downloaded in different formats and be reprojected, edited, etc.

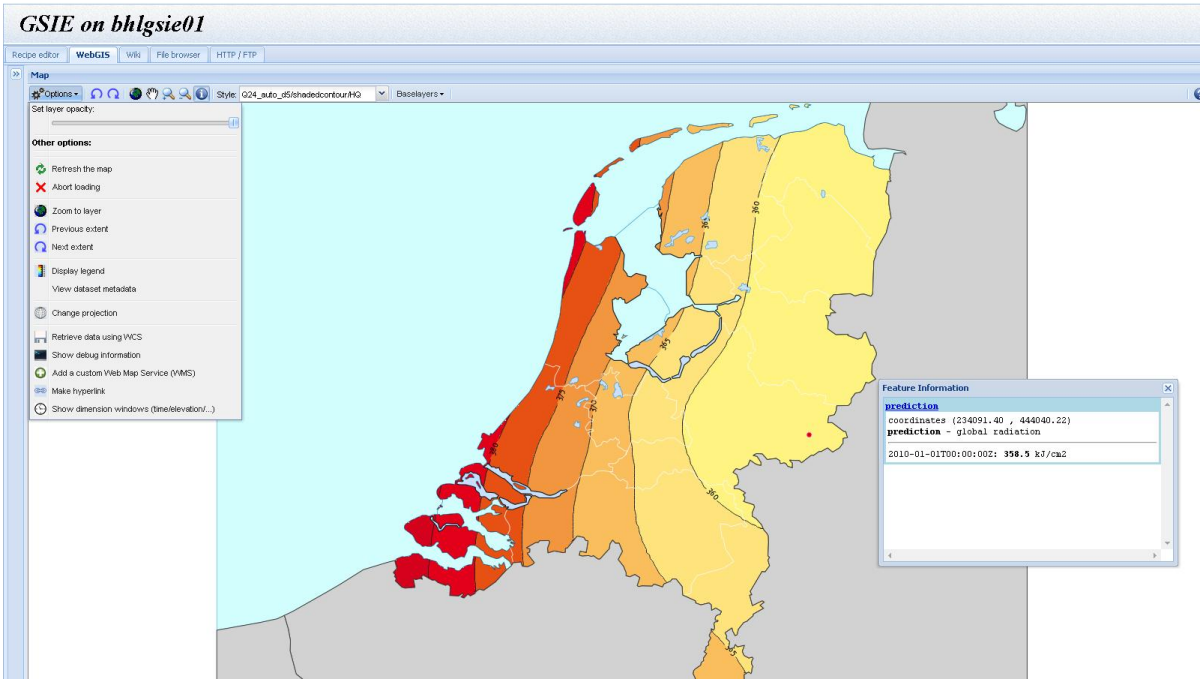


Figure 7: The GSIE web mapping service from the KNMI.

.

3.1.1 Data assimilation

For the exploration of combining in-situ data with satellite data different interpolation techniques had to be analysed. There are several interpolation methods that allow the use of auxiliary data. Kriging with external drift is an example of this (Sluiter, 2008). However data can also be combined on different ways. A Bias interpolation (Journée et al., 2010) method could also be made to compare differ-

ences between ground measurements and satellite measurements and take these differences into account when interpolating the station values.

3.1.2 Interpolation

Interpolation is a process used to determine the variable of an unmeasured location by using known values of measured locations. Each method has its advantages and disadvantages (Hiemstra et al., 2011). Interpolation can however be divided in several classes.

The first division can be made in whether a method is deterministic or probabilistic. Deterministic methods create an interpolated surface by only using the geometric characteristics of the measured points. Probabilistic methods use these points as well but assume that there is a random element in the interpolation. This type of interpolation allows the user to add a variance into the interpolation method and it will provide a statistical significance of the results it gives.

The second division in interpolation methods can be made whether a method is a global or local interpolater. Global interpolation uses one function to apply on the entire field of interest. Local interpolators use several different functions for the entire field or part of the field. Global interpolators often gives a smooth map.

The last division that can be made is if a interpolation method is an exact or inexact interpolation. Exact interpolations assume that values on which the interpolation is based are correct and will return the same values as measured at the locations of the measurement stations/locations. Inexact interpolations assume that there are uncertainties in the measured values and will therefore not return the exact value as the input data at the measured location (Sluiter, 2009).

The simplest form of interpolation is taking the mean of all the measurements and apply this to all unknown locations. However for this research that would have been to simplistic. Different interpolation methods should be explored and tested for there accuracy.

There are however several characteristics that can be expected when using any interpolation method:

1. If there is a dense dataset of sample points and they are spread uniformly over the area of interest. A fairly good interpolation will be performed, no matter which method is used.

2. If the data is clustered in several locations with large distances between them, the output of the interpolation will be unreliable, no matter which method is used.

3. Due to the fact that all interpolation methods average out values from measurements, it is expected that high values will be underestimated and low values will be overestimated.

In the next section some of the most used and known interpolation methods will be explained. These methods are: nearest neighbour, inverse distance weighting, splines and Kriging. Besides these interpolation methods, two more methods for combining satellite data with in-situ measurements will be explained. The mean bias correction and the interpolated bias correction.

Nearest neighbour interpolation (NN) is a relative simple interpolation method. NN is also known as Thiessen polygon or Voronoï interpolation (Sluiter, 2008). NN takes the value of the closest measured location and allocates it to the unknown point. This is a very fast and mathematically simple interpolation method. However the results are very simplistic and do not look realistic in most cases. NN interpolation is also very sensitive for the amount of input data. The more measurements available the better the result. (Sluiter, 2008). For this research NN will not be used due to the relative low amount of measurement locations.

Inverse distance weighting (IDW) interpolation is an exact interpolation method that continues on the basis of the NN interpolation. IDW allows more than just one measurement to influence the value of a location. The influence of the measurement on the unknown location is determined by the distance from the unknown location to the measurement. The further away the measurement the less influence it has on the determination of the value of the unknown location. The value of an unknown location (Z_u) is given by equation 11.

$$\hat{\mathrm{Z}}_{\mathrm{i}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \omega(\mathrm{x_i})\mathrm{Z}(\mathrm{x_i})}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \omega(\mathrm{x_i})} \tag{11}$$

Where:

- $\hat{Z}_i$ = The estimated value.
- $\omega(x_i)$ = The weight of location (x_i).
- $Z(x_i)$ = The measured value at location (x_i).

The weight $\omega(x_i)$ that a location receives is determined by the distance to the measured points and the power that determines how fast weights drop with an increasing distance, see equation 12.

$$\omega(\mathrm{x_i}) = \|\mathrm{x_i} - \mathrm{x_0}\|^{-\mathrm{p}} \tag{12}$$

Where:

- $\| * \|$ = The Euclidean distance between the unknown location and the measurement.
- $-p$ = The power that determines how fast weights drop.

IDW interpolation is a fast and easy to implement method. The user can change the $-p$ parameter to influence the output and create a better result. However this interpolation method does not allow the user to add a secondary data source and was therefore not sufficient for the combination of satellite data with in situ data in this paper. However IDW was used to explore the interpolated bias method, since it was a relative simple but exact method. IDW is already a widely used method in interpolation of meteorological data and was therefore chosen to be used for the interpolated bias interpolation (Hiemstra et al., 2011; Sluiter, 2008).

Splines interpolates a surface by applying a set of polynomials trough the observations. The polynomials are often of a third order degree. Lowering the degree makes the interpolation more general and simplistic. Having a too high degree of polynomials can create errors in the data set due to the high amount of oscillations that can occur.

For splines interpolation it is important to determine whether to apply the polynomials to a global pattern or a local pattern. The difference between local and global patterns is visualized in fig 8.

To get the best accuracy with splines a cost function can be applied. A cost function will minimize the amount of bending while optimizing the accuracy of

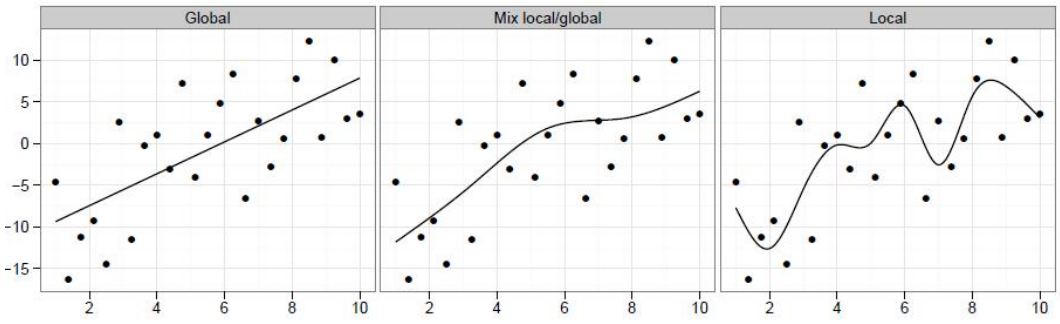


Figure 8: global vs local patterns in splines interpolation (Hiemstra et al., 2011).

.

the polynomial at the measured locations (Hiemstra et al., 2011; Sluiter, 2008). Higher errors at a measured location have a higher cost. The goal is to keep the cost as low as possible by keeping the biases at the measured locations as small as possible.

For this research splines was not the most interesting interpolation technique to use. This due to the fact that the amount of versions of splines allowing a secondary data set to be used is limited and very complex. However the concept of the spline interpolation was of great importance due to the fact that the current maps of global radiation data are interpolated with Thin plate splines (TPS). TPS interpolation is assumed to be a good method to interpolate monthly and yearly climate elements (Sluiter, 2008). Several climatologic and meteorological elements are currently interpolated at the KNMI using TPS. Not all climatological and meteorological data used for the interpolations comes with the same spatial and temporal resolution. So TPS might not have been the optimal interpolation technique for every element. However since it gave the best results on average, the KNMI chose to use TPS for the evaporation and radiation datasets. One of the data sets which is interpolated using TPS is the Makkink evaporation. Since Makkink evaporation is determined by temperature and most importantly global radiation, TPS is used for global radiation mapping to keep the spatial patterns the same. (Sluiter, 2012).

Kriging is a geo-statistical interpolation method that assumes that spatial variation, of the to be interpolated attribute, is often not able to be described with a simple function.

Ordinary Kriging (OK) is the basic form of Kriging. OK uses weights, which are described by a variogram, to optimize the interpolation values. OK uses the following equation, equation 13, as an assumption to calculate the unknown value of a variable at an unmeasured location.

$$\hat{\mathrm{Z}}_{\mathrm{i}} = \mathrm{m(x)} + \epsilon'(\mathrm{x}) + \epsilon\text{”} \tag{13}$$

Where:

- $\hat{Z}_i$ = The variable at location x.
- m(x) = A function describing a structural component of Z at location x.
- ϵ’ = A random spatially correlated component.
- ϵ” = A random non-spatially correlated term.

When the structural components have been accounted for. The semi-variance can be calculated. This explains the correlation the residuals have with each other. A semi-variogram is computed using equation 14.

$$\mathrm{Y_u(h)} = \frac{1}{2\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \left\{ \mathrm{z(x_i)} - \mathrm{z(x_i + h)} \right\}^2 \tag{14}$$

Where:

- $Y_u(h)$ = The semi-variance.
- n = The number of point pairs of the sample data z separated by distance h.

The semi-variogram (figure 9) gives the user information about several components: the sill, the range and the nugget. The sill is the maximum value the semi-variogram reaches. The range is the distance at which the sill is reached. This means that from this distance on, sample points will no longer influence the predicted value at a certain location. The nugget is the error or noise in the data. It assumes that if a location is measured more then once, different values will be given, this difference is explained by the nugget.

When the optimum weight is obtained by the semi-variogram, the expected value at an unmeasured location can be calculated using equation 15.

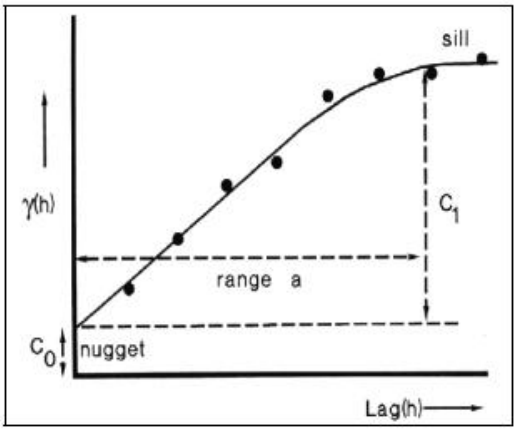


Figure 9: Semivariogram (Sluiter, R. 2009).

.

$$\hat{Z}_i = \sum_{i=1}^{n} \lambda_i * z(x_i) \tag{15}$$

Where:

- $\hat{Z}_i$ = The expected value at location x_0.
- λ_i = The optimal weight of location i.

This method is known as OK. There is a variant called: "Simple Kriging". It works the same as OK however, it makes use of known mean. This makes it a slightly better version of OK but it is often difficult to derive the mean (Sluiter. R. 2009).
When Kriging is used as an interpolation method, the user has the possibility to make use of blocks. When blocks are set, Kriging predicts the blocks mean values. This smoothens the map and often gives a better look to the map. However when blocks are set, the original values of the observations are not returned at the observation points. This leads to a mismatch of the original observed value and the value obtained from the map (Sluiter, 2012). For this research blocks have not been used since the aim was to get exact maps with high resolution instead of smoothend maps.

Universal Kriging/Kriging with a trend/Kriging with external drift (KED) is a different variant of OK. KED makes use of a secondary data source which has a high spatial resolution and is closely correlated to the main attribute to predict the output values of the interpolation (Journée et al., 2010). In this case it used in-situ measurements as the main input source and the satellite image as a secondary data source. The satellite image had a relative high spatial resolution and was most likely correlated with the in-situ data because they both gave information about the same attribute. The satellite image was used as a trend in the interpolation of the in-situ data. At the KNMI KED is used for several datasets. When using KED several parameters can be set or adjusted. At the KNMI KED is performed using a spherical or exponential variogram model and the nugget of the variogram is set to 0. The nugget is set to 0 to ensure that Kriging returns the in-situ measured values at the stations location, since it is possible to assume that this value is correct. The other variogram parameters are set automatically by using an autofit function.

Mean Bias correction (MB) is not really an interpolation technique but a method that could show potentials for improving the current resolution of global radiation maps. Journée et al., from the royal meteorological institute of Belgium used this technique to improve their global radiation maps. It turned out that even though it is a very simplistic method the results were only slightly less good in respect to the more complex and detailed interpolation techniques (Journée et al., 2010). Therefore it should be explored in this research.

The method is based on the idea that the satellite has an error and that this error is constant over the entire area. So by looking at the differences between the satellite measurements and the ground measurements and taking the average of this we can apply this difference on all other unmeasured locations.

The method works as followed, equation 16 and 17:

$$\hat{Z}_i = \delta q(x_i) \tag{16}$$

With

$$\delta = \frac{\sum_{i=1}^{n} Q(x_i)}{\sum_{i=1}^{n} q(x_i)} \tag{17}$$

Or it can be used as an additive function, function 18 and 19:

$$\hat{Z}_i = q(x_i) + \Delta \tag{18}$$

With

$$\Delta = \frac{1}{\mathrm{N}} \left(\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{Q(x_i)} - \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{q(x_i)} \right) \tag{19}$$

Where:

- $\hat{Z}_i$ = The estimated value at location (x_i).
- x_i = The location of interest.
- $Q(x_i)$ = The in-situ measured radiation at location (x_i).
- $q(x_i)$ = The satellite measured radiation at location (x_i).

Journée et al., already described that this is not the best method to get accurate radiation data. However since it gave better results as expected and since it is a very easy to apply method testing it for the Netherlands was worth the effort (Journée et al, 2010).

Interpolated bias correction (IB) is also a combination method that was explored rather than an interpolation method. The concept of this method was the same as that of the MB correction. However in this case, the bias that is found between the measurement stations and satellite observations is interpolated with equation 20 and equation 21. With this method we no longer assume that the error the satellite captured is uniform but varies from location. It was possible to use different interpolation techniques on the bias values. Journée et al. used IDW interpolation on the bias. However Kriging or Splines could also work. (Journée et al., 2010).

$$\hat{\mathrm{Z}}_{\mathrm{i}} = \delta \mathrm{q(x_i)} \tag{20}$$

With

$$\delta = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{W_i} \frac{\mathrm{Q(x_i)}}{\mathrm{q(x_i)}} \tag{21}$$

Or it can be used as an additive function, function 22 and 23:

$$\hat{\mathrm{Z}}_{\mathrm{i}} = \mathrm{q(x_i)} + \Delta \tag{22}$$

With

$$\Delta = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{W_i}(\mathrm{Q(x_i)} - \mathrm{q(x_i)}) \tag{23}$$

Where:

- $\hat{Z}_i$ = The estimated value at location (x_i).
- x_i = The location of interest.
- $Q(x_i)$ = The in-situ measured radiation at location (x_i).
- $q(x_i)$ = The satellite measured radiation at location (x_i).
- W_i = The interpolation weights obtained by the IDW interpolation.

Table 2 gives a short overview of the interpolations and methods explained above and the data they use for the interpolation.

Interpolation method	Abbreviation	Main interpolation parameter	Auxilary data
Nearest Neighbour	NN	in-situ measurements	No
Inverse Distance Weighting	IDW	in-situ measurements	No
Thin Plate Splines	TPS	in-situ measurements	No
Ordinary Kriging	OK	in-situ measurements	No
Kriging with External Drift	KED	in-situ measurements	Yes
Mean Bias correction	MB	in-situ measurements	Yes
Interpolated Bias correction	IB	in-situ measurements	Yes

Table 2: The different interpolations and methods explained and the data they use.

.

3.1.3 Data validation

To see if the interpolation techniques or methods used to get values at unmeasured locations performed a proper job, validation was needed. There are several ways to perform a validation of the data.

Data splitting is one method to validate the interpolation. Before the interpolation takes place. The input data is divided into two different sets. One set that is used to perform the interpolation and one control set that is used to validate the interpolation. Once the interpolation is done the control set is taken and

the value of the measured location is compared to the value of the same location in the interpolation. By looking at the interpolated value and the real value out of the control set an assessment can be made on how good of a performance the interpolation gave (Sluiter, 2008). Data splitting was an option for this research due to the fact that a background trend in the form of a satellite image was used which could have a big influence on the output of the interpolation. However it could also have had a big influence on the interpolation quality since the meteorological stations are used as the main input and only 32 stations were available. Therefore a different validation method was preferred but this one was not excluded.

Cross validation is a more statistical approach to validate the data. Cross validation uses all input stations to evaluate the interpolation. It does this by a "leave one out" technique. A station was left out in the interpolation process and its value was predicted using the other observations. The predicted value was then compared with the measured value at that stations location. This step was repeated for every station to test the model. By doing this a set of residuals was obtained from each stations location. Statistics can then be performed on these residuals to test how good the interpolation performed. Several statistics can be computed with the cross-validation residuals.

The R^2 value. The R^2 is calculated as shown in equation 24:

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SS_e}}{\mathrm{SS_{tot}}} = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{Z_i} - \hat{\mathrm{Z}}_\mathrm{i})^2}{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{Z_i} - \bar{\mathrm{Z}}_\mathrm{i})^2} \tag{24}$$

- n = The number of observations.
- Z_i = The measured global radiation at location i.
- $\hat{Z}_i$ = The estimated value at location i.
- $\bar{Z}$ = The mean global radiation.
- SS_e = The residual sum of squares ($\sum_{i=1}^{n}(Z_i - \hat{Z}_i)^2$).
- SS_{tot} = The total sum of squares ($\sum_{i=1}^{n}(Z_i - \bar{Z}_i)^2$).

The R^2 value tells the user how much of the variance in the data can be explained by the model. The R^2 value is normally located between 0 and 1. The best value to obtain is 1. This would indicate that the model can account for all the

variation that is obtained by running the model. The R^2 is often used to compare different models with each other. R^2 is used in the previous version of the global radiation maps of the Netherlands and was therefore used again as a comparison (Hiemstra et al., 2011).

The root mean square error (RMSE), equation 25. The RMSE is a measure of the difference between the observations and predicted values. It tells the user how big the magnitude of the errors is. Lower RMSE values are preferred to be obtained since this means that the errors in the model are small.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\hat{\mathrm{Z}}_{\mathrm{cv,i}} - \mathrm{Z}_{\mathrm{i}})^2} \tag{25}$$

Where:

- $\hat{Z}_{cv,i}$ = The estimate from the cross validation at location i.
- Z_i = The measured global radiation at location i.

To see if the interpolation generally over or underestimates the values the mean error can be used. The Mean error (ME) can be calculated using equation 26.

$$\mathrm{ME} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\hat{\mathrm{Z}}_{\mathrm{cv,i}} - \mathrm{Z}_{\mathrm{i}}) \tag{26}$$

The RMSE and the ME are relatively easy methods to compare different interpolation methods with each other. However it is only a measurement of relative performance. When comparing two different methods it could turn out that the RMSE and the ME of one method are very low compared to the other. However in reality the method could still perform very bad. To be able to say something about the real performance of an interpolation method other statistics have to be looked at.

$$\mathrm{RMSE}_{\mathrm{sd}} = \frac{\sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\hat{\mathrm{Z}}_{\mathrm{cv,i}} - \mathrm{Z}_{\mathrm{i}})^2}}{\sqrt{\frac{1}{\mathrm{n}-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{Z}_{\mathrm{i}} - \bar{\mathrm{Z}})^2}} \tag{27}$$

The $RMSE_{sd}$ (equation 27) divides the cross-validation RMSE by the standard deviation of the observed measurements. This method compares the interpolation against the observed mean ($\bar{Z}$) if it would be interpolated.

A similar method can be used to make the ME say something about the real performance of the interpolation. This can be done by taking the ME equation and dividing it by the mean of the observations ($\bar{Z}$). This provides a scale that allows the user to compare the ME with the scale of the observations (the mean). The ME_{mean} is calculated by using equation 28

$$\mathrm{ME_{mean}} = \frac{\frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} (\hat{\mathrm{Z}}_{\mathrm{cv,i}} - \mathrm{Z_i})}{\bar{\mathrm{Z}}} \tag{28}$$

With these statistics known it is possible to evaluate the different interpolation methods. It is possible to evaluate them relative compared to each other with the RMSE and ME and it is possible to judge them on there performance against reality by using the $RMSE_{sd}$ and ME_{mean} (Hiemstra et al., 2011).

Other methods to get a value to compare models with each other are the scaled NRMSE and MAPE. The NRMSE is the normalized root mean squared error. It divides the RMSE by the range of the observations, see equation 29. It is often expressed as a percentage. Lower values indicate that there is less variance in the residuals.

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} (\hat{\mathrm{Z}}_{\mathrm{i}} - \mathrm{Z_i})^2}}{\mathrm{x_{max}} - \mathrm{x_{min}}} \tag{29}$$

The Mean Absolute Percentage Error (MAPE) method expresses the error as a percentage. It takes the measured in-situ value minus the predicted value and divides it by the measured in-situ value. For more observations these values can be summed up. To get the percentage it needs to be multiplied by 100

$$\mathrm{MAPE} = \frac{100\%}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \frac{\mathrm{Z_i} - \hat{\mathrm{Z}}_{\mathrm{i}}}{\mathrm{Z_i}} \tag{30}$$

The MAPE statistic shows how much the residuals deviate from the original values. Therefore it is a very good statistic to look at since it gives a sense of how good the model predicts reality (Hyndman et al., 2006).

Random points between the measurement stations could be analyzed and compared with the interpolations and satellite image to get a better understanding of the data in- and output. Since there were no ground control points to validate the errors of the interpolation methods, the different outputs and satellite products were compared against each other. Doing so gave a better understanding on how

the satellite images stood on their own, how the satellite images influenced the interpolation and how the interpolation methods performed compared to these images. If the interpolations were accurate at the locations of the in-situ measurement stations it was possible to assume that these interpolations are just as accurate in-between the stations. By making this assumption it was possible to see if the satellite images followed the same pattern as the interpolated output and the same could be assumed the other way around. This would mean that; if the interpolated output followed the same pattern as the satellite image (this does not mean it had to return the same value), then it was possible to assume that the interpolation method made good use of the trend that could be found in the satellite product. On the other hand, if the values of the satellite image came close to the values of the interpolation we could assume that the satellite image consisted of accurate measurements. Due to the fact that the validation of the satellite products already proves that they perform an accurate form of measurements this method was a sort of double check.

This method also gave the opportunity to compare the satellite image values with those of the TPS method which didn't use the satellite data as an input at all.

For this paper the R^2 and MAPE were used. R^2 was used to get a first comparison between the different interpolation techniques. The R^2 made it relatively easy to compare different models and it was a method that was used before within the KNMI. This allows a comparison with the previous global radiation dataset.

MAPE was used on the cross validation. It looked at the percentage difference of a station when it was initially left out in the interpolation. With this method a very clear and honest statistics was obtained that gave a good overview on how the interpolation performed.

To get an even better understanding of the data set a data split was used where 8 random stations were left out when the interpolation was performed. Afterwards these stations were put back in and the bias between the interpolation and the stations was analysed. This gave an even better overview on how the interpolation behaved and performed.

The data in between the stations was compared with each other at 60 different locations. This was done to see how the interpolated maps and satellite images stood against each other on locations that are not measured and known exactly.

Expert judgement wasl also used as a validation method. It was not statistically sound, however it was of great help when analyzing results. Interpolation is a pure mathematically process which does not account for physical or biological or any

other process that could have an influence on the interpolated value in the real world. Experts on radiation data could help analyze the map and find patterns that should or should not be there. A discussion with an expert on radiation patterns could lead to a better understanding and could be taken into account when choosing an interpolation method or setting a parameter in an interpolation equation. The expert that has been consulted was R. Sluijter, a climate expert at the KNMI.

3.1.4 Spatial patterns

As a result of a discussion with the climate expert, when creating a map of global radiation in the Netherlands, several patterns are to be expected.

1. Due to the slower warming of water, less clouds were expected to form directly above water bodies during spring and summer. Once the air moves over land which warms faster, clouds are expected to form. Due to this a cloud free coast was expected with clouds forming further in land. Therefore especially in spring and summer time more radiation is expected along the coast due to the lack of clouds.

2. Less radiation was expected above the Veluwe and the Utrechtse Heuvelrug. These areas in the middle of the Netherlands are characterized by a higher elevation (appendix G). The higher elevation canl cause orographic lift of air, resulting in the formation of clouds due to the lowering temperature by increasing elevation. The clouds will block out radiation due to a higher albedo, leading to a lower radiation value. This pattern can also be observed when looking at precipitation maps of the Netherlands. Most precipitation falls in this area, see figure 10.

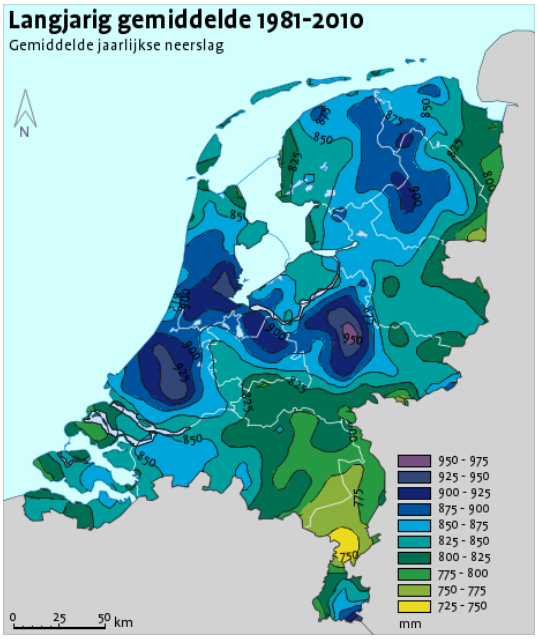


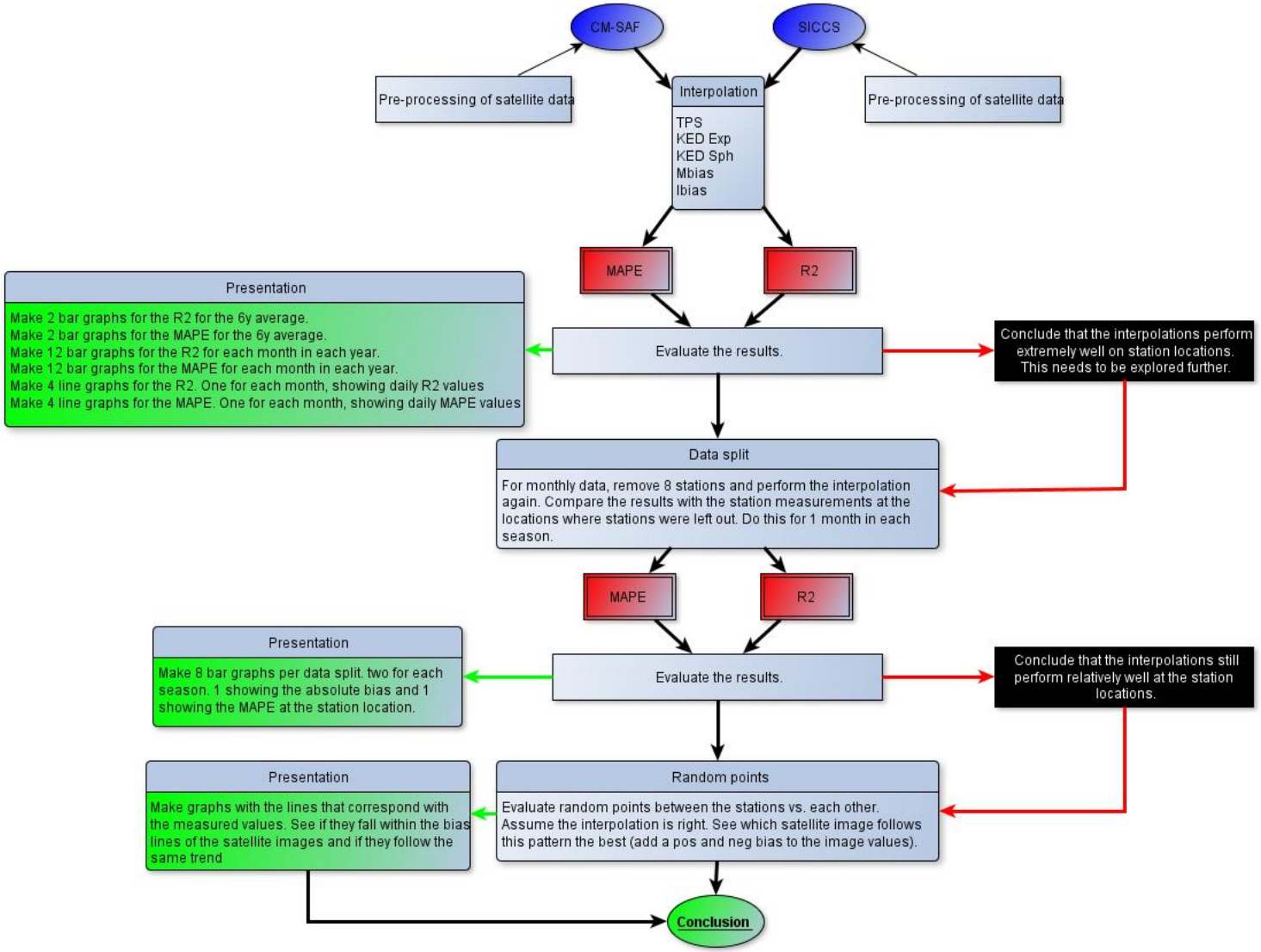
Figure 10: Average precipitation in the Netherlands

.

3. In Spring and Summer a general west-east gradient can be detected with the highest radiation values in the west along the coast.

4. In Autumn and Winter a general north-west gradient can be tetected with the highest radiation values in the south. This pattern is observed due to the differences in the lenght of the day.

With interpolation and validation methods known, the data could be processed. In general the workflow shown on the next page can be followed to process the data and evaluate the output.


CM-SAF
SICCS
Pre-processing of satellite data
Interpolation
TPS
KED Exp
KED Sph
Mbias
Ibias
Pre-processing of satellite data
MAPE
R2
Presentation
Make 2 bar graphs for the R2 for the 6y average.
Make 2 bar graphs for the MAPE for the 6y average.
Make 12 bar graphs for the R2 for each month in each year.
Make 12 bar graphs for the MAPE for each month in each year.
Make 4 line graphs for the R2. One for each month, showing daily R2 values
Make 4 line graphs for the MAPE. One for each month, showing daily MAPE values
Evaluate the results.
Conclude that the interpolations perform extremely well on station locations. This needs to be explored further.
Data split
For monthly data, remove 8 stations and perform the interpolation again. Compare the results with the station measurements at the locations where stations were left out. Do this for 1 month in each season.
MAPE
R2
Presentation
Make 8 bar graphs per data split. two for each season. 1 showing the absolute bias and 1 showing the MAPE at the station location.
Evaluate the results.
Conclude that the interpolations still perform relatively well at the station locations.
Presentation
Make graphs with the lines that correspond with the measured values. See if they fall within the bias lines of the satellite images and if they follow the same trend
Random points
Evaluate random points between the stations vs. each other. Assume the interpolation is right. See which satellite image follows this pattern the best (add a pos and neg bias to the image values).
Conclusion

4 Results

Cross validation

Five interpolation methods were used (TPS, MB, IB, KED-SPH & KED-EXP). The interpolations were performed for the average radiation over 6 years. For each individual month of every year and for every single day from April 1st 2010 untill July 31st 2010.

For the 6 year average, the R^2 value of the IB was lower (0.19 for the CM-SAF product and 0.43 for the SICCS product) than that of the other interpolation methods (0.62 to 0.77 for the CM-SAF product and 0.49 to 0.62 for the SICCS product)(figure: 11).

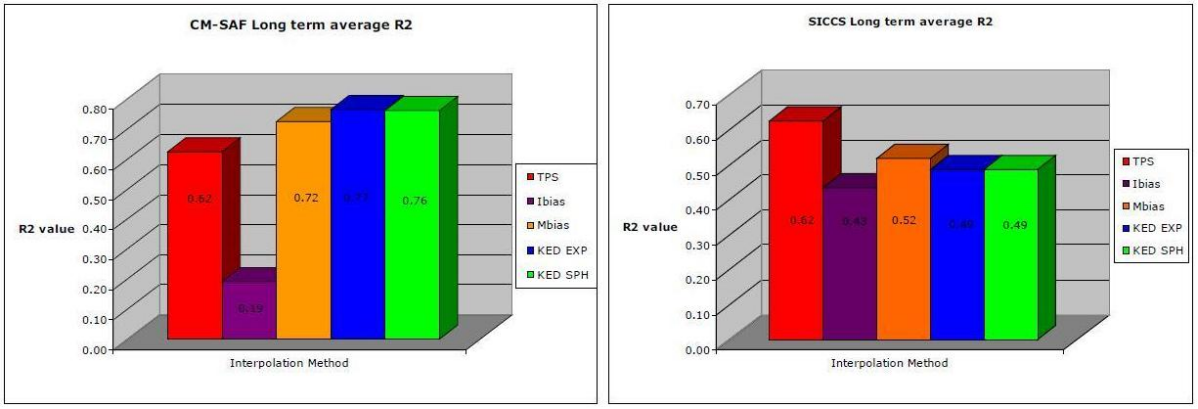


Figure 11: R^2 values for the 6 year average interpolation. Left shows the results for the CM-SAF product and right shows the results for the SICCS product.

The MAPE value of the IB performed just as good or even better than the other interpolation methods (1.14% for the CM-SAF product versus 1.03% to 1.39% and 1.12% for the SICCS product versus 1.34% to 1.42%).

The highest average MAPE found for the CM-SAF product was 1.39% when using the TPS interpolation method.

The MAPE's in the SICCS product of the KNMI were a little higher in general with a maximum error of 1.42% for both of the KED interpolations (Figure 12).

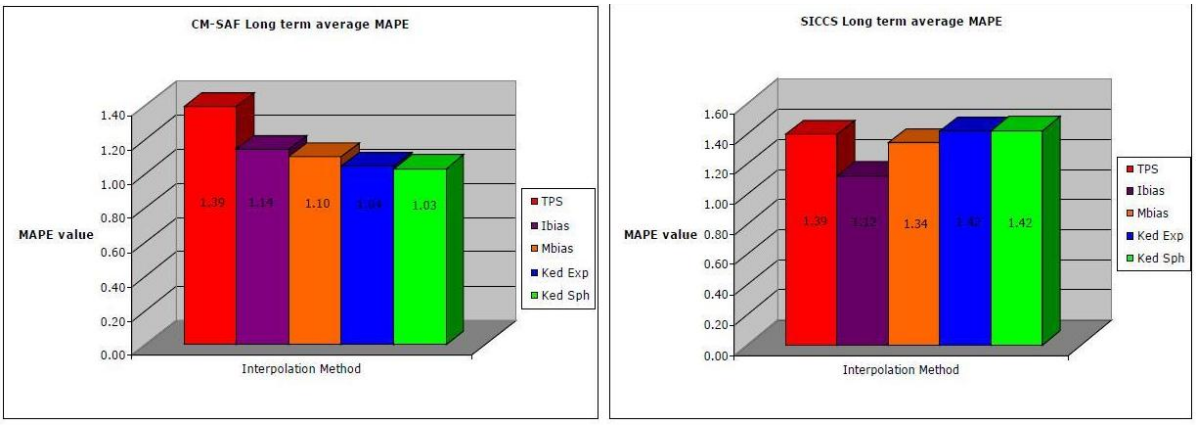


Figure 12: MAPE values for the 6 year average interpolation. Left shows the results for the CM-SAF product and right shows the results for the SICCS product.

When looking at the output maps it was possible to see a clear distinction between interpolation methods that use auxilary data and methods that didn't (figure 13 and 14. The TPS method, which was not using auxilary data returns a relative smooth map with no local variation. All interpolation methods that did use auxilary data show local variation according to expected patterns.

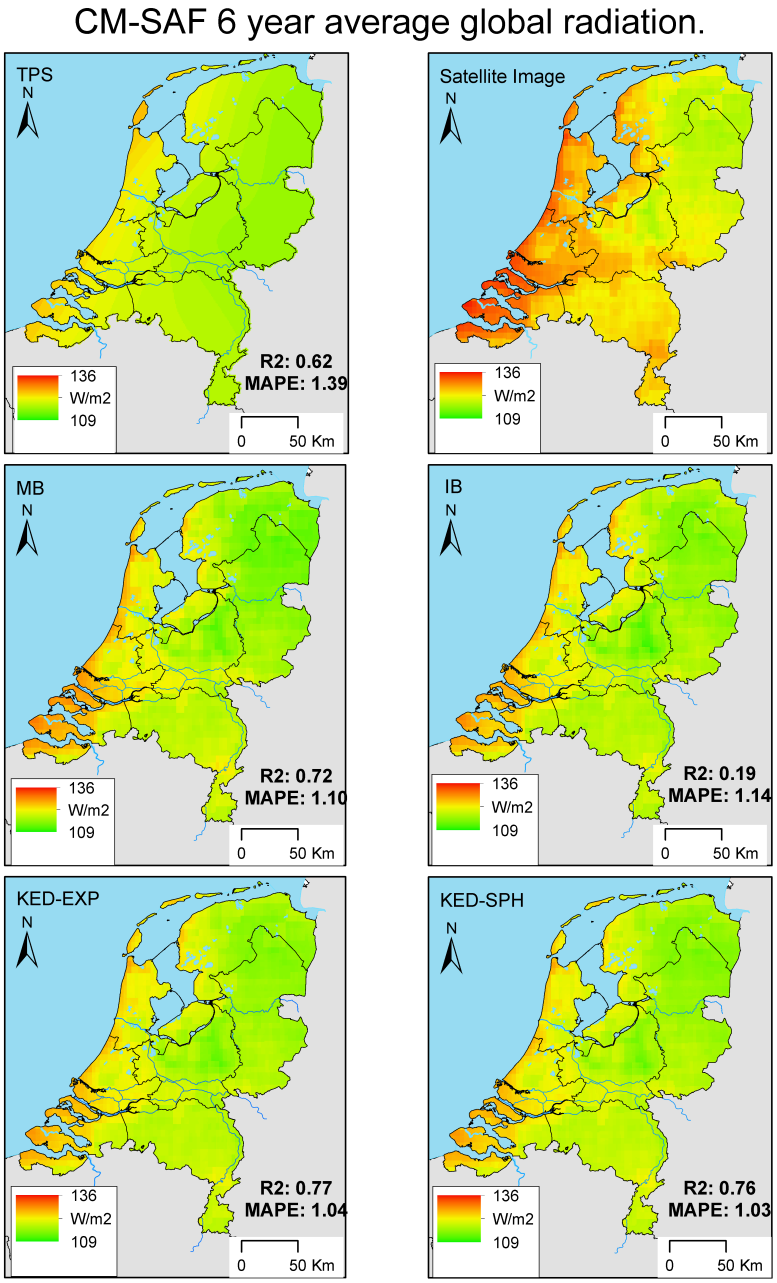


Figure 13: CM-SAF 6 year average interpolation output.

.

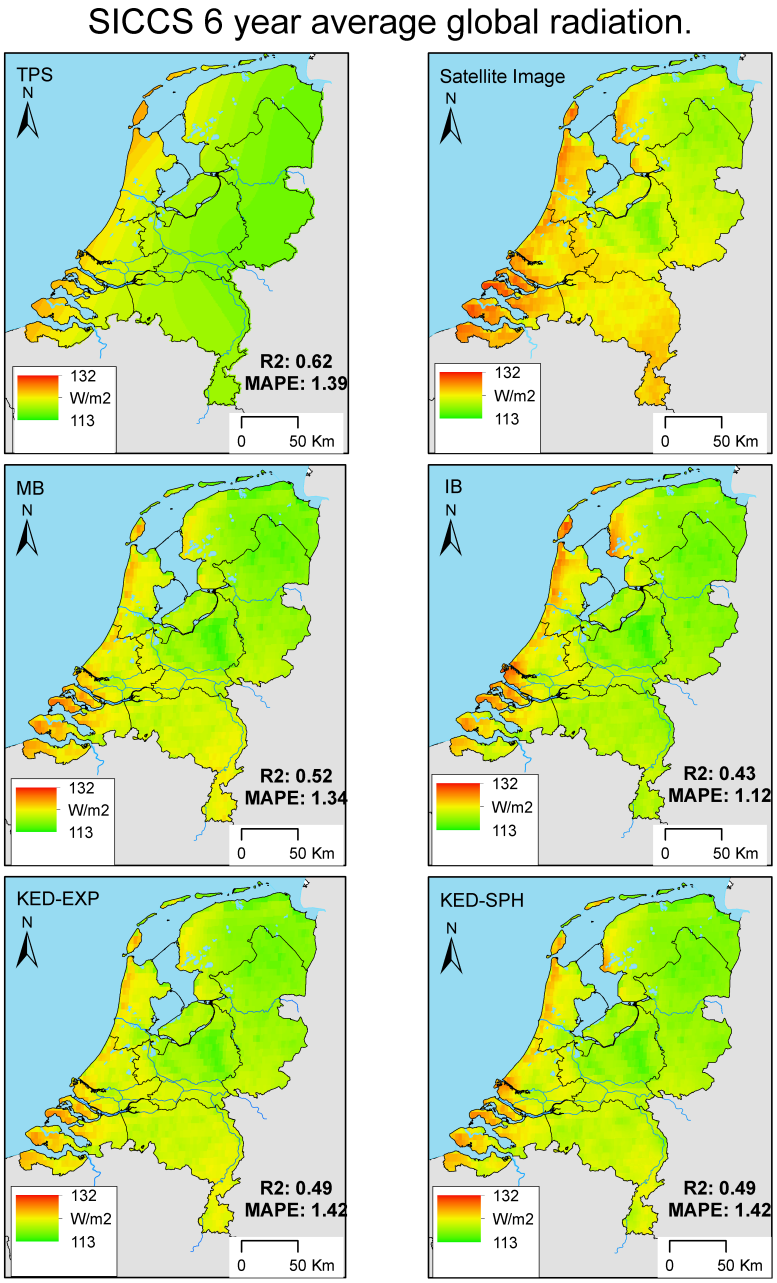


Figure 14: SICCS 6 year average interpolation output.

.

Looking at the monthly interpolation methods it was possible to see some more variation in the results (figure) The CM-SAF products R^2 was fairly stable for the TPS and KED interpolations. The R^2 of the IB and the MB interpolations were changing quite a bit, ranging from -126.41 to 0.94. Especially the MB had a big difference in R^2 values. The SICCS product showed this same trend however the variation was less extreme than in the CM-SAF product. The variation in this product varied from -5.39 to 0.96. See figures 15 to 20

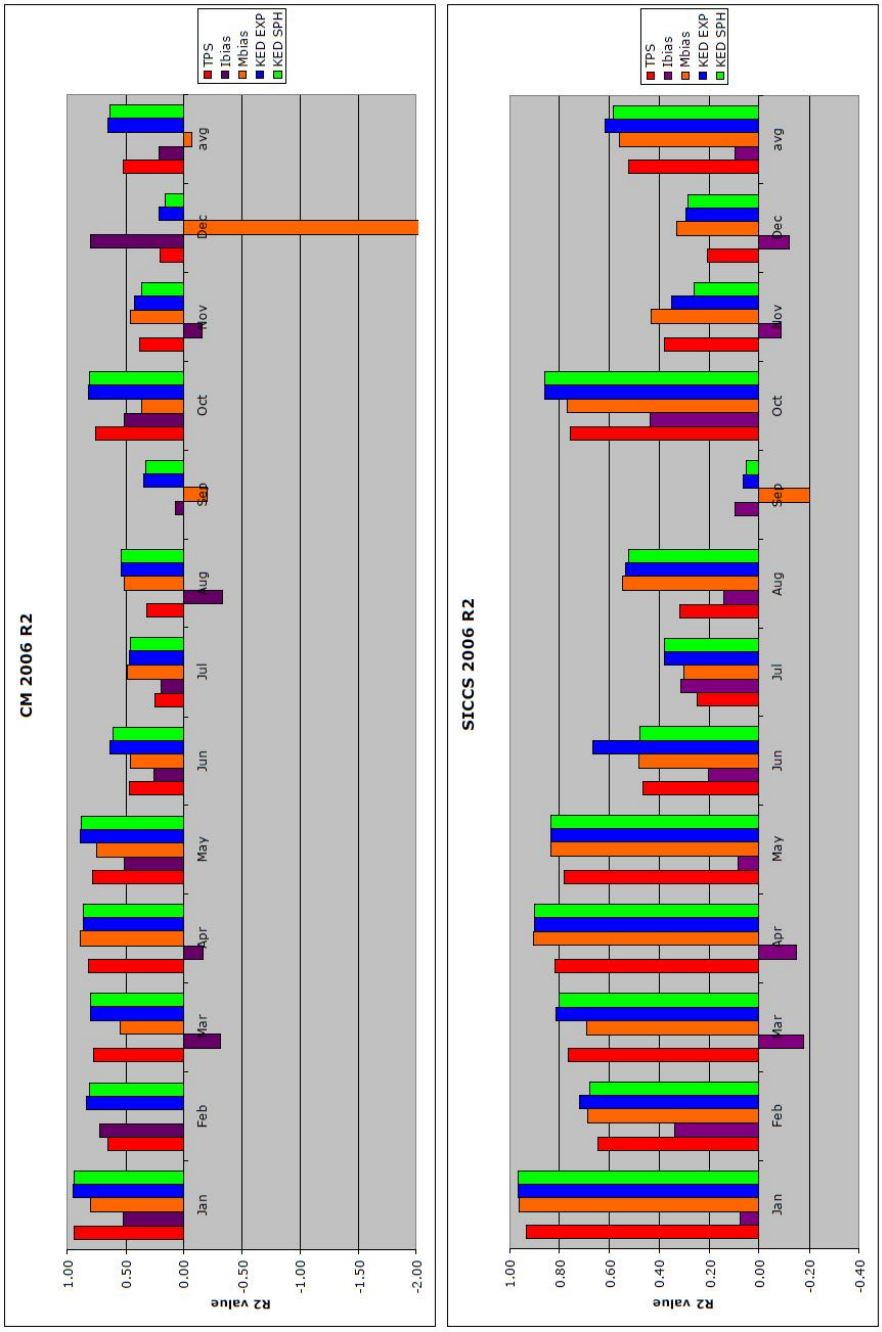


Figure 15: R^2 values for 2006 for both products.

.

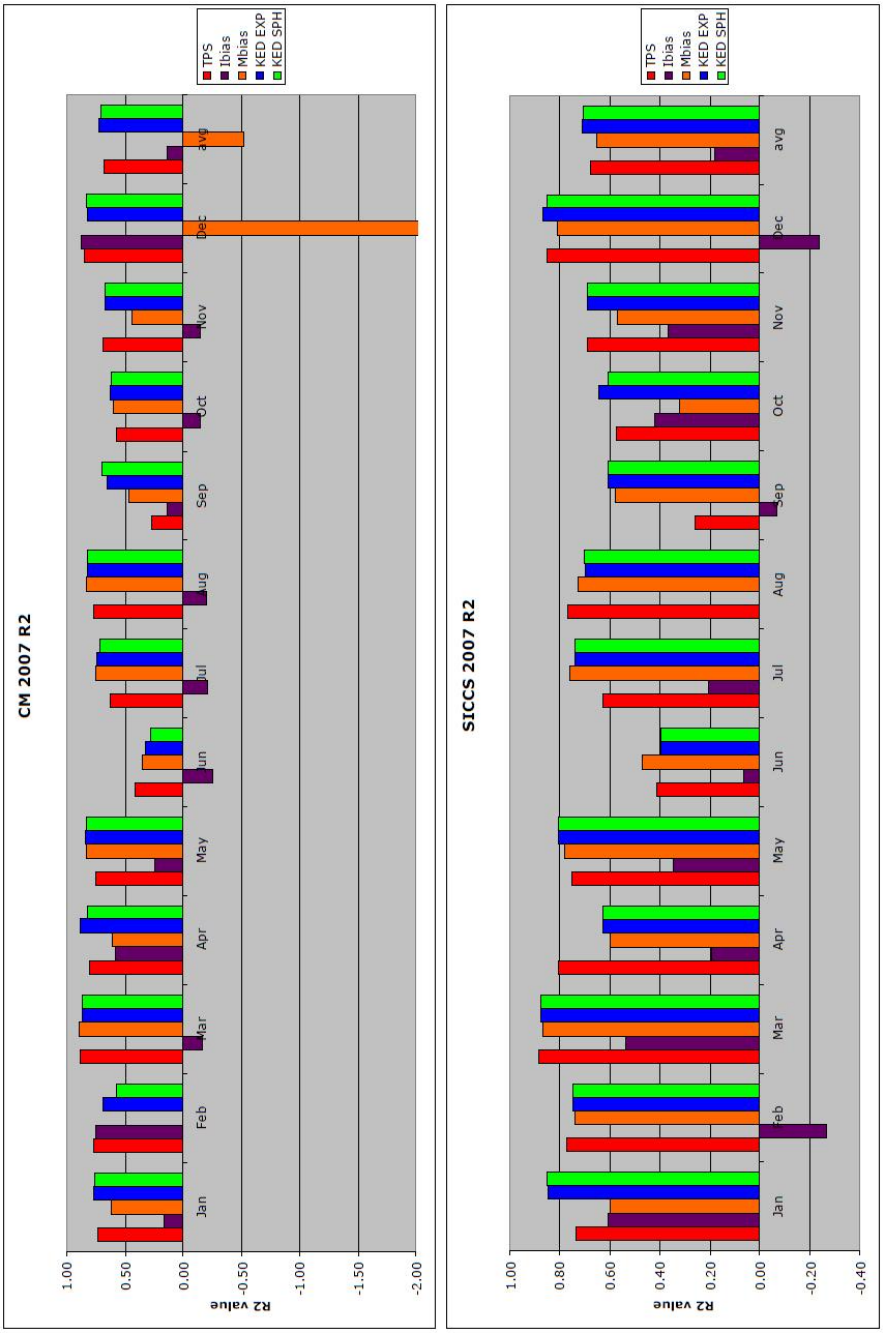


Figure 16: R^2 values for 2007 for both products.

.

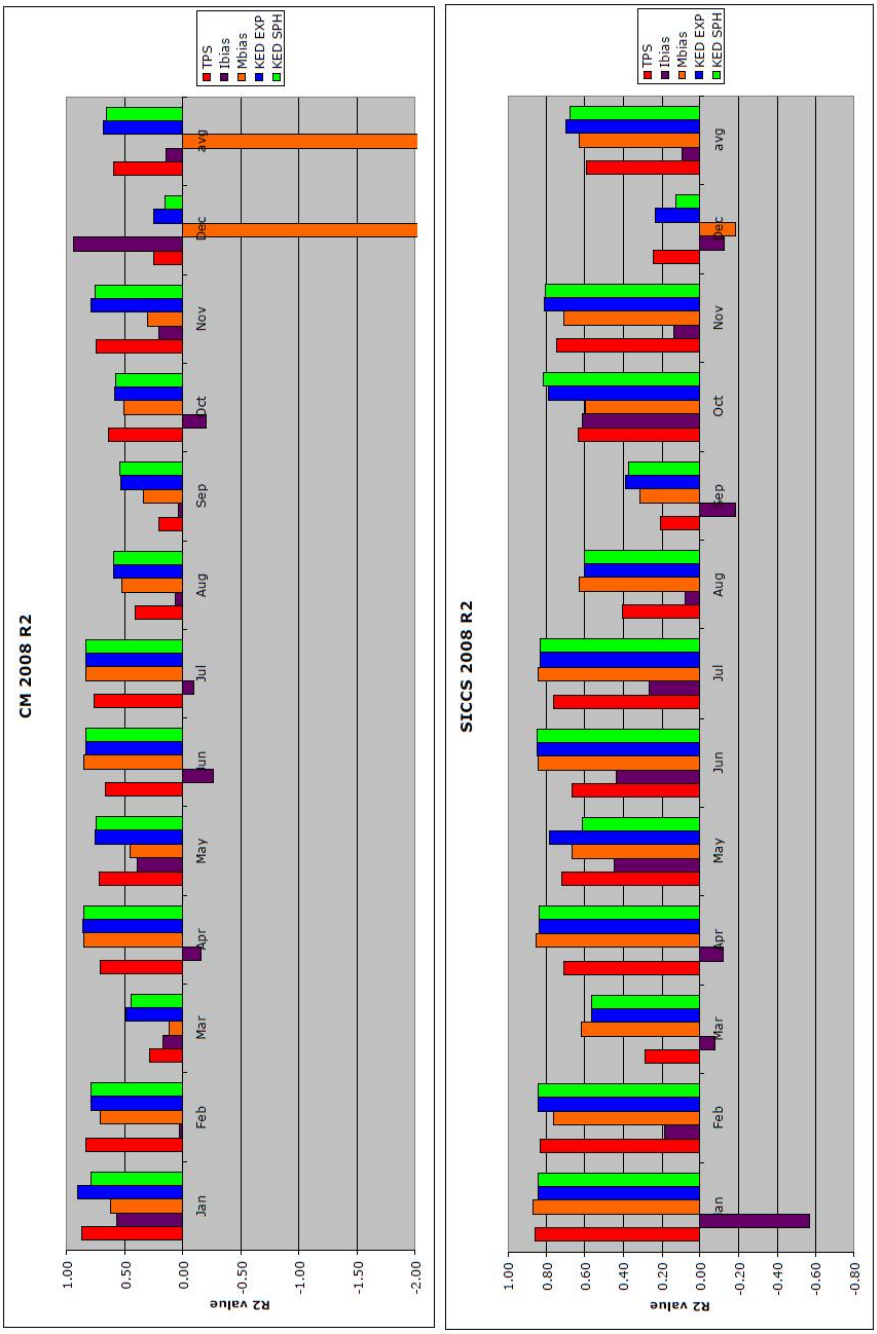


Figure 17: R^2 values for 2008 for both products.

.

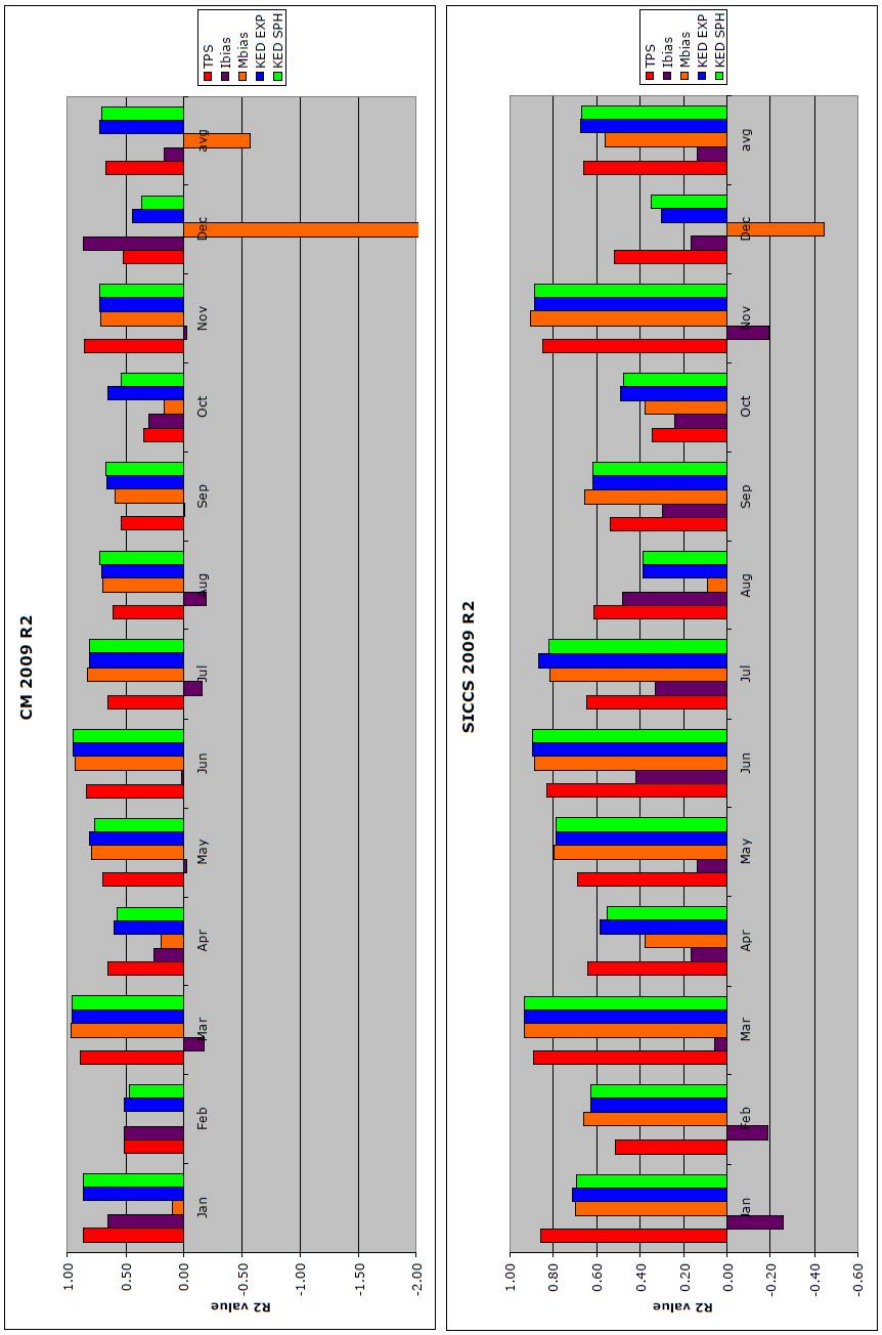


Figure 18: R^2 values for 2009 for both products.

.

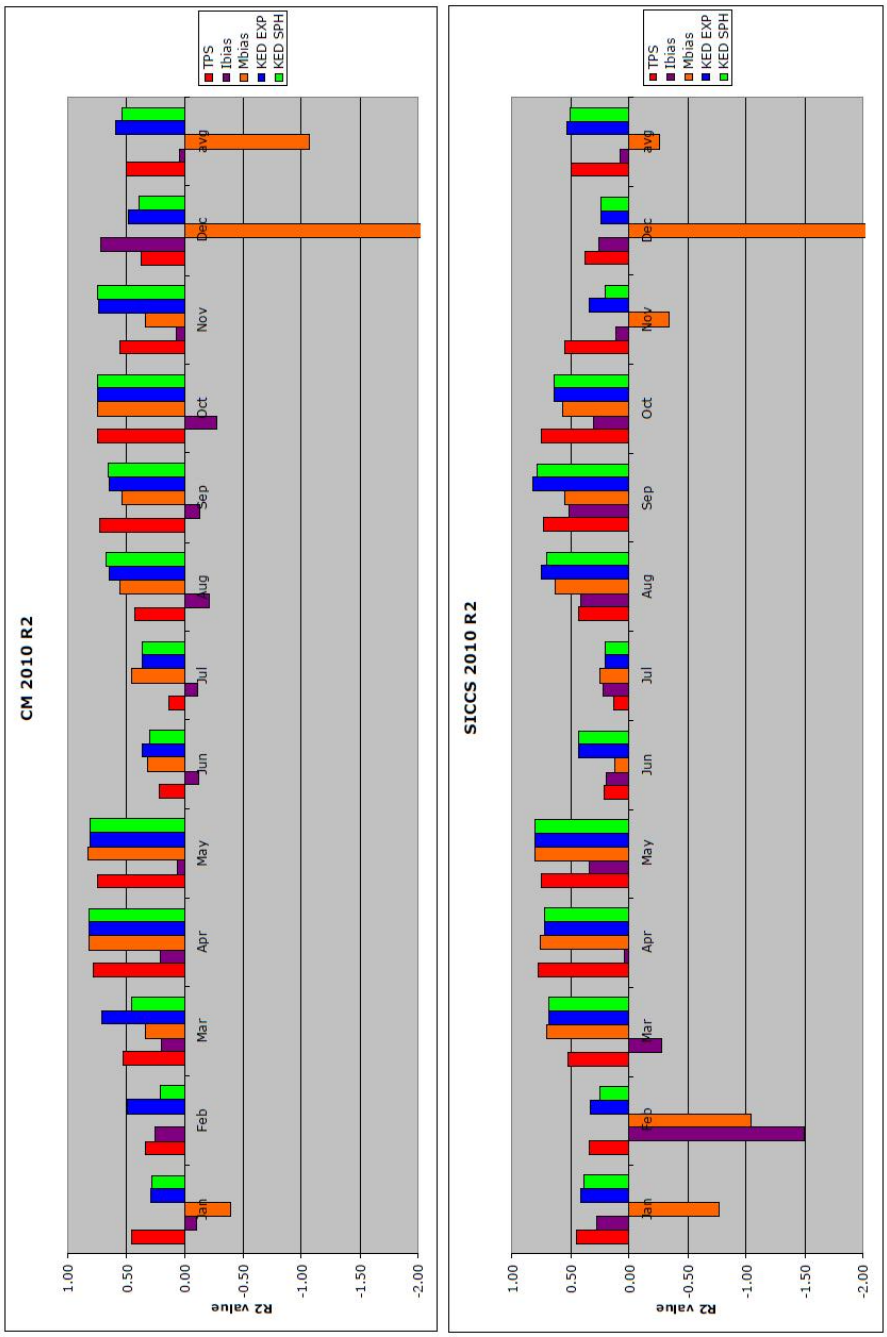


Figure 19: R^2 values for 2010 for both products.

.

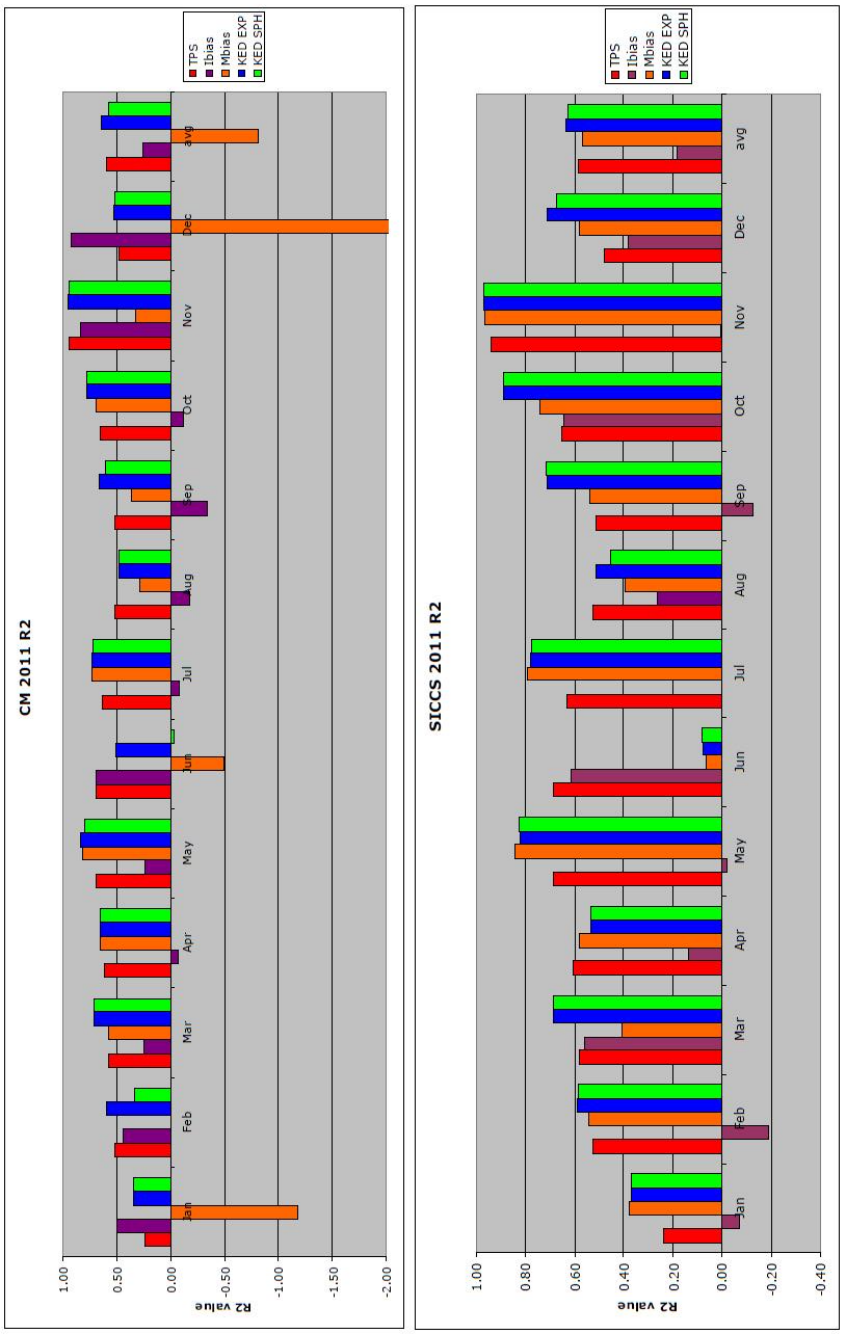


Figure 20: R^2 values for 2011 for both products.

.

The MAPE was low for all interpolation methods. The highest MAPE was observed in January 2010 in the CM-SAF product using the IB method and was 11.62%. The average errors for each year are shown in table 3. As shown in the table the highest average MAPE's were found for the MB and IB interpolations in the CM-SAF product. The MAPE's of TPS and KED interpolation methods were very low on the cross validated interpolations (3.02% at maximum).

CM-SAF	2006	2007	2008	2009	2010	2011
TPS	3.01	2.47	2.41	2.69	3.00	2.61
MB	3.54	3.18	2.98	3.12	4.23	3.03
IB	5.68	6.03	6.21	5.18	5.32	5.84
KED-Exp	2.63	2.36	2.09	2.59	2.76	2.40
KED-Sph	2.80	2.44	2.25	2.62	2.90	2.58
SICCS						
TPS	3.01	2.47	2.41	2.69	3.00	2.61
MB	2.60	2.25	2.12	2.56	4.40	2.10
IB	2.68	2.49	2.17	2.74	4.54	2.46
KED-Exp	2.67	2.27	2.00	2.52	2.92	2.21
KED-Sph	2.80	2.29	2.06	2.55	3.02	2.25

Table 3: Average MAPE values for the interpolations per year

.

These low mean absolute percentage error (MAPE) values indicated that the interpolations predicted the global radiation very well when using a cross validation method. This ment that the values obtained by the interpolations were very accurate and therefore all interpolations could be used. Figure 21 to 26 show the MAPE for each month for both products.

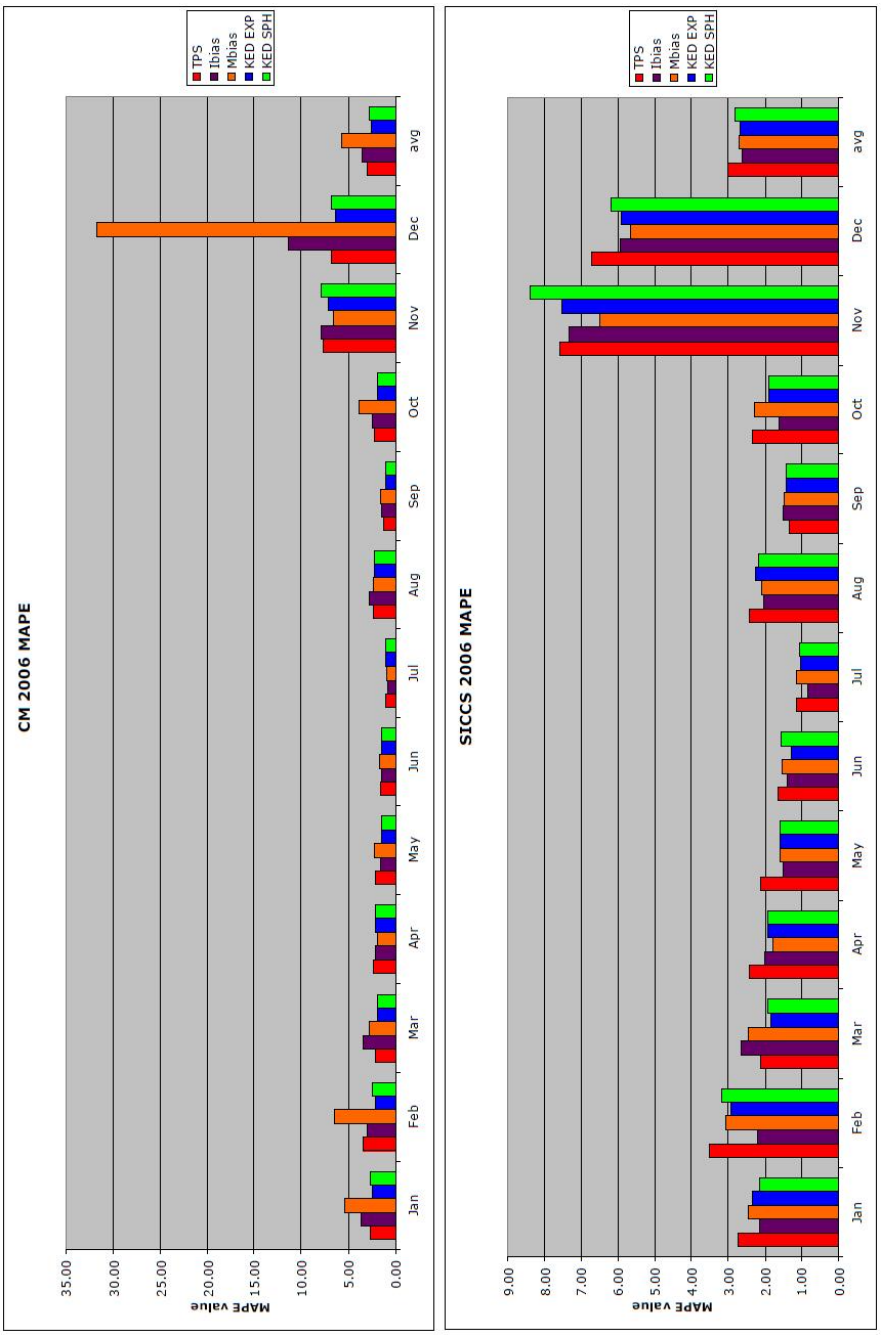


Figure 21: MAPE values for 2006 for both products.

.

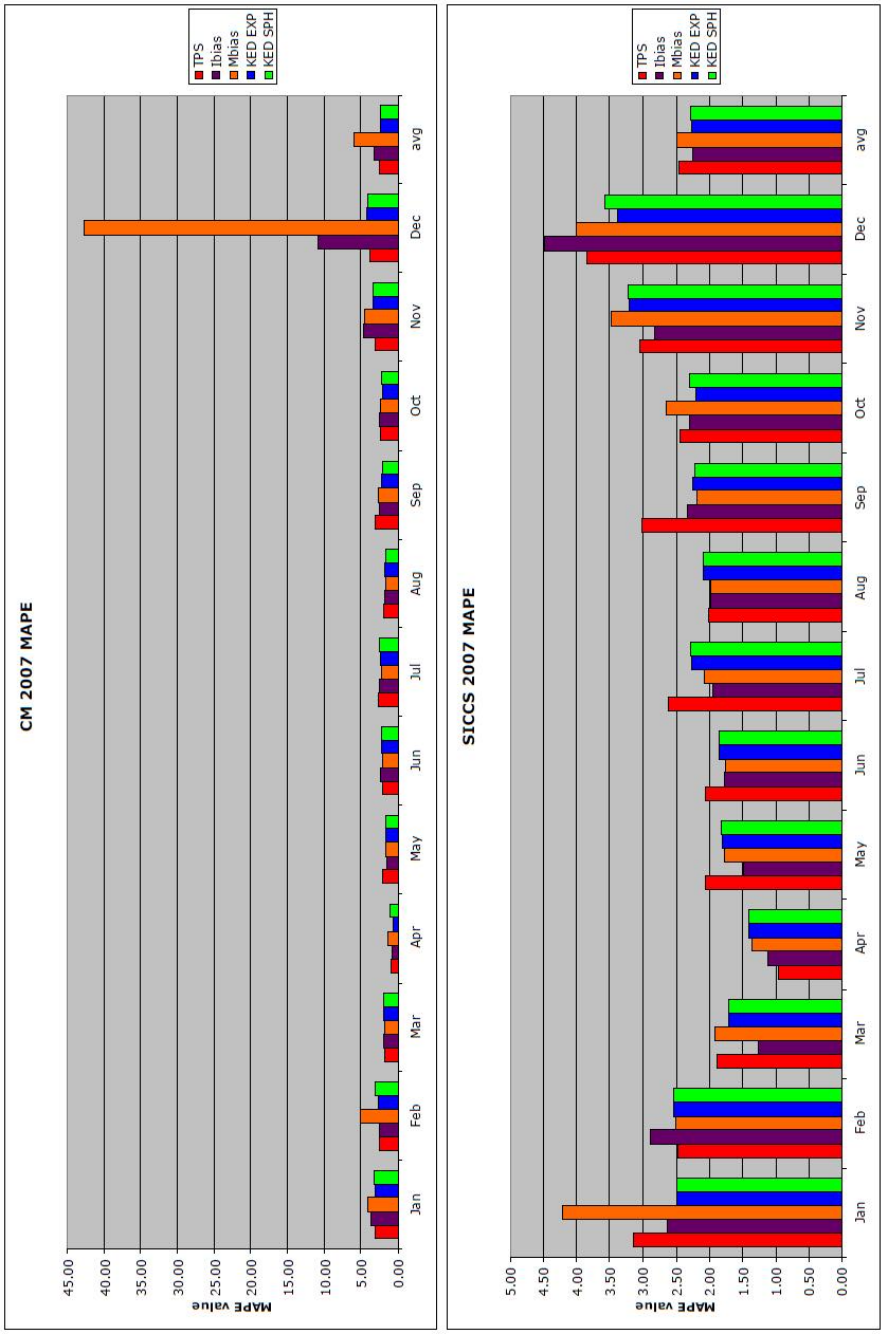


Figure 22: MAPE values for 2007 for both products.

.

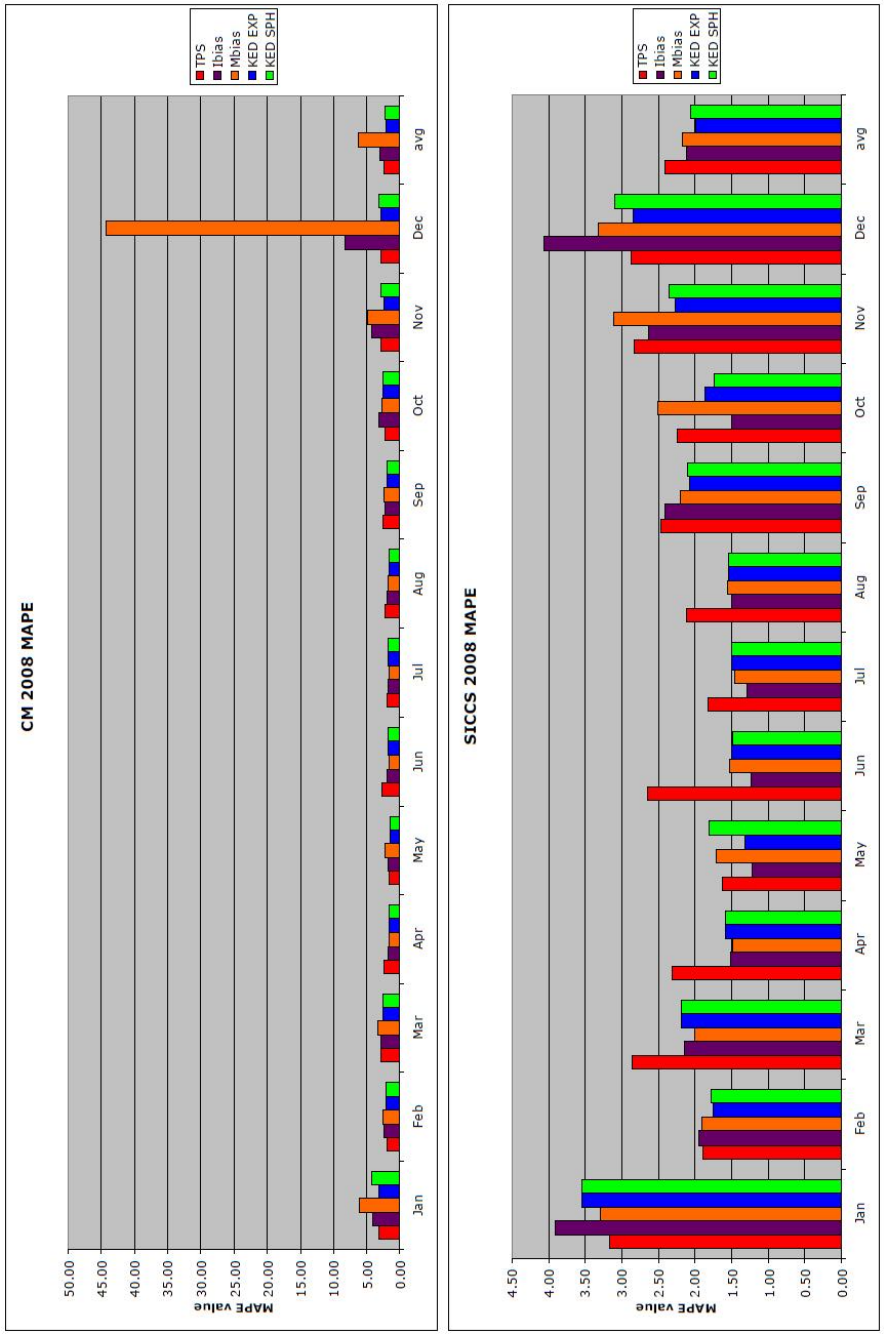


Figure 23: MAPE values for 2008 for both products.

.

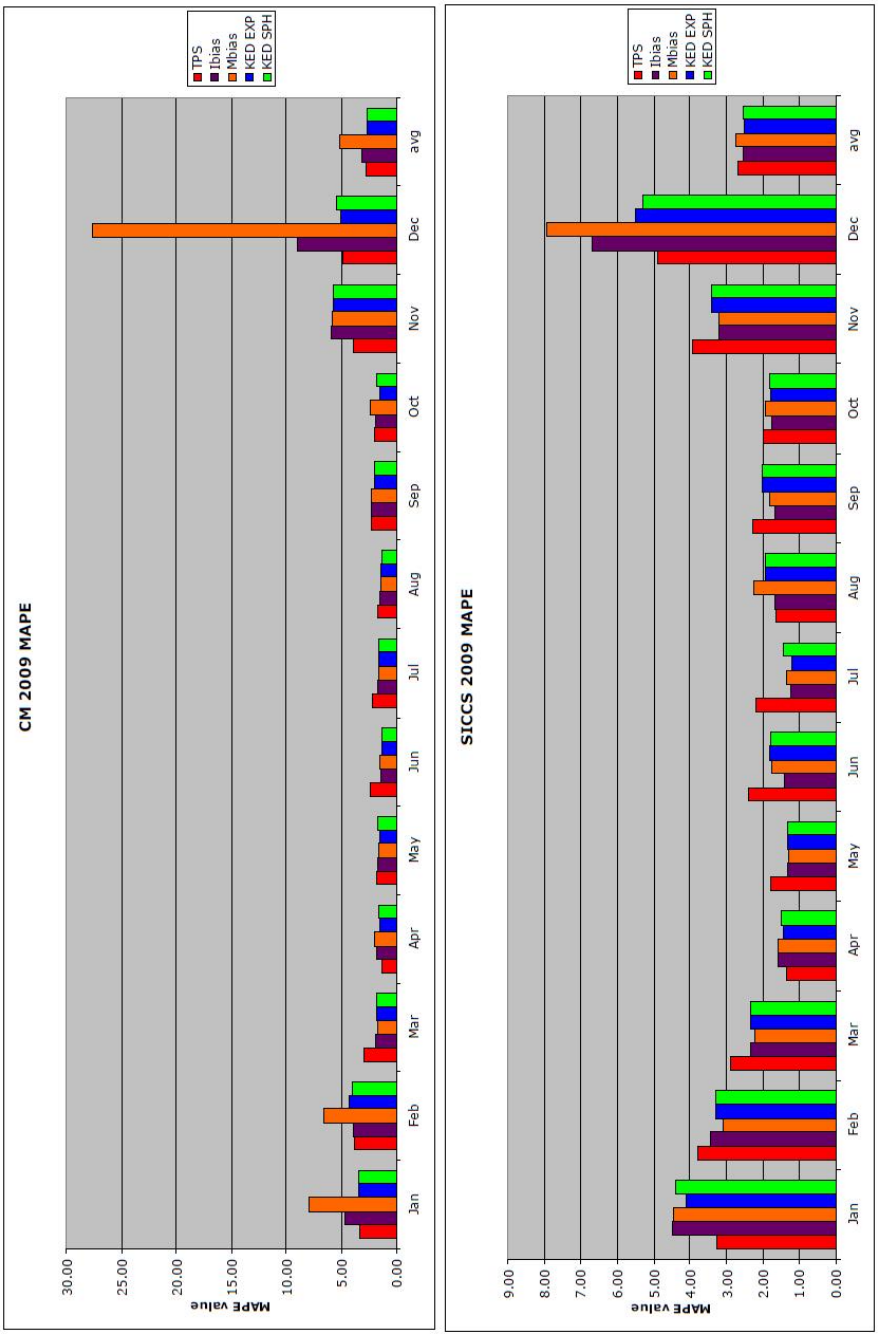


Figure 24: MAPE values for 2009 for both products.

.

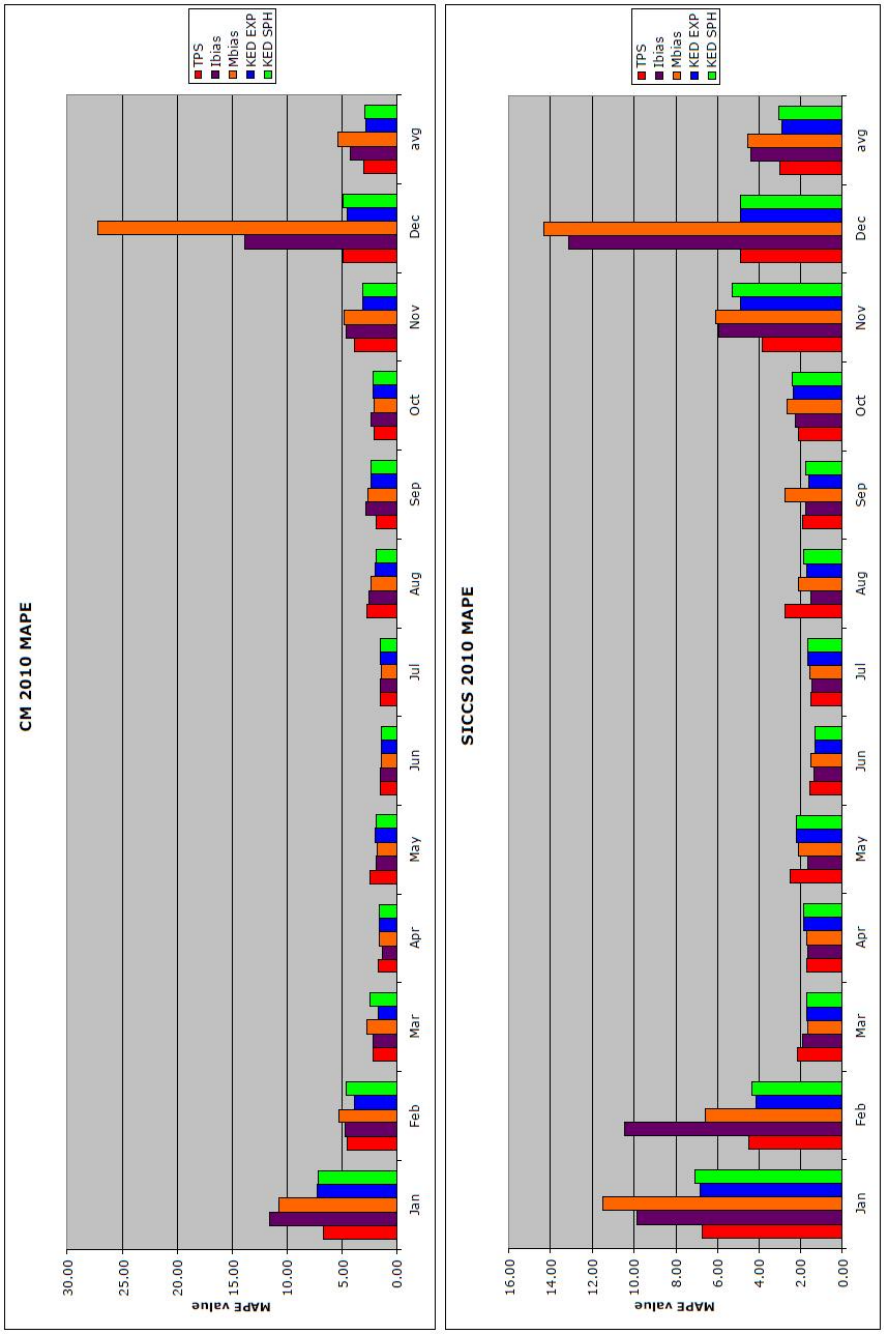


Figure 25: MAPE values for 2010 for both products.

.

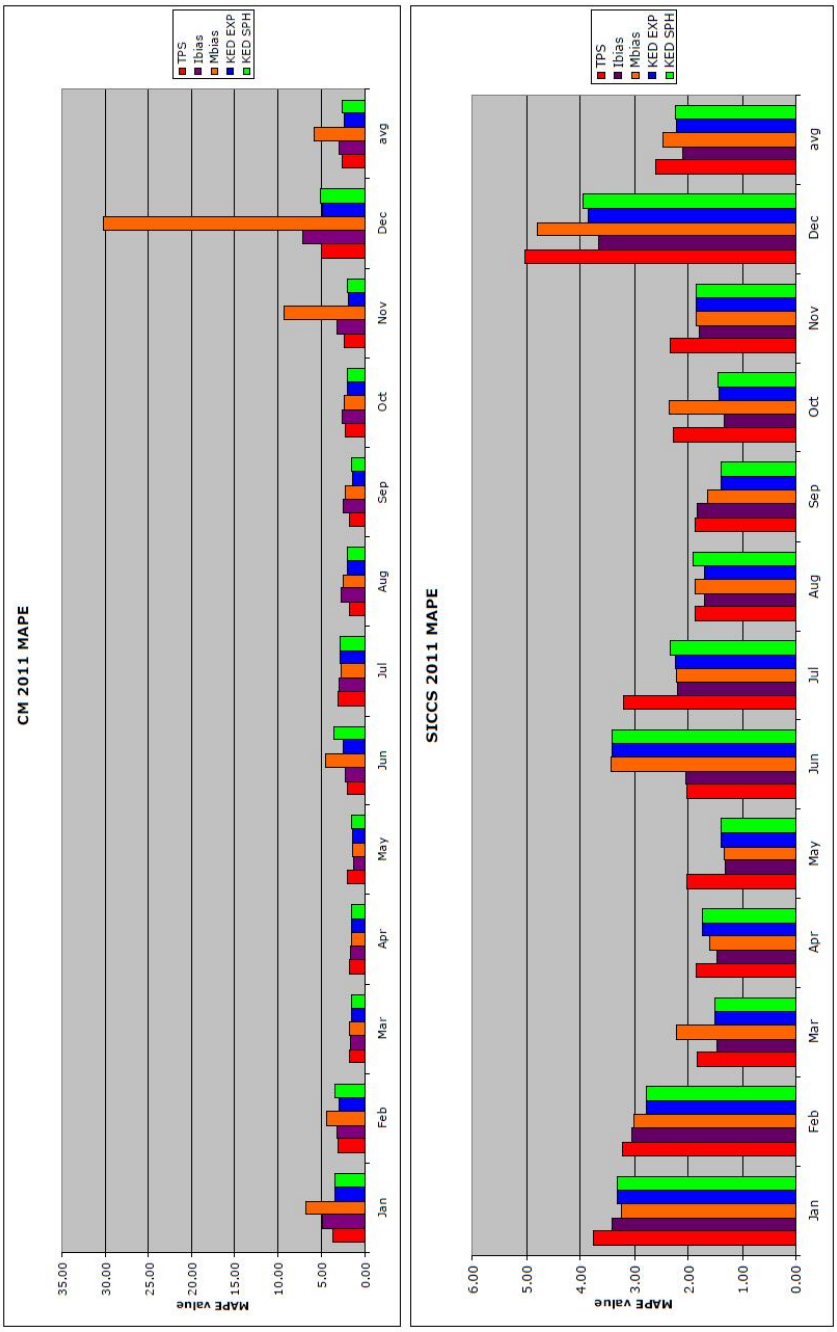


Figure 26: MAPE values for 2011 for both products.

.

Data Split

Due to the fact that all interpolation methods performed relatively well, especially on the station locations, other forms of validation tests were performed. In a next evaluation test, 8 stations were left out as a form of data splitting. 8 stations is 1/4th of the data which was relatively much due to the fact that only 32 observations were available. Figure 27 shows which stations were left out. First a data split with stations spread across the country was performed. After that a data split with station on the western side of the country was performed. This was done to see how the interpolation performed if the data was not evenly distributed across the country.

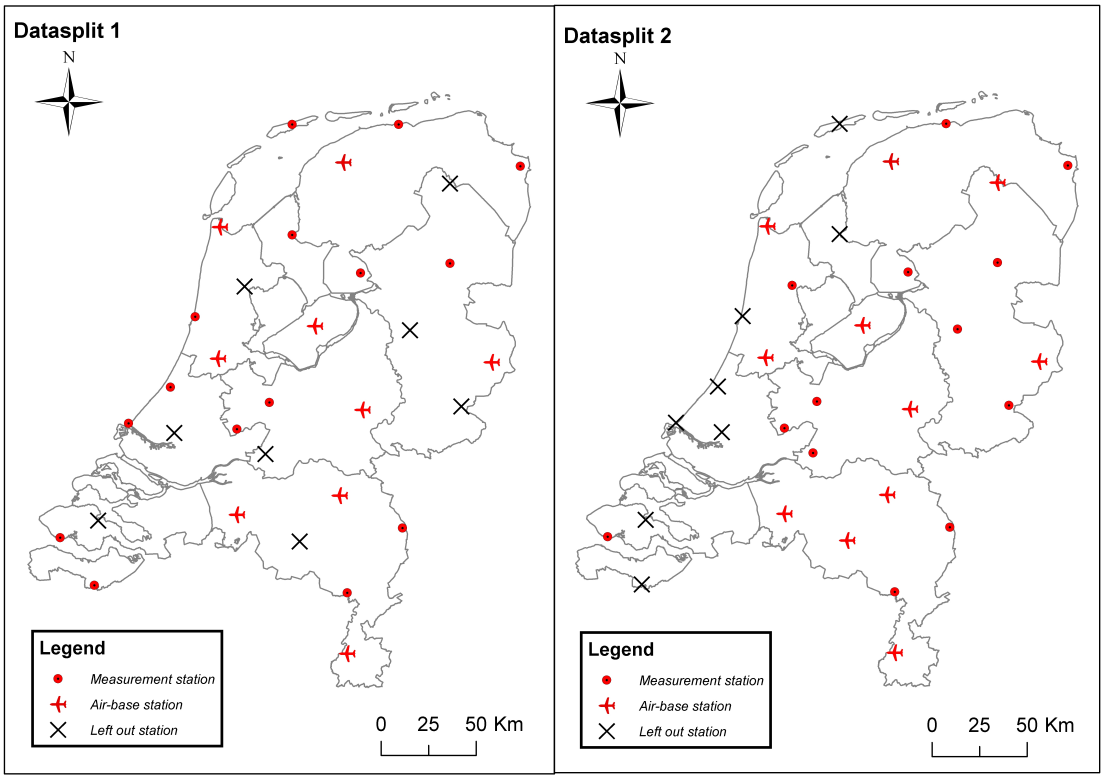


Figure 27: Stations that were left out in the data split are shown with a cross. The left map is the map for the first data split and the right map is for the second data split.

.

When looking at the results of this data split (figure 28 to 31) it seemed that the interpolated values yet again come very close to the in-situ measurements of the stations. The maximum MAPE was 7% in January for the CM-SAF product. This error was found in the MB interpolation method.

Comparing the products with each other showed that the SICCS product performed a little better on average. The difference between the products were however extremely small with differences less than 1 to 2%.

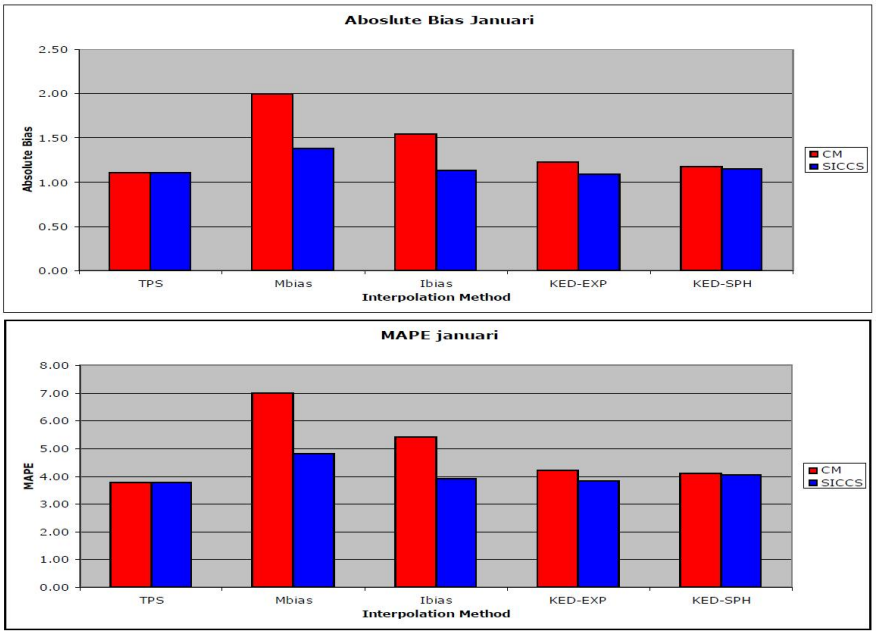


Figure 28: Results for the first data split in Januari.

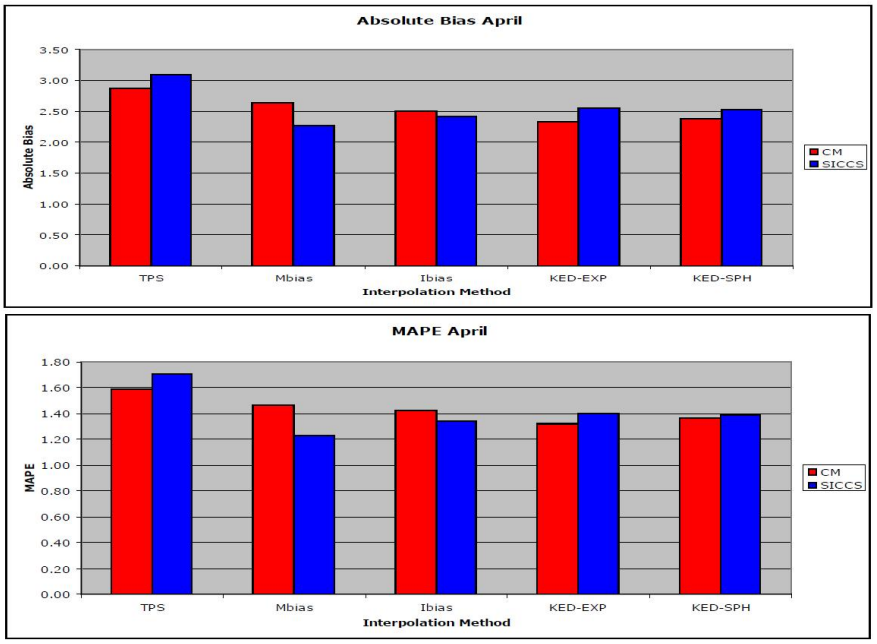


Figure 29: Results for the first data split in April.

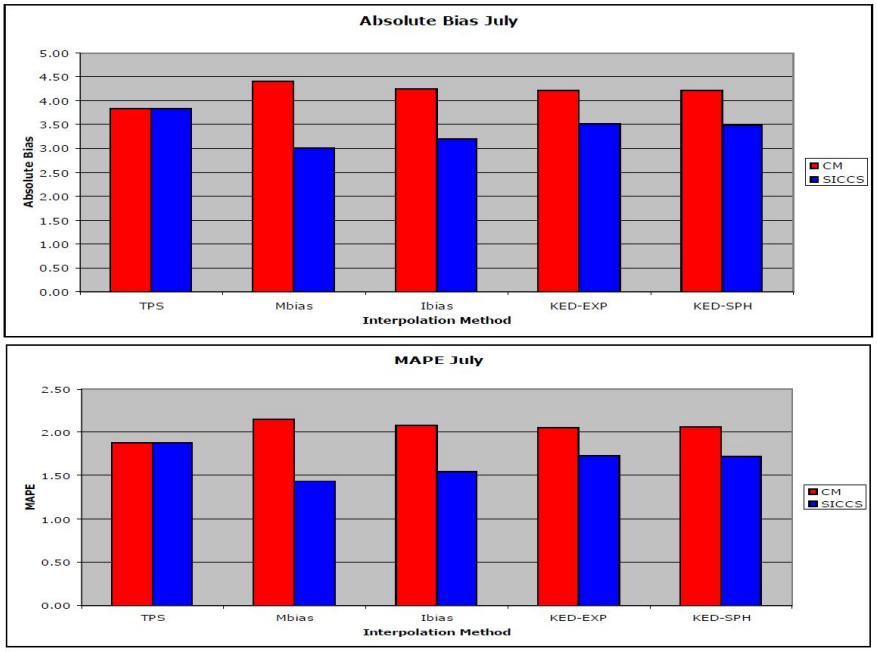


Figure 30: Results for the first data split in July.

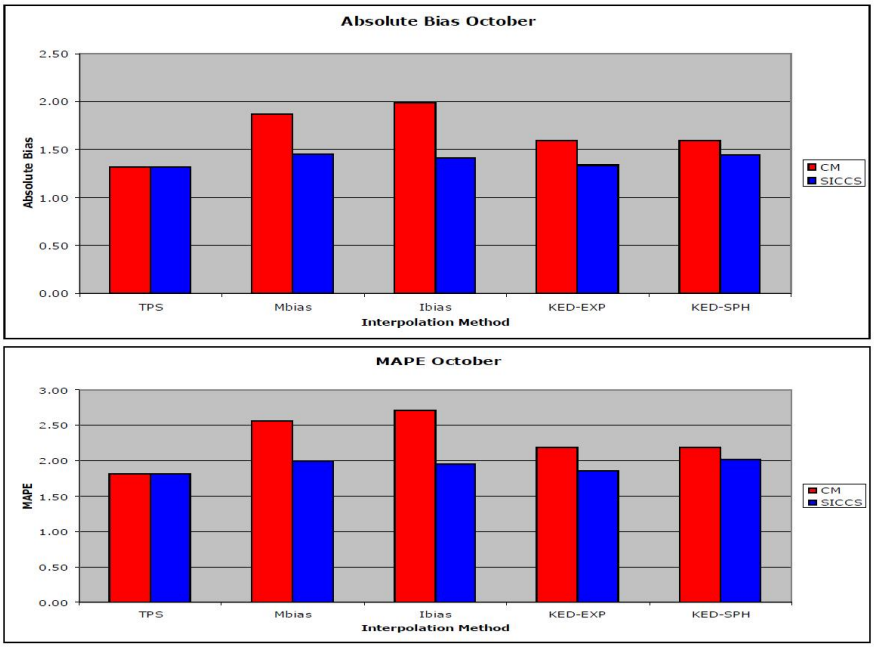


Figure 31: Results for the first data split in October.

The average absolute bias and MAPE of the second data split are presented in figure 32 to 35.The errors of the interpolation did start to increase now. The biggest error was yet again found in January, in the MB interpolation for the CM-SAF product with an average MAPE of 9.27%.

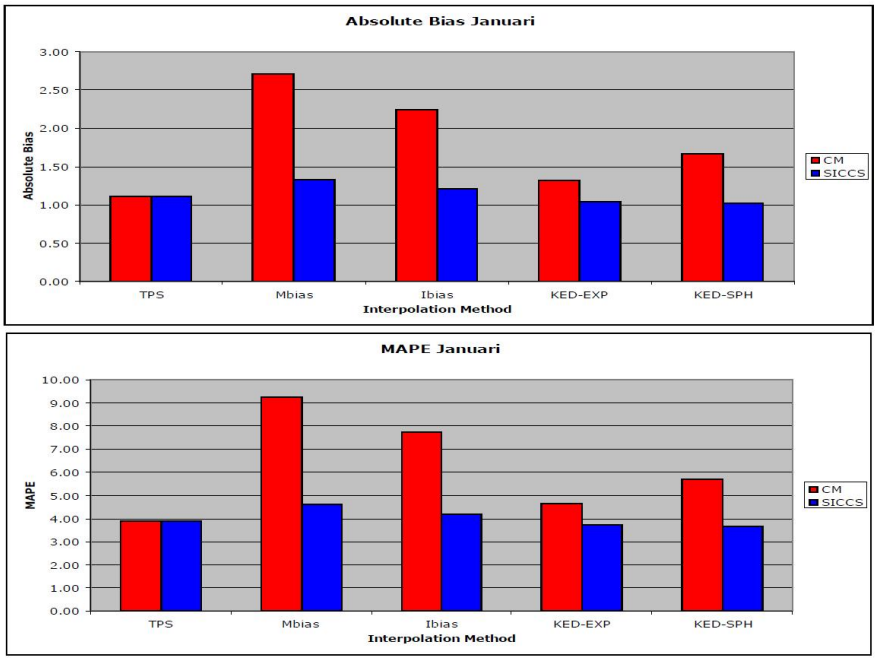


Figure 32: Results for the second data split in Januari.

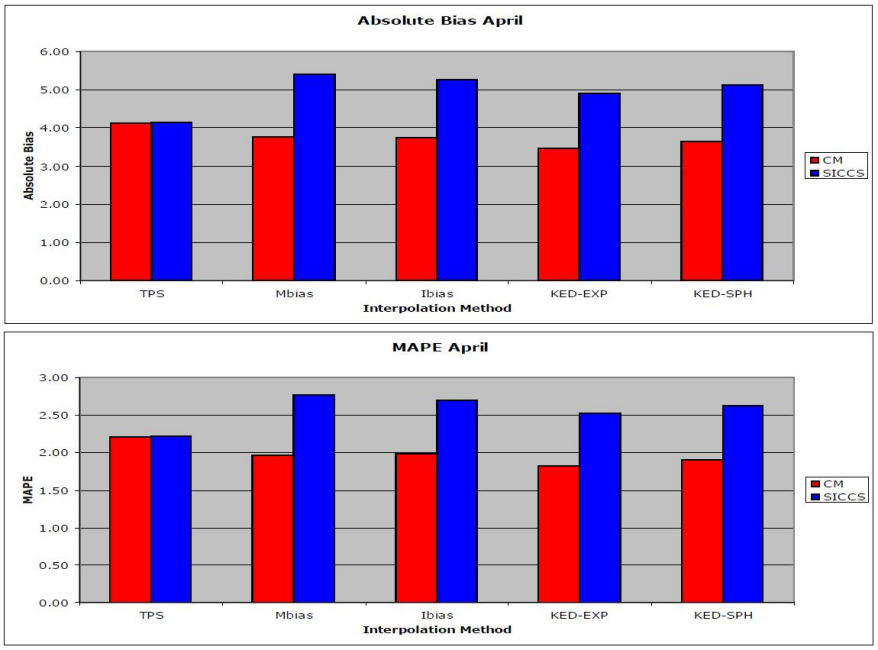


Figure 33: Results for the second data split in April.

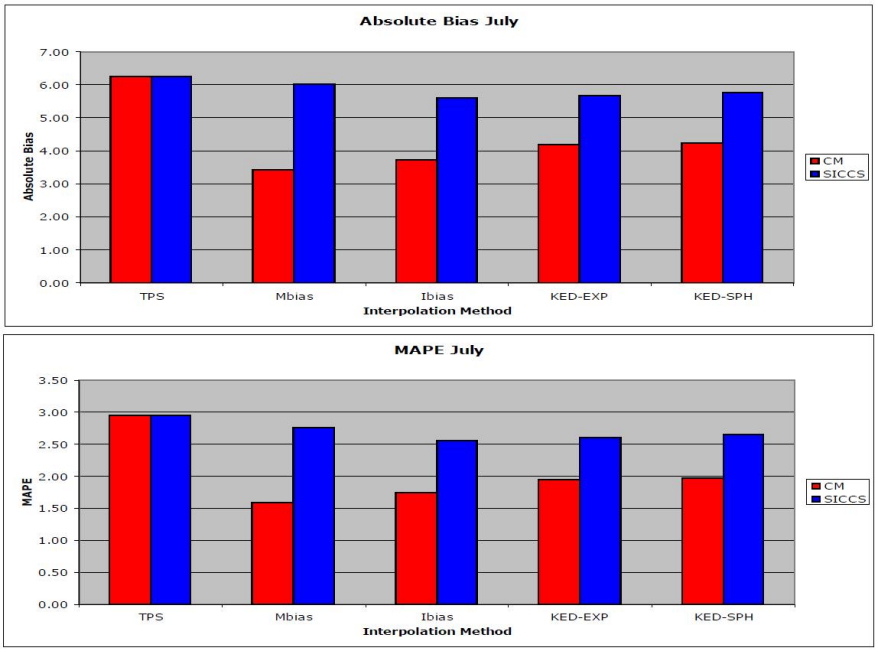


Figure 34: Results for the second data split in July.

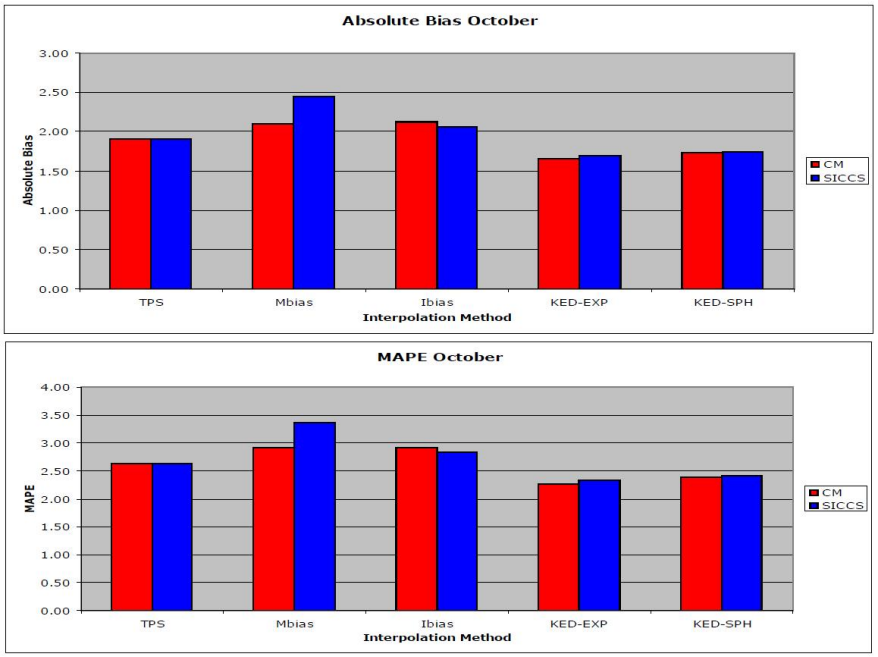


Figure 35: Results for the second data split in October.

Random Points

Since the interpolation kept performing this well at the station locations a validation on random locations has been performed.

For each year January, April, July and October were used as input data. On this way each season in each year was represented. The output values for each interpolation method at the random locations were then compared with the value of the satellite images. Figure 36 shows the locations of the random points used to compare the values.

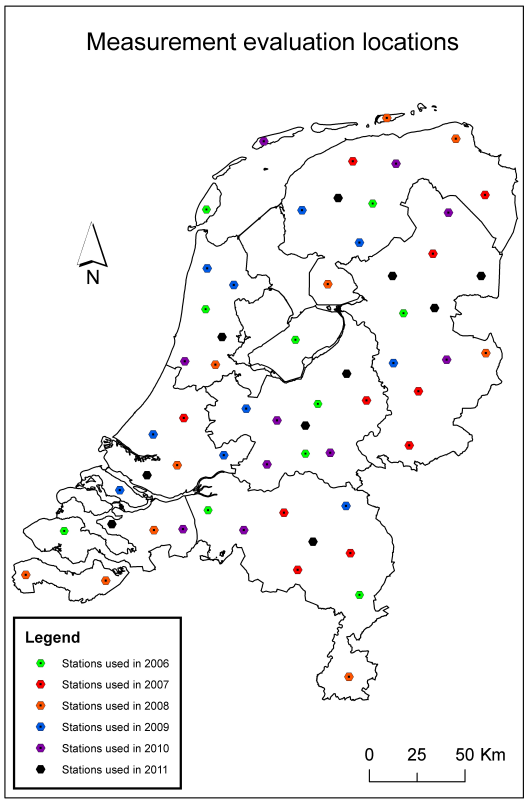


Figure 36: Random locations used between stations to evaluate the satellite image versus the interpolations.

.

The results are presented as line graphs below (figure 45 to 52). The black line represents the values obtained by the interpolation. The red lines represent the satellite image including the bias which was allowed/observed according to the validation of the satellite product. This means that the CM-SAF product has a positive bias of 4 W/m^2 and a negative bias of 4 W/m^2 which adds up to the target of 8 W/m^2 (CM-SAF 2013). For the SICCS product the positive bias was set to 3.5 W/m^2 and the negative bias to minus 3.5 W/m^2, which adds up to the observed bias of 7 W/m^2 (Gruell et al., 2013). In this case the assumption was made, that the bias that both products account for are evenly spread on both the positive and negative side and fall within the optimal target accuracy of the product.

If the interpolation performed well the black line should stay between the red lines. In that case it did not pass the accepted bias (assuming the satellite values are correct). In order to make it easier to asses the output, trend lines were added. The orange trend lines are the trends that correspond with the positive and negative bias. The green line is the trend line corresponding with the interpolation. All trend lines are third order polynomials. Third order polynomials were chosen because they do capture the variation in the data without creating to many extremes on the edges. To evaluate the data, the trend line from the interpolation method should stay between the trend lines from the satellite bias. If the line stays exactly in the middle the interpolation performs best.

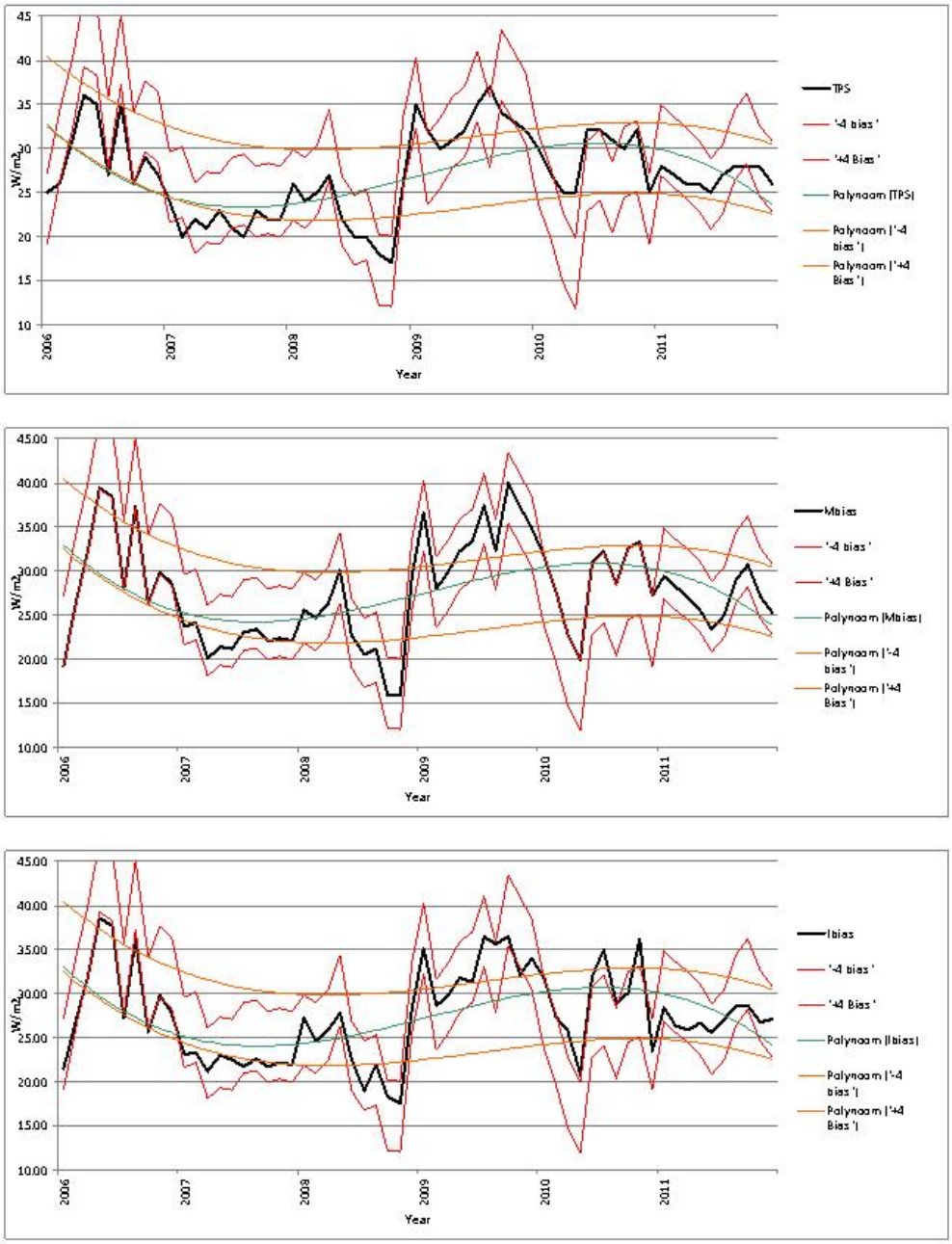


Figure 37: Random point analysis Januari CM-SAF.

.

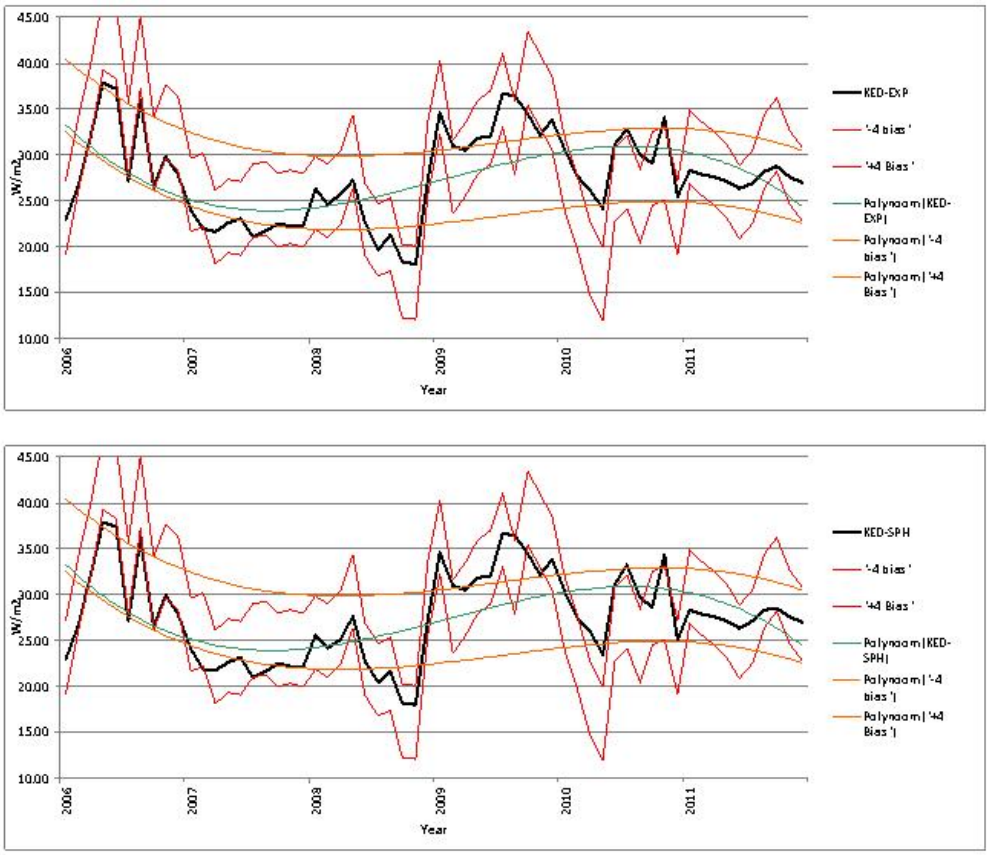


Figure 38: Random point analysis Januari CM-SAF.

.

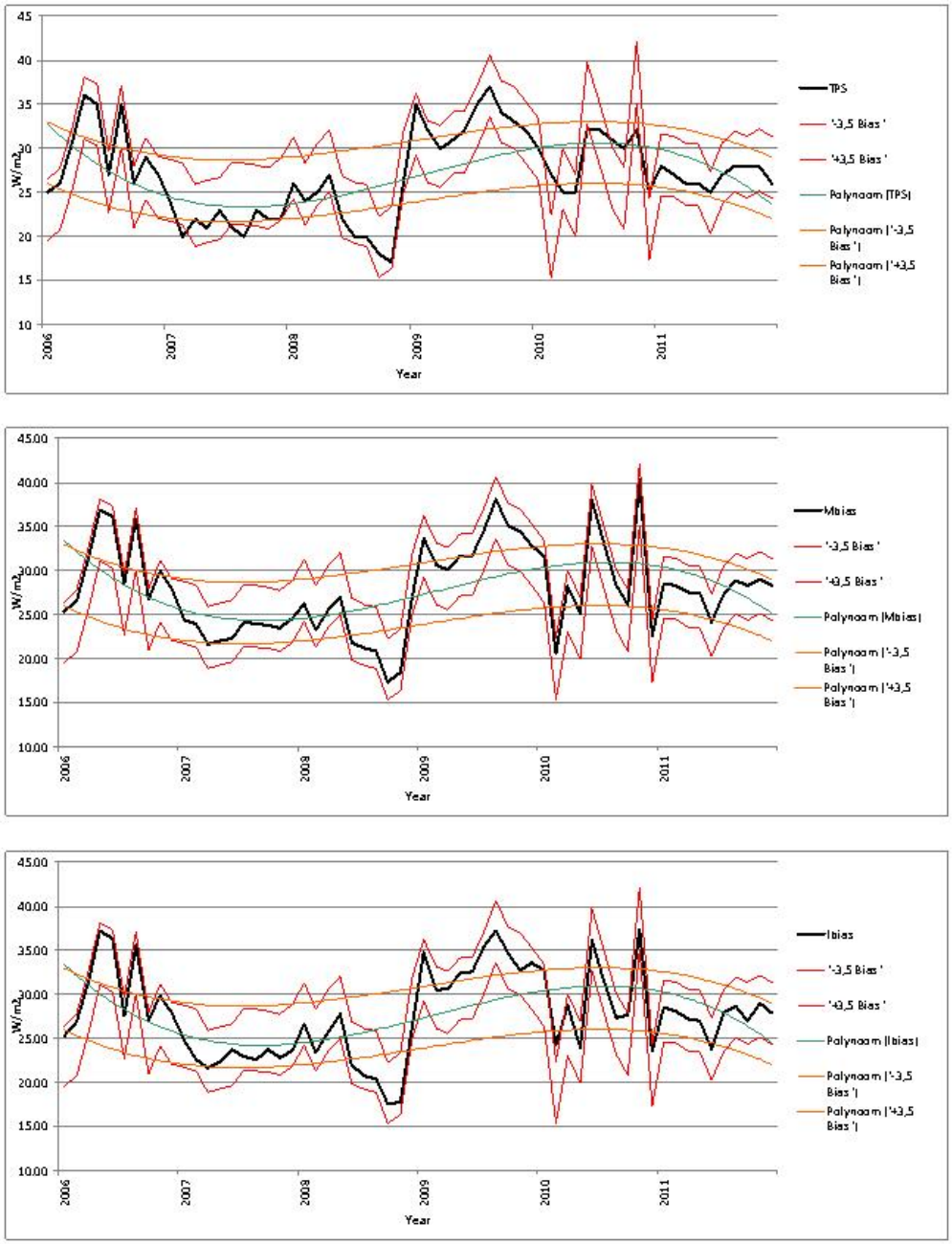


Figure 39: Random point analysis Januari SICCS.

.

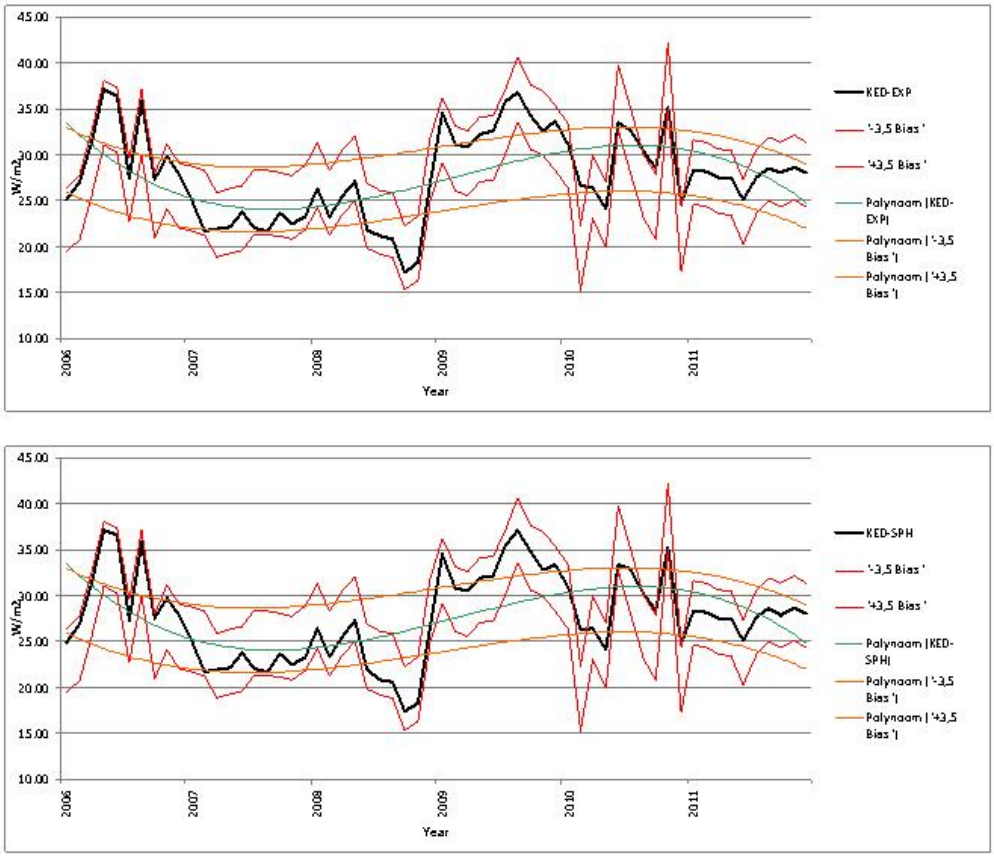


Figure 40: Random point analysis Januari SICCS.

.

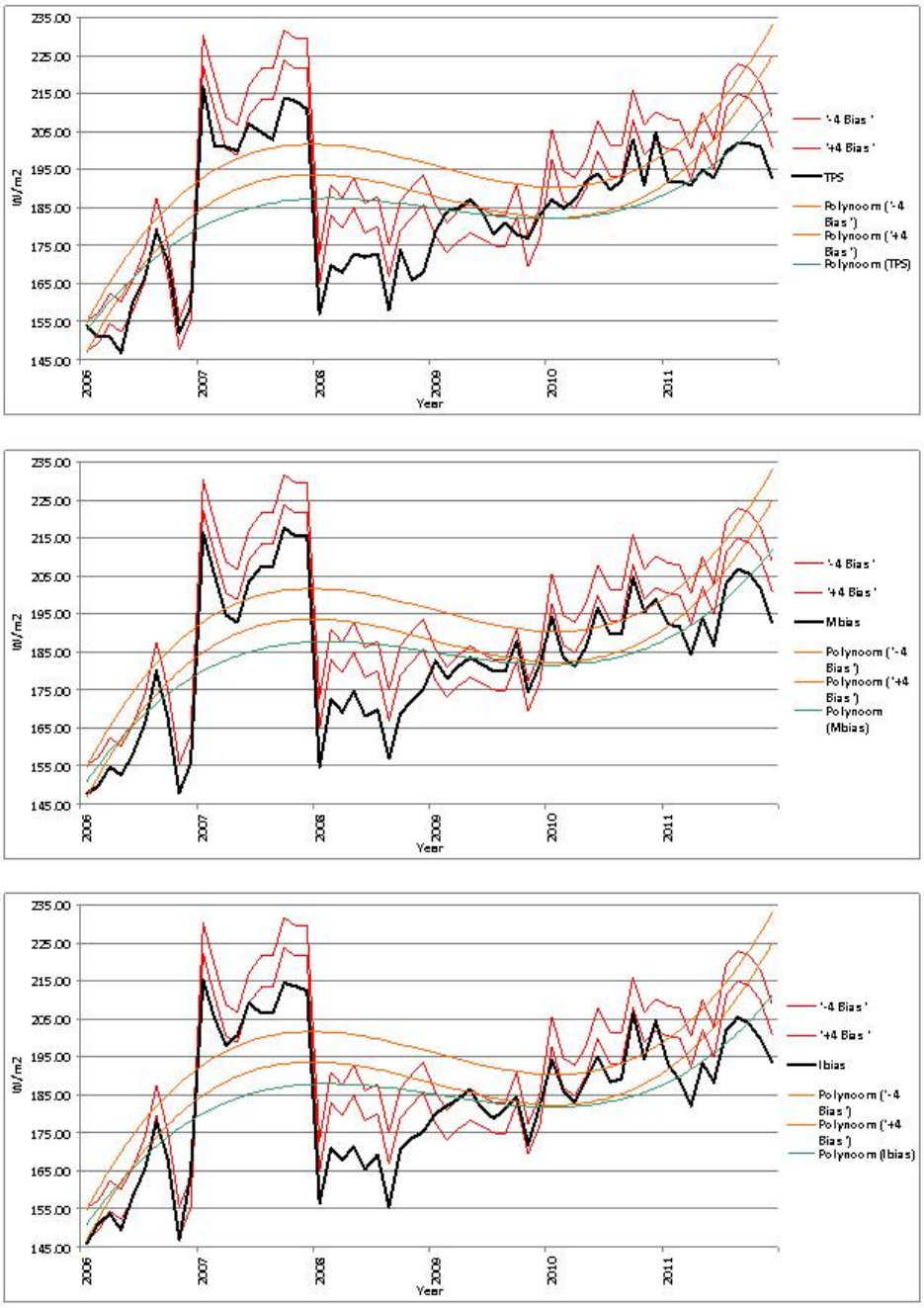


Figure 41: Random point analysis April CM-SAF.

.

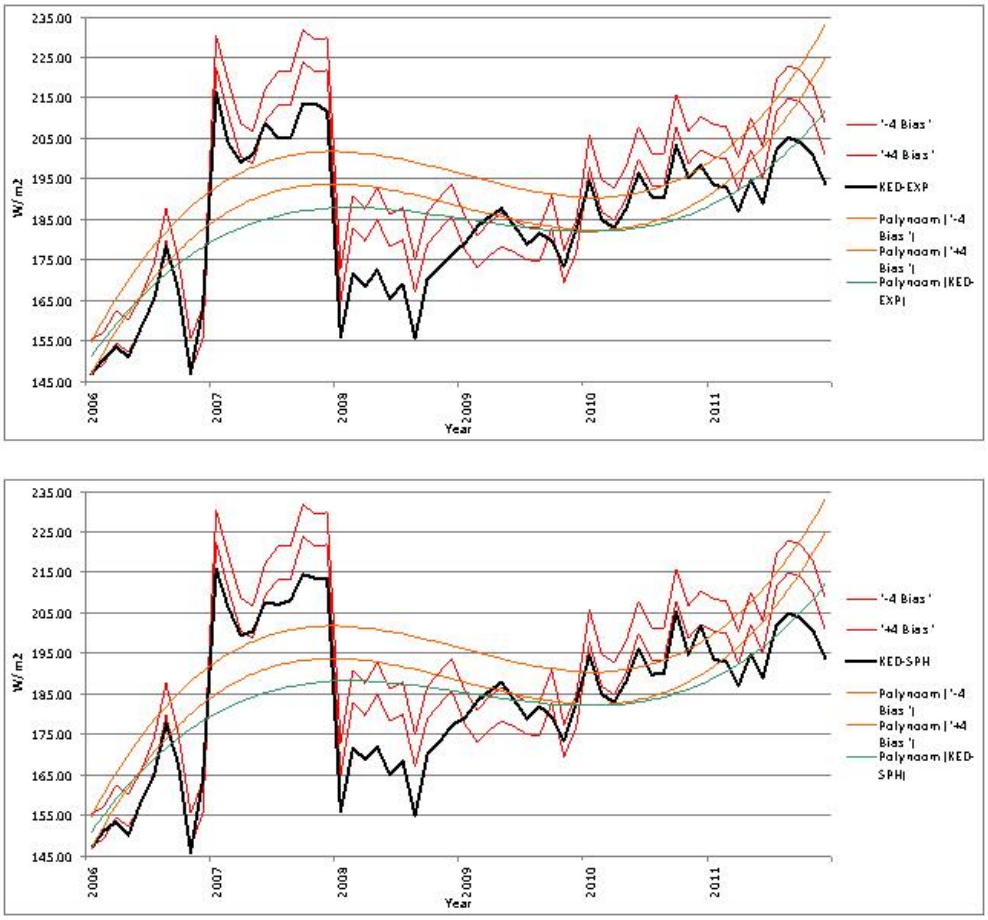


Figure 42: Random point analysis April CM-SAF.

.

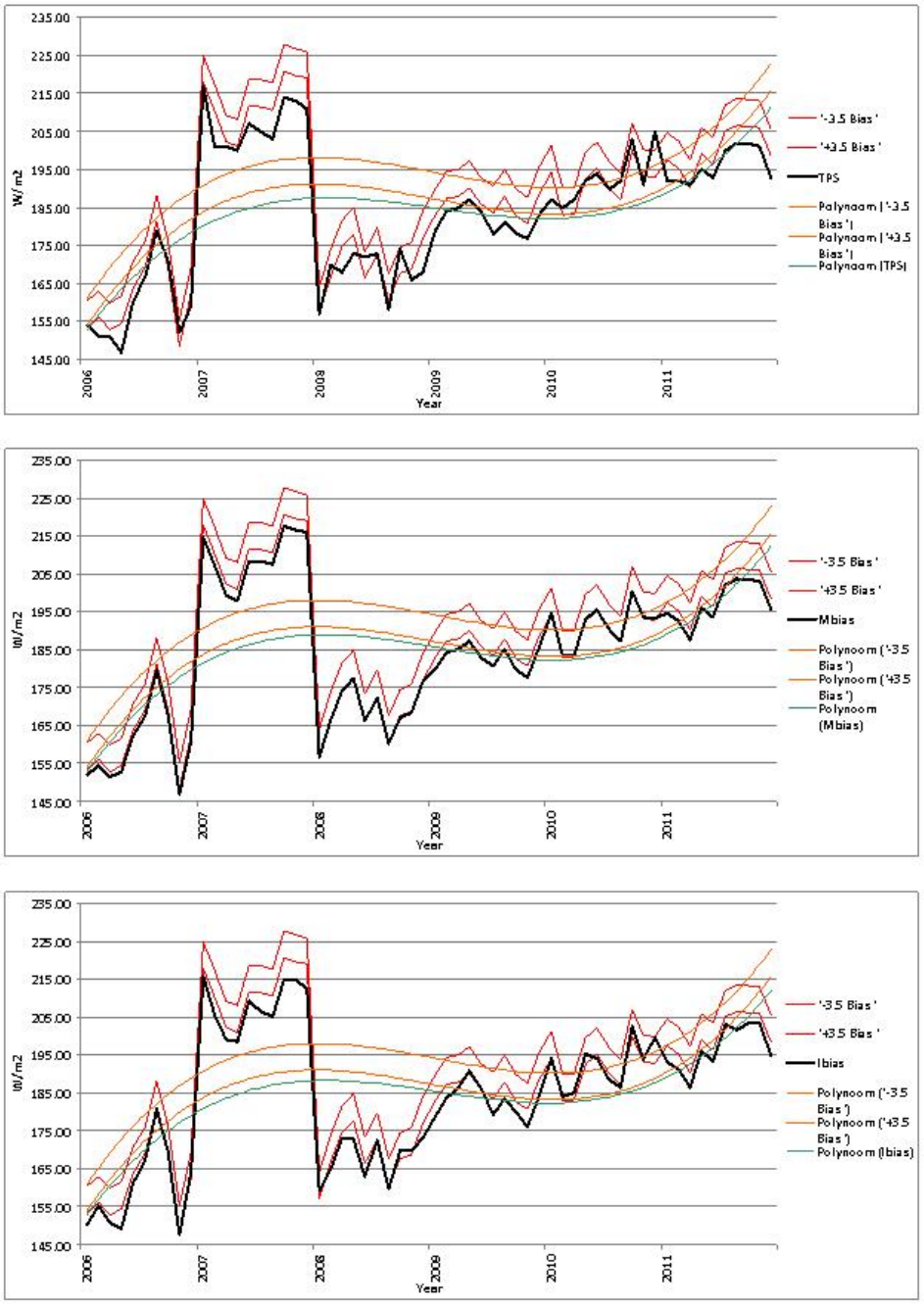


Figure 43: Random point analysis April SICCS.

.

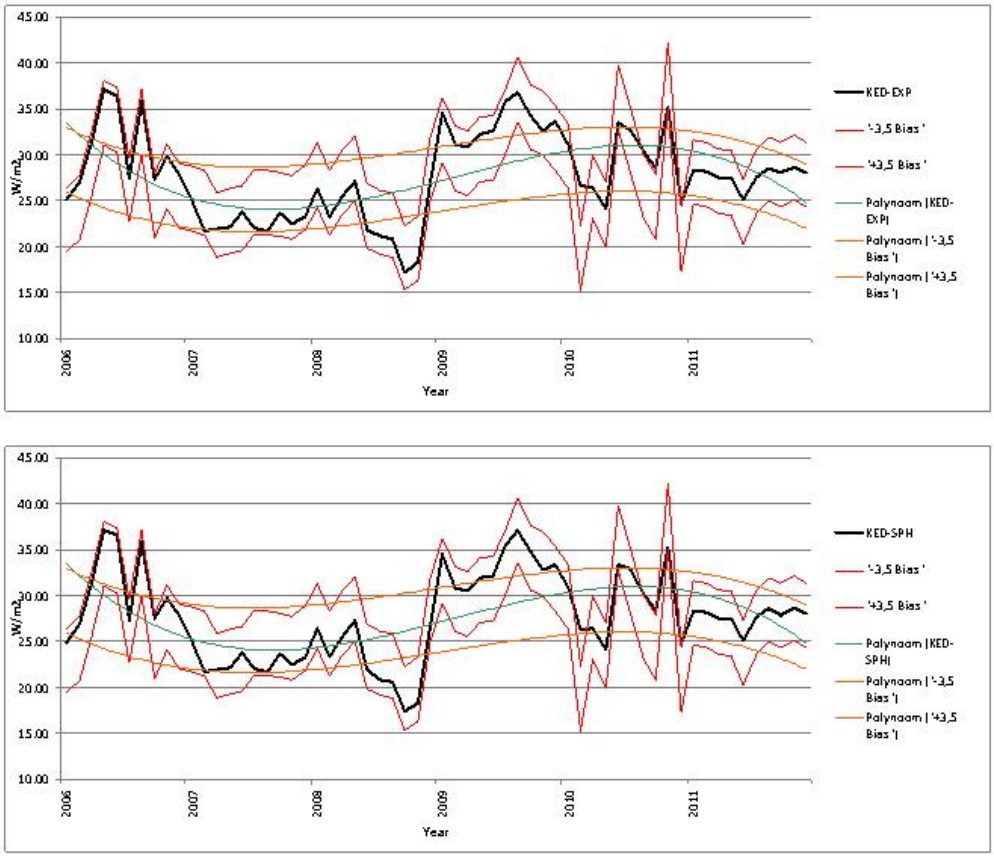


Figure 44: Random point analysis April SICCS.

.

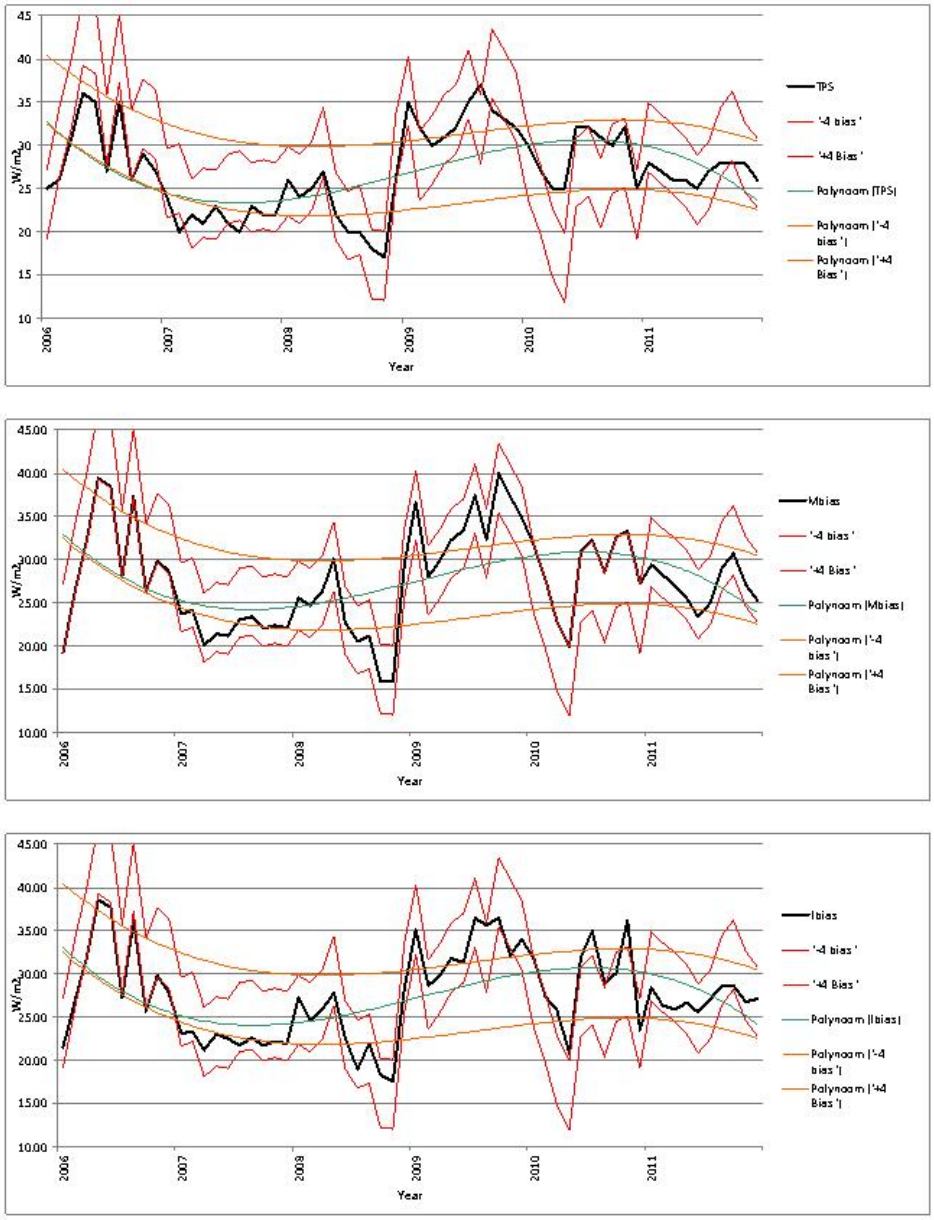


Figure 45: Random point analysis July CM-SAF.

.

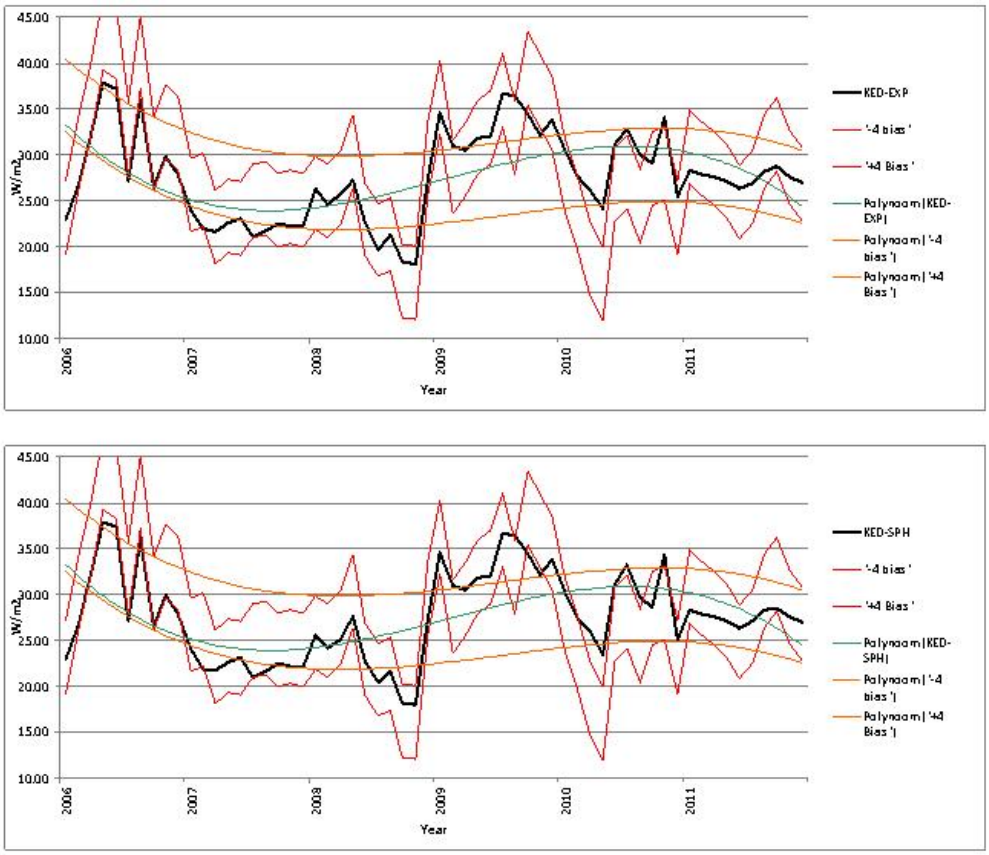


Figure 46: Random point analysis July CM-SAF.

.

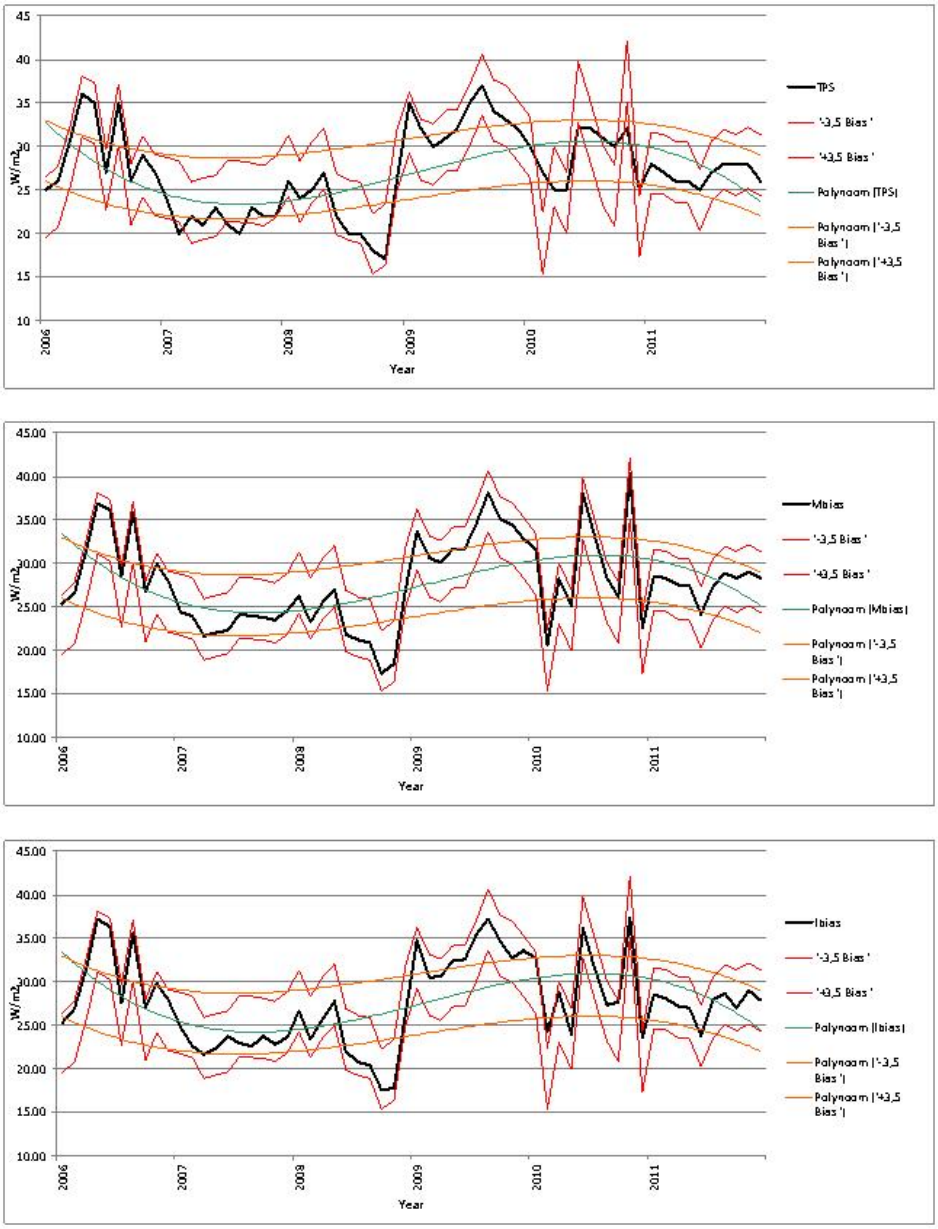


Figure 47: Random point analysis July SICCS.

.

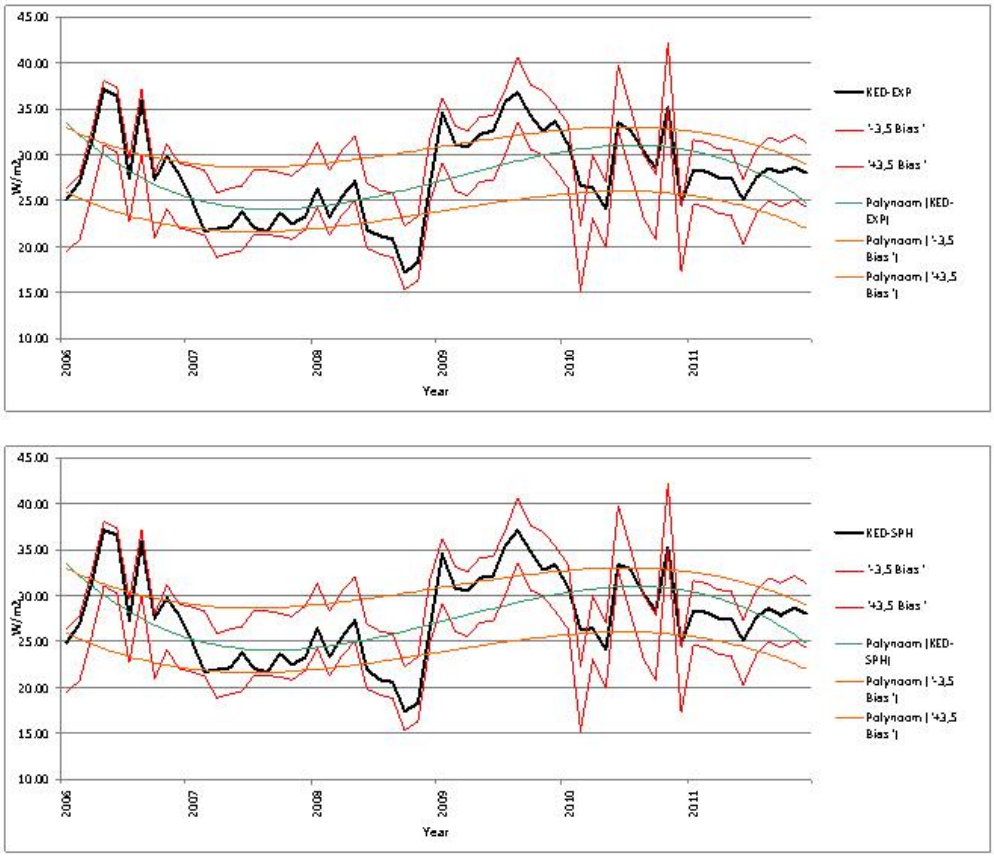


Figure 48: Random point analysis July SICCS.

.

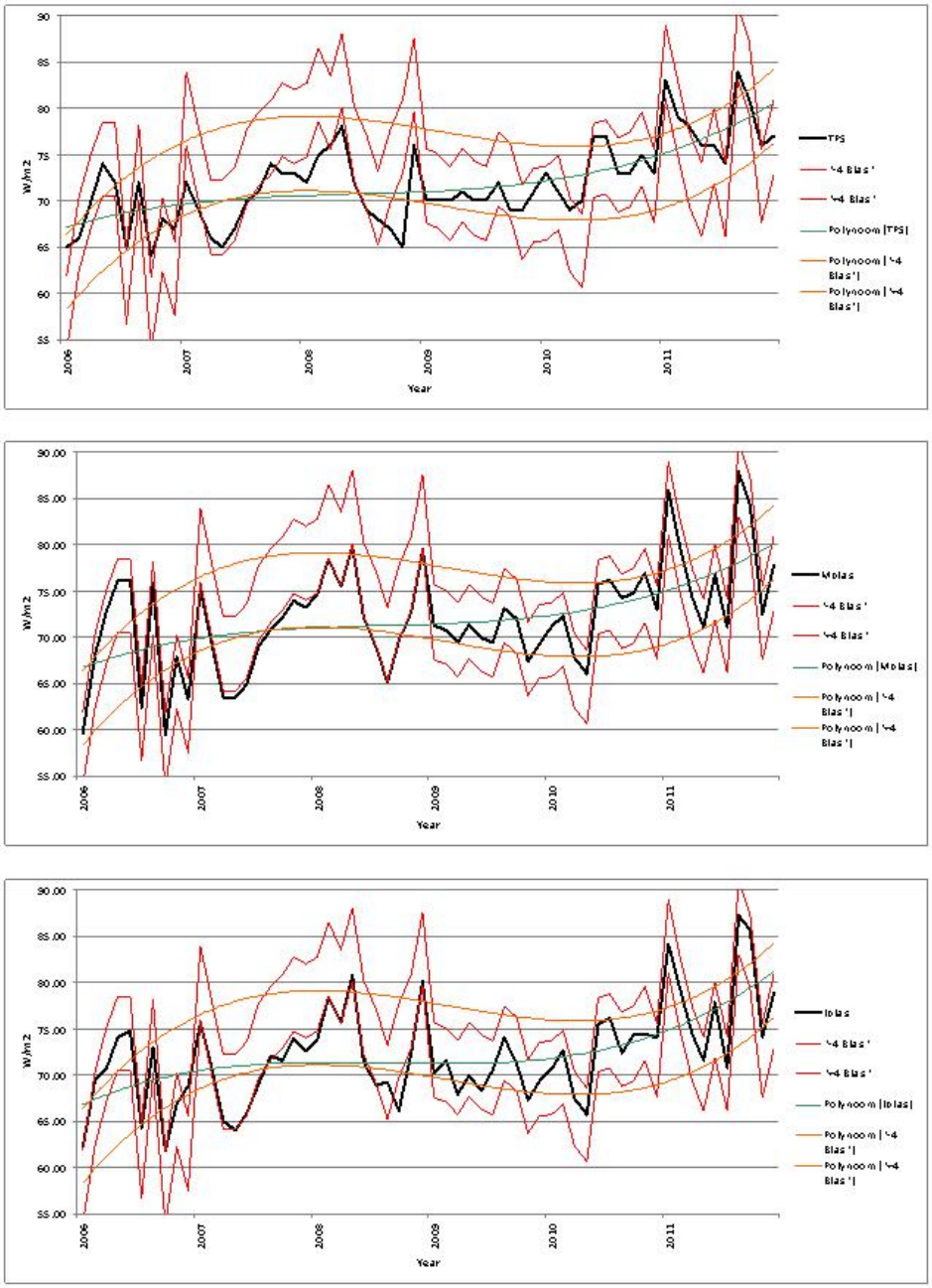


Figure 49: Random point analysis October CM-SAF.

.

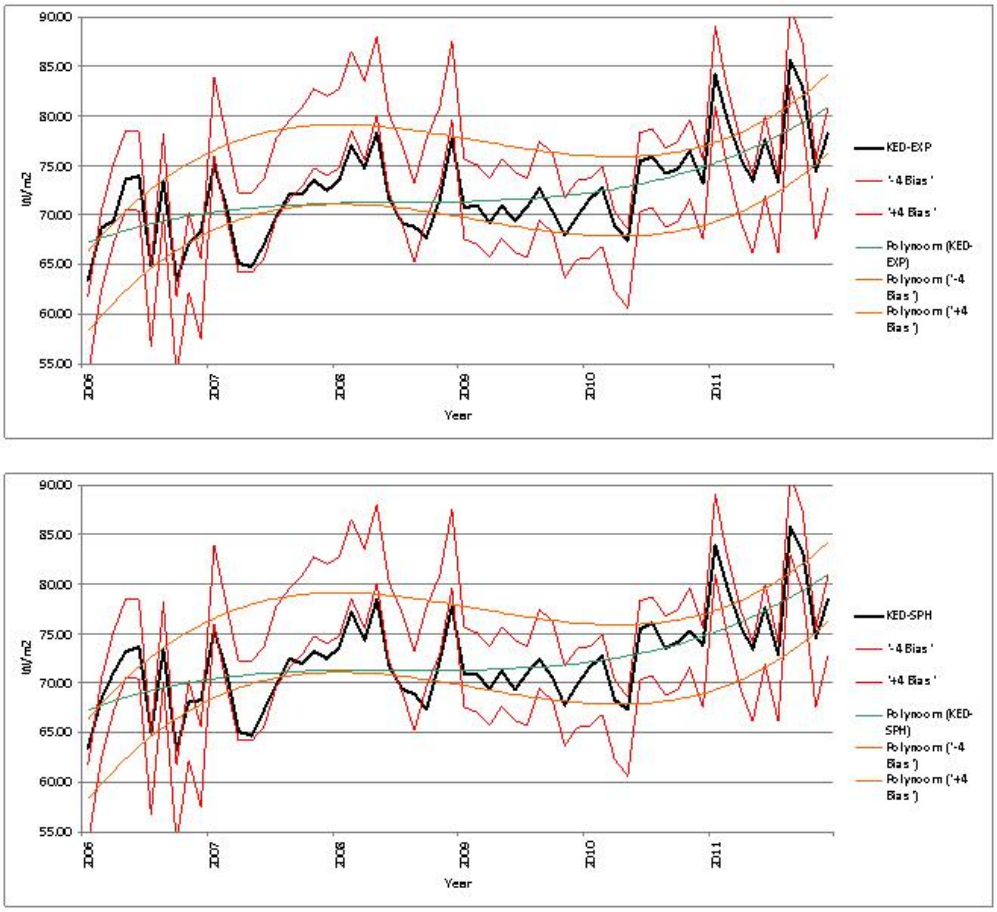


Figure 50: Random point analysis October CM-SAF.

.

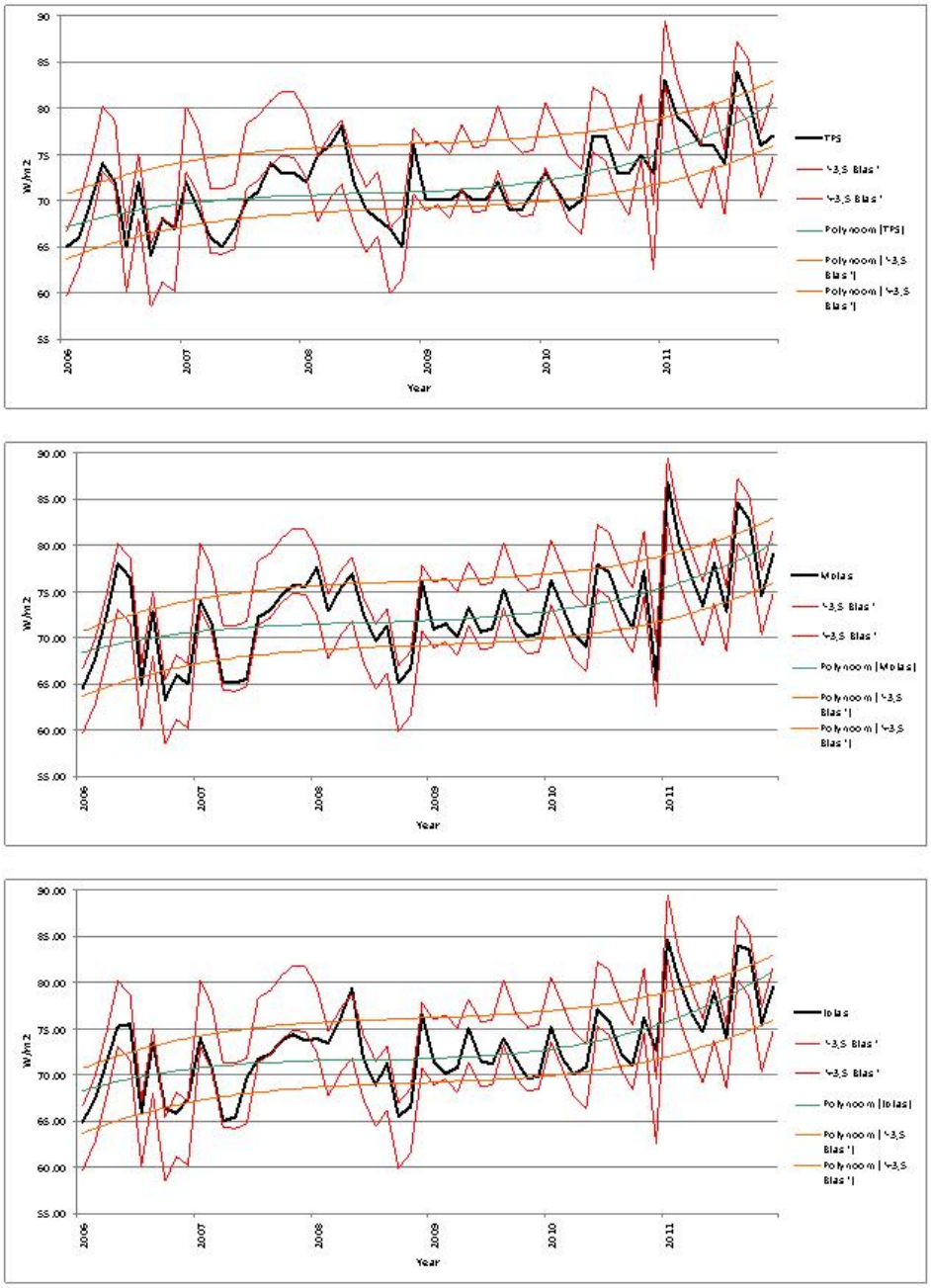


Figure 51: Random point analysis October SICCS.

.

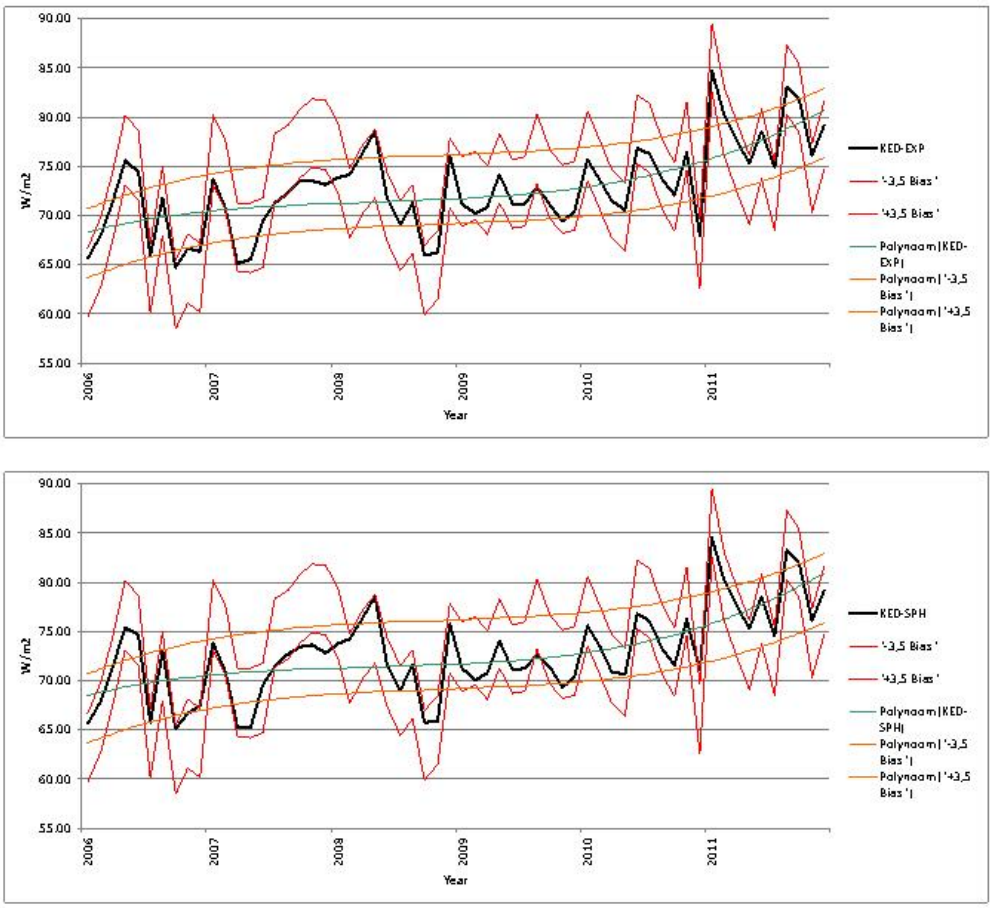


Figure 52: Random point analysis October SICCS.

.

The CM-SAF product seemed to perform very well again. All methods that used the satellite image as input for the interpolation method followed the pattern found in the image. This means that the interpolation methods made good use of the satellite image as a background trend. However, TPS did not use the satellite image as an input source. Looking at the pattern obtained by the TPS interpolation method it is possible to see that it does fit the pattern of the satellite image quite well. The pattern is less extreme than that of the other interpolation methods but it is definitely visible. This indicates that TPS performs very well in-between the measurement stations as well.

It did turn out that the CM-SAF satellite image overestimates the radiation values between the stations. All interpolation methods used the in-situ station measurements as main input and the values obtained at these locations fit that observed by the station. Since it was already proven that these values obtained at the station locations were very accurate we could assume the values in-between the stations should fit this accuracy as well. In the graphs that show the data between the stations we see that the values tend to be on the lower side of the negative bias. This means that the interpolation method lowered the values found in the satellite images when it interpolated the map. This indicated that the values in the satellite images alone were too high. This pattern was especially observed in the spring and summer months which is displayed in table 4. The SICCS product showed the same pattern. However, the graphs from this product seemed to have a better correlation. When looking at the graphs it is possible to see that the satellite image values and the interpolation values followed the same pattern. The satellite images did yet again overestimates the values compared to the interpolated values. However the bias in this product was smaller than that of the CM-SAF product. This indicated that the SICCS product on its own seems to be better in quality (table 4).

Januari	Avg SIS value	CM-SAF	Avg Bias
TPS	27.15	28.20	1.05
MB	27.70	28.20	0.50
IB	27.57	28.20	0.62
Ked Exp	27.65	28.20	0.54
Ked Sph	27.61	28.20	0.58

Januari	Avg SIS value	SICCS	Avg Bias
TPS	27.15	27.38	0.23
MB	27.88	27.38	-0.50
IB	27.80	27.38	-0.42
Ked Exp	27.80	27.38	-0.42
Ked Sph	27.76	27.38	-0.39

April	Avg SIS value	CM-SAF	Avg Bias
TPS	184.08	191.61	7.52
MB	183.91	191.61	7.70
IB	183.97	191.61	7.64
Ked Exp	184.09	191.61	7.52
Ked Sph	184.23	191.61	7.38

April	Avg SIS value	SICCS	Avg Bias
TPS	184.08	190.15	6.07
MB	184.93	190.15	5.22
IB	184.68	190.15	5.47
Ked Exp	184.94	190.15	5.21
Ked Sph	184.95	190.15	5.20

July	Avg SIS value	CM-SAF	Avg Bias
TPS	215.38	227.72	12.33
MB	215.46	227.72	12.25
IB	215.17	227.72	12.55
Ked Exp	215.39	227.72	12.33
Ked Sph	215.32	227.72	12.40

July	Avg SIS value	SICCS	Avg Bias
TPS	215.38	221.94	6.56
MB	216.75	221.94	5.19
IB	216.59	221.94	5.36
Ked Exp	216.68	221.94	5.27
Ked Sph	216.65	221.94	5.29

October	Avg SIS value	CM-SAF	Avg Bias
TPS	71.97	72.98	1.01
MB	72.11	72.98	0.87
IB	72.19	72.98	0.79
Ked Exp	72.32	72.98	0.67
Ked Sph	72.33	72.98	0.65

October	Avg SIS value	SICCS	Avg Bias
TPS	71.97	72.89	0.93
MB	66.77	72.89	0.20
IB	64.58	72.89	0.11
Ked Exp	62.93	72.89	0.23
Ked Sph	64.95	72.89	0.18

Table 4: Table showing the average values for each interpolation and the average of the satellite image. Left is for the CM-SAF product and right for the SICCS product.

.

The montly output maps showed more local variation than the 6 year average. This was expected due to the fact that the 6 year map used average values over a longer time period, removing the local patterns that could be observed in smaller areas. Figure 53 to 58 show three examples from the montly output maps for both the CM-SAF product as the SICCS product. In these examples it was possible to see that TPS does show the same average pattern as the other interpolation methods that made use of auxiliary data. However the interpolation methods that did use auxiliary data show more local variation, visualizing expected patterns like the lower radiation above the Veluwe for example. Comparing the producs with each other showed that in general the resolution of the SICCS product is higher. The CM-SAF product returned a more gridded map compared to the SICCS product.

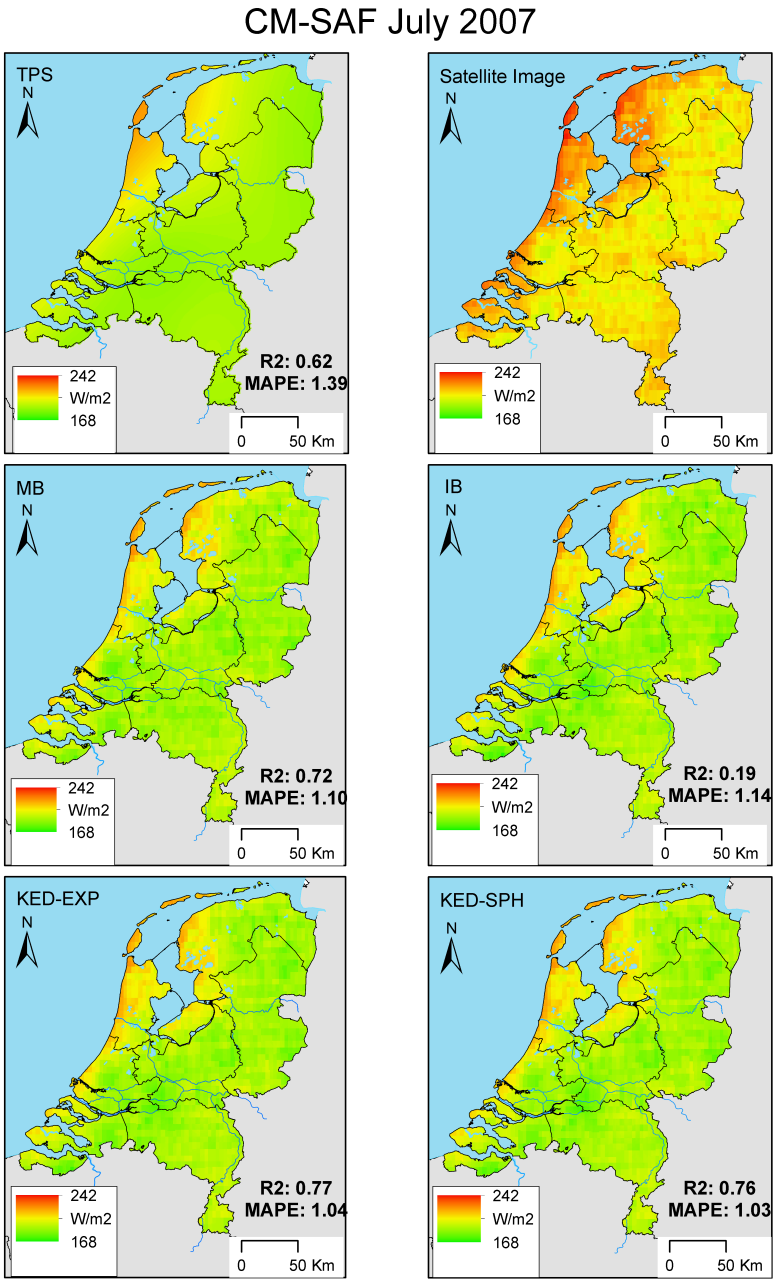


Figure 53: CM-SAF results for July 2007.

.

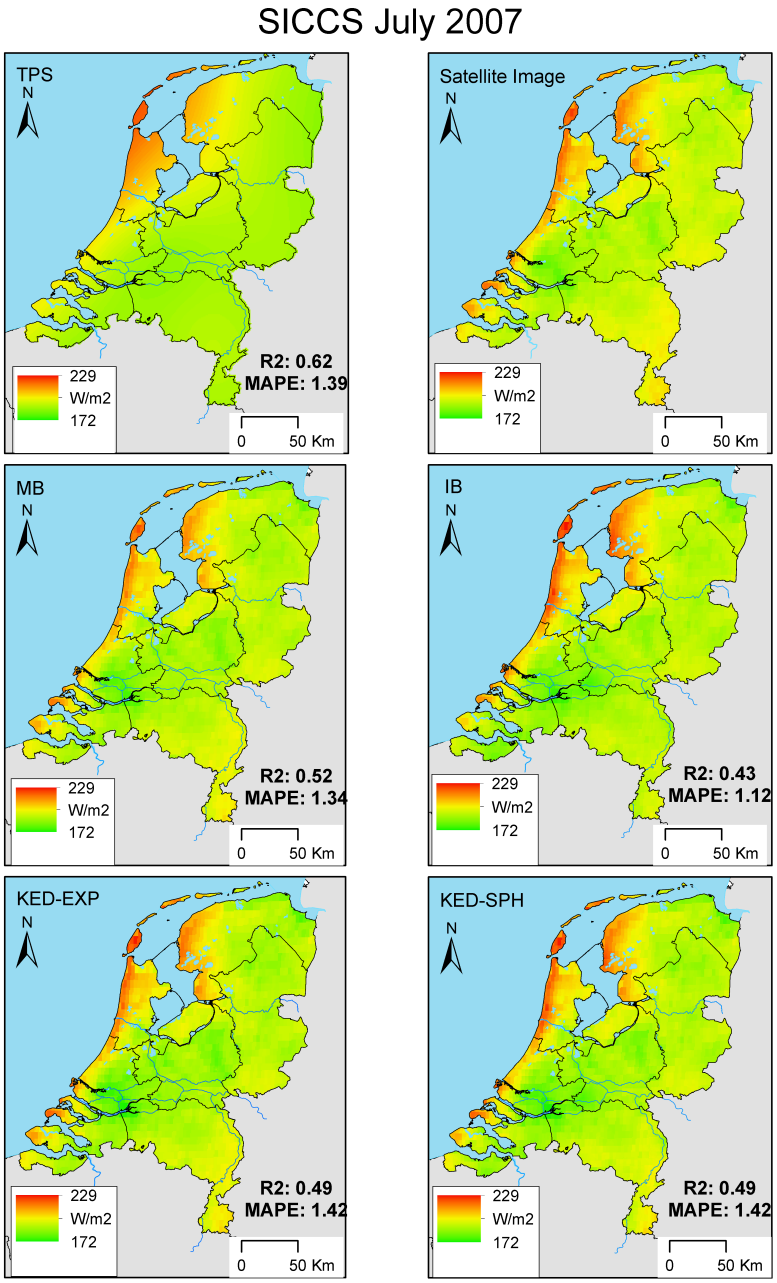


Figure 54: SICCS results for July 2007.

.

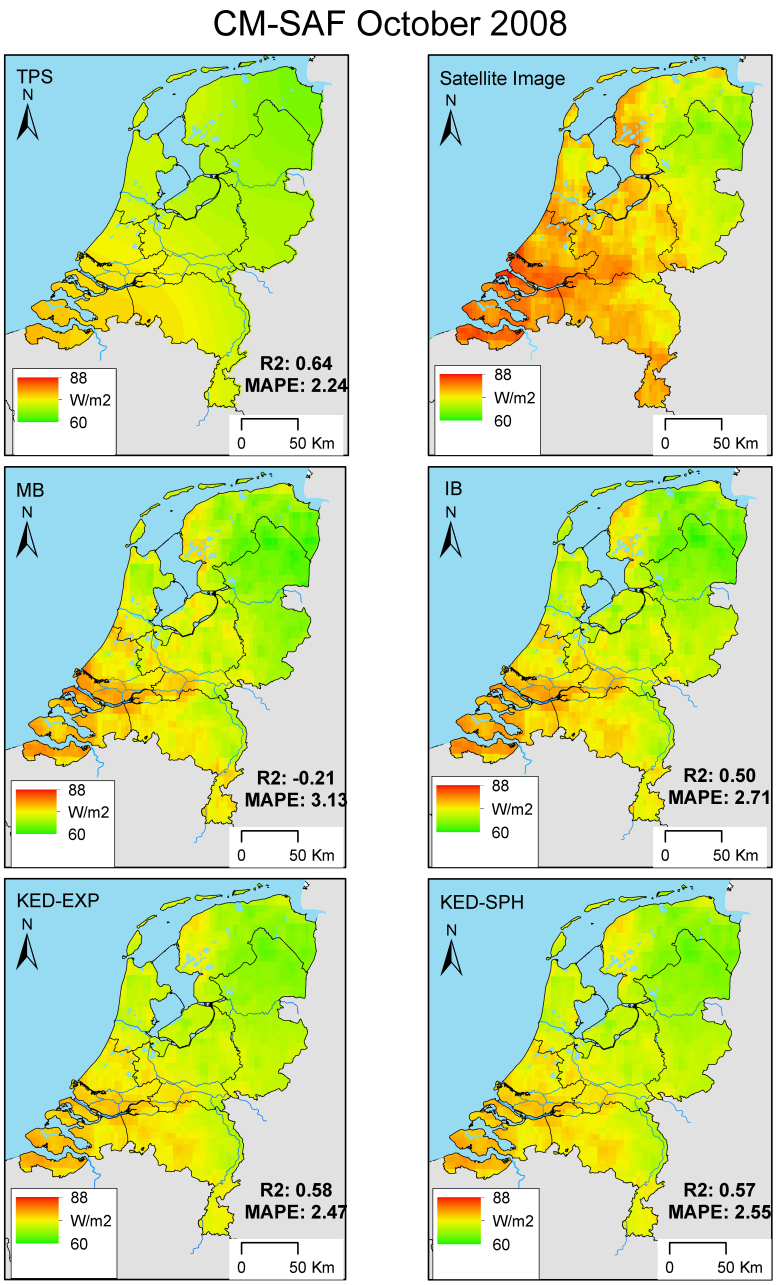


Figure 55: CM-SAF results for October 2008.

.

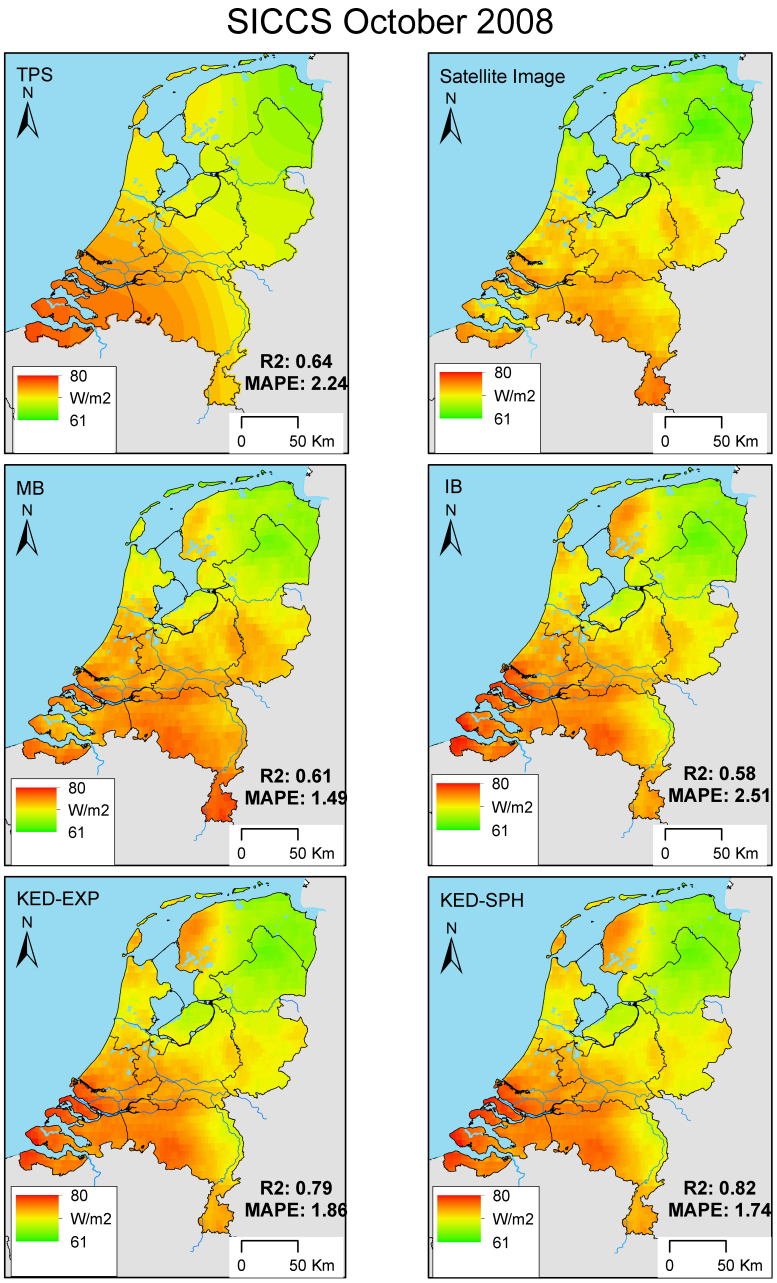


Figure 56: SICCS results for October 2008.

.

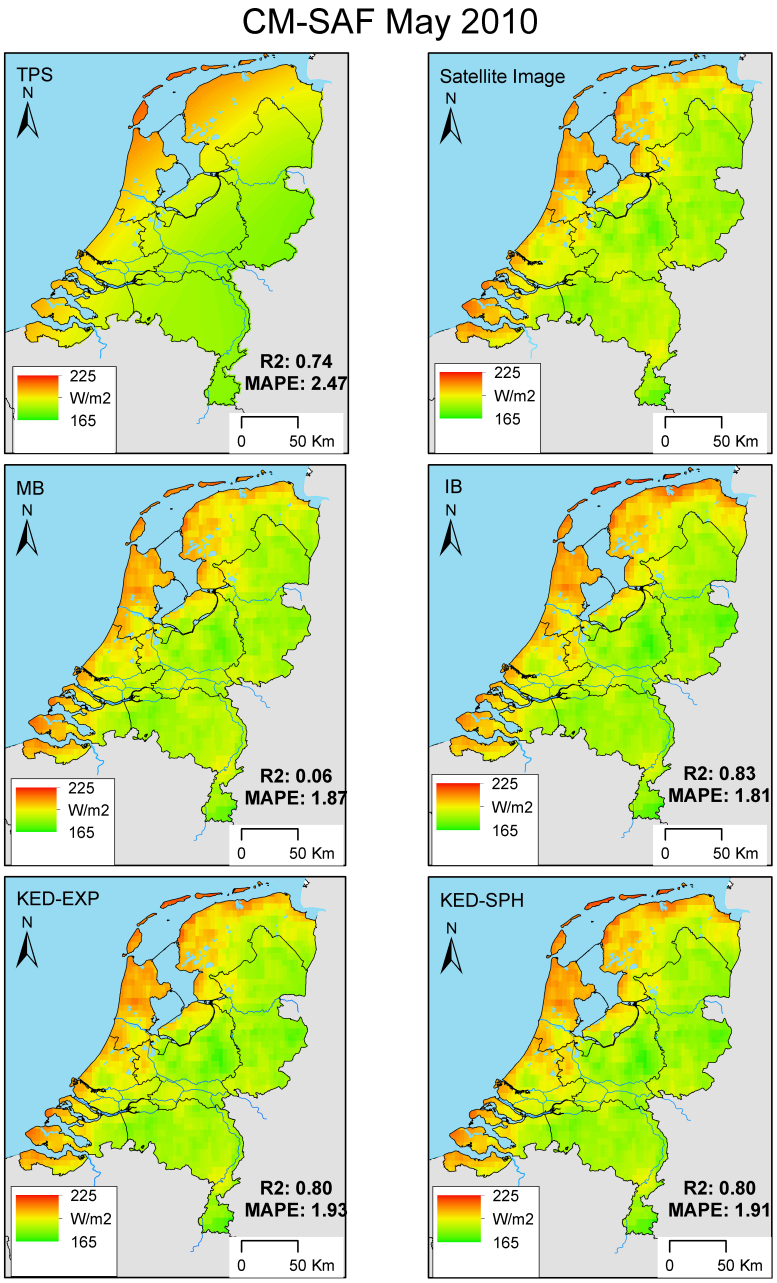


Figure 57: CM-SAF results for May 2010.

.

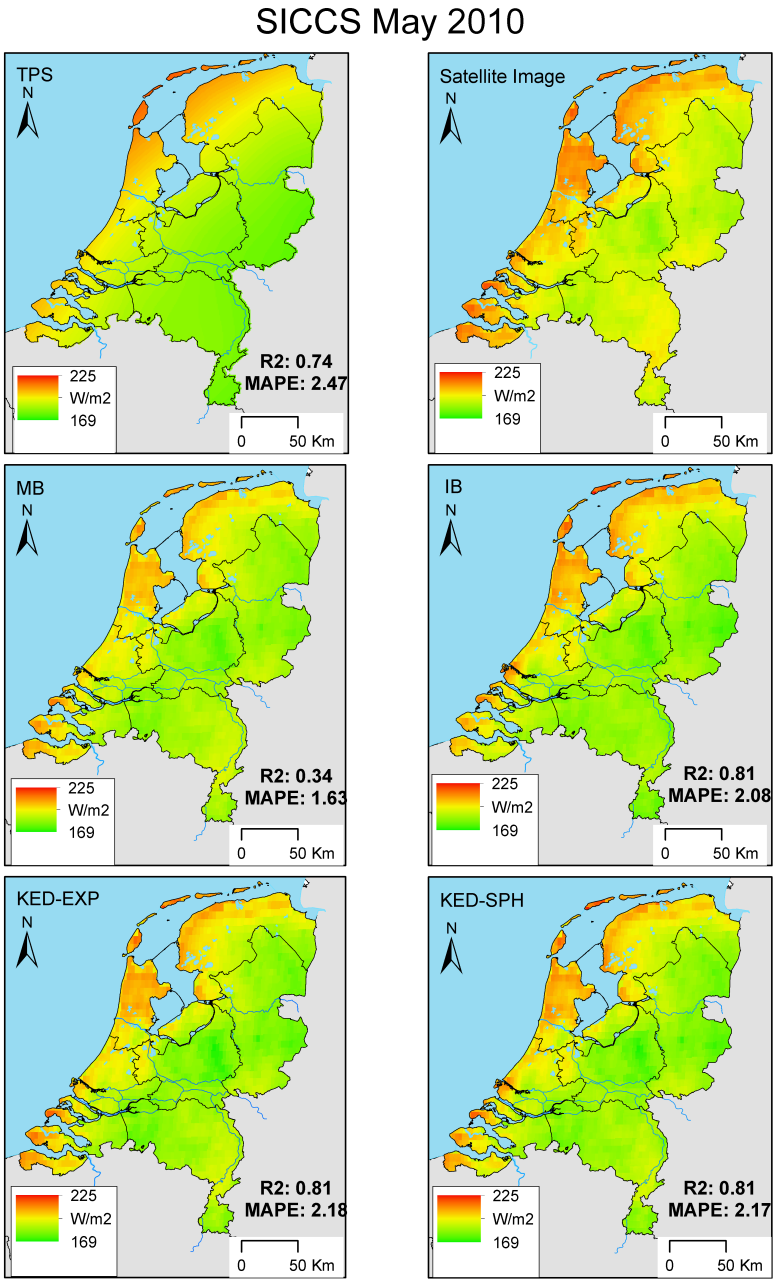


Figure 58: SICCS results for May 2010.

.

Daily data

Since all interpolations performed very well on a monthly and long time average scale, daily data was analyzed as an extra option. For the daily interpolation only the months April to July in 2010 were analyzed for this research. These months were chosen since they are the most interesting when it comes down to radiation. Radiation values are higher during these months and products and processes that use radiation are therefore more interesting in these periods. The interpolations were performed with the same methods as for the monthly and long term yearly average. The average R^2 and MAPE showed that interpolations that made use of satellite data returned better results than those who didn't use auxilary data (see table 5 and table 6).

R2	April	May	June	July
TPS	0.53	0.57	0.46	0.62
MB	0.30	0.75	0.45	0.55
IB	-0.08	-0.18	-0.09	-0.14
KED-EXP	0.65	0.75	0.58	0.62
KED-SPH	0.62	0.72	0.57	0.60

MAPE	April	May	June	July
TPS	6.01	10.70	5.68	5.72
MB	5.60	8.06	5.41	5.83
IB	0.00	0.01	0.00	0.00
KED-EXP	0.52	0.95	0.40	0.36
KED-SPH	0.11	0.13	0.08	0.08

Table 5: The average R^2 and MAPE for the daily interpolations on the CM-SAF product.

.

R2	April	May	June	July
TPS	0.53	0.57	0.46	0.62
MB	0.57	0.84	0.55	0.75
IB	0.03	-0.08	0.02	0.03
KED-EXP	0.73	0.83	0.70	0.78
KED-SPH	0.73	0.83	0.69	0.78

MAPE	April	May	June	July
TPS	6.01	10.70	5.68	5.72
MB	4.08	6.03	3.91	3.89
IB	0.00	0.00	0.00	0.11
KED-EXP	0.43	0.86	0.28	0.40
KED-SPH	0.07	0.11	0.06	0.06

Table 6: The average R^2 and MAPE for the daily interpolations on the SICCS product.

.

For the CM-SAF product, TPS had the highest average MAPE of 10.7% this was significantly higher than that of the KED and IB interpolations. The MB had a maximum average error of 8.06% which was also quite high compared to the KED and IB. The IB had the lowest errors but this went together with a very poor R^2 which was also observed in montly interpolations.

The SICCS product results showed the same trend as the CM-SAF product. TPS had the highest MAPE followed by the MB. The IB performed the best when only looking at the MAPE but had a very unpredictable R^2 yet again.

To get a better understanding on how the actual R^2 and MAPE behaved from day to day results are presented in line graphs (figure 59 to 66). These graphs show the obtained values for each day. Here it is possible to see that the R^2 value of the IB performed worse than the other interpolation methods. However the MAPE graphs show a different trend. As expected from the average values the IB and KED interpolations perform significantly better than the TPS and MB interpolations. It is clear from the graphs that especially the TPS method was not able to capture the variation in global radiation on a daily scale as good as the other interpolation methods.

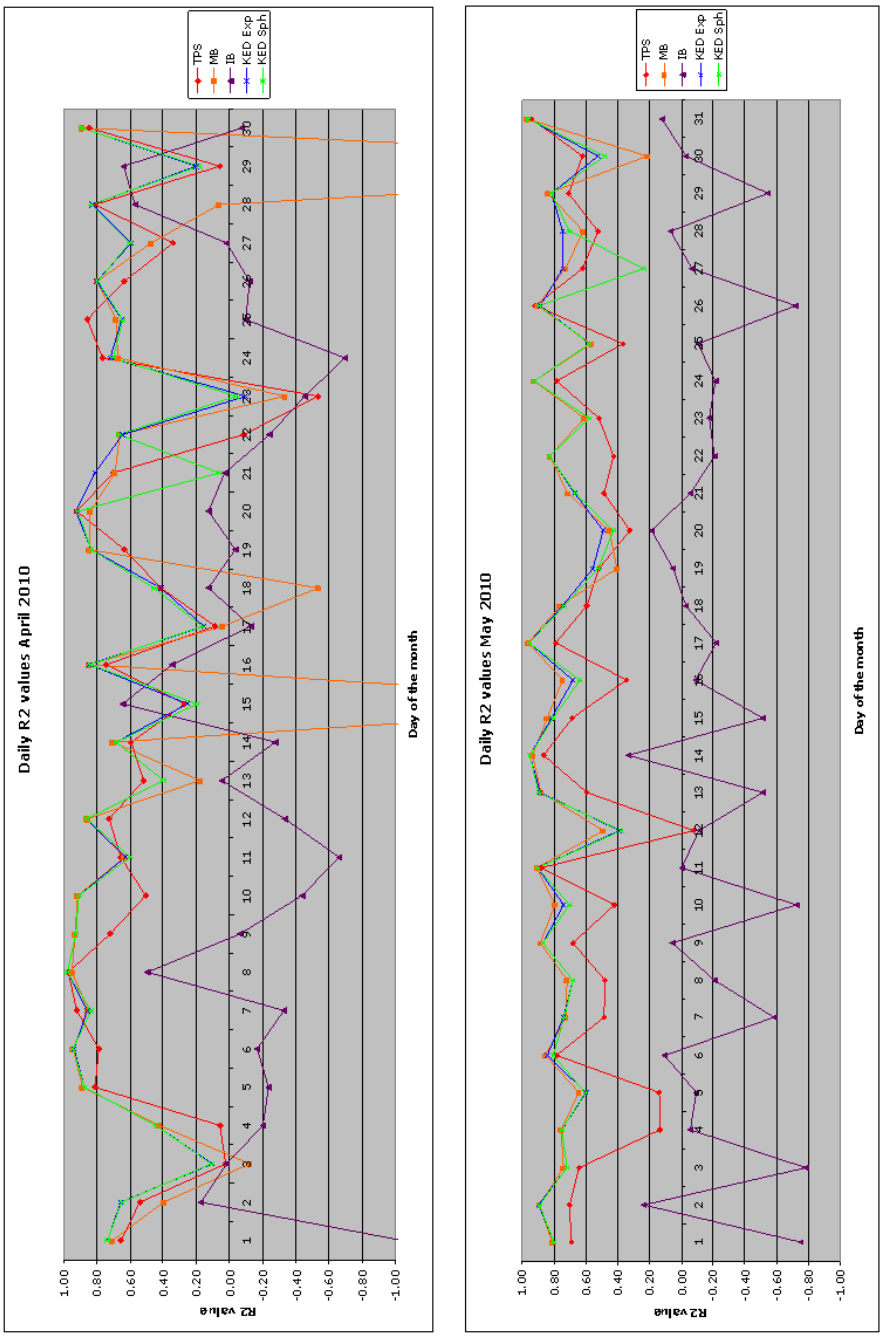


Figure 59: Daily R^2 values for April and May 2010 using the CM-SAF product.

.

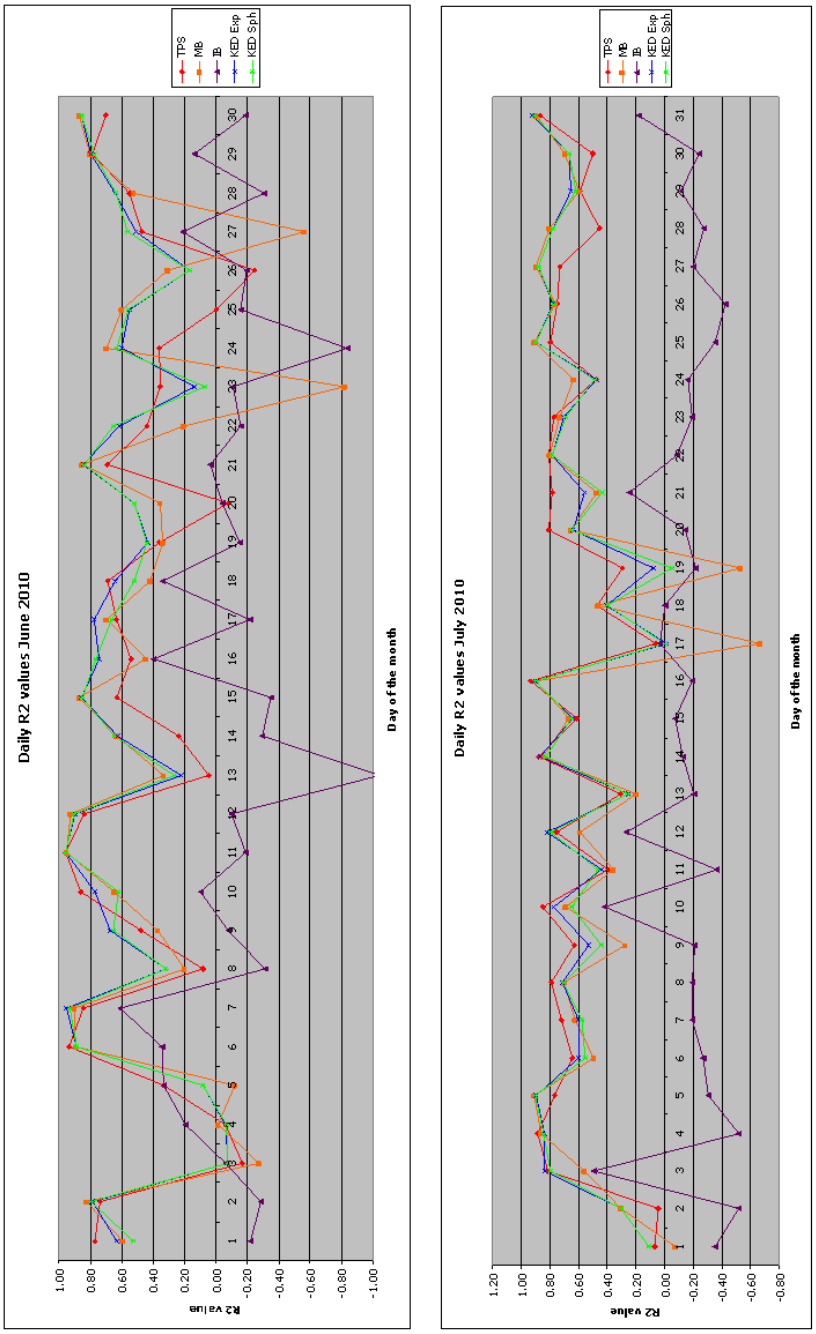


Figure 60: Daily R^2 values for June and July 2010 using the CM-SAF product.

.

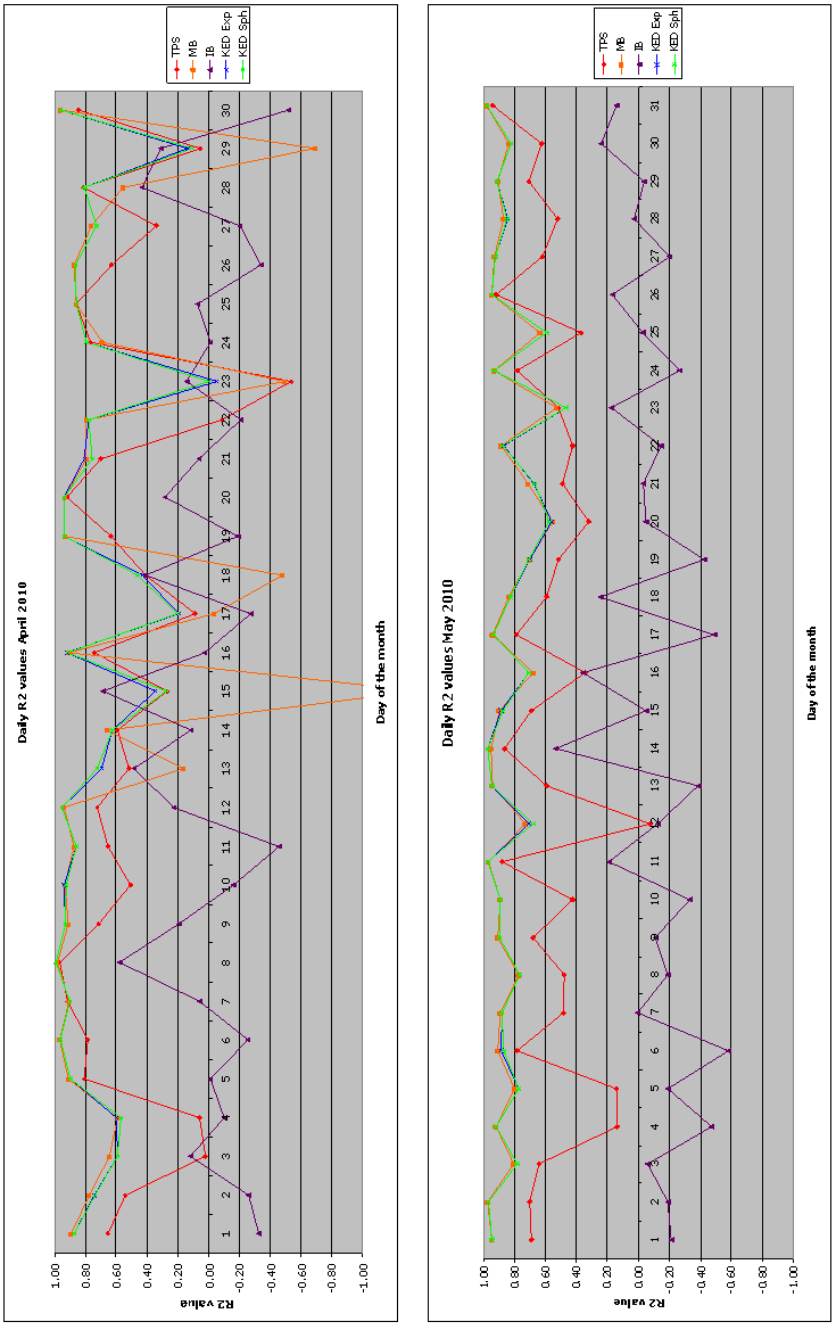


Figure 61: Daily R^2 values for April and May 2010 using the SICCS product.

.

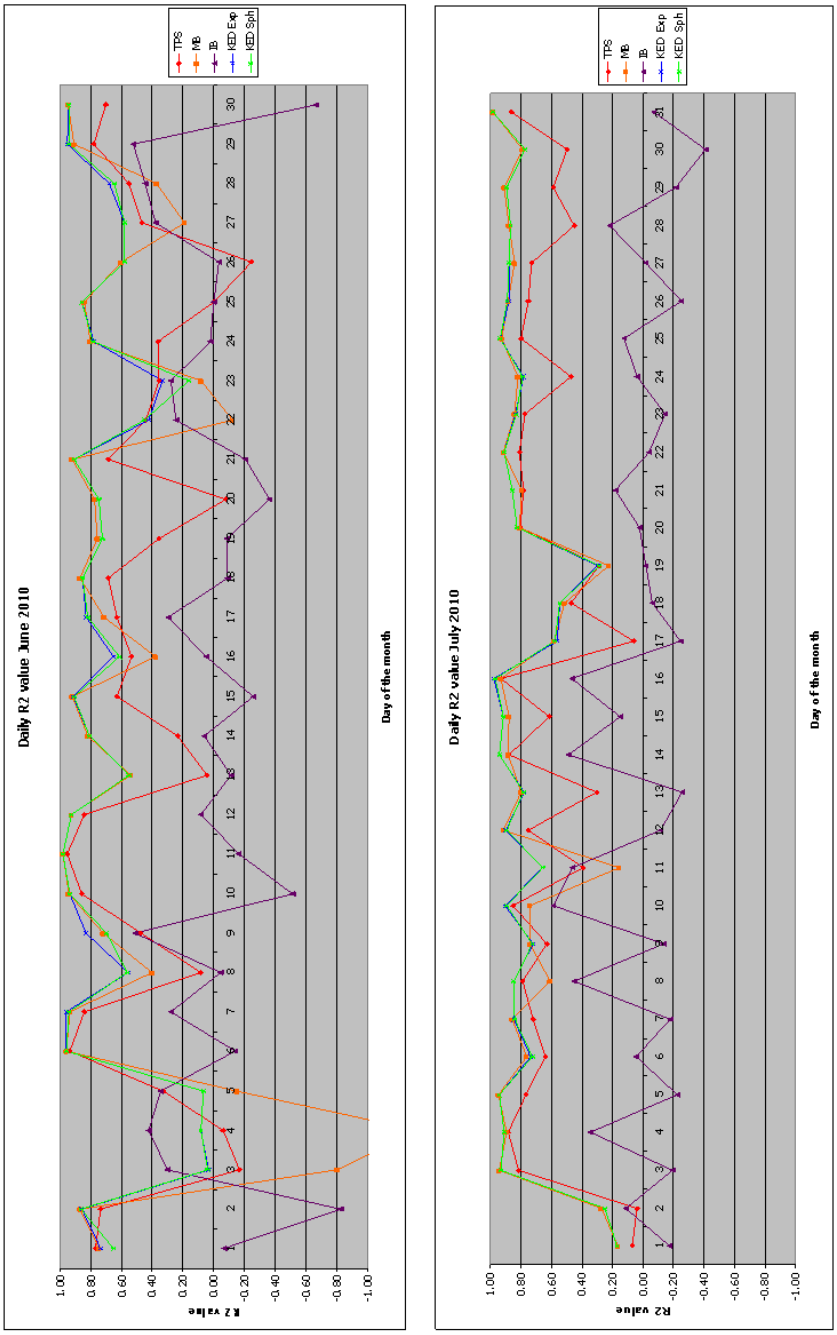


Figure 62: Daily R^2 values for June and July 2010 using the SICCS product.

.

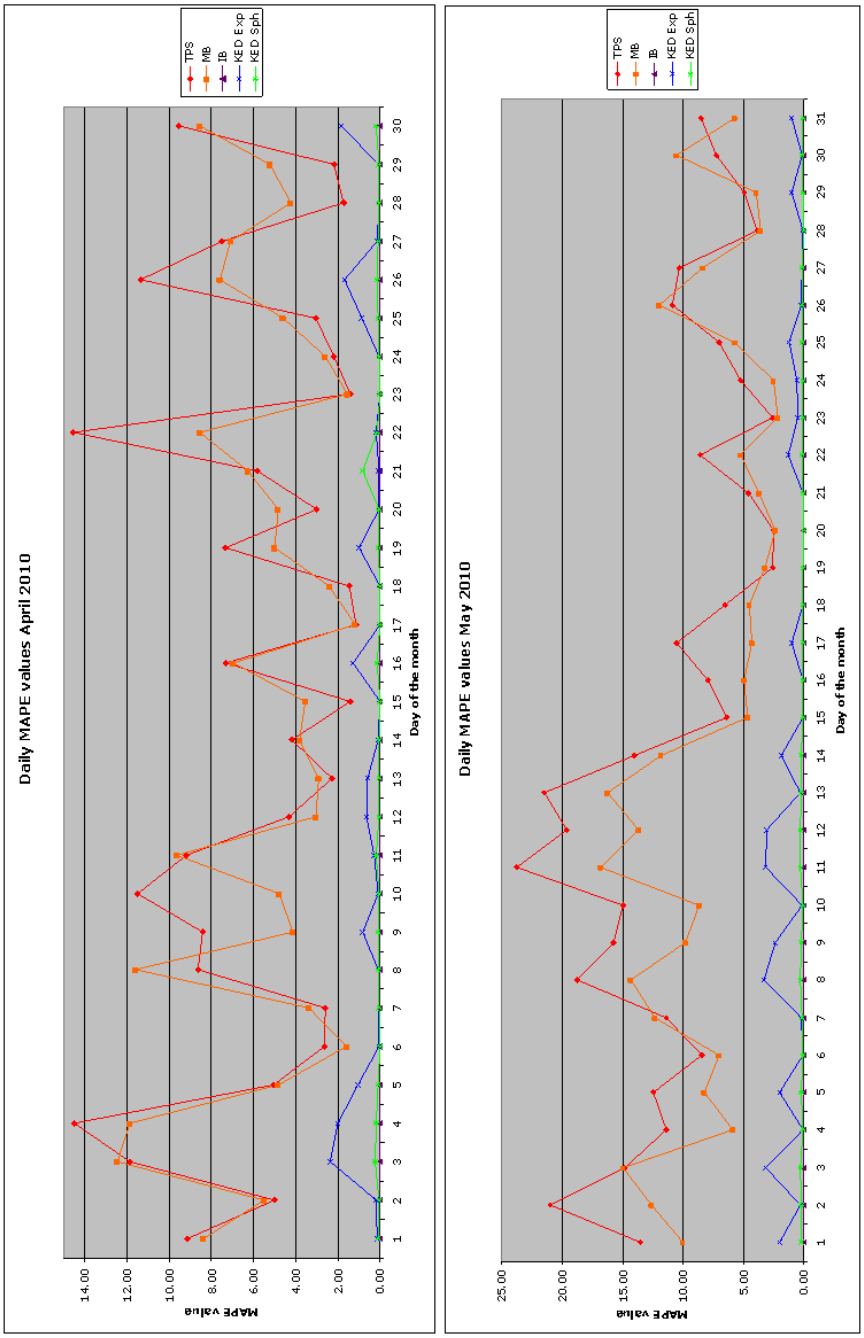


Figure 63: Daily MAPE values for April and May 2010 using the CM-SAF product.

.

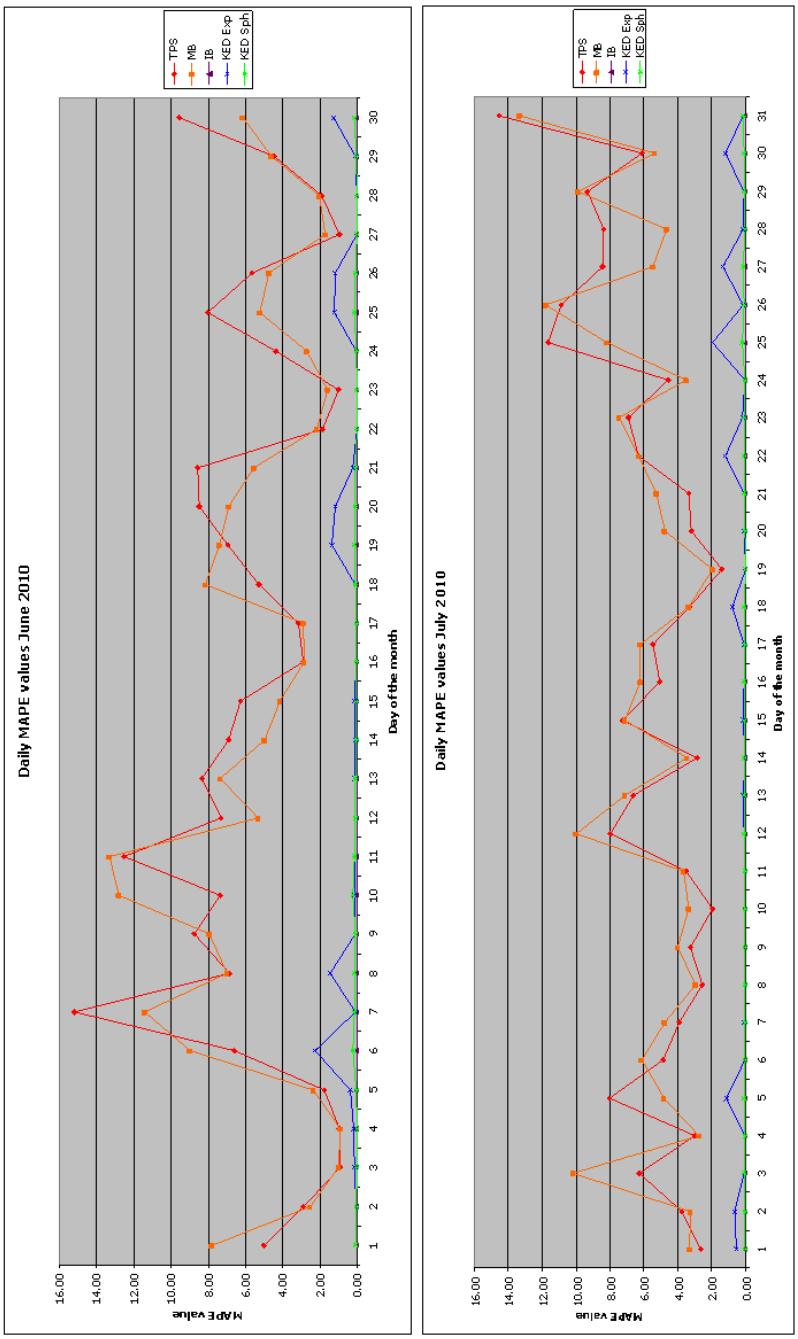


Figure 64: Daily MAPE values for June and July 2010 using the CM-SAF product.

.

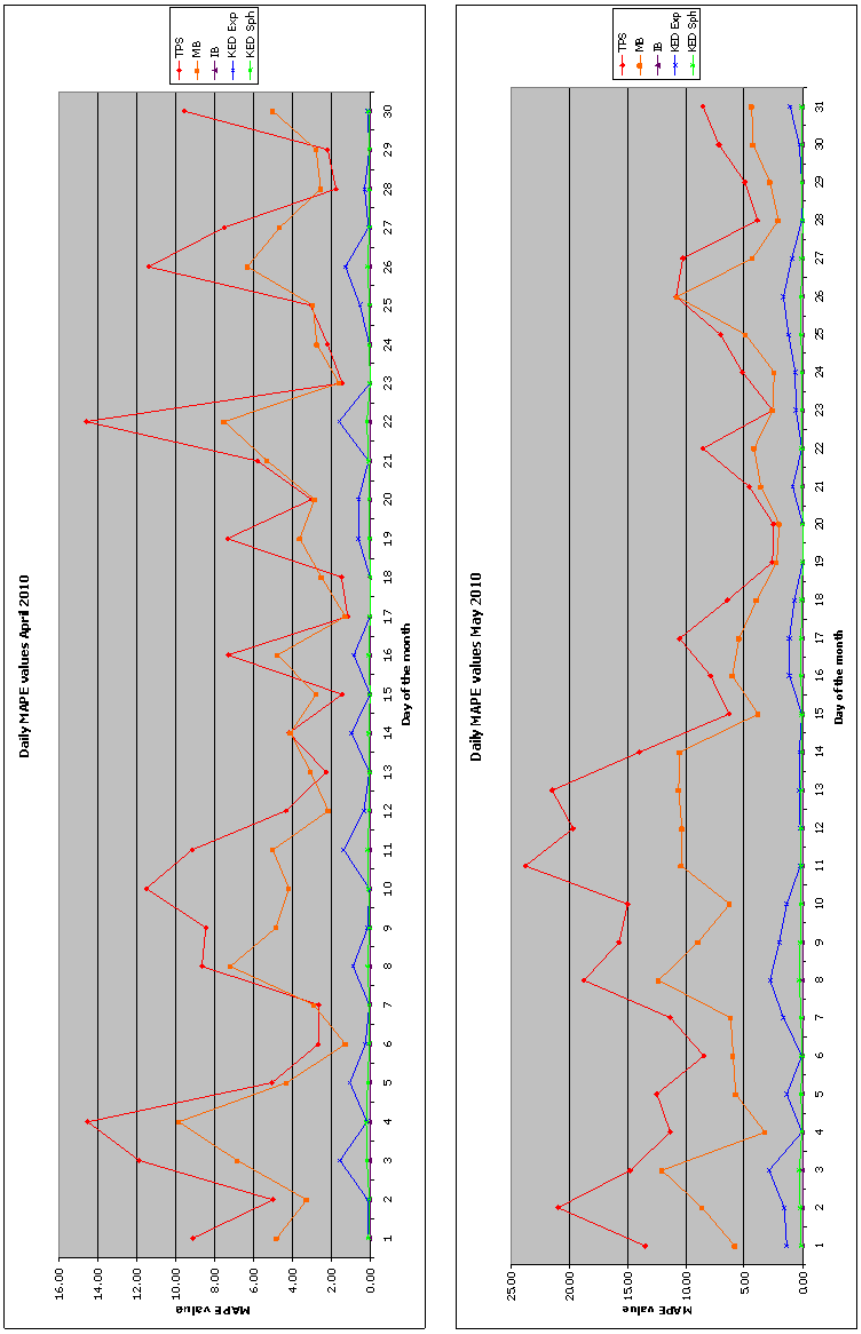


Figure 65: Daily MAPE values for April and May 2010 using the SICCS product.

.

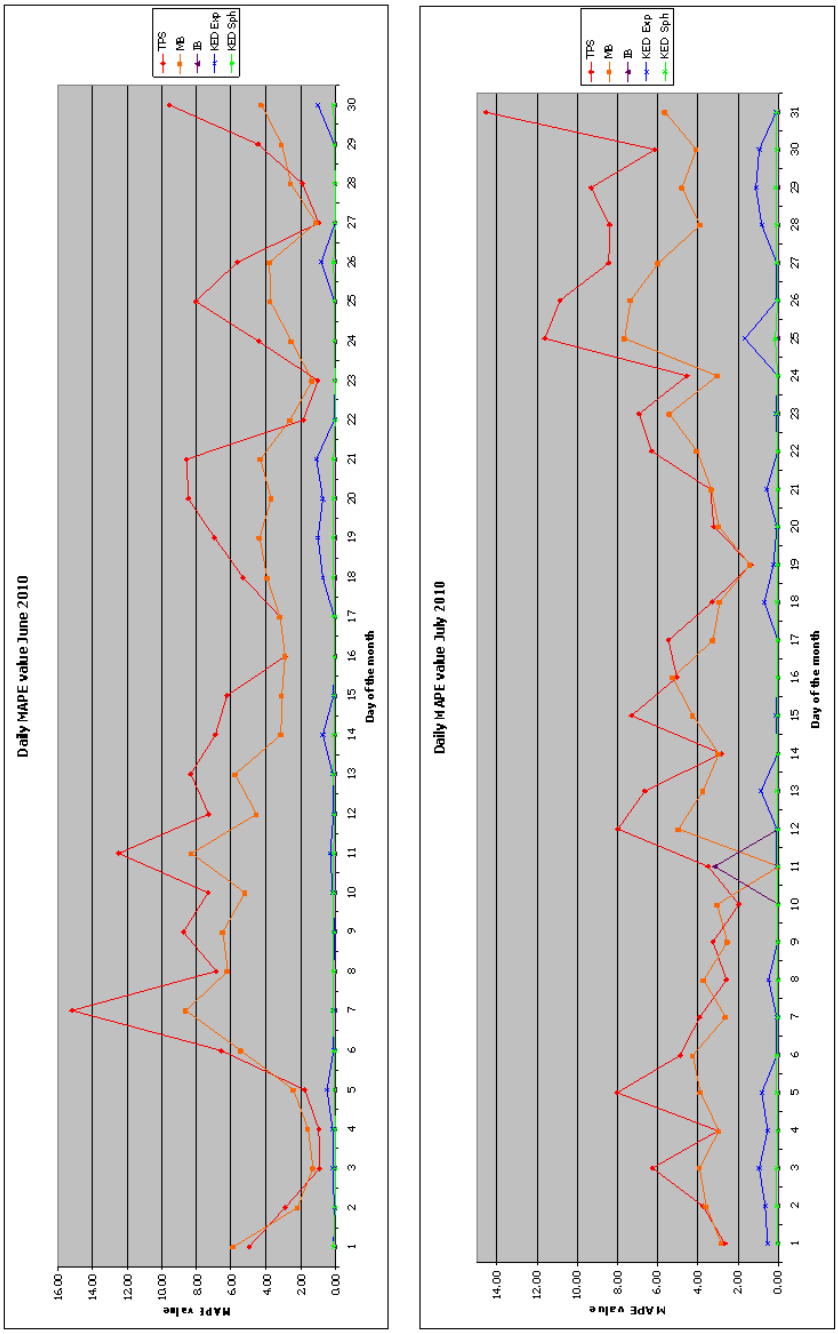


Figure 66: Daily MAPE values for June and July 2010 using the SICCS product.

.

When comparing the satellite measurements with the station measurements it was possible to see that the interpolation methods adjusted the values found in the satellite image to match the measurements made by the stations on the ground. Table 7 shows how big the bias was between the satellite image and the ground measurements for the CM-SAF product. The optimal accuracy of the CM-SAF product for daily data was an error smaller than 20 W/m^2 (CM-SAF 2013). As shown in the table, this was not always obtained by the satellite image on its own. Therefore using the satellite image as a stand alone product is not a valid option and interpolation is needed.

Date	Avg Error	Min Error	Max Error
1-Apr	-8.14	-1.04	-37.68
11-Apr	-12.33	-1.56	-40.68
21-Apr	3.65	0.68	43.95
30-Apr	-3.71	1.34	-28.64

Date	Avg Error	Min Error	Max Error
1-May	-1.28	0.57	43.83
11-May	9.38	1.72	64.49
21-May	-2.33	-0.61	-36.84
31-May	-16.61	-0.27	-37.35

Date	Avg Error	Min Error	Max Error
1-Jun	3.16	-1.46	-30.99
11-Jun	-37.28	-1.73	-70.59
21-Jun	-8.42	0.48	-55.48
30-Jun	-27.46	0.97	-89.63

Date	Avg Error	Min Error	Max Error
1-Jul	12.91	-1.36	47.50
11-Jul	-18.90	-0.35	-43.80
21-Jul	-15.10	1.33	-48.46
31-Jul	-30.45	-5.95	-61.00

Table 7: The bias in W/m^2 showing the station value minus the satellite image of the CM-SAF product, for 4 different days in 4 months in 2010.

.

The same could be said about the SICCS product. It would be unwise to use the satellite image on its own at this point. Table 8 shows the average bias and the minimum and maximum bias between the in-situ measurements and satellite measurements. Although the bias was smaller compared to the CM-SAF product it was still relatively high to use as a product on its own. Especialy when the over or underestimation can easly be accounted for by using interpolation as used above.

Date	Avg Error	Min Error	Max Error
1-Apr	0.62	-0.23	23.25
11-Apr	-10.76	3.09	-31.97
21-Apr	2.07	-0.06	-26.74
30-Apr	0.28	-0.27	-20.59

Date	Avg Error	Min Error	Max Error
1-May	-2.68	0.00	-24.32
11-May	-4.85	0.35	-29.34
21-May	-10.56	-0.69	-46.59
31-May	-12.20	-1.28	-27.55

Date	Avg Error	Min Error	Max Error
1-Jun	15.23	-1.02	37.30
11-Jun	-15.67	-0.05	-35.51
21-Jun	-8.73	2.42	-38.89
30-Jun	-10.73	1.36	-32.31

Date	Avg Error	Min Error	Max Error
1-Jul	-0.42	-0.14	23.67
11-Jul	-14.79	1.09	-43.43
21-Jul	-2.87	0.18	17.33
31-Jul	-12.00	-3.01	-21.60

Table 8: The bias in W/m^2 showing the station value minus the satellite image of the SICCS product, for 4 different days in 4 months in 2010.

.

In figure 67 to 72 three examples of daily interpolations are presented for both products. The results have already shown that the error's for the IB and KED interpolations were smaller than those of the IB and TPS interpolations. Visually all the interpolation methods gave more details than the TPS interpolation. More local patterns and variations can be observed in both products.

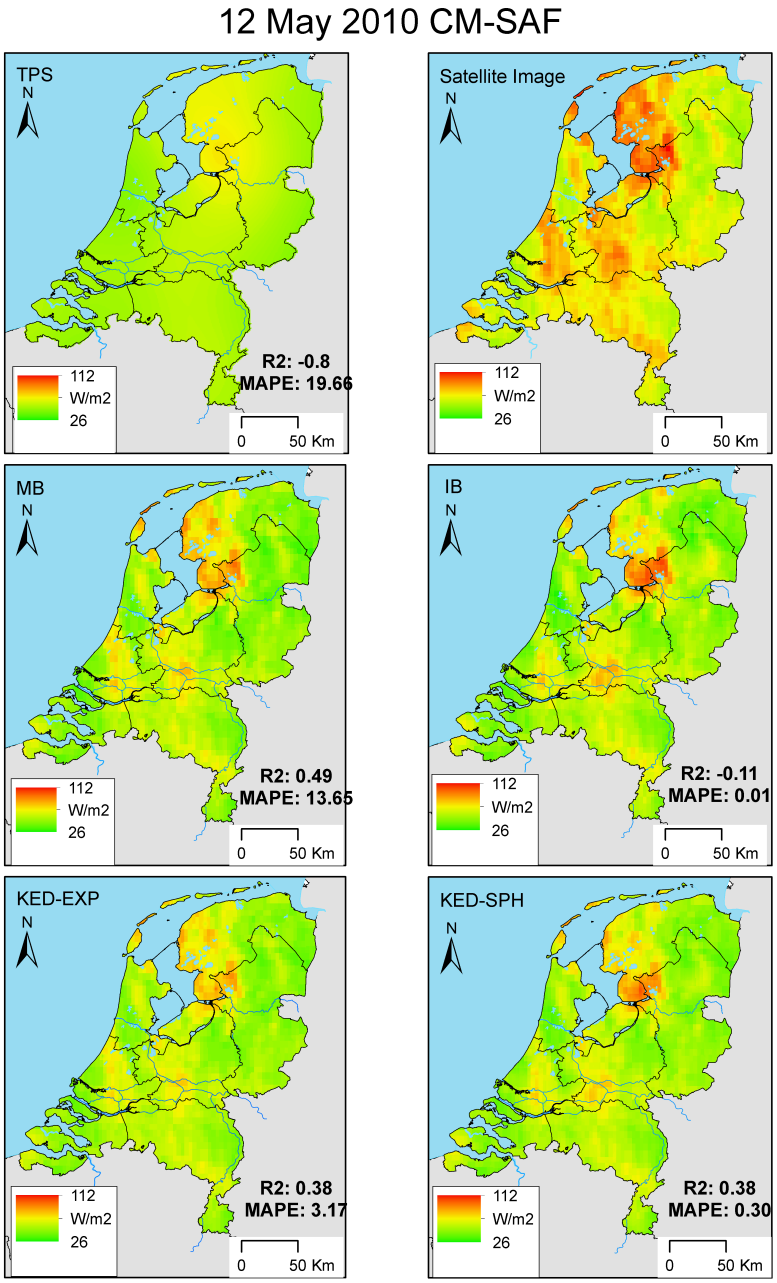


Figure 67: CM-SAF results for the 12th of May 2010.

.

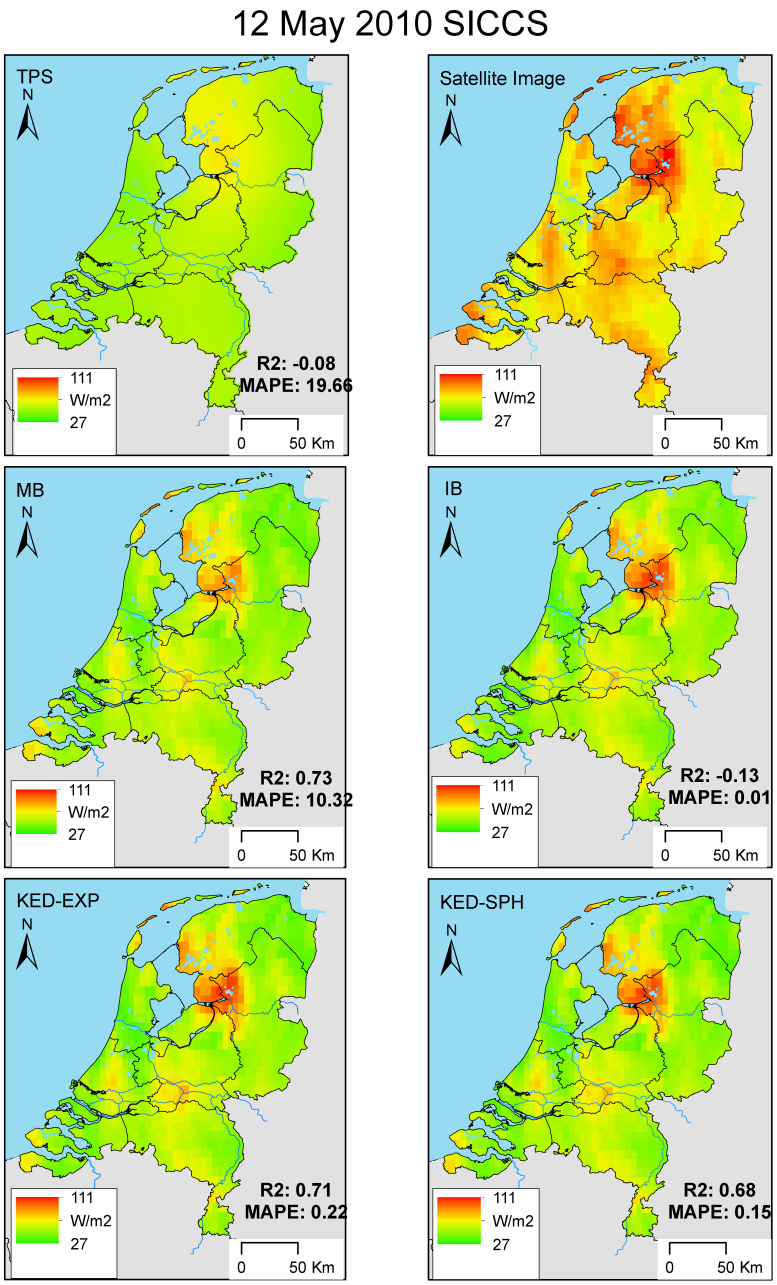


Figure 68: SICCS results for the 12th of May 2010.

.

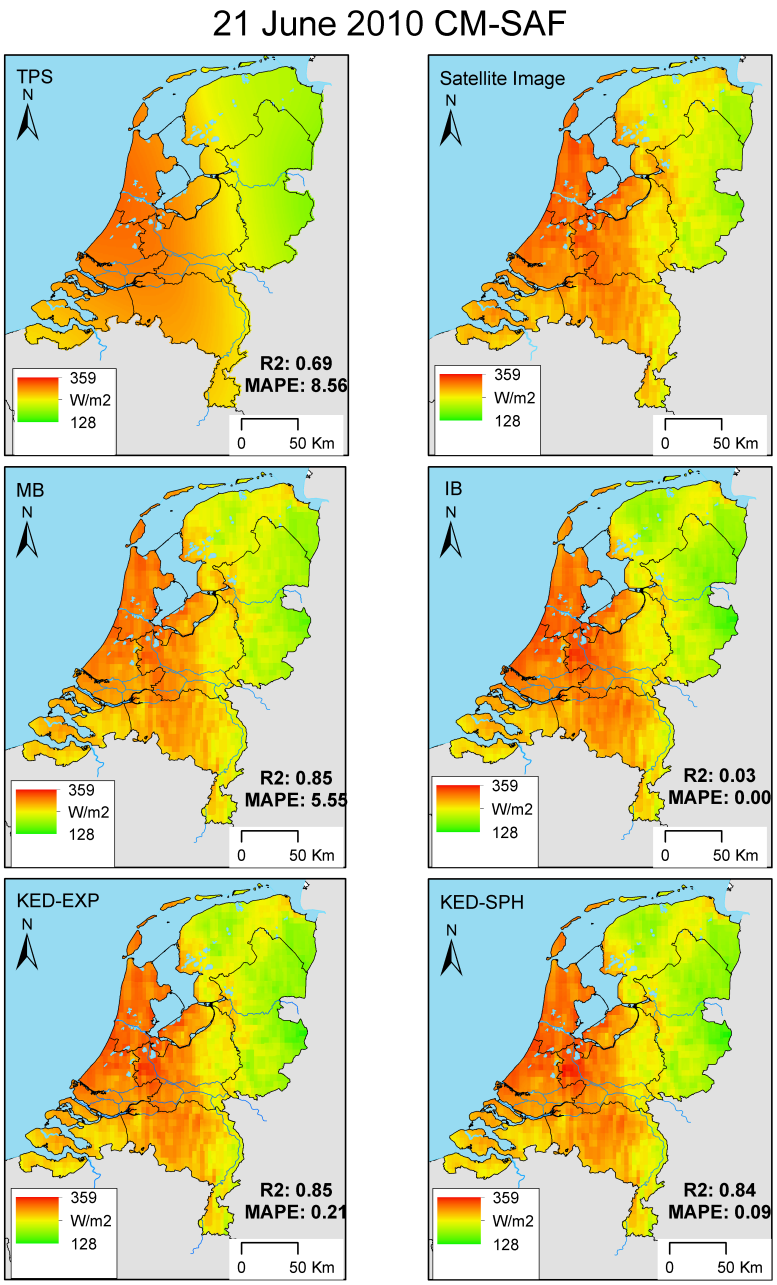


Figure 69: CM-SAF results for the 21th of June 2010.

.

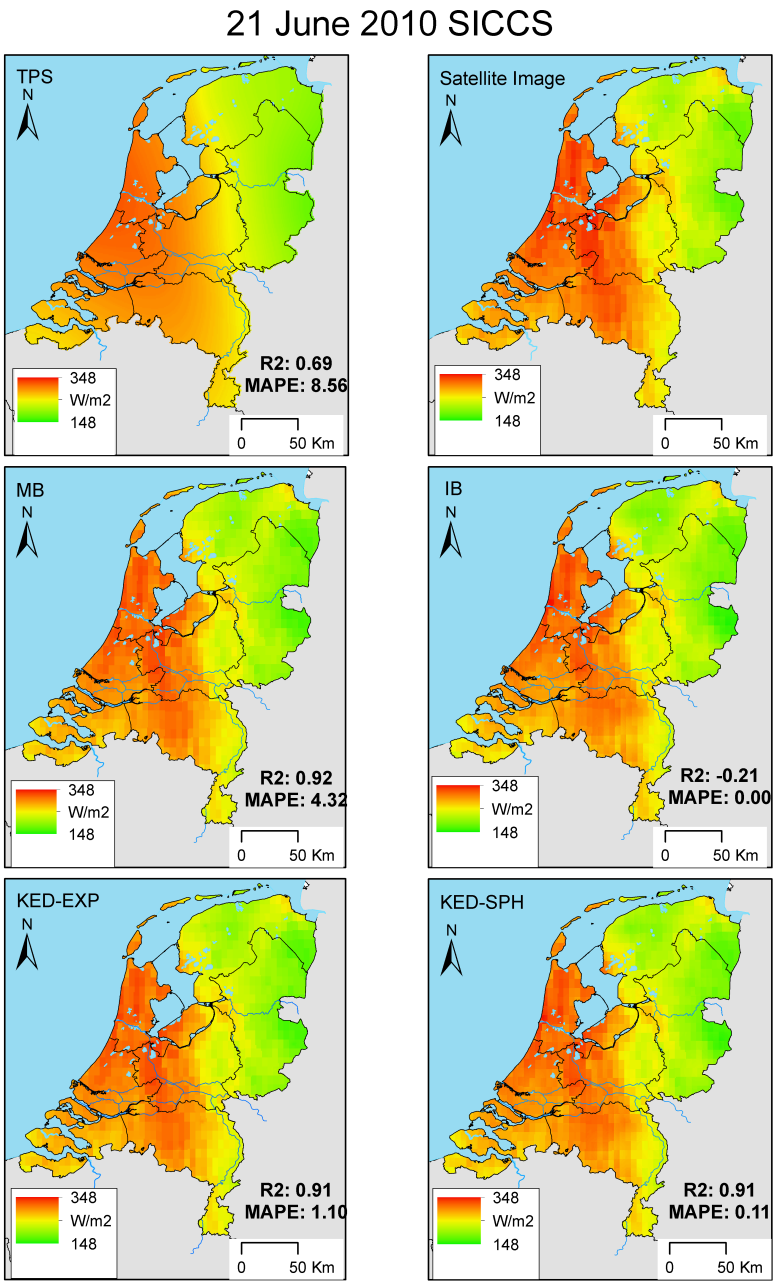


Figure 70: SICCS results for the 21th of June 2010.

.

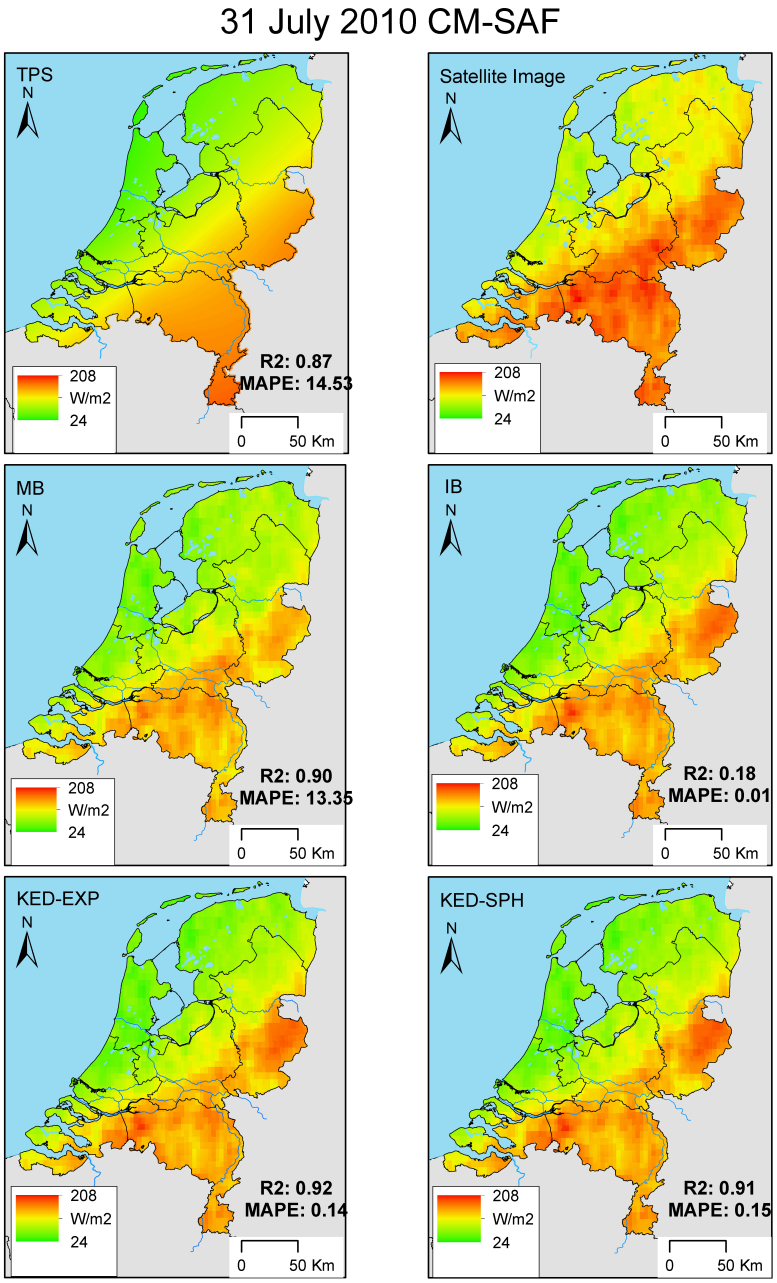


Figure 71: CM-SAF results for the 31th of July 2010.

.

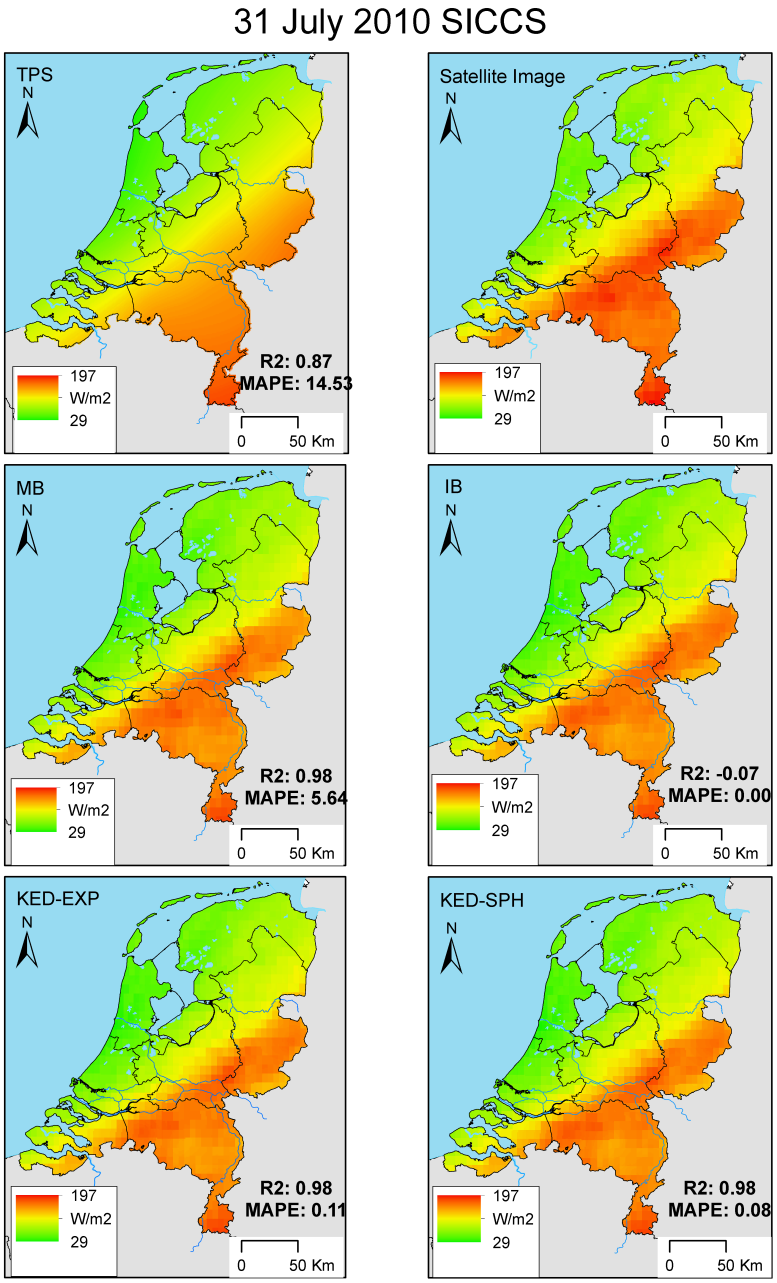


Figure 72: SICCS results for the 31th of July 2010.

.

5 Discussion and Conclusion

5.0.5 Discussion

The 6 year average seemed to perform fairly well, no matter which interpolation method or input source was used (figure 11 and 12). The lowest R^2 values were found for the IB which means that the model couldn't predict or explain the residuals as well as in the other methods. The highest average MAPE found for the CM-SAF product was 1.39% for TPS. This was a very low value, especially considering the fact that the maximum observed in-situ value was 127 W/m^2. This means that at maximum, the interpolation on the cross validation was 1.77 W/m^2 off. This is well under the 2-3% that the WMO uses as a standard to classify measurement equipment as secondary standard (WMO, 2008).

The SICCS product of the KNMI had a maximum error of 1.42% (which is, at maximum a bias of 1.8 W/m^2) it is also still under the targets of the WMO and falls within the bias observed by the validation of the SICCS product (Greuell et al., 2013).

The unstable R^2 that was observed in the monthly MB and IB could be explained by the quality of the satellite image. In the winter the solar angle for the Netherlands can be quite low. This makes it hard for the satellite to capture the data with high precision. Due to this some satellite images were not complete or had extremely low values in the northern half of the Netherlands. These low or missing values had a big influence on the mean bias and the interpolated bias, creating artefacts in the interpolation and therefore show bad results. The reason why these interpolation methods were effected most is due to the fact that the values that are interpolated are directly influenced by the difference between the in-situ measurements and the satellite measurements. If one of these measurements was not represented the interpolation would be less accurate. The other interpolation methods only used the auxilary data as a trend or not at all. Their interpolation on the in-situ measurements were therefore not or less effected.

In the satellite images it is possible to see why the SICCS product performed better then the CM-SAF product fig 73. The SICCS images were in general more complete than those of the CM-SAF. Since this pattern was visible in December every year the SICCS product seemed to be a little better to use if the IB or MB interpolation method is chosen as the most optimal interpolation method.

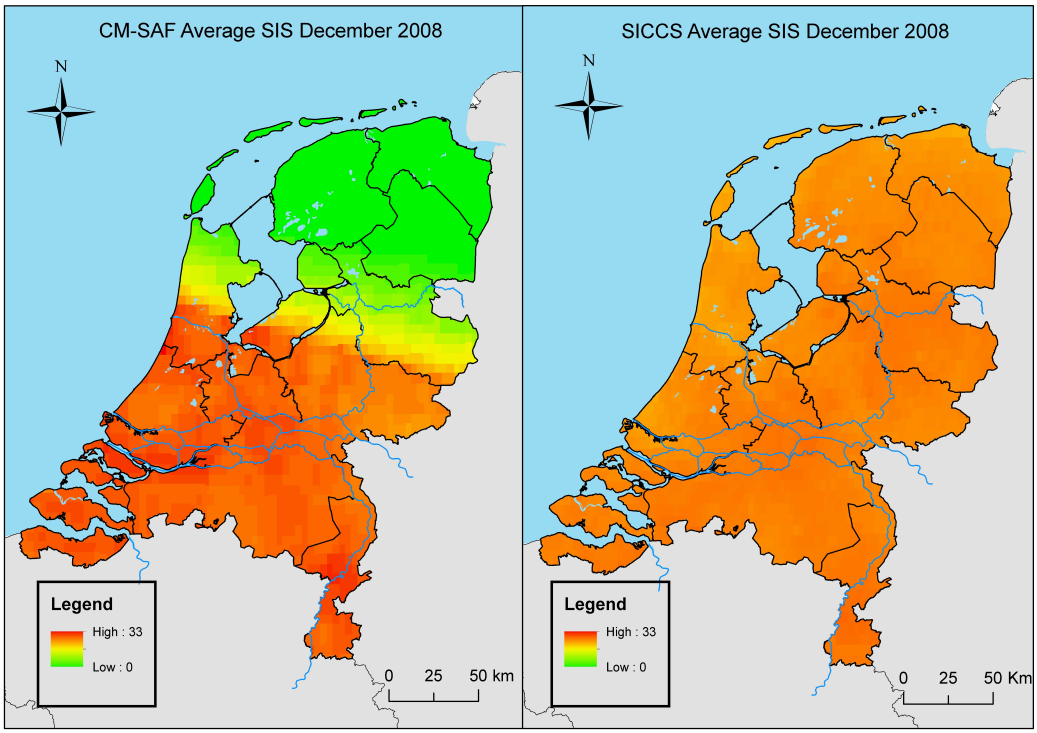


Figure 73: CM and SICCS satellite image for December 2008. The values of the CM-SAF image are 0 in the northern part of the Netherlands.

.

The only time the MAPE exceeded 11.62% was when the satellite images were not complete or had values of 0 as explained above. A MAPE of 11.62% is still relatively low. This error was found in January 2010 where radiation reaches maximum values of 35.87 W/m^2. This means that the error was still only $\pm 4.17 W/m^2$. Which was yet again under the target of the CM-SAF product.

Besides that, the MAPE of all other interpolation methods was very low on the cross validated interpolations (3.03% at maximum). The secondary standard of the WMO is set at 2-3% for measuring equipment. This means that the monthly MAPE of the combination of the satellite image and the in-situ measurements still came close to, or reach this target.

The data split that was used to analyze how the interpolations perform with less in-situ measurements showed that there were changes in MAPE values. The maximum MAPE of 7% which was observed is not as bad as it seems. The rea-

son this error was relatively high is due to low global radiation values that were present in January. The radiation in the satellite images ranged from 11 to 45 W/m^2. Therefore 7% was still a relative low. When looking at the absolute bias this corresponded with a value of 2.01 W/m^2, which was well below the CM-SAF optimal target bias of less than $8W/m^2$ (CM-SAF, 2013). Besides this particular error it was possible to see that all interpolation methods still perform very well. This would mean that even with less in-situ measurements the interpolation methods still manage to predict the global radiation values with high accuracy. Though this theory was only plausible so far when the in-situ observations were spread evenly around the country.

The increase in the MAPE in the second data split to 9.27% was yet again a relatively small increase. As explained above this value still only corresponded with an absolute bias of 2.71 W/m^2. An interesting thing was that in the first data split the SICCS product performed better when looking at the MAPE. In the second data split it was the other way around. Besides the January month, the CM-SAF product performed a better interpolation with stations missing at a concentrated area. Though the difference between the products MAPE was on a scale smaller then 2%.

The final analysis with random points in-between the stations showed that it is possible to come to the following conclusions:

1. All interpolation methods that used the satellite products as auxiliary data in their interpolation made good use of the pattern found in the satellite product. The interpolation method almost follows the exact same pattern as the satellite image.

2. All interpolation methods, including TPS, came very close to the satellite measurements when it comes down to absolute values. The satellite images seem to overestimate the global radiation but this is an overestimation which is often still within the allowed/target accuracy.

The results obtained confirm the findings of R. Sluiter that concluded that TPS was a good interpolation method for monthly and yearly climate elements (Sluiter, 2008).

For the daily interpolations TPS and the MB interpolation performed worse than the IB and KED interpolations. The reason why the TPS and the MB performed worse lies in the fact that the variation in global radiation on a daily scale is larger than that of the monthly and long term yearly data. For the daily data the differences in incoming global radiation between stations can be 4 to 10 times higher while this variation was very small in monthly data due to the fact that these maps were obtained by taking the average of daily measurements. Due to the high variation, larger errors were to be expected when performing a crossvalidation with TPS since one station can make a bigger difference when variation is bigger. This increase in variation also made it harder for the MB interpolation to get an accurate average error that could be accounted for in the satellite image.

The variation in the map, which causes the TPS and MB interpolation to perform worse, can be explained by weather conditions. For daily data weather conditions play an important role. The weather conditions on a daily basis are very random compared to a month or year average. One cloud above a measurement station can block out a big proportion of the potential incoming global radiation. This can cause high variation between different locations depeding on the local weather. The KED and IB methods performed a lot better on a daily basis. This is because these methods do account for the bigger variation and random patterns that are present due to weather conditions. Both KED and IB use the trend in the satellite image but alter the values in such a way that they correspond with the measurements made by the stations. By doing so the interpolation returned the correct values at the stations location and it keeps the trend from the satellite image. Therefore variation and the expected random patterns caused by weather can be observed in the results.

More importantly for this research, they are in agreement with the results obtained by Journée et al. (2010) in their research *"Improving the spatio-temporal distribution of surface solar radiation data by merging ground and satellite measuremetns".*

In this research they come to the following conclusion: *"The best merging performance was equally obtained by kriging technique (i.e., kriging with external drift) and by an adjustment of the SAF products with the spatially interpolated bias between stations and SAF data. The distribution of surface solar radiation inferred by merging ground and SAF data was systematiccaly more accurate than when us-*

ing each data source seperatly."

In the research performed by Journée et al. they looked at data with a high temporal resolution. This is in agreement with the results found in this research. The KED and IB method perform better when the temporal resolution is increasing. They also point out that this is caused due to the fact that by combining the data the strenghts from both products is used to its optimal. The satellite image contributes to the spatial distribution of the solar irradiance which is largely influenced by clouds and the in-situ measurements have their accuracy as a strong point. The combination of these strenghts and the coverage of the satellite data leads to a better product (Journée et al., 2010).

The results that were obtained by Journée et al. are represented in figure 74. The figure shows the cross-validated mean bias error (MBE), mean absolute error (MAE) and the root mean square error (RMSE) for all the interpolation/merging methods they used. These statistics are based on 2 years of data (2008 and 2009). In these results it is possible to see that the differences between the errors in the interpolation techniques are small.

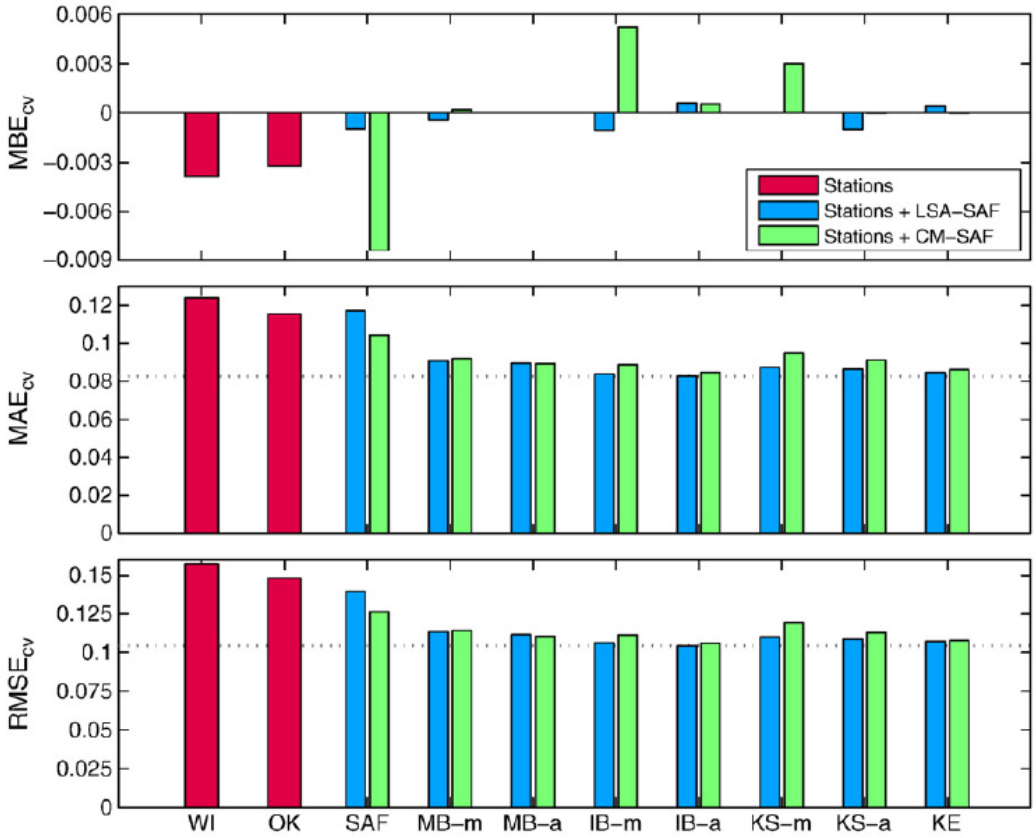


Figure 74: The cross-validated mean bias error (MBE), mean absolute error (MAE) and the root mean square error (RMSE) for the; Weighted interpolation (WI), Ordinary Kriging (OK), SAF-product (SAF), Mean bias correction (multiplicative adjustment) (MB-m), Mean bias correction (additive adjustment) (MB-a), Interpolated bias correction (multiplicative adjustment) (IB-m), Interpolated bias correction (additive adjustment) (IB-a), Ordinary Kriging with multiplicative satellite -based correction (KS-m), Ordinary Kriging with additive satellite-based correction (KS-a) and Kriging with external drift (KE) (Journée et al., 2010).

Expert Judgement

Since it turned out that both satellite products and all interpolation methods performed well when it comes down to absolute errors on a monthly and long term temporal resolution, it was hard to judge them for the real quality they give the user. Therefore as a last validation a discussion with the climate expert (R. Sluijter) was held to analyze patterns in radiation.

The discussion lead to the following findings:

1. The interpolation methods that made use of the satellite images create a more realistic map of radiation in the Netherlands. Spatial patterns of forests in the center of the Netherlands are made visible by using satellite data. This is not the case when performing a TPS interpolation. The same can be said about the patterns visible in Friesland (a province in the North of the Netherlands). More radiation is to be expected on the West coast of Friesland. This pattern is better represented in interpolations using satellite images as auxiliary data.

2. In the winter months strange or unexpected patterns are visible. This is caused by the lack of incoming solar radiation on shorter days combined with random (less predictable) weather processes. This is visible for all interpolation methods.

3. The SICCS product has a higher visual resolution with more visible variation.

4. For the daily data, adding auxilary data to the interpolation results in better maps. Although it isn't possible to see known patterns due to chaotic weather conditions the values returned are realistic and the chaotic patterns due to weather are expected to give this kind of output. The daily maps that do contain a certain pattern correspond with past weather conditions that are known from a different databases. Figure 75 shows the daily incoming global radiation on the 12th of May compared to the solar duration on that day, obtained from the KNMI database.

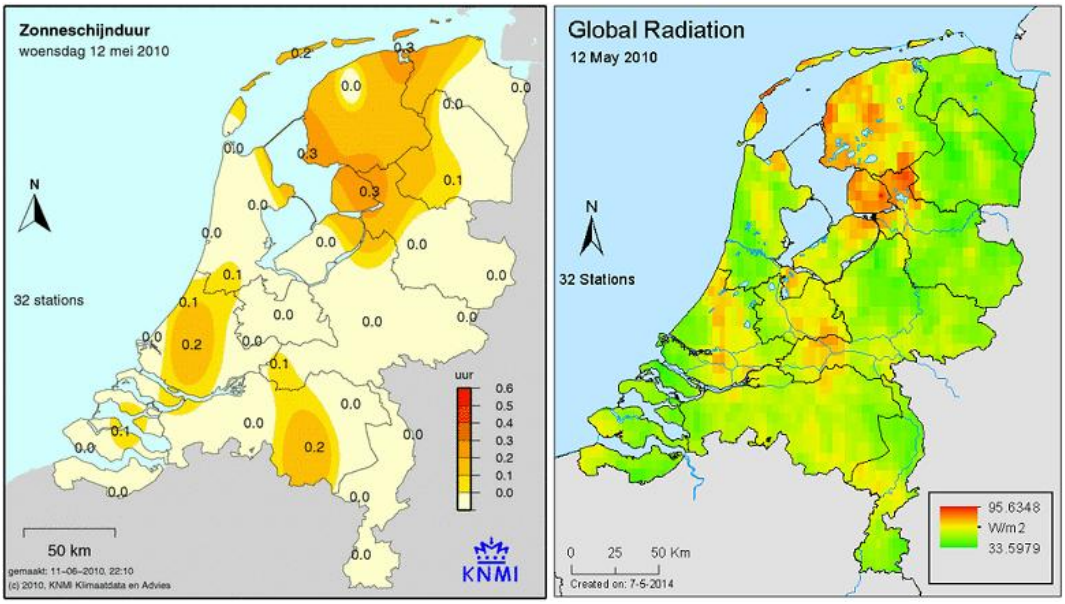


Figure 75: A comparisment between the solar duration (left) and the daily global radiation (right) for the 12th of May. The solar duration map was created with a TPS interpolation using 32 in-situ measurements. The global radiation map was created with a KED-EXP interpolation using 32 in-situ measurements and the CM-SAF product as auxilary data.

.

5.0.6 Conclusion

The results showed that all interpolation methods, depending on the temporal resolution, performed well at the locations of the in-situ measurement stations. Therefore all interpolation methods have their uses and cannot be classified as unusable. The output of all interpolation methods returned the pattern that would be expected when analyzing radiation in the Netherlands. This was especially visible when interpolating radiation on average timescales of one month or longer.

In-between the stations it is possible to see that the interpolation methods that used auxiliary satellite data follow the patterns observed in the satellite images better than that of TPS. Though it is possible to conclude that this difference was not big enough to say that one interpolation method turns out to be better than the rest. This is however only true for the longer term averages and montly data. On a visual side it is of course possible to make a choice for the "best" product. The reason for this is due to the fact that the stations are used to set the amount

of incoming global radiation and the satellite data is used to describe the patterns. This leads to maps with local observable patterns and a higher resolution.

In general the following conclusions can be made:

Long term average: All interpolations performed almost equally well here. The R^2 of 0.19 (CM-SAF) and 0.43 (SICCS) for the IB was lower and was therefore more doubtful as a model. However, the absolute errors were just as small as in the other interpolation methods. There was no significant difference between the two different satellite products. Therefore the choice of product and interpolation method is only dependent on the requirement specifications of the user.

Monthly data: TPS and both the KED interpolations performed the best in general. The IB interpolation seemed to be less predictable when looking at the R^2 values. When looking at the MAPE's it turned out that the MB interpolation scores significantly higher. This was mainly due to the December month where radiation values were low or satellite data was not complete. This pattern could especially be observed in the CM-SAF product where the December months lacked data. Therefore it could be said that for monthly data the SICCS product would be a better choice and KED or TPS should be used as interpolation method.

Daily data: Here it was possible to say that TPS and MB perform worse than KED and IB. TPS was not able to capture the chaotic patterns that can be observed on a daily basis due to weather conditions. The density of the in-situ measurement stations is to low to capture the variation that is present inbetween the stations. The big variation of the incoming solar radiation and the big differece between satellite observations and in-situ observations also make it hard to find a mean error in the satellite image. This leads to an inaccurate map using the MB method.

Due to these circomstances it is possible to say that the KED and IB interpolations would be the most optimal interpolation method on a daily basis. Especially the KED due to its stable R^2 compared to IB. Productwise, SICCS would be a better choise due to the smaller differences and errors in the satellite images and interpolations.

To come back and answer the research questions;

1. Is Kriging with external drift the best interpolation method, as expected?
2. Is the physically detailed SICCS product a better underlying data source than the CM-SAF product?

First off, Kriging with external drift is not necessarily the best interpolation method. It is one of the most stable methods and it does capture the trend of the satellite image very well. But, all other interpolation methods perform relatively equal when it comes down to absolute errors for the 6 year average and monthly data. It is possible to say though, that Kriging with external drift would be a safer choice to use in the months where radiation values are low and satellite images limited (in the Netherlands this would be in December and January). For daily data KED interpolation will result it more accurate maps compared to TPS and MB. KED is able to capture the chaotic weather conditions and therefore give more accurate results than TPS or MB interpolation. The advantage of using KED compared to using satellite images on their own would be the fact that Kriging accounts for the bias found in the satellite images. KED adjusts the values in the satellite images to match those of the in-situ measurements thereby reducing the errors in the product without reducing spatial resolution.

Secondly, the detailed SICCS product is not necessarily a better product to use as auxiliary data when interpolating global radiation in the Netherlands. The differences between the product are to small and change to much to say if one product is always better than the other. However for the December month the physically detailed SICCS product is more stable and for this month it would be wise to use the SICCS over the CM-SAF product.

For the daily data however the SICCS product would be a better choice. The differences between the SICCS product and the CM-SAF product after using interpolation are not that high. However, the satellite image on its own performs better when looking at the bias between the image and the in-situ measurements and would therefore be a more stable choice.

6 Outlook

Both of the anomalies/artefacts above the water bodies near and in the Netherlands are not a problem for global radiation maps in this paper. GSIE automatically masks the output maps with an overlay of the Netherlands and its water bodies. Hereby cutting out these two observed problems. However when the radiation maps are used as input for other models the data is not cut away and can lead to artefacts in the model.

Since it turns out that the satellite images on their own perform a very good job in predicting global radiation values it could be possible to immediately use the satellite images as global radiation maps instead of using an interpolation method. This would be true for atleast the long time average maps and monthly maps. The SICCS product is immediately available at the KNMI since it is produced here. The CM-SAF product is available without costs and could be used. However since these images seem to be less complete in the winter months it might not perform well enough as a stand alone product. To see if this is possible more research should be performed. Analysis have to be made on how good the images perform over longer periods of time and under different conditions. Also satellite equipment needs to be calibrated and corrected and both the models used to predict satellite global radiation measurements are dependent on input from auxiliary data themselves.

Further research in the improvement of global radiation maps in the Netherlands can be done. Since it turned out elevation has an influence on cloud formation a DEM (Digital Elevation Model) could be taken as auxiliary data to obtain an even more detailed model. Pons et al., (2008) have already shown that it is possible to predict global radiation by using a DEM. Other auxiliary data such as a vegetation or albedo map could also possibly improve the quality of radiation maps. The Veluwe area (which also has a higher elevation) and the centre of the Netherlands show lower radiation values on monthly and long term averages. These areas are characterised by forests. However to come to definite conclusions about these input sources more research has to be performed.

7 Bibliography

Ackerman, S.A., Heidinger, A., Foster, M.J., Maddux, B. 2013. Satellite Regional cloud climatology over the Great Lakes. Remote Sensing 5: 6223-6240.

Arvizu, D., P. Balaya, L. Cabeza, T. Hollands, A. Jger-Waldau, M. Kondo, C. Konseibo, V. Meleshko,W. Stein, Y. Tamaura, H. Xu, R. Zilles, 2011: Direct Solar Energy. In IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation [O. Edenhofer, R. Pichs-Madruga, Y. Sokona, K. Seyboth, P. Matschoss, S. Kadner, T. Zwickel, P. Eickemeier, G. Hansen, S. Schlmer, C. von Stechow (eds)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (2011).

Cano, D., Monget, J.M., Albuisson, M., Guillard, H., Regas, N., Wald, L. 1985. A method for the determination of the global solar radiation from meteorological satellite data. Solar Energy Vol. 37, No. 1:31-39.

Cubasch, U., D. Wuebbles, D. Chen, M.C. Facchini, D. Frame, N. Mahowald, and J.-G. Winther, 2013: Introduction. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Diabaté, L., Moussu, G., Wald, L. 1989. Description of an operational tool for determining global solar radiation at ground using geostationary satellite images. Solar energy Vol. 42. No. 3: 201-207.

Duffie, J.A., Beckman, W. 1994. Equation solving in solar energy engineering courses. Solar-annual meeting- American solar energy society. Solar '94-proceedings of the American solar energy society annual conference. 492-497. ISBN: 0-89553-166-6.

Greuell, W., Meirink. J.F., Wang. P. 2013. Retrieval and validation of global, direct, and diffuse irradiance derived from SEVIRI satellite observations. Journal of geophysical research: Atmospheres, Vol. 118: 2340-2361.

Hiemstra, P., Sluiter, R. 2011. Interpolation of Makkink evaporation in the Netherlands. Technical report; TR-327. KNMI, de Bilt, the Netherlands, version 1.0.

Huashan, L., Yongwang, L., Xianlong, W., Weibin, M., Liang, Z. 2010. Solar constant values for estimating solar radiation. Energy 36: 1785-1789.

Hyndman, R.J., Koehler. A.B. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22: 679-688.

Journée, M., Bertrand, C. 2010. Improving the spatio-temporal distribution of solar radiation data by merging ground and satellite measurements. Remote sensing of environment, 114: 2692-2704.

Journée, M., Stöckli. R., Bertrand, C. 2012. Sensitivity to spatio-temporal resolution of satellite-derived daily surface solar irradiation. Remote sensing letters, 3:4: 315-324.

Mueller, R.W., Matsoukas, C., Grazki, A., Behr, H.D., Hollmann, R., 2009. The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance - A LUT based eigenvector hybrid approach. Remote Sensing of Environment, 113: 1012-1024.

Noia, M., Ratto, C.F., Festa. R. 1993. Solar irradiance estimation from geostationary satellite data: I. Statistical models. Solar Energy Vol. 51 No. 6:449-456.

Noia, M., Ratto, C.F., Festa. R. 1993. Solar irradiance estimation from geostationary satellite data: II. Physical models. Solar Energy Vol. 51 No. 6:457-465.

Otani, K., Kurokawa, K., Tsuda, I., Saitoh, T., Horigome, T. 1994. Estimation of ground albedo by GMS images for solar irradiation monitoring. Solar Energy Materials and Solar Cells 35: 395-400.

Perez, R., Seals, R., Steward, R., Zelenka, A. 1994. Using satellite-derived insolation data for the site/time specific simulation of solar energy systems. Solar Energy Vol. 53 No. 6:491-495.

Perez, R., Seals, R., Zelenka, A. 1996. Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data. Solar Energy Vol. 60 No. 2:89-96.

Pons, X., Ninyerola, M. 2008. Mapping a topographic global solar radiation model implemented in a GIS and refined with ground data. International journal of climatology. 28: 1821-1834.

Schmetz, J., Pili. P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., et al. 2002. An introduction to Meteosat Second Generation (MSG). Bulletin of the American Meteorological Society. 83, 977-992.

Schuurmans, J.M., Bierkens, M.F.P., Pebesma, E.J. 2007. Automatic prediction of high resolution daily rainfall fields for multiple extents: The potential of operational radar. Joernal of hydrometeorology 8. 1204-1224.

Sluijter, R., Leenaers, H., Camarasa, M., 2010. De bosatlas van het klimaat. Noordhoff Uitgevers B.V.

Sluiter, R. 2008. Interpolation methods for climate data literature review. Internal report. KNMI. R and D Information and observation technology, version 1.0. (2009).

Sluiter, R. 2012. Interpolation methods for the climate atlas. Technical report. KNMI. De Bilt, The Netherlands, version 1.0.

Tarpley, J.D. 1979. Estimating incident solar radiation at the surface from geostationary satellite data. Journal of applied meteorology. Vol. 18: 1172-1181.

Tovar, H. F., Baldasano, J. M. 2001. Solar radiation mapping from NOAA AVHRR data in Catalonia, Spain. Journal of applied meteorology Vol. 40: 1821-1835.

WMO, 2008. Guide to Meteorological Instruments and Methods of Observation. WMO-No. 8, 7th edition. World Meteorological Organization, Geneva, Switzerland.

Yorukoglu, M., Celik, A.N. 2005. A critical review on the estimation of daily

global solar radiation from Sunshine duration. Energy Conversion and Management 47: 2441-2450.

Websites and downloadable content

CM-SAF satellite product. Retrieved 20 May, 2014, from:
(http://wui.cmsaf.eu/safira/action/viewProduktDetails?id=20909_20913).

ESA. Retrieved 20 May, 2014, from:
(http://www.esa.int/Our_Activities/Observing_the_Earth/MSG_overview).

EUMETSAT. Retrieved 20 May, 2014, from: (http://www.eumetsat.int/website/home/Satellites/LaunchesandOrbits/SatelliteOrbits/index.html).

EUMETSAT SAF. Retrieved 20 May, 2014, from: (http://www.eumetsat.int/website/home/Satellites/GroundSegment/Safs/index.html).

EUMETSAT MTG. Retrieved 19 May, 2014, from: (http://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/MTGDesign/index.html).

EUMETSAT CM-SAF. 2013. CM SAF Algorithm Theoretical Basis Document, MSG Surface Radiation, Edition 1. Retrieved 27 May, 2014, from:
(http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/atbd__pdf/SAF__CM__DWD__ATBD__SEVIRI__RAD__1.1,templateId=raw,property=publicationFile.pdf/SAF_CM_DWD_ATBD_SEVIRI_RAD_1.pdf).

EUMETSAT CM-SAF. 2012. CM SAF Cloud, Albedo, Radiation dataset, AVHRR-based, Edition 1. Surface Radiation Products. Algorithm Theoretical Basis Document. Retrieved 27 May, 2014, from: (http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/atbd__pdf/SAF__CM__DWD__ATBD__GAC__CLD__1.1,templateId=raw,property=publicationFile.pdf/SAF_CM_DWD_ATBD_GAC_CLD_1.pdf).

EUMETSAT CM-SAF. 2013. Validation Report, MSG Surface Radiation, Edition1. Retrieved 27 May, 2014, from:
(http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/pdf/SAF_

_CM__DWD__VAL__SEVIRI__RAD__1.2,templateId=raw,property=publicationFile.pdf/SAF_CM_DWD_VAL_SEVIRI_RAD_1.pdf).

Kipp & Zonen. Pyranometers. Brochure. Delft, The Netherlands. Retrieved 27 May, 2014, from: (http://www.kippzonen.com/DownloadCategory/5/Pyranometers).

Kipp & Zonen. Instruction Manual Pyranometer Pyranometer CM 11/CM 14. 1992. Delft, The Netherlands. Retrieved 27 May, 2014, from: (http://www.kippzonen.com/DownloadCategory/5/Pyranometers).

KNMI. 2005. Handboek Waarnemingen, 7 Straling. version april 2005. Retrieved 27 May, 2014, from: (http://www.knmi.nl/samenw/hawa/download.html)(In Dutch).

NASA. Retrieved 20 May, 2014, from: (http://www.nasa.gov/audience/forstudents/5-8/features/what-is-a-satellite-58.html#.Uvx88mJ5NDQ).

NASA. Landsat 7 Science Data Users Handbook. Retrieved 27 May, 2014, from: (http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_Handbook.pdf).

Nc-climate. Retrieved 20 June, 2014, from: (https://www.nc-climate.ncsu.edu/edu/k12/.vegetation).

R-Project. Retrieved 20 May, 2014, from: (http://www.r-project.org/).

SPOT. Retrieved 20 May, 2014, from: (http://www.cnes.fr/web/CNES-en/1420-technical-feat
php).

Personal communication

Climate Expert 2014, R. Sluijter. Climatologist at KNMI.

A Metadata

The research area is the entire Netherlands. Figure 76 shows the location of the Netherlands within Europe. The metadata used for all the maps of the Netherlands is described on the next page.

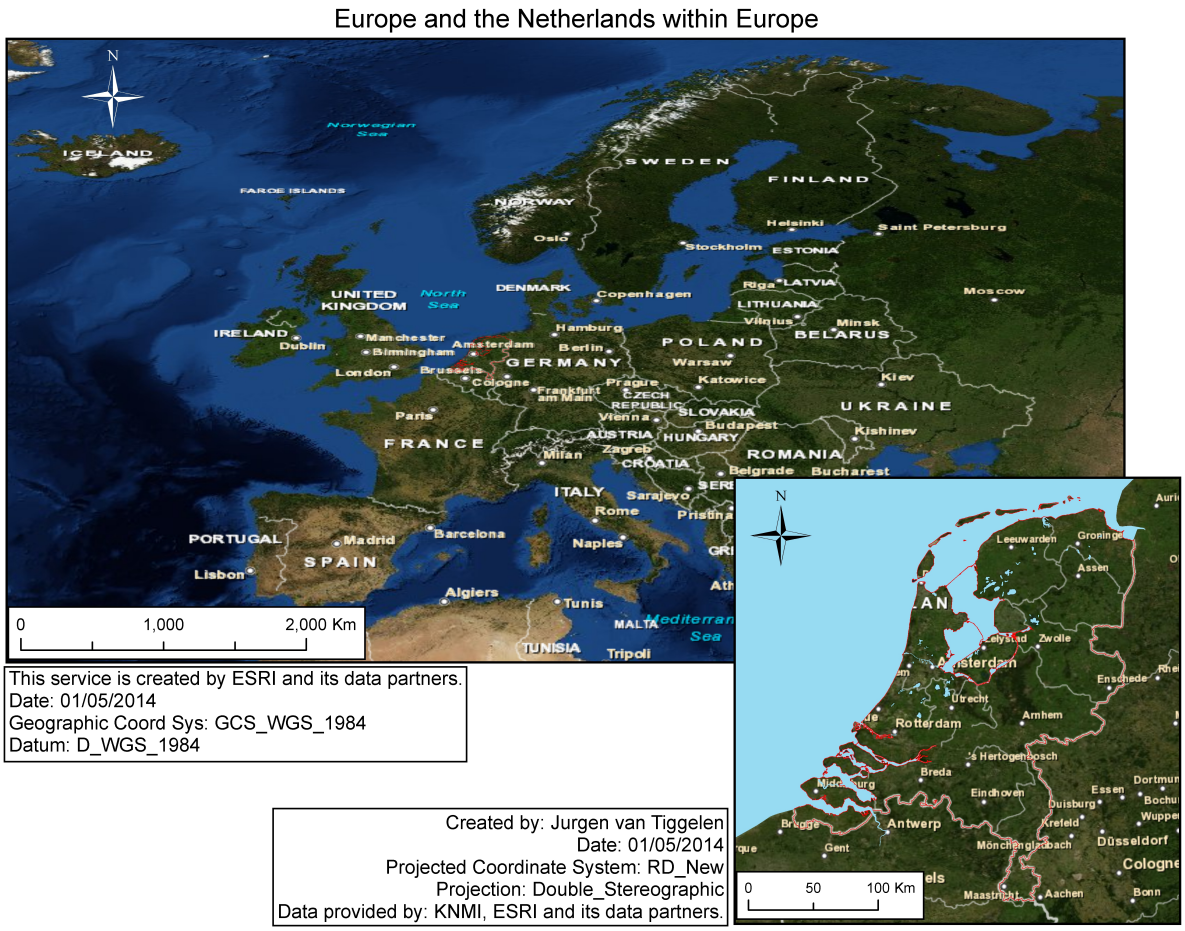


Figure 76: The Netherlands is hightlighted with a red border within Europe.

.

Metadata:

The output maps all have the same extend. They come with the following standard:

Left bottom corner (m) : (0, 300000)

Right top corner (m): (300000, 640000)

Cell size (m): (1000, 1000)

Rows/collumns : (340, 300)

For the CM-SAF product:

Maps created by: Jurgen van Tiggelen

Date: April 2014

Data provided by: CM-SAF and the KNMI.

Projected coordinate system: RD_New

Projection: Double_Stereographic

False_Easting: 155000.000000

False_Northing: 463000.000000

Central_Meridian: 5.387639

Scale_Factor: 0.999908

Latitude_Of_Origin: 52.156161

Linear Unit: Meter

Geographic coordinate system: GCS_Amersfoort

Datum: D_Amersfoort

For the SICCS product:

Maps created by: Jurgen van Tiggelen

Date: April 2014

Data provided by: The KNMI.

Projected coordinate system: RD_New

Projection: Double_Stereographic

False_Easting: 155000.000000

False_Northing: 463000.000000

Central_Meridian: 5.387639

Scale_Factor: 0.999908

Latitude_Of_Origin: 52.156161

Linear Unit: Meter

Geographic coordinate system: GCS_Amersfoort

Datum: D_Amersfoort

B R-script for global radiation using TPS

```
#REMOVE ALL OBJECTS
rm(list=ls(all=TRUE))

#SET WORKING DIRECTORY
setwd('F:/Data/Testmap')

#Current system date: "2014-01-28T08:12:17Z"
#Date used for the interpolation:
ISO8601Time="2010-01-01T00:00:00Z"

#LOAD PACKAGES library(methods)
library(sp)
library(gstat)
library(automap)
library(grid)
library(spam)

#LOAD VALIDATION AND TPS
source ("./inputdata/doTps.r")
source ("./inputdata/crossvalidate.r")

#Grid definitions
gridTopology = GridTopology(cellcentre.offset=c(0+1000/2,300000+1000/2), cellsize=c(1000,1000), cells.dim=c(300,340))
gridDefinition =SpatialGrid(gridTopology, proj4string = CRS(as.character(NA)))
c = coordinates(gridDefinition)

#LOAD IN-SITU DATA
var = read.table("./inputdata/2006/0612.dat",header=TRUE) #Querry output
var$VARIABLE = var$VARIABLE*10000
#var$VARIABLE = var$VARIABLE/31550399.99 #6YEAR AVERAGE
#var$VARIABLE = var$VARIABLE/31536000 #YEAR
#var$VARIABLE = var$VARIABLE/31622400 #LEAP YEAR (2008)
#var$VARIABLE = var$VARIABLE/2678400 #31 DAYS
#var$VARIABLE = var$VARIABLE/2592000 #30 DAYS
#var$VARIABLE = var$VARIABLE/2505600 #29 DAYS
#Var$VARIABLE = var$VARIABLE/2419200 #28 DAYS
coordinates(var) = ~RD_LOCATION_X+RD_LOCATION_Y

#SET AMOUNT OF STATIONS USED
StNr = nrow(var)

#READ IN MAKS GRID
nl.inputdata = read.asciigrid("./inputdata/wn_maskbuffer_001.asc")

#Get indexes of stations in the grid
stationIndicesInGrid=over(var,gridDefinition)

#Apply fixed coordinate system on mask map
gridded(nl.inputdata)=TRUE;
nl.grd = data.frame(mask = over(gridDefinition,nl.inputdata), xc = c[, 1], yc = c[, 2])
coordinates(nl.grd) = ~xc+yc
gridded(nl.grd) = TRUE
nl.grd = as(nl.grd, "SpatialGridDataFrame")
fullgrid(nl.grd) = TRUE
```

```
#Start TPS process
# TPS preprocess data
nl_tps.grd = nl.grd
fullgrid(nl_tps.grd) = FALSE

# TPS Standard code
lambda_fixed=0.004 # set TPS lambda smoothing parameter, NA for autofit
tps = doTps(VARIABLE ~1, var, nl_tps.grd, addFit = TRUE, debug.level = 1, lambda = lambda_fixed)
lambdafit=tps$fit$lambda
#write.table(lambdafit, "./output/lambda.txt", row.names=FALSE, col.names=FALSE)

#TPS Cross validation
tps.cv = crossvalidate(VARIABLE ~1, var, func = "doTps", debug.level = 0, lambda = lambda_fixed)

teller = sum(tps.cv$residual^2)
noemer = sum((var$VARIABLE - mean(var$VARIABLE))^2)
tps.r2 = 1 - teller/noemer
tps_meanvar = mean(tps.cv$residual)
tps_maxvar = max(tps.cv$residual)
tps_minvar = min(tps.cv$residual)
tps_sdvar = sd (tps.cv$residual,na.rm=TRUE)
cv_TPS = data.frame(tps.r2, tps_minvar, tps_maxvar, tps_meanvar, tps_sdvar)
#write.table(cv_TPS, "./output/TPS_R2.txt", row.names=FALSE, col.names=TRUE)

# Cut off at zero (no negative EV allowed)
result = tps$krige_output
result$var1.pred = pmax(result$var1.pred,0)

# Grid output
write.asciigrid(result, "./output/prediction.asc", attr = "var1.pred", na.value = -9999)

# Data output
# Stations output
# create a dataframe with the station numbers and their corresponding indexes in the grid
stationAndIndex = data.frame(stationIndicesInGrid,var$STN,var$VARIABLE)
#remove stations which do not fall within the new grid
stationAndIndex = na.omit(stationAndIndex );

# StationFields #
#Fill all default values with 0
stationField=data.frame(data.frame(gridDefinition ,0))
#Fill in the stations
stationField[stationAndIndex$stationIndicesInGrid,3]=stationAndIndex$var.STN
#Create a SpatialDataGrid
coordinates(stationField)=~s1+s2
gridded(stationField) = TRUE

# StationValues #
#Fill all default values with 0
stationValues=data.frame(data.frame(gridDefinition ,0))
#Fill in the stations values
stationValues[stationAndIndex$stationIndicesInGrid,3]=stationAndIndex$var.VARIABLE
#Create a SpatialDataGrid
coordinates(stationValues)=~s1+s2
gridded(stationValues) = TRUE
```

```
# Grid output
write.asciigrid(stationValues, "./output/stationvalues.asc",na.value = 0)
write.asciigrid(stationField, "./output/stations.asc",na.value = 0)

#MORE STATISTICS
#Mean Error TPS
ME = ((1/StNr)*((sum(abs(tps.cv$residual)))))
MEmean = (ME/(mean(tps.cv$observed)))

#Root mean squared error
RMSE = (sqrt((1/StNr) * ((sum((tps.cv$var1.pred - tps.cv$observed)^2)))))
sdvar = ((tps.cv$observed - (mean(tps.cv$observed)))^2)
sdvar = sum(sdvar)
sdvar = ((1/(StNr-1))*sdvar)
sdvar = sqrt(sdvar)
RMSEsd = (RMSE/sdvar)
NRMSE = (RMSE/((max(var$VARIABLE)-(min(var$VARIABLE)))))

#MAPE
tps.cv$MAE = abs((tps.cv$residual/tps.cv$observed))
MAEsum = sum(tps.cv$MAE)
MAPE = (MAEsum*(100/StNr))
#TPS_errors = data.frame(ME, MEmean, RMSE, RMSEsd, NRMSE, MAPE)
TPS_errors = data.frame(tps.r2, MAPE)
write.table(TPS_errors, "./output/TPS_Errors.txt", row.names=FALSE, col.names=TRUE)
```

C R-script for global radiation using MB

```
#REMOVE ALL OBJECTS
rm(list=ls(all=TRUE))

#SET WORKING DIRECTORY
setwd('F:/Data/Testmap')

#Current system date: "2014-01-28T08:12:17Z"
#Date used for the interpolation:
ISO8601Time="2010-01-01T00:00:00Z"

#LOAD PACKAGES library(methods)
library(sp)
library(gstat)
library(automap)
library(grid)
library(spam)

#LOAD VALIDATION AND TPS
source ("./inputdata/doTps.r")
source ("./inputdata/crossvalidate.r")

#Grid definitions
gridTopology = GridTopology(cellcentre.offset=c(0+1000/2,300000+1000/2), cellsize=c(1000,1000), cells.dim=c(300,340))
gridDefinition =SpatialGrid(gridTopology, proj4string = CRS(as.character(NA)))
c = coordinates(gridDefinition)

#LOAD SATELLITE IMAGE
sis.grd = read.asciigrid("./inputdata/2006/0612.asc",colname="sis")

#LOAD IN-SITU DATA
var = read.table("./inputdata/2006/0612.dat",header=TRUE) #Querry output
var$VARIABLE = var$VARIABLE*10000
#var$VARIABLE = var$VARIABLE/31550399.99 #6YEAR AVERAGE
#var$VARIABLE = var$VARIABLE/31536000 #YEAR
#var$VARIABLE = var$VARIABLE/31622400 #LEAP YEAR (2008)
```

```
#var$VARIABLE = var$VARIABLE/2678400 #31 DAYS
#var$VARIABLE = var$VARIABLE/2592000 #30 DAYS
#var$VARIABLE = var$VARIABLE/2505600 #29 DAYS
#Var$VARIABLE = var$VARIABLE/2419200 #28 DAYS
coordinates(var) = ~RD_LOCATION_X+RD_LOCATION_Y

#SET AMOUNT OF STATIONS USED
StNr = nrow(var)

#READ IN MAKS GRID
nl.inputdata = read.asciigrid("./inputdata/wn_maskbuffer_001.asc")

#Get indexes of stations in the grid
stationIndicesInGrid=over(var,gridDefinition)

#Apply fixed coordinate system on mask map
gridded(nl.inputdata)=TRUE;
nl.grd = data.frame(mask = over(gridDefinition,nl.inputdata), xc = c[, 1], yc = c[, 2])
coordinates(nl.grd) = ~xc+yc
gridded(nl.grd) = TRUE
nl.grd = as(nl.grd, "SpatialGridDataFrame")
fullgrid(nl.grd) = TRUE

# Calculate the mean bias
diff.grd = read.asciigrid("./output/stationValues.asc",colname="sis") #read in the stationvalues as a grid
diff.grd$sis = (diff.grd$sis-sis.grd$sis) #calculate the difference between the ground observations and satellite
observations
Mbias = mean(diff.grd$sis, na.rm=TRUE) #Find the mean of the differences

#Create and write the grid
Mbiasgrid = sis.grd #make a new grid with the same extend as the other grids
Mbiasgrid$sis = 0 #Change the values in the grid to 0 so its an "empty" grid
Mbiasgrid$sis = (sis.grd$sis+Mbias) #fill in the grid by using the satellite values and add the mean bias error

#Output
#write.table(Mbias, "./output/Mean_Bias.txt", row.names=FALSE, col.names=TRUE)
write.asciigrid(Mbiasgrid, "./output/MeanBias.asc", attr = "sis", na.value = -9999)
```

#Stations output

stations = data.frame(stationIndicesInGrid,varSTN,varVARIABLE)

#remove stations which do not fall within the new grid

stations = na.omit(stations);

stations$predicted = (Mbiasgrid$sis[stations$stationIndicesInGrid]) #load predicted values into overview

stations$residual = (stations$var.VARIABLE-stations$predicted) #calculate the residuals

MbiasStat = stations

Statistics

teller = $sum(MbiasStat\$residual^2)$

noemer = $sum((var\$VARIABLE - mean(var\$VARIABLE))^2)$

Mbias.r2 = 1 - teller/noemer

Mbias_meanvar = mean(MbiasStat$residual)

Mbias_maxvar = max(MbiasStat$residual)

Mbias_minvar = min(MbiasStat$residual)

Mbias_sdvar = sd (MbiasStat$residual,na.rm=TRUE)

cv_Mbias = data.frame(Mbias.r2, Mbias_minvar, Mbias_maxvar, Mbias_meanvar, Mbias_sdvar)

#write.table(cv_Mbias, "./output/MbiasStats.txt", row.names=FALSE, col.names=TRUE)

#Mean Error MB

ME = ((1/StNr)*(sum(abs(MbiasStat$residual)))) #Calculate the mean error

MEmean = (ME/(mean(MbiasStat$var.VARIABLE))) #calculate the mean error mean

#Root mean squared error

RMSE =$(sqrt((1/StNr) * ((sum((MbiasStat\$predicted - MbiasStat\$var.VARIABLE)^2)))))$ #Calculate the Root mean Square error

RMSEsd = (RMSE/Mbias_sdvar)

NRMSE = (RMSE/((max(MbiasStat$var.VARIABLE)-(min(MbiasStat$var.VARIABLE)))))

#MAPE

MbiasStat$MAE = abs((MbiasStat$residual/MbiasStat$var.VARIABLE))

MAEsum = sum(MbiasStat$MAE)

MAPE = (MAEsum*(100/StNr))

#MBias_errors = data.frame(ME, MEmean, RMSE, RMSEsd, NRMSE, MAPE)

MBias_errors = data.frame(Mbias.r2, MAPE)

write.table(MBias_errors, "./output/Mbias_Errors.txt", row.names=FALSE, col.names=TRUE)

D R-script for global radiation using IB

```
#REMOVE ALL OBJECTS
rm(list=ls(all=TRUE))

#SET WORKING DIRECTORY
setwd('F:/Data/Testmap')

#Current system date: "2014-01-28T08:12:17Z"
#Date used for the interpolation:
ISO8601Time="2010-01-01T00:00:00Z"

#LOAD PACKAGES library(methods)
library(sp)
library(gstat)
library(automap)
library(grid)
library(spam)

#LOAD VALIDATION AND TPS
source ("./inputdata/doTps.r")
source ("./inputdata/crossvalidate.r")

#Grid definitions
gridTopology = GridTopology(cellcentre.offset=c(0+1000/2,300000+1000/2), cellsize=c(1000,1000), cells.dim=c(300,340))
gridDefinition =SpatialGrid(gridTopology, proj4string = CRS(as.character(NA)))
c = coordinates(gridDefinition)

mxdidw=120000 # maxdist IDW
mxdkrige=Inf # maxdist Krige

#LOAD SATELLITE IMAGE
sis.grd = read.asciigrid("./inputdata/2006/0612.asc",colname="sis")

#LOAD IN-SITU DATA
var = read.table("./inputdata/2006/0612.dat",header=TRUE) #Querry output
var$VARIABLE = var$VARIABLE*10000
```

```
#var$VARIABLE = var$VARIABLE/31550399.99 #6YEAR AVERAGE
#var$VARIABLE = var$VARIABLE/31536000 #YEAR
#var$VARIABLE = var$VARIABLE/31622400 #LEAP YEAR (2008)
#var$VARIABLE = var$VARIABLE/2678400 #31 DAYS
#var$VARIABLE = var$VARIABLE/2592000 #30 DAYS
#var$VARIABLE = var$VARIABLE/2505600 #29 DAYS
#Var$VARIABLE = var$VARIABLE/2419200 #28 DAYS
coordinates(var) = ~RD_LOCATION_X+RD_LOCATION_Y

#SET AMOUNT OF STATIONS USED
StNr = nrow(var)
InterBias = var

#READ IN MAKS GRID
nl.inputdata = read.asciigrid("./inputdata/wn_maskbuffer_001.asc")

#Get indexes of stations in the grid
stationIndicesInGrid=over(var,gridDefinition)

#Apply fixed coordinate system on mask map
gridded(nl.inputdata)=TRUE;
nl.grd = data.frame(mask = over(gridDefinition,nl.inputdata), xc = c[, 1], yc = c[, 2])
coordinates(nl.grd) = ~xc+yc
gridded(nl.grd) = TRUE
nl.grd = as(nl.grd, "SpatialGridDataFrame")
fullgrid(nl.grd) = TRUE

#create empty raster to work on
Ibiasgrid = sis.grd #make a new grid with the same extend as the other grids
Ibiasgrid$sis = 0 #Change the values in the grid to 0 so its an "empty" grid

#Stations output
# create a dataframe with the station numbers and their corresponding indexes in the grid
stations2 = data.frame(stationIndicesInGrid,var$STN,InterBias$VARIABLE)
#remove stations which do not fall within the new grid
stations2 = na.omit(stations2 );
stations2$predicted = (sis.grd$sis[stations2$stationIndicesInGrid]) #load predicted values into overview
```

```
stations2$residual = (stations2$InterBias.VARIABLE-stations2$predicted) #calculate the residuals
InterbiasStat = stations2

InterBias$VARIABLE = InterbiasStat$residual #set the residuals as the variable to be interpolated

#Interpolate the bias
idw = idw(VARIABLE~1, InterBias, Ibiasgrid, maxdist=mxdidw, na.action=na.pass)

idw.cv = crossvalidate(VARIABLE~1, InterBias, debug.level = 0)
teller = sum(idw.cv$residual^2)
noemer = sum((InterBias$VARIABLE - mean(InterBias$VARIABLE))^2)
idw.r2 = 1 - teller/noemer
idw_sph.cv = idw.cv

#fill in the data
Ibiasgrid$sis = (sis.grd$sis+idw$var1.pred) #fill in the grid by using the satellite values and add the mean bias
error

write.asciigrid(Ibiasgrid, "./output/Ibias.asc", attr = "sis", na.value = -9999)

#Stations output
#create a dataframe with the station numbers and their corresponding indexes in the grid
stations3 = data.frame(stationIndicesInGrid,InterBias$STN,InterBias$VARIABLE)
#remove stations which do not fall within the new grid
stations3 = na.omit(stations3 );

stations3$predicted = (Ibiasgrid$sis[stations3$stationIndicesInGrid]) #load predicted values into overview
stations3$residual = (var$VARIABLE-stations3$predicted) #calculate the residuals
IbiasStat = stations3

#Statistics
Ibias_meanvar = mean(IbiasStat$residual)
Ibias_maxvar = max(IbiasStat$residual)
Ibias_minvar = min(IbiasStat$residual)
Ibias_sdvar = sd (IbiasStat$residual,na.rm=TRUE)

cv_Ibias = data.frame(idw.r2, Ibias_minvar, Ibias_maxvar, Ibias_meanvar, Ibias_sdvar)
```

#write.table(cv_Ibias, "./output/IbiasStats.txt", row.names=FALSE, col.names=TRUE)

#Mean Error TPS

ME = ((1/StNr)*((sum(IbiasStat$predicted-var$VARIABLE)))) #Calculate the mean error

MEmean = (ME/(mean(var$VARIABLE))) #calculate the mean error mean

#Root mean squared error

RMSE = $(sqrt((1/StNr) * ((sum((IbiasStat\$predicted - var\$VARIABLE)^2)))))$ #Calculate the Root mean Square error

RMSEsd = (RMSE/Ibias_sdvar)

NRMSE = (RMSE/((max(var$VARIABLE)-(min(var$VARIABLE)))))

IbiasStat$MAE = abs((IbiasStat$residual/var$VARIABLE))

MAEsum = sum(IbiasStat$MAE)

MAPE = (MAEsum*(100/StNr))

#IBias_errors = data.frame(ME, MEmean, RMSE, RMSEsd, NRMSE, MAPE)

IBias_errors = data.frame(idw.r2, MAPE)

write.table(IBias_errors, "./output/Ibias_Errors.txt", row.names=FALSE, col.names=TRUE)

E R-script for global radiation using KED-Exp

```
#REMOVE ALL OBJECTS
rm(list=ls(all=TRUE))

#SET WORKING DIRECTORY
setwd('F:/Data/Testmap')

#Current system date: "2014-01-28T08:12:17Z"
#Date used for the interpolation:
ISO8601Time="2010-01-01T00:00:00Z"

#LOAD PACKAGES library(methods)
library(sp)
library(gstat)
library(automap)
library(grid)
library(spam)

#LOAD VALIDATION AND TPS
source ("./inputdata/doTps.r")
source ("./inputdata/crossvalidate.r")

#Grid definitions
gridTopology = GridTopology(cellcentre.offset=c(0+1000/2,300000+1000/2), cellsize=c(1000,1000), cells.dim=c(300,340))
gridDefinition =SpatialGrid(gridTopology, proj4string = CRS(as.character(NA)))
c = coordinates(gridDefinition)

mxdkrige=Inf # maxdist Krige

#LOAD SATELLITE IMAGE
sis.grd = read.asciigrid("./inputdata/2006/0612.asc",colname="sis")

#LOAD IN-SITU DATA
var = read.table("./inputdata/2006/0612.dat",header=TRUE) #Querry output
var$VARIABLE = var$VARIABLE*10000
#var$VARIABLE = var$VARIABLE/31550399.99 #6YEAR AVERAGE
```

```
#var$VARIABLE = var$VARIABLE/31536000 #YEAR
#var$VARIABLE = var$VARIABLE/31622400 #LEAP YEAR (2008)
#var$VARIABLE = var$VARIABLE/2678400 #31 DAYS
#var$VARIABLE = var$VARIABLE/2592000 #30 DAYS
#var$VARIABLE = var$VARIABLE/2505600 #29 DAYS
#Var$VARIABLE = var$VARIABLE/2419200 #28 DAYS
coordinates(var) = ~RD_LOCATION_X+RD_LOCATION_Y

#SET AMOUNT OF STATIONS USED
StNr = nrow(var)

#READ IN MAKS GRID
nl.inputdata = read.asciigrid("./inputdata/wn_maskbuffer_001.asc")

#Get indexes of stations in the grid
stationIndicesInGrid=over(var,gridDefinition)

#Apply fixed coordinate system on mask map
gridded(nl.inputdata)=TRUE;
nl.grd = data.frame(mask = over(gridDefinition,nl.inputdata), xc = c[, 1], yc = c[, 2])
coordinates(nl.grd) = ~xc+yc
gridded(nl.grd) = TRUE
nl.grd = as(nl.grd, "SpatialGridDataFrame")
fullgrid(nl.grd) = TRUE

# Overlay functions
sis.ov=overlay(sis.grd,var)
# Copy the values to Var
var$sis=sis.ov$sis

ked = autoKrige(VARIABLE~sis, var, sis.grd, maxdist=mxdkrige, model = c("Exp"), na.action=na.pass, fix.values=c(0,NA,NA),
miscFitOptions = list(merge.small.bins = FALSE))

# Krige Cross validation
ked.cv = autoKrige.cv(VARIABLE~sis, var, model = c("Exp"),maxdist=mxdkrige,fix.values=c(0,NA,NA), mis-
cFitOptions = list(merge.small.bins = FALSE))
teller = sum(ked.cv$krige.cv_output$residual^2)
```

noemer = $sum((var\$VARIABLE - mean(var\$VARIABLE))^2)$

ked_exp.r2 = 1 - teller/noemer

ked.zscoremean = mean(ked.cv$krige.cv_output$zscore)

ked.zscore.var = var(ked.cv$krige.cv_output$zscore)

cv_exp = data.frame(ked_exp.r2,ked.zscoremean, ked.zscore.var)

var_cv = ked.cv

#write.table(cv_exp, "./output/ked_exp_cv.txt", row.names=FALSE, col.names=TRUE)

FORCE GRID CELLS TO BE SQUARE

ked=ked$krige_output

slot(slot(ked, "grid"), "cellsize") = rep(mean(slot(slot(ked, "grid"), "cellsize")), 2)

Calculate differences at observation points

predicted= overlay (ked,var)

var$predicted = predicted$var1.pred

var$difference = (var$VARIABLE - var$predicted)

difmin = min (var$difference,na.rm=TRUE)

difmax = max (var$difference,na.rm=TRUE)

difmean = mean (var$difference,na.rm=TRUE)

difsd = sd (var$difference,na.rm=TRUE)

output = data.frame(difmin,difmax,difmean,difsd)

#write.table(output, "./output/ked_exp_pointdifference.txt", row.names=FALSE, col.names=TRUE)

Grid output

write.asciigrid(ked, "./output/ked_exp_prediction.asc", attr = "var1.pred", na.value = -9999)

write.asciigrid(ked, "./output/ked_exp_variance.asc", attr = "var1.var", na.value = -9999)

#Mean Error KED

ME = ((1/StNr)*((sum(var$difference)))) #Calculate the mean error

MEmean = (ME/(mean(var$VARIABLE))) #calculate the mean error mean

#Root mean squared error

RMSE = ($sqrt((1/StNr) * ((sum((var\$predicted - var\$VARIABLE)^2)))))$ #Calculate the Root mean Square error

sdvar = $((var\$VARIABLE - (mean(var\$VARIABLE)))^2)$

sdvar = sum(sdvar)

sdvar = ((1/(StNr-1))*sdvar)

sdvar = sqrt(sdvar)

```
RMSEsd = (RMSE/sdvar)
RMSEsd = (RMSE/difsd)
NRMSE = (RMSE/((max(var$VARIABLE)-(min(var$VARIABLE)))))

var$MAE = abs((var$difference/var$VARIABLE))
MAEsum = sum(var$MAE)
MAPE = (MAEsum*(100/StNr))

#KED_EXP_errors = data.frame(ME, MEmean, RMSE, RMSEsd, NRMSE, MAPE)
KED_EXP_errors = data.frame(ked_exp.r2, MAPE)
write.table(KED_EXP_errors, "./output/KED_EXP_Errors.txt", row.names=FALSE, col.names=TRUE)
```

F R-script for global radiation using KED-Sph

```
#REMOVE ALL OBJECTS
rm(list=ls(all=TRUE))

#SET WORKING DIRECTORY
setwd('F:/Data/Testmap')

#Current system date: "2014-01-28T08:12:17Z"
#Date used for the interpolation:
ISO8601Time="2010-01-01T00:00:00Z"

#LOAD PACKAGES library(methods)
library(sp)
library(gstat)
library(automap)
library(grid)
library(spam)

#LOAD VALIDATION AND TPS
source ("./inputdata/doTps.r")
source ("./inputdata/crossvalidate.r")

#Grid definitions
gridTopology = GridTopology(cellcentre.offset=c(0+1000/2,300000+1000/2), cellsize=c(1000,1000), cells.dim=c(300,340))
gridDefinition =SpatialGrid(gridTopology, proj4string = CRS(as.character(NA)))
c = coordinates(gridDefinition)

mxdkrige=Inf # maxdist Krige

#LOAD SATELLITE IMAGE
sis.grd = read.asciigrid("./inputdata/2006/0612.asc",colname="sis")

#LOAD IN-SITU DATA
var = read.table("./inputdata/2006/0612.dat",header=TRUE) #Querry output
var$VARIABLE = var$VARIABLE*10000
#var$VARIABLE = var$VARIABLE/31550399.99 #6YEAR AVERAGE
```

```
#var$VARIABLE = var$VARIABLE/31536000 #YEAR
#var$VARIABLE = var$VARIABLE/31622400 #LEAP YEAR (2008)
#var$VARIABLE = var$VARIABLE/2678400 #31 DAYS
#var$VARIABLE = var$VARIABLE/2592000 #30 DAYS
#var$VARIABLE = var$VARIABLE/2505600 #29 DAYS
#Var$VARIABLE = var$VARIABLE/2419200 #28 DAYS
coordinates(var) = ~RD_LOCATION_X+RD_LOCATION_Y

#SET AMOUNT OF STATIONS USED
StNr = nrow(var)

#READ IN MAKS GRID
nl.inputdata = read.asciigrid("./inputdata/wn_maskbuffer_001.asc")

#Get indexes of stations in the grid
stationIndicesInGrid=over(var,gridDefinition)

#Apply fixed coordinate system on mask map
gridded(nl.inputdata)=TRUE;
nl.grd = data.frame(mask = over(gridDefinition,nl.inputdata), xc = c[, 1], yc = c[, 2])
coordinates(nl.grd) = ~xc+yc
gridded(nl.grd) = TRUE
nl.grd = as(nl.grd, "SpatialGridDataFrame")
fullgrid(nl.grd) = TRUE

# Overlay functions
sis.ov=overlay(sis.grd,var)
# Copy the values to Var
var$sis=sis.ov$sis

# Kriging
ked = autoKrige(VARIABLE~sis, var, sis.grd, maxdist=mxdkrige, model = c("Sph"), na.action=na.pass, fix.values=c(0,NA,NA),
miscFitOptions = list(merge.small.bins = FALSE))

# Krige Cross validation
ked.cv = autoKrige.cv(VARIABLE~sis, var, model = c("Sph"),maxdist=mxdkrige,fix.values=c(0,NA,NA), mis-
cFitOptions = list(merge.small.bins = FALSE))
```

teller = $sum(ked.cv\$krige.cv\_output\$residual^2)$

noemer = $sum((var\$VARIABLE - mean(var\$VARIABLE))^2)$

ked_sph.r2 = 1 - teller/noemer

ked.zscoremean = mean(ked.cv$krige.cv_output$zscore)

ked.zscore.var = var(ked.cv$krige.cv_output$zscore)

cv_sph = data.frame(ked_sph.r2,ked.zscoremean, ked.zscore.var)

#write.table(cv_sph, "./output/ked_sph_cv.txt", row.names=FALSE, col.names=TRUE)

FORCE GRID CELLS TO BE SQUARE

ked=ked$krige_output

slot(slot(ked, "grid"), "cellsize") = rep(mean(slot(slot(ked, "grid"), "cellsize")), 2)

Calculate differences at observation points

predicted= overlay (ked,var)

var$predicted = predicted$var1.pred

var$difference = (var$VARIABLE - var$predicted)

difmin = min (var$difference,na.rm=TRUE)

difmax = max (var$difference,na.rm=TRUE)

difmean = mean (var$difference,na.rm=TRUE)

difsd = sd (var$difference,na.rm=TRUE)

output = data.frame(difmin,difmax,difmean,difsd)

#write.table(output, "./output/ked_sph_pointdifference.txt", row.names=FALSE, col.names=TRUE)

Grid output

write.asciigrid(ked, "./output/ked_sph_prediction.asc", attr = "var1.pred", na.value = -9999)

write.asciigrid(ked, "./output/ked_sph_variance.asc", attr = "var1.var", na.value = -9999)

#Mean Error KED

ME = ((1/StNr)*((sum(var$difference)))) #Calculate the mean error

MEmean = (ME/(mean(var$VARIABLE))) #calculate the mean error mean

#Root mean squared error

RMSE = $(sqrt((1/StNr) * ((sum((var\$predicted - var\$VARIABLE)^2)))))$ #Calculate the Root mean Square error

sdvar = $((var\$VARIABLE - (mean(var\$VARIABLE)))^2)$

sdvar = sum(sdvar)

sdvar = ((1/(StNr-1))*sdvar)

sdvar = sqrt(sdvar)

```
RMSEsd = (RMSE/sdvar)
RMSEsd = (RMSE/difsd)
NRMSE = (RMSE/((max(var$VARIABLE)-(min(var$VARIABLE)))))

var$MAE = abs((var$difference/var$VARIABLE))
MAEsum = sum(var$MAE)
MAPE = MAEsum*(100/StNr))

#KED_SPH_errors = data.frame(ME, MEmean, RMSE, RMSEsd, NRMSE, MAPE)
KED_SPH_errors = data.frame(ked_sph.r2, MAPE)
write.table(KED_SPH_errors, "./output/KED_SPH_Errors.txt", row.names=FALSE, col.names=TRUE)
```

G Elevation in the Netherlands

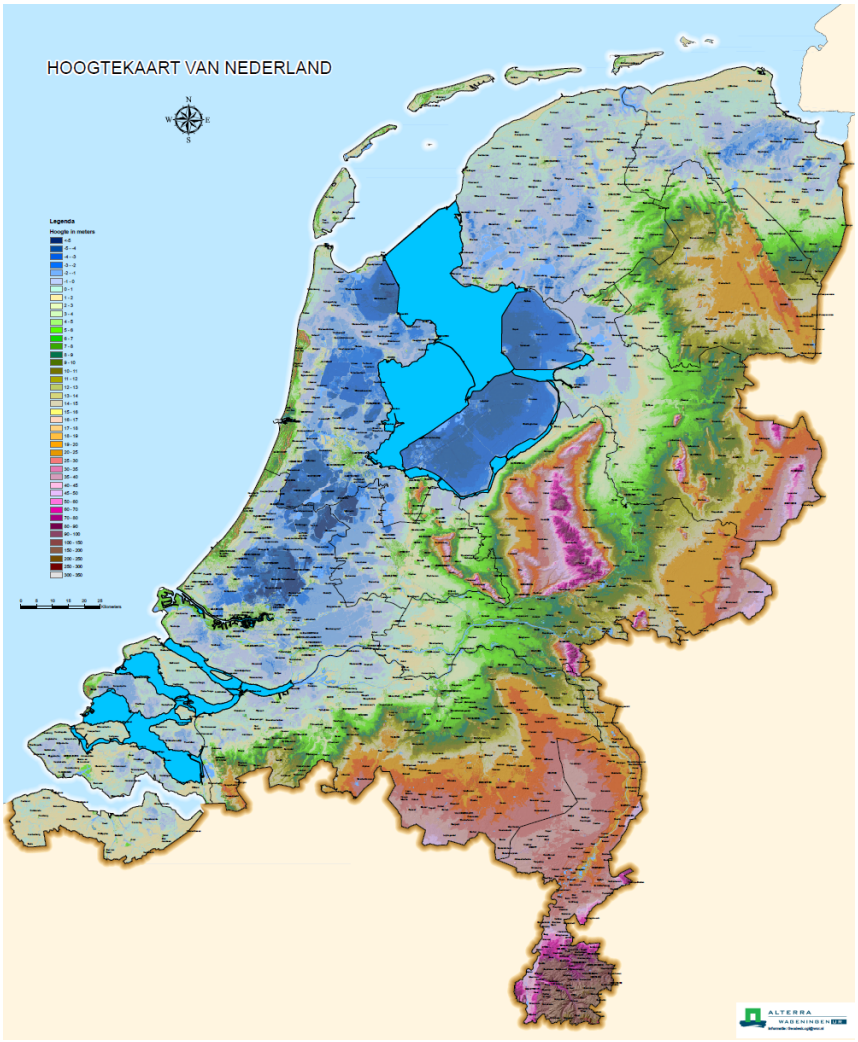


Figure 77: Elevation in the Netherlands. The pink/purple zone in the middle of the Netherlands is the Veluwe with the Utrechtse Heuvelrug just to the west of it. The blue lake north-west of the Veluwe is the IJselmeer.

Institutionen för naturgeografi och ekosystemvetenskap, Lunds Universitet.

Student examensarbete (Seminarieuppsatser). Uppsatserna finns tillgängliga på institutionens geobibliotek, Sölvegatan 12, 223 62 LUND. Serien startade 1985. Hela listan och själva uppsatserna är även tillgängliga på LUP student papers (www.nateko.lu.se/masterthesis) och via Geobiblioteket (www.geobib.lu.se)

The student thesis reports are available at the Geo-Library, Department of Physical Geography and Ecosystem Science, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden. Report series started 1985. The complete list and electronic versions are also electronic available at the LUP student papers (www.nateko.lu.se/masterthesis) and through the Geo-library (www.geobib.lu.se)

285 Cansu Karsili (2013) Calculation of past and present water availability in the Mediterranean region and future estimates according to the Thornthwaite water-balance model
286 Elise Palm (2013) Finding a method for simplified biomass measurements on Sahelian grasslands
287 Manon Marcon (2013) Analysis of biodiversity spatial patterns across multiple taxa, in Sweden
288 Emma Li Johansson (2013) A multi-scale analysis of biofuel-related land acquisitions in Tanzania - with focus on Sweden as an investor
289 Dipa Paul Chowdhury (2013) Centennial and Millennial climate-carbon cycle feedback analysis for future anthropogenic climate change
290 Zhiyong Qi (2013) Geovisualization using HTML5 - A case study to improve animations of historical geographic data
291 Boyi Jiang (2013) GIS-based time series study of soil erosion risk using the Revised Universal Soil Loss Equation (RUSLE) model in a micro-catchment on Mount Elgon, Uganda
292 Sabina Berntsson & Josefin Winberg (2013) The influence of water availability on land cover and tree functionality in a small-holder farming system. A minor field study in Trans Nzoia County, NW Kenya
293 Camilla Blixt (2013) Vattenkvalitet - En fältstudie av skånska Säbybäcken
294 Mattias Spångmyr (2014) Development of an Open-Source Mobile Application for Emergency Data Collection
295 Hammad Javid (2013) Snowmelt and Runoff Assessment of Talas River Basin Using Remote Sensing Approach
296 Kirstine Skov (2014) Spatiotemporal variability in methane emission from an Arctic fen over a growing season – dynamics and driving factors
297 Sandra Persson (2014) Estimating leaf area index from satellite data in deciduous forests of southern Sweden
298 Ludvig Forslund (2014) Using digital repeat photography for monitoring the regrowth of a clear-cut area
299 Julia Jacobsson (2014) The Suitability of Using Landsat TM-5 Images for Estimating Chromophoric Dissolved Organic Matter in Subarctic Lakes
300 Johan Westin (2014) Remote sensing of deforestation along the trans-Amazonian highway
301 Sean Demet (2014) Modeling the evolution of wildfire: an analysis of short term wildfire events and their relationship to meteorological variables
302 Madelene Holmblad (2014). How does urban discharge affect a lake in a

recreational area in central Sweden? – A comparison of metals in the sediments of three similar lakes

303 Sohidul Islam (2014) The effect of the freshwater-sea transition on short-term dissolved organic carbon bio-reactivity: the case of Baltic Sea river mouths

304 Mozafar Veysipanah (2014) Polynomial trends of vegetation phenology in Sahelian to equatorial Africa using remotely sensed time series from 1983 to 2005

305 Natalia Kelbus (2014) Is there new particle formation in the marine boundary layer of the North Sea?

306 Zhanzhang Cai (2014) Modelling methane emissions from Arctic tundra wetlands: effects of fractional wetland maps

307 Erica Perming (2014) Paddy and banana cultivation in Sri Lanka - A study analysing the farmers' constraints in agriculture with focus on Sooriyawewa D.S. division

308 Nazar Jameel Khalid (2014) Urban Heat Island in Erbil City.

309 Jessica, Ahlgren & Sophie Rudbäck (2014) The development of GIS-usage in developed and undeveloped countries during 2005-2014: Tendencies, problems and limitations

310 Jenny Ahlstrand (2014) En jämförelse av två riskkarteringar av fosforförlust från jordbruksmark – Utförda med Ekologgruppens enkla verktyg och erosionsmodellen USPED

311 William Walker (2014) Planning Green Infrastructure Using Habitat Modelling. A Case Study of the Common Toad in Lomma Municipality

312 Christiana Marie Walcher (2014) Effects of methane and coastal erosion on subsea-permafrost and emissions

313 Anette Fast (2014) Konsekvenser av stigande havsnivå för ett kustsamhälle- en fallstudie av VA systemet i Beddingestrand

314 Maja Jensen (2014) Stubbrytningens klimatpåverkan. En studie av stubbrytningens kortsiktiga effekter på koldioxidbalansen i boreal barrskog

315 Emelie Norhagen (2014) Växters fenologiska svar på ett förändrat klimat - modellering av knoppsprickning för hägg, björk och asp i Skåne

316 Liisi Nõgu (2014) The effects of site preparation on carbon fluxes at two clear-cuts in southern Sweden

317 Julian Will (2014) Development of an automated matching algorithm to assess the quality of the OpenStreetMap road network - A case study in Göteborg, Sweden

318 Niklas Olén (2011) Water drainage from a Swedish waste treatment facility and the expected effect of climate change

319 Wösel Thoresen (2014) Burn the forest - Let it live. Identifying potential areas for controlled forest fires on Gotland using Geographic Information System

320 Jurgen van Tiggelen (2014) Assimilation of satellite data and in-situ data for the improvement of global radiation maps in the Netherlands.

321 Sam Khallaghi (2014) Posidonia Oceanica habitat mapping in shallow coastal waters along Losinj Island, Croatia using Geoeye-1 multispectral imagery.