front cover

Development of an optimal data screening method for a climate field reconstruction based on data assimilation

Master thesis
Faculty of Science, University of Bern

handed in by

Sabrina Thaler Minke

2021

Supervisor

Dr. Jörg Franke

Co-Supervisor

Prof. Dr. Stefan Brönnimann

Contents

Abstract

In the attempt of producing reanalyzes that extend back in time to understand the climate of the past and its variations four data screening approaches are proposed and analysed. In specific the treatment of observations of temperature, precipitation and sea-level-pressure that lay in the same grid box is researched. Each variable (temperature, precipitation and sea-level-pressure) is researched separately. The assimilation of all the observations, of their average, of the choice of a random observation per grid box and by using a Buddy Check algorithm that helps eliminating suspect observations are looked into. The skill is assessed with the Mean Squared Error Skill Score ($MSESS$) and the Pearson correlation skill score. The results show that the only variable that has a difference in the $MSESS$ between the proposed screening method and the assimilation of all the observations is the precipitation. The Pearson correlation coefficient does not show any significant difference. By further analysing the data of the precipitation we found out that the data is not normally distributed, which does not produce ideal results for Kalman filters. The conclusion is that the use of these screening methods does not make the reconstruction overwhelmingly better. The results suggest that the Buddy Check algorithm is not helpful for a more skilful reconstruction but further research in this topic is needed.

1 Introduction

Climate change is increasingly gaining awareness and confronting society with new challenges. As a consequence, a better understanding of the climate of the past, as well as the climate variability and extreme events is of vital importance to predict and prepare for extreme weather events in the future (Brönnimann, Rohr, et al. 2020).

A good reconstruction of the climate of the past is vital to understand how climate variability evolved in time. The longer we go back in time, the fewer instrumental records (direct sources) are available to produce climate reconstructions. Reconstructions of past climate conditions, when instrumental data are available just at a small set of locations, may be derived from indirect sources like tree rings, ice cores, etc.. However, These indirect sources (called paleoclimatology proxies) have information on the climate of the past just at annual resolution or lower. To enhance the resolution of these records different types of proxy records are combined (Valler, Franke, Brugnara, et al. 2020).

The method described above is an empirical method to reconstruct climate, which relates changes in climate variables to changes in paleoclimatology proxies. Another method used is the dynamical method, which constrains simulations of the past climate with reconstructed external forcings. Both these methods have their strengths and weaknesses. Bhend et al. 2012 proposes to directly assimilate proxy data into climate model simulations to minimize the weaknesses of the approaches.

A major role in the differences between climate reconstruction is the difference in input data (Franke, Valler, et al. 2020). For this reason quality control of the data is an important tool to eliminate rough errors in the analysis of meteorological data. Random errors are errors that are normally small but influence all meteorological data. Rough errors, in contrast to random errors, often are big errors that occur in a small set of the data. This is why it is important to address and try to eliminate rough errors from any analysis that uses meteorological data. Since the advent of numerical weather prediction this task has become particularly important (Collins and Gandin 1990).

Data assimilation wants to provide the best estimate of the true state of the system, in this case the climate, and it is called the analysis. As stated above, many problems arise when we want to produce an analysis of the climate of the past. One of these problems is the density of the observations. Liu and Rabier 2002 argues that if the observation density increases, the quality of the analysis decreases. He explains that the reason behind this is that when the data set is too large the error correlations are neglected and thus the quality decreases.Therefore, the ”thinning” of the data is a method of selection of observations in the locations where the observations are very dense, so as not to loose any valuable information.

1.1 Meteorological reanalysis

Meteorological reanalyses are products of data assimilation, *analyses*, these are produced using just a single version of a data assimilation system (Dee, Uppala, et al. 2011). Dee, Uppala, et al. 2011 claims that there is an emerging need for accurate representations of the climate of the past. Especially the representation of variability on interannual and decadal time scales.

As of today, progress in the reanalyses has been made and Compo et al. 2011 was able to produce a reanalysis spanning the twentieth century (20CR). Further extension in the past, using direct observations, is limited because the observations are missing and paleoclimate proxies may have to be used.

Compo et al. 2011 uses a conventional approach to create a reanalysis with 6-hourly resolution for the 20th century (20CR). Franke, Brönnimann, et al. 2017 uses another approach to extend the reanalysis of Compo et al. 2011 further back in time. Franke produces a data set that has seasonal to monthly resolution. The model he uses assimilates all variables, including the monthly mean observations as well as annually resolved paleodata. This analysis offers a better understanding of the past climate because it's possible to extrapolate model variables which are not assimilated, e.g. assimilation of pressure fields which can lead to a better understanding of the corresponding temperature anomalies.

Variational data assimilation methods have also been used to reconstruct the climate of the past. Kobayashi et al. 2015 has tried to do a reanalysis (JRA-55) which covers the last 55 years using a variational technique (4D-Var). One of the elements which improved the reconstruction of the JRA-55 is using a quality control which excludes observations which are inconsistent. The result is an improved reanalysis but sill with some limitations. The ERA-5 reanalysis also uses 4D-Var which, as in the case of Kobayashi, resulted in improved versions of their reanalysis but again with some hindrances (Hersbach et al. 2020). So as to understand how the reanalysis that are produced today came to be, it is interesting to know how they evolved in time.

1.2 Extension of reanalysis in the past

In 1819, 365 daily maps of pressure contours of the year 1783 were constructed by Brandes, which marked the first generated retrospective analysis (Monmonier 1999 cited in Compo et al. 2011). After Brandes' reanalysis many others followed. However, just in the early 1990s did the first cross-disciplinary studies come out, which attempted at bringing together reconstructions from diverse proxies at hemispheric-scale. Unfortunately, they did not receive much public attention at the time (Jones et al. 2009).

Nowadays, the state of the atmosphere at any particular time is estimated through the collection and analysis of data from many measurement platforms. However, this type of analysis has become possible just since the advent of the objective methods which measure the state of the atmosphere. In fact, until the mid-twentieth century, analyses of the state of the atmosphere were subjective and relied on the experience of the meteorologist who was making the analysis (Compo et al. 2011).

Before that, modern 3D data sets extended from 1948 to the present day. Many of them utilized variational assimilation techniques, of which the most widely used is 3D-Var. Yet, 3D-Var has certain shortcomings: a. reanalyses over longer periods, b. show spurious long-term trends and c. understating extratropical storm track variability. To work around these hindrances, more advanced data assimilation methods were studied, e.g. 4D-Var assimilation or the Ensemble Kalman Filter. 20CR generated reliable reanalyses of the earlier periods using only surface observations applied to ensemble Kalman filtering methods, which is very similar to the method of Franke, Brönnimann, et al. 2017 which is the EKF400 (Compo et al. 2011).

1.3 EKF400

Franke, Brönnimann, et al. 2017 exploited the advantages of reanalysis to create the first monthly resolved paleo-reanalysis for the period 1600-2005 using statistical climate field reconstruction. Franke used the Ensemble Kalman fitting (EKF) assimilation method covering the past 400 years at a monthly resolution with 2° resolution globally, which is why he called it EKF400 (Valler, Franke, Brugnara, et al. 2020).

Kalman filter is a method which falls into covariance-based approaches and are used for the assimilation of paleoclimate proxies (Brönnimann, Franke, et al. 2013). Bhend et al. 2012 used Kalman filtering techniques applied in pseudo-proxy experiments with promising results. Compo et al. 2011 and Slivinski et al. 2021 also used the Kalman filter to produce reconstructions with good results. An extensive explanation on the Ensemble Kalman Filter is given by Evensen 2003 in which he also explains how to make the computation of a global analysis cost effective.

Valler, Brugnara, et al. 2020 in her paper uses the EKF400 to assimilate the precipitation. She found that when assimilating the precipitation with the EKF in arid regions, the results were not optimal. The distribution of the precipitation observations in these areas is, as expected, significantly different from a normal distribution. In the other hand, when assimilating the precipitation with the EKF in the mid-latitudes, where the hypothesis of normality is less often rejected, the skill is much better. Consequently, it is probable that the assimilation of observations with a Gaussian distribution may yield optimal results.

The EKF400 uses atmospheric-only general circulation model simulations. It blends 30 ensemble members generated with ECHAM5.4 general circulation model with different types of observations (early instrumental measurements, documentary data, proxy records) (Bhend et al. 2012).

The improvements of this reanalysis compared to others is assessed using the Pearson correlation coefficient (corr) and the mean squared error skill score (MSESS). Temperature, precipitation and sea-level-pressure are the main variables that are assimilated. The results of the assimilation with the newest version (v. 2.0) of the EKF show very good skill in the regions where most of the observations are located, i.e. Northern Hemisphere (Valler, Franke, Brugnara, et al. 2020). However, the analysis suffers from a reduced availability of instrumental observations going further into the past. For precipitation, the assimilation causes an increase in variability and shifts the absolute values further away from the instrumental data (Valler, Franke, Brugnara, et al. 2020).

Precipitation showed negative skill in some regions of Europe, Asia and North America, however EKF400 mostly has the correct sign of the precipitation anomaly but not the correct amount. In version 2 the assimilation of precipitation was improved, especially in the extratropical Northern Hemisphere in the October- March period (Valler, Franke, Brugnara, et al. 2020). The uncertainty of the reconstruction remains rather large in the low latitudes.

Observations in the EKF400 are quality checked before they are assimilated. If Proxy records and instrumental measurements are more than five standard deviations away from their mean in the current time step they are excluded. Further screening in the EKF400 consists in assessing whether proxy record has a significant climatic signal by doing an F- Test with null-hypothesis that (all of the regression coefficients are equal to zero (rejecting the proxies with with p¡0.05) (Valler, Franke, Brugnara, et al. 2020). Additionally tree-rings proxies are screened to remove any duplicate record.

With this research, I will try to produce a more skilful reconstruction with, between others, an algorithm that screens the data and eliminates observations which do not improve the reconstruction quality. The objective of this screening is to thin the data in the regions where there are many observations.

1.4 Research question

The research question of this thesis will be at improving EKF400 to increase our understanding of the past decadal variations. In my research “development of an optimal data screening method for a climate field reconstruction based on data assimilation” I will employ classical analyses as well as make other tests, e.g. considering just the observations that are in a prescribed range and accepted as ”buddies” with the help of the Buddy Check algorithm to make an informed proposal for an improved EKF400 as suggested by Dee, Rukhovets, et al. 2001.

The reminder of the thesis is organized in the following way. In chapter 2, the methods are presented as well as the skill score used to estimate the quality of the methods. In chapter 3, all the results are presented. Finally, in the last chapter, a discussion of the results is presented and conclusions is drawn. An outlook is also given at the end of the thesis.

2 Methods

The first step in finding an optimal data screening method for a climate field reconstruction based on data assimilation was to familiarize myself with the assimilation system. Therefore, three tests for the screening of the data with different conditions were be performed. The first experiment was a paleo-reanalysis, or reconstruction, using all the observations. In the second experiment I averaged all the observations within one grid point. The third experiment was to pick the first data point per grid point in an arbitrary fashion.

These experiments were conducted over a data assimilation system that is already existing, the EKF400. The EKF400 will be described in the next chapter (Chapter 2.1). I tested the performance of the tests mentioned before and the adaptive Buddy Check algorithm with the EKF400 taking inspiration from Compo et al. 2011.

2.1 EKF400

This thesis investigates, among other screening methods, an adaptive quality control for an optimal data assimilation. The adaptive quality control, called ”Buddy Check”, was implemented to an already existing data assimilation system, the EKF400, which was built to study past decadal climate variations. The proposed data screening have the goal to improve the quality of the EKF400.

Franke, Brönnimann, et al. 2017 explains that the EKF400 is a 30-member Ensemble that uses the Kalman Fitting method. The EKF400 is able to assimilate different types of observations (instrumental temperatures, pressure observations, etc.) using a kalman filtering technique. This is the reason for the acronym EKF in the EKF400. The analysis covers the last 400 years (1600 to 2005) in a monthly time resolution, and is thus called EKF400. In this thesis the only type of observation which will be used are instrumental observations of temperature, precipitation and sea-level-pressure. As Franke, Brönnimann, et al. 2017 describes, ”the model is compared to CCC400 global mean land temperature with instrumental global mean temperatures from the CRUTEM4 (Osborn and Jones 2014) and found good agreement.” Additionally, Documentary and proxy data often have only interannual or decadal variability. To overcome this problem anomalies for the 70-year period around the current year were assimilated so as to have a multi-decadal to centennial-scale variability which are a function of the model simulations to external forcings (Valler, Franke, Brugnara, et al. 2020).

The EKF400 is a reanalysis, which means it combines information from numerical models and observations. Both the numerical model calculation as well as the observations have errors. To minimize the errors anomalies for the 70-year period were assimilated as explained above. In addition, the model is constrained by boundary conditions. In the end the assimilation calculated the most likely state of the atmosphere given all errors (Franke, Brönnimann, et al. 2017).

The data set in the EKF400 is composed of monthly data. The assimilation step, however, is bi-annual. Meaning that the assimilation is divided in the growing seasons that influence the tree-ring measurements. These are boreal summer (April-to-September) and boreal winter (October-to-March).

To test the skill of the Buddy Check, just the temperature, precipitation and sea-level-pressure were assimilated, one at a time in separated experiments. The motivation for this choice is to research the performance of the single variables so as to make an educated choice in how to combine them in an assimilation in the future. Another reason behind this choice is that the less variables we assimilate, the less computational time it will take to perform the calculations. If the assimilation of these three variables with the Buddy Check would show positive results, the next step would be to assimilate more variables together and make other tests.

An example of the treatment of the errors are the exclusion of outliers. Anomalies are excluded if, at the current time step, they differ more than four standard deviations from the mean of a compared dataset (CCC400) (Franke, Brönnimann, et al. 2017).

2.2 Data screening approach

The research goal of this master thesis, as stated in the introduction, is to propose a screening/selection approach for a data assimilation method that will allow to remove observations, which do not improve the reconstruction quality. In specific, the problem of how to treat grid cells that contain multiple observations will be addressed. It also has to be pointed out that all the assimilations are made for the period 1951-1981.

A similar approach was used by Compo et al. 2011 and Tingley et al. 2012. The Pearson correlation coefficient and the mean squared error skill score as a method to assess the improvement of the quality of the reconstructions were also used for the EKF400 in Valler, Franke, and Brönnimann 2019.

2.2.1 All

The first assimilation to be computed was the assimilation of all observations (ALL). This assimilation will be computed making no changes nor adding anything to the EKF400, all the observations were assimilated with no averaging and no data screening with this setup.

Tingley et al. 2012 particularly underlines the importance of comparing results to the original model output. With this assimilation the aim was to have a data set with which we can compare and analyse the skill of the experiments of data screening with other settings.

2.2.2 Average

The first approach in solving the problem of the treatment of multiple observations in the same grid cell was to average them. It is a simple procedure and the objective of averaging (AVG) was to reduce the background noise and thus improve the signal-to-noise ratio. This approach is already used in the EKF400 as described in Franke, Brönnimann, et al. 2017.

2.2.3 First

In order to take independent samples from the analysis one observation was selected randomly for every grid cell (FIRST). This procedure also minimizes computer storage and allows for faster assimilation times because less observations are assimilated (Wilks 2011).

2.2.4 Buddy Check

Dee, Rukhovets, et al. 2001 notes that there is a problem in the exclusion of an observation using the quality check of the standard deviation explained in the Introduction (Chapter 1), whenever the local variability is larger than usual, it tends to reject too many observations. one possible solution for this problem would be use an adaptive quality check. Which is an approach that will be investigated with this thesis.

Local variability can be larger than usual in the advent of extreme events. In the EKF400 though, local variability won't be larger than usual because of extreme events because the observation is assimilated as a monthly mean. Which is to say that all the local variability caused by an extreme events won't weigh too much on the monthly mean.

Just like in the analysis of Compo et al. 2011 the EKF400 already has uncertainty estimators and error elimination, however, to try and produce an even better reconstruction by eliminating suspect observations, an additional adaptive quality control method is tested.

The adaptive Buddy Check tries to help with the elimination of suspect observations, as described in the research paper by Dee, Rukhovets, et al. 2001. The idea behind the Buddy Check is to rule out contaminated data from genuine observation about the atmosphere. It's called "adaptive Buddy Check" because the algorithm adapts the prescribed error estimates during each iteration, depending on the available density and values of data.

As a general consideration the Buddy Check can be thought of as a statistical test, with null hypothesis that "all residuals are jointly normally distributed with zero mean and known covariance (Dee, Rukhovets, et al. 2001)". The rejection of the null hypothesis suggests that there are contaminated data in the dataset. It is important to emphasize, however, that the criteria for the exclusion of an observation depend on the type of the observation (Dee, Rukhovets, et al. 2001).

The focus of writing a formula for an adaptive Buddy Check algorithm is to decide when an observation is corrupt. An observation will be flagged corrupt if it's not consistent with the null hypothesis. A probability p that the observation is consistent with the null hypothesis must be computed. If p is smaller than the significance level (δ) the null hypothesis must be rejected because at least part of it is probably corrupt. After the conclusion that at least part of the data is corrupt we nevertheless need a technique to point out which observations were corrupt and which not. A simple approach is to take each suspect observation and build the marginal distribution in respect to the analysis. "All marginal distributions of a normal distribution are themselves normal (Jazwinski 1970, p. 44 theorem 2.12 cited in Dee, Rukhovets, et al. 2001)".

To decide if an observation is discarded or not we considered each *i*th observation (y_i^O), one after the other, as it gets assimilated. The observation that is currently checked, will be examined in relation to the their buddies. The buddies are all the observations that lie in a predefined radius D from the currently assimilated observation. Almost all these experiments were performed with the same distance for the observations to be considered as "Buddies". Compo et al. 2011 used a radius of 1000 km but in this thesis a smaller radius was chosen. The reason behind this choice is that the weather in a 1000 km scale can be varied and cause the buddies to contradict each other. The effect would be the exclusion of valid observations. For this reason the radius chosen for most of the experiments is 200 km. One test was also made for the assimilation of the temperature with a radius of 500 km, to see if the results are better since the variables may have different spatial correlation.

The numerator of B_i in equation 1 represents the mean-square observation departure from $\overline{\mathbf{x_i}}^a$ and the denominator represents the mean square departure from the first guess $\overline{\mathbf{x}}^b$.

$$B_i = \frac{\frac{1}{K}\sum_{\substack{K=1\\K\neq 1}}^{K}(H\overline{\mathbf{x_i}}^a - y_k^O)^2}{\frac{1}{K}\sum_{\substack{K=1\\K\neq 1}}^{K}(H\overline{\mathbf{x}}^b - y_k^O)^2} \tag{1}$$

where H is the operator that translates the raw data (observations) into the model variables and K is the number of buddies. B_i represents how the currently assimilated observation improves the fit of the analysis, according to the neighboring, already assimilated, observations.

Equation 1 can result as greater than, smaller than or equal 1. If B_i is smaller than 1, the currently assimilated observation improves the fit of the analysis. This implies that we want to keep the observation as it may be genuine.

On the other hand, B_i is greater than 1, in the other hand, the observation degrades the analysis and it is not in agreement with the neighboring stations. It is thus marked as obsolete and it is discarded.

Additionally, if B_i equals 1, signifies that the observation does not worsen the fit of the analysis but also that it doesn't improve it. The choice was to assimilate the observations that resulted in a B_i equal to 1.

A flowchart summarising the basic steps of the Buddy Check algorithm is given in Figure 1.

2.2.5 Monthly observations

EKF400 has a monthly resolution with a half-yearly assimilation step (Franke, Brönnimann, et al. 2017). This implies that we have monthly data for all the variables assimilated (precipitation, temperature, sea-level-pressure) with a monthly ensemble mean.

It is important to mention that the data for the validation is not completely independent. The reason behind this is that the goal is to assimilate all the available data (Franke, Brönnimann, et al. 2017). The too high correlations and MSESS that may come to place because of this do not matter in the framework of this thesis, because the experiments are evaluated as relative differences between other experiments.

2.2.6 Shortcomings of the Buddy Check

Dee, Rukhovets, et al. 2001 also identifies some weaknesses of the Buddy Check algorithm. The algorithm rejects all the available observations collectively if they all fail the background check. The proposed solution to this problem is applying a more tolerant background check to isolated observations. Another challenge, which has to be taken into consideration, is that the algorithm may accept a corrupt observations which agree with the background. It has to be noted that this adaptive quality control represents a small portion of the quality controls performed on the dataset, as Dee, Rukhovets, et al. 2001 notes himself on the paper. These shortcomings will be taken into consideration in the evaluation of the results.

2.3 Metrics of skill

Finally, it is important to have an estimate of the quality of the output of the different screening methods. All the experiments were separated in order to see which settings performs better.

To test which experiment has a higher skill we subtracted the skill of the single experiments to the skill of the assimilation of all the observations making no screening of the data with the average, first or Buddy Check (Chapter 2.2.2, Chapter 2.2.3, Chapter 2.2.4 to the skill of the assimilation of all dataset (Chapter 2.2.1).

2.3.1 Correlation skill score

As proposed by Franke, Brönnimann, et al. 2017 and Bhend et al. 2012 the first skill used is the Pearson correlation coefficient, also called correlation, to evaluate the covariability of data sets.

The correlation skill score, in contrast to the $MSESS$ explained in Chapter 2.3.2, is a measure of linear association between the observations and the forecast. It has to be highlighted that the correlation is a measure of potential skill only. The reason for this is that correlation does not take into account the biases in the forecast (conditional and unconditional) (Murphy 1988).

Correlation was calculated between the absolute values of the ensemble mean of the analysis and the reference series at each grid point (Franke, Brönnimann, et al. 2017).

2.3.2 Mean squared error skill score $MSESS$

In addition to the correlation a good measure for the accuracy of ensemble forecasting is the mean squared error skill score ($MSESS$), as suggested by Bhend et al. 2012 and Wilks 2011. Jolliffe and Stephenson 2011 starts to explain the MSESS with the mean squared error (Equation 2).

$$MSE = \frac{1}{n}\sum_{K=1}^{n}(y_k - o_k)^2 \tag{2}$$

Where y denotes the forecast at an individual point and time and o describes the corresponding value of the observation. The footer k describes the number of the observation at a particular time and n the number of observations in the grid (Jolliffe and Stephenson 2011).

This is the evaluation of the forecast at a single time. Jolliffe and Stephenson 2011 states that the MSE is tolerant towards small errors and penalizes larger errors. Indeed Franke, Valler, et al. 2020 argues that this skill score punishes a wrong amplitude in variability, making it harder in respect to the correlation coefficient to have positive results, When the MSE is equal to 0 than the forecast is perfect, the more the number increases the more discrepancies there are between the forecast and the observations (Wilks 2011).

In the EKF400 the MSE is used as a skill score. Murphy 1988 explains that a generic skill score is calculated using a forecast of interest, a reference forecast and the observations. In this case if we base the skill score in the MSE, it will look like equation 3.

$$MSESS = 1 - \frac{MSE}{MSE_{ref}} \tag{3}$$

With MSE as the MSE of the EKF400 and MSE_{ref} as the MSE of the reference. In the case of the EKF400 the reference is the CRU TS 3.10 instrumental data presented in Harris et al. 2014.

The $MSESS$ will have the value 1 when the forecast is perfect. If, in our case, the EKF400 euqals the reference forecast than the result of the $MSESS$ will be 0. Negative skill when the accuracy of the EKF400 is greater than the accuracy of the reference, the more positive the greater the accuracy. A $MSESS$ of 0 signifies that the accuracy of the EKF400 is equal to the reference and a negative value means that the EKF400 has a smaller accuracy compared to the reference (Wilks 2011).

The mean squared error is referred to also as mean squared error in the presentation of the results.

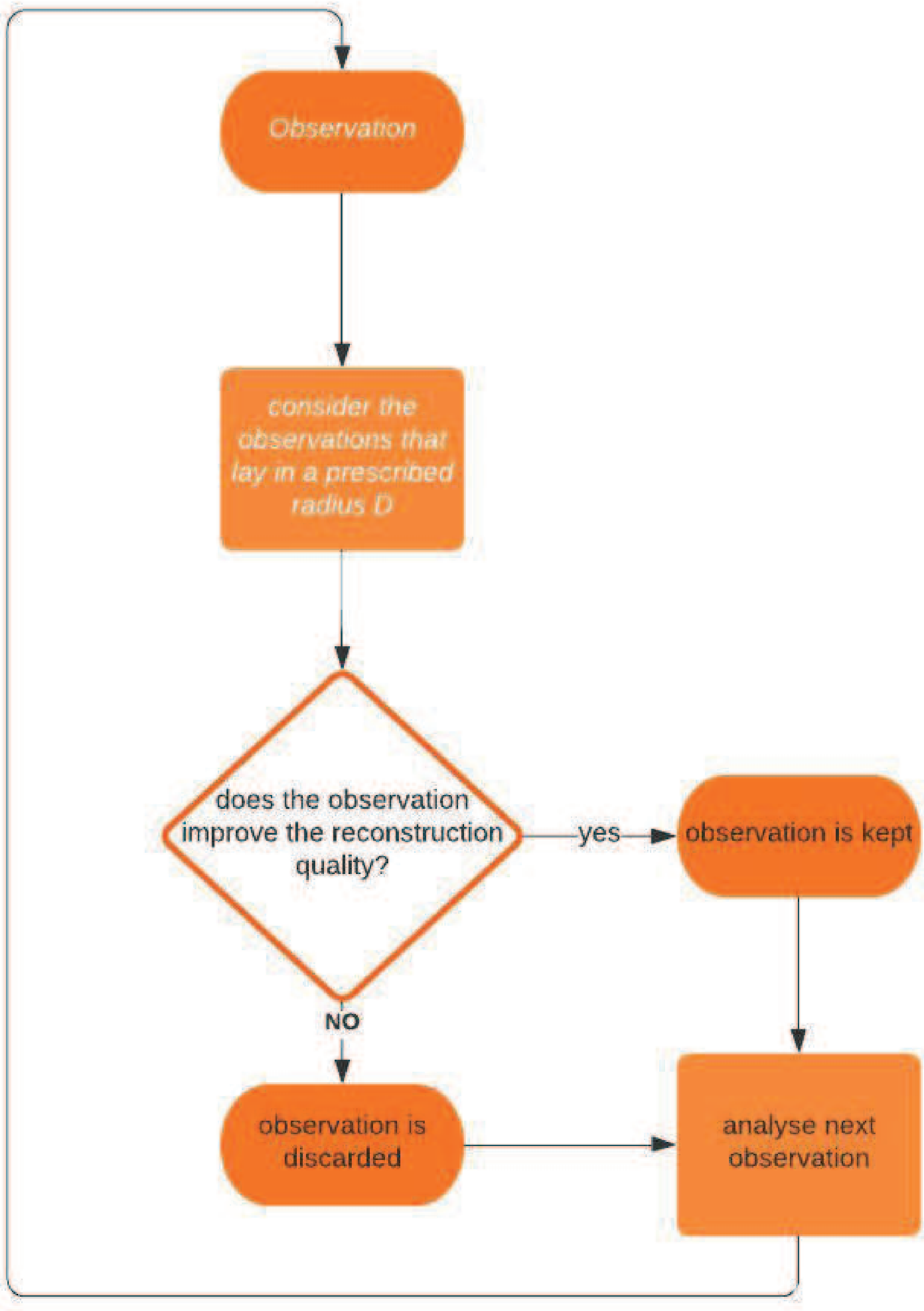

Figure 1: Flowchart of the main steps of the Buddy Check algorithm.

3 Results

Since I am searching for the best assimilation method, I tried different possible screening methods and combinations. In this chapter, I will be presenting the results of the analysis using the methods described in Chapter 2.

I started the analysis by assimilating the available data using three different basic screening methods. As explained in the Introduction section (Chapter 1), these screening methods are used to screen all the available data sets (ALL), the average of the data set (AVG) in the same grid box, or just randomly selected observations of the data set (FIRST) in one grid box. After these initial data assimilations, the algorithm of the Buddy Check (Chapter 2.2.4) is tested (BC). All the results, as explained in the Methods section, are exposed as seasonal results: for the winter season (October-to-March) and for the summer season (April-to-September) apart from the medians presented in Table 2 and in Table 3, the time series of the climate explorer (Figure 30 and Figure 31), the tables of the summary of excluded stations (Table 4, Table 5 and Table 6) and lastly the tables of excluded and included observations sorted by decades (Figure 44, Figure 45 and Figure 15). An overview of all performed experiments conducted in this thesis is given in Table 1. Almost all these experiments are performed with the same distance for the observations to be considered as "Buddies" (D = 200 km) as explained in the Introduction, one experiment for the assimilation of the temperature was conducted with a distance of D = 500 km.

Table 1: Summary of the experiments performed in this thesis all conducted for the period 1951-1981. The name of the experiment describes the settings of the experiment. Variable assimilated refers to which variable was assimilated: TEMP for Temperature, PRECIP for precipitation and SLP for Sea-level-pressure. The Description refers to how multiple data in the same grid cell was treated: ALL for the original setup, AVG averages multiple observations in the same grid cell, FIRST picks randomly one observation from the grid cell and BC uses the algorithm of the Buddy Check

Name	Variable assimilated	Description
NOBC_JUSTtemp_all_Sabrina_51-81	TEMP	ALL
NOBC_JUSTtemp_avg_Sabrina_51 − 81	TEMP	AVG
NOBC_JUSTtemp_first_obs Sabrina_51 − 81	TEMP	FIRST
BC_D200_JUSTtemp_all_Sabrina_51 − 81	TEMP	BC
BC_D500_JUSTtemp_all_Sabrina_51 − 81	TEMP	BC
NOBC_JUSTprecip_all_Sabrina_51 − 81	PRECIP	ALL
NOBC_JUSTprecip_avg_Sabrina_51 − 81	PRECIP	AVG
NOBC_JUSTprecip_first_Sabrina_51 − 81	PRECIP	FIRST
BC_D200_JUSTprecip_all_Sabrina 51 − 81	PRECIP	BC
NOBC_JUSTslp_all_Sabrina_51 − 81	SLP	ALL
NOBC_JUSTslp_avg_Sabrina_51 − 81	SLP	AVG
NOBC_JUSTslp_first_Sabrina_51 − 81	SLP	FIRST
BC_D200_JUSTslp_all_Sabrina 51 − 81	SLP	BC

In the next sections I will be using the term observation, this term encompasses all the available data, spanning from documentary data to instrumental measurements. All the analyses were computed for the years 1951-1981.

3.1 All

The experiments conducted in this thesis will be executed with different setups for data screening to search for the optimal solution. In order to assess the experiments the original setup is assimilated and used as basis for the comparisons.

The first assimilation to be done is therefore the assimilation of all the observations that are available in EKF400, making no averages and with no thinning of the data nor using the Buddy Check algorithm to screen the data. The aim of this assimilation is to have a data set that makes it possible to compare all the other assimilations with the original settings.

3.2 distribution of the data

Box plots are simple tools that provide an overview on the distribution of the data (Wilks 2011). These plots will be used to compare the data sets of the entire assimilation.

To compare and evaluate the assimilation methods two verification measures were used as explained in the Methods section: mean squared error skill score and the correlation coefficient. The distributions of the assimilations are summarized with the help of box plots with a separation of the two growing seasons (October-March and April-September).

A summary of the medians of the correlation coefficient is given in Table 2 where it can be seen that the differences between the correlation coefficients is minimal. As explained before these medians are not the medians of seasonal data, the are the medians for the whole year.

Table 2: Summary of the medians of the Pearson correlation coefficient for all the experiments for the period 1951-1981 for summer and winter combined

Median of the correlation coefficients			
Description	Temperature	Precipitation	Sea-level-pressure
ALL	0.24	0.19	0.26
AVG	0.24	0.17	0.26
FIRST	0.24	0.18	0.26
BC	0.24	0.19	0.26

So as to give an overview of the distribution of the experiments, boxplots of all the experiments are built (Figure 2, Figure 3, Figure 4 and Figure 5). The boxplots are divided in boxplots of the Pearson correlation skill score and the MSESS. The data used is the data for the entire period that was studied (1951-1981). The observations of both hemispheres are taken in consideration when plotting the boxplots, the reason is that even if there are much more observations in the Southern Hemisphere, thanks to the figures of the number of excluded and included observations (Figure 13, Figure 14 and Figure 15), we can see that there are many observations of precipitation in Australia and South Africa. The comparison was done for the entire data set for all the variables.

Thanks to the visualisation box plots of the correlation coefficient we see indeed that there is no differences between the experiments, nor there is a difference between the two growing seasons (Figure 2). The correlation coefficient is always positive and has an almost identical range for every assimilation. A zoomed version of Figure 2 can be seen in Figure 3, where more details are visible. Thanks to this visualisation we can see that the winter season has slightly more skill than the summer season.

The box plots of the mean squared error skill score (reduction of error) show a skill that is very close to 0 (Figure 4). Thanks to a more zoomed version of the same box plots the quantiles can be seen with more precision (Figure 5). The outliers were left out for a better visualisation of the quantiles. We can see a difference between the quantiles of the assimilation. The assimilation of the summer season produces a smaller range than the assimilation of the winter season. The ranges between the variables are also different. It can be noted that the experiments with the largest range are when sea-level-pressure is assimilated (up to 0.175). Temperature and precipitation have smaller ranges than the sea-level-pressure in the MSESS. As in the box plots of the correlation coefficient we can see that the skill score is always positive but almost zero.

A table that summarizes the medians of the mean squared error skill score is given in Table 3. It is important to remind that these are medians for the whole year, they are not seasonal medians. Thanks to this table and the boxplots, it is clear that the difference between the experiments in the MSESS is practically 0. The reason behind this is that the MSESS punishes the errors in amplitude, as will be seen more in depth in the next chapters.

Table 3: Summary of the medians of the mean squared error skill score for all the experiments for the period 1951-1981 for summer and winter combined

Median of the mean squared error skill scores			
Description	Temperature	Precipitation	Sea-level-pressure
ALL	0.00	0.00	0.01
AVG	0.00	0.00	0.01
FIRST	0.00	0.00	0.01
BC	0.00	0.00	0.01

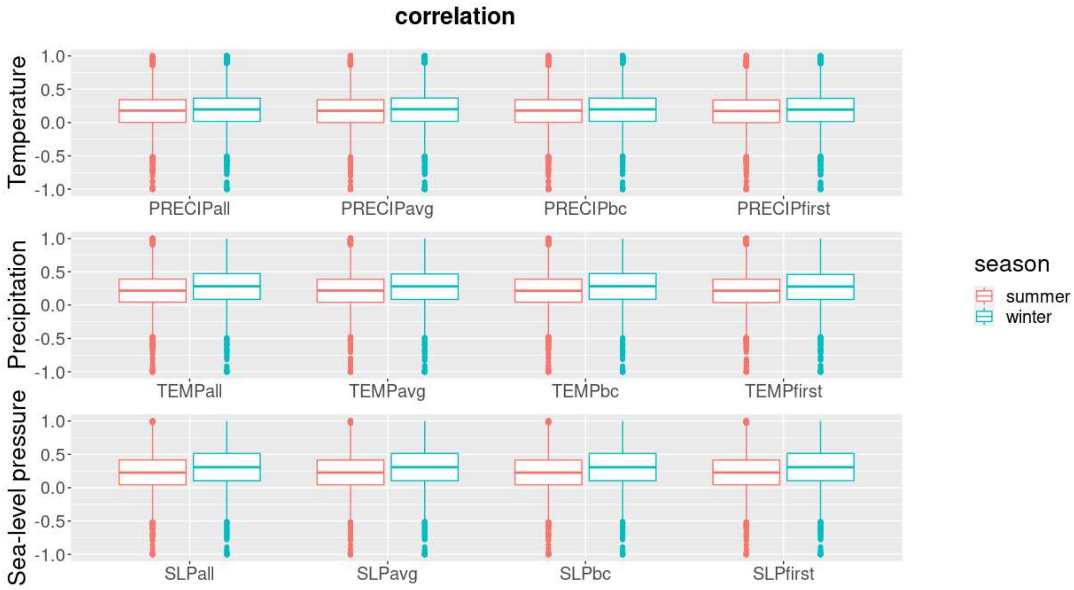

Figure 2: Distribution of correlation coefficient values for all observations of the experiments performed in this thesis for the period 1951-1981. Red colors indicate the correlation coefficient for the boreal summer season and blue colors indicate the correlation coefficient for the boreal winter season

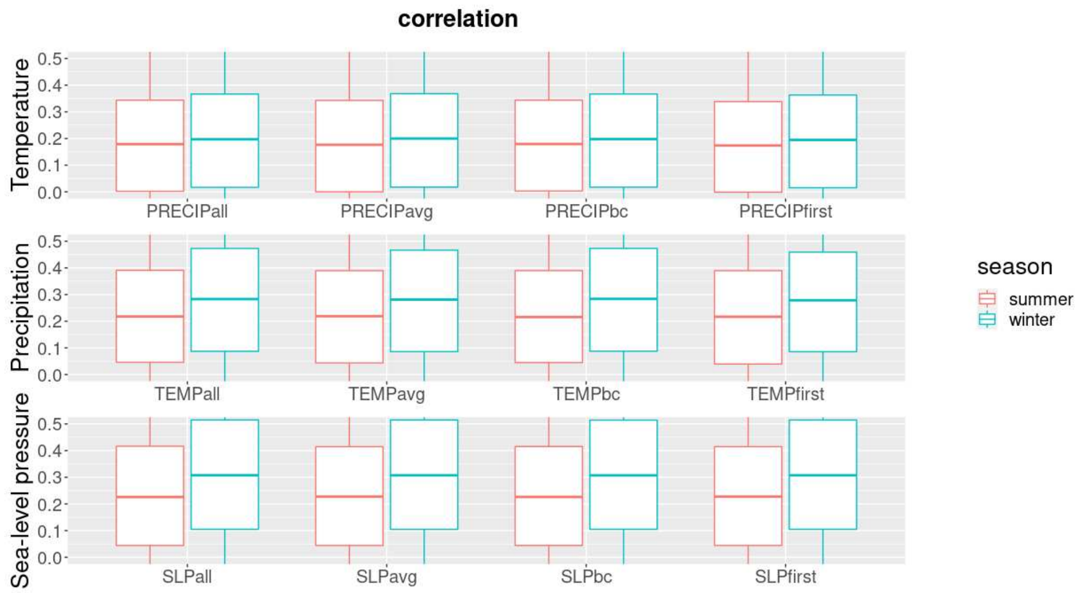

Figure 3: Zoomed version of Figure 2. The Whiskers extend beyond the margins of the figure, as can be seen in the original plot (Figure 2).

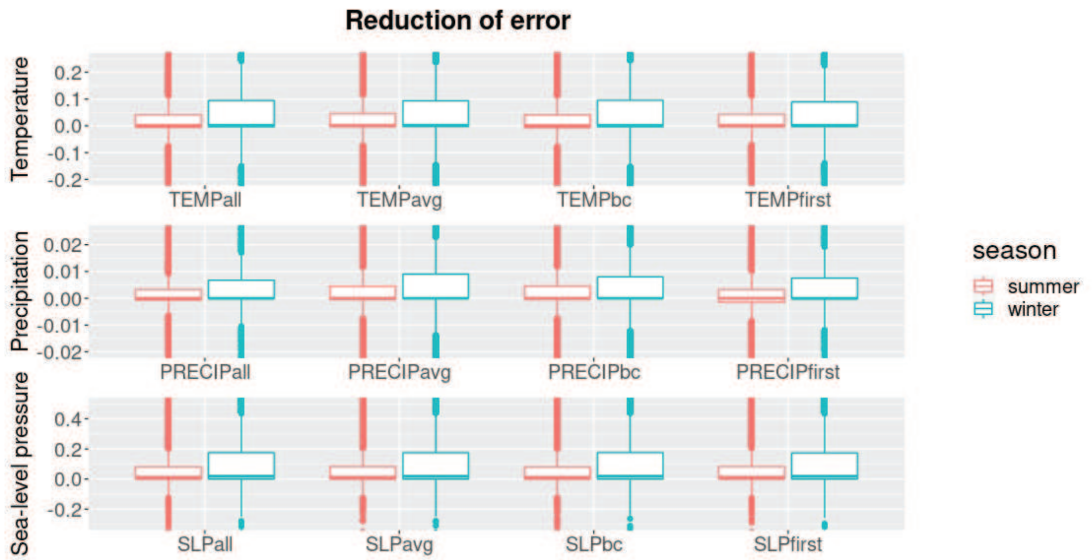

Figure 4: Distribution of reduction of error (mean squared error) values for all observations of the experiments performed in this thesis for the period 1951-1981. Red colors indicate the correlation coefficient for the boreal summer season and blue colors indicate the correlation coefficient for the boreal winter season. The plot was cropped, there were other outliers.

Figure 5: Zoomed version of Figure 4. The Whiskers extend beyond the margins of the figure, as can be seen in the original plot (Figure 4).

3.3 Differences in correlation coefficient

As stated in Chapter 3.1, to determine the quality of the screening method, we compare the experiments with the original setup.

To compare the experiments and assess the quality of the analyses with the new setups, the difference between each of the experiments and the assimilation of all observations (Chapter 3.1) was made.

Two skill scores are used to evaluate the quality of these datasets, as explained in the Methods section: the correlation coefficient and the Mean Squared Error skill score ($MSESS$). In this chapter the results of the correlation coefficient are presented. The results of the $MSESS$, in the other hand, are shown in Chapter 3.4.

3.3.1 Average

As a first experiment, if there was more than one observation in one grid box, the observations were averaged as explained in Chapter 2.2.2. The goal is to reduce the signal-to-noise ratio of the results.

The difference between the correlation coefficient of averaging the data in the same grid box and the correlation coefficient of the original setup for the temperature and the sea-level-pressure is not very high (up to -0.37 for the average of the temperature but in a very small area). For the precipitation, though, the correlation coefficient gets a little higher (up to -0.81). The assimilation of the average of the temperature shows slight negative differences of the correlation

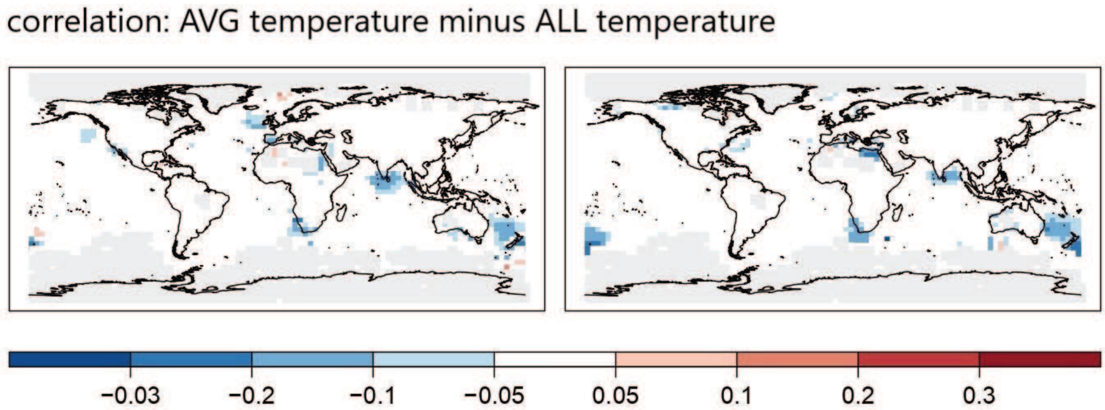

Figure 6: correlation coefficient difference between the assimilation of the average of all the observations in the grid boxes of the temperature (NOBC_JUSTtemp_avg_Sabrina_51-81) and the original assimilation of just the temperature (NOBC_JUSTtemp_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

coefficient with the original setup (Figure 6). These very small differences are of magnitude of less than 0.2.

The results of the difference in the correlation coefficient between the assimilation of the average precipitation per grid box and the original setup for the assimilation of the precipitation shows different results. The differences here go up to -0.81 and 0.53 in some regions (Figure 7). In the season October-March a significant negative difference is present over Afghanistan. Other spurious differences occur between India, Central Asia and East Europe. While in the season April-September there are positive and negative differences in the area around the Caucasus.

The assimilation of just the sea-level-pressure shows, thanks to the correlation coefficient, no differences between the assimilation of the average of the observations in one grid box for the sea-level-pressure and the original setup. The results of the differences of both the assimilations for temperature and sea-level-pressure are so low that it can be said that they equal zero, and thus are negligible, everywhere (Figure 8).

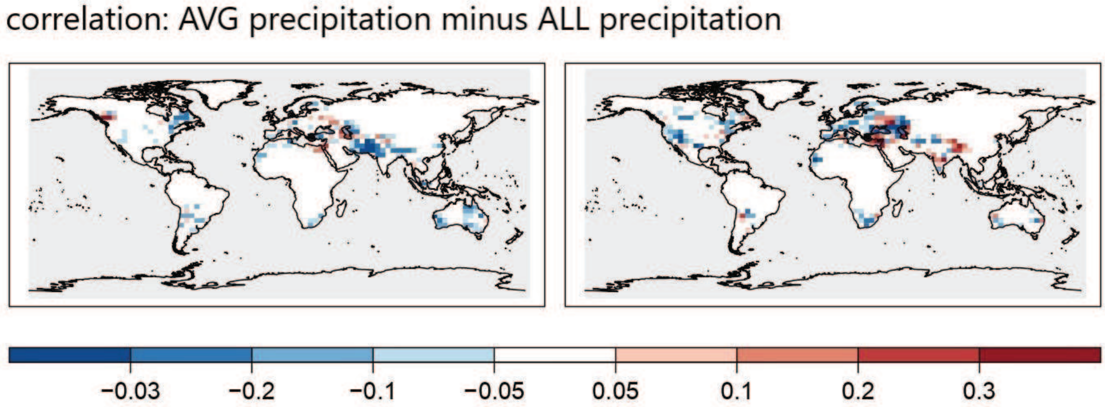

Figure 7: correlation coefficient difference between the assimilation of the average (NOBC_JUSTprecip_avg_Sabrina_51-81) of all the observations in the grid boxes of the precipitation and the original assimilation of just the precipitation (NOBC_JUSTprecip_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

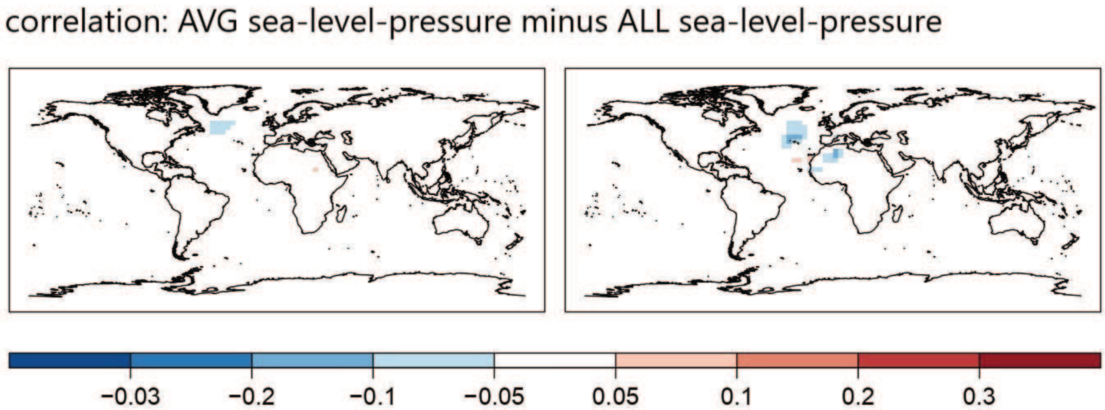

Figure 8: correlation coefficient difference between the assimilation of the average of all the observations in the grid boxes of the sea-level-pressure (NOBC_JUSTslp_avg_Sabrina_51-81) and the original assimilation of just the sea-level-pressure (NOBC_JUSTslp_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

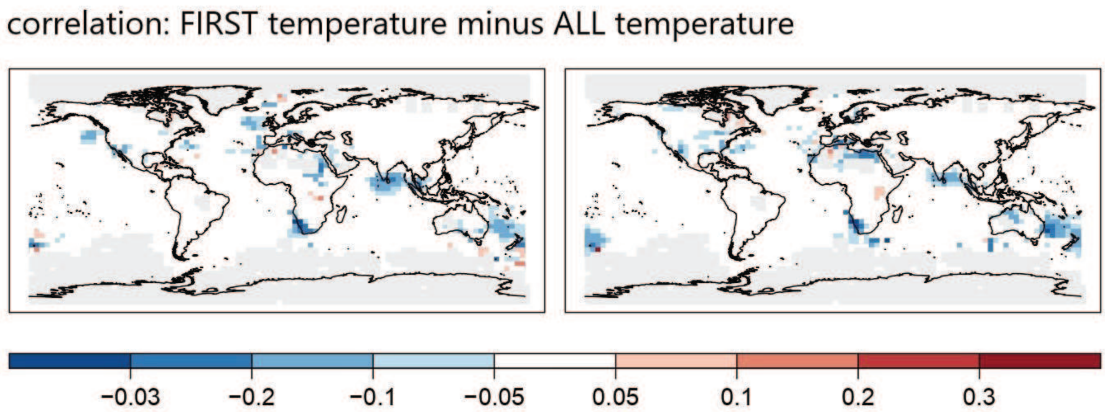

Figure 9: correlation coefficient difference between the random assimilation of a precipitation observation in a grid box if there were more observations in a grid box(NOBC_JUSTtemp_first_Sabrina_51-81) and the original assimilation of just the temperature (NOBC_JUSTtemp_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

3.3.2 First

With the next experiment, if there were more observations in the same grid box just one random observation was chosen as a representative for the whole grid box.

Regarding the results, they are similar to the results of the assimilation of the average of the grid box (Chapter 3.3.1) but show higher numbers in the range of the difference of the temperature as well as in the precipitation (-0.57 to 2.00 for the temperature and -0.69 to 2.00 for the precipitation). Even if the range of the difference in correlation coefficient of the temperature is a little higher, the figures of the random assimilation of an observation, if there were more observations in one grid box, of the temperature and the sea-level-pressure are very similar to the figures in Chapter 3.3.1 (Figure 9 and Figure 11).

However, the figures for the assimilation of the precipitation show different results (Figure 10). It has to be reminded that, when looking at the correlation coefficient for the assimilation of the average (Figure 7), the precipitation was also the only variable with results that are different from zero.

The results of the difference in correlation coefficient, between assimilating the precipitation using a random observation per grid box and the assimilation of the precipitation of a random observation per grid box for the original setup, show negative differences almost everywhere on earth (Figure 10). There are a few positive areas (East Africa, Afghanistan) but the overwhelming majority are negative results for the difference in correlation coefficient for this setup.

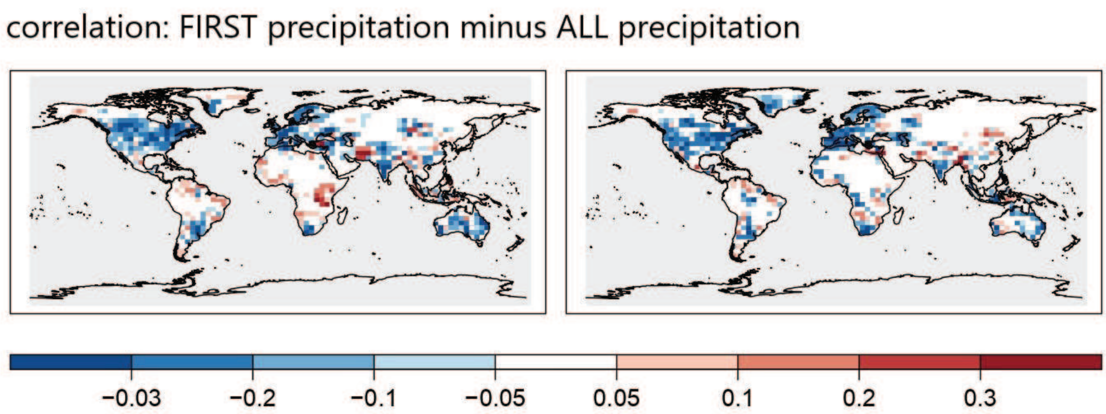

Figure 10: correlation coefficient difference between the random assimilation of a precipitation observation in a grid box if there were more observations in a grid box(NOBC_JUSTprecip_first_Sabrina_51-81) and the original assimilation of just the precipitation (NOBC_JUSTprecip_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

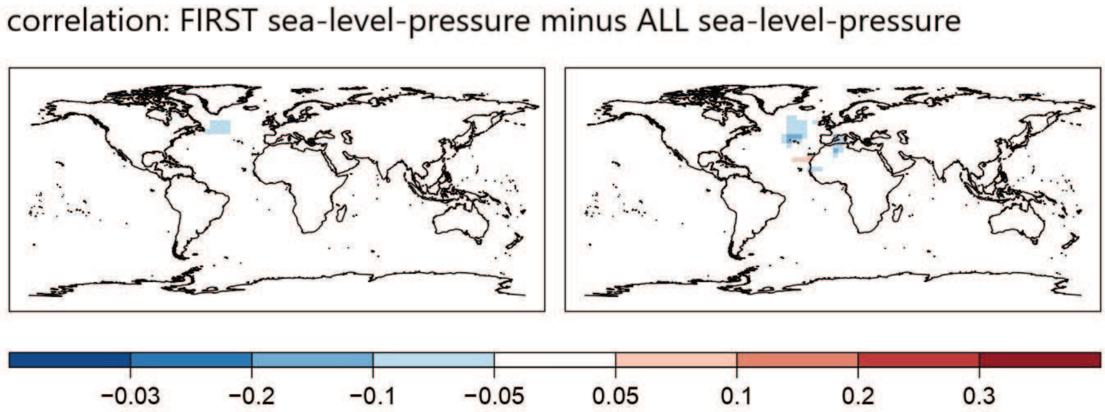

Figure 11: correlation coefficient difference between the random assimilation of a sea-level-pressure observation in a grid box if there were more observations in a grid box (NOBC_JUSTslp_first_Sabrina_51-81) and the original assimilation of just the sea-level-pressure (NOBC_JUSTslp_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

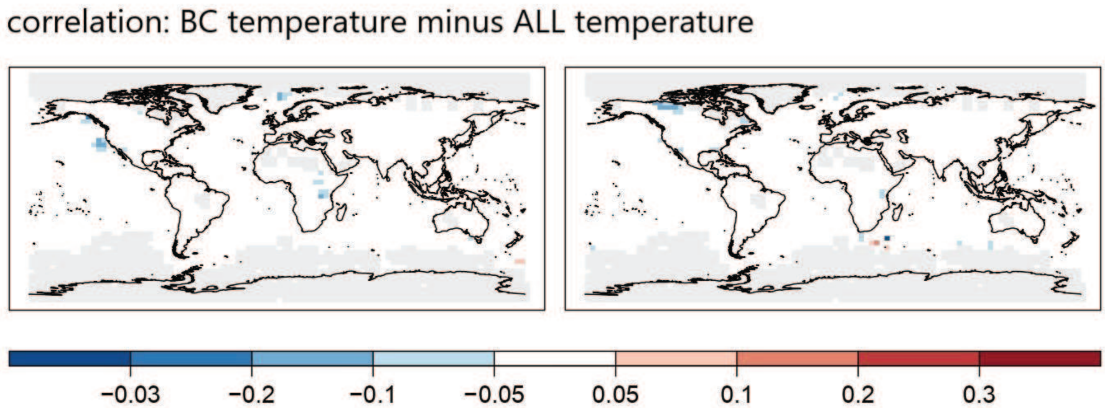

Figure 12: correlation coefficient difference between the assimilation with the Buddy Check algorithm of all the observations in the grid boxes of the temperature (BC_D200_JUSTtemp_all_Sabrina_51-81) and the original assimilation of just the temperature (NOBC_JUSTtemp_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

3.3.3 Buddy Check

The last method tested for data screening is a method which uses the implementation of the Buddy Check algorithm.

To justify the use of a D = 200 km radius used for the "buddies" figures with all the included and excluded observations for the assimilation of temperature (Figure 13), precipitation (Figure 14) and sea-level pressure (Figure 15) are produced. It is clear from these figures that there are some regions where the observations are tightly packed and that the radius of 200 km is justified. However there are also many regions where there are no observations at all, or regions where just a few observations are available. In these regions the Buddy Check may not be helpful.

The presentation of the included and excluded observations on two separate maps show that most of the observations for all three variables temperature, precipitation, sea-level-pressure (included and excluded) are in India, South America, Eastern Australia and South Africa (Figure 13, Figure 14 and Figure 15). These are also the regions where, in the skill score differences, we have unsurprisingly the most difference in the skill scores.

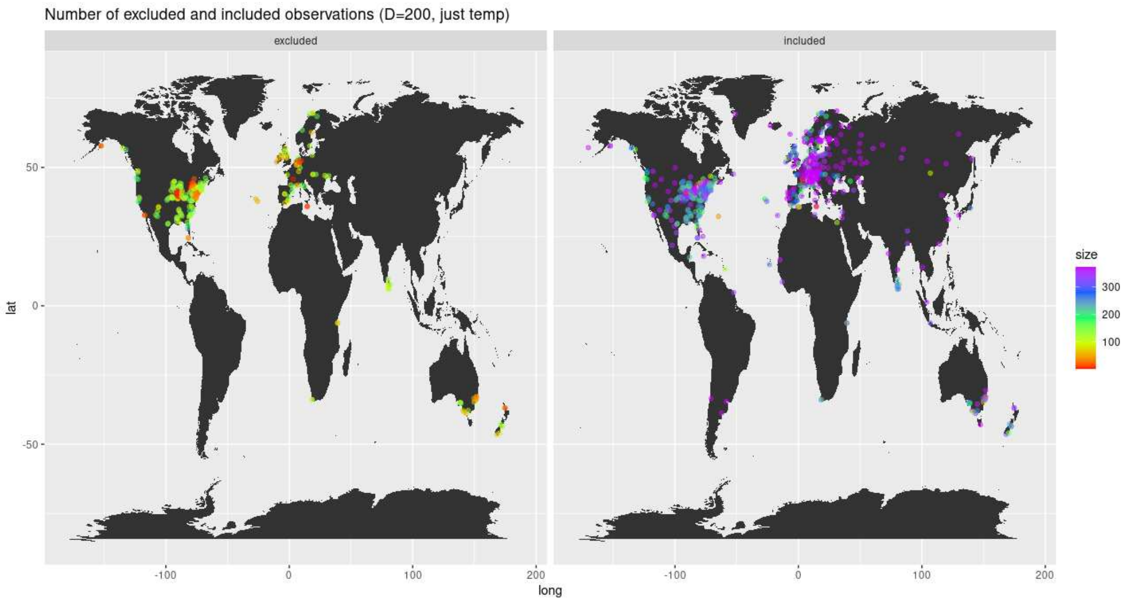

Figure 13: Cumulative number of excluded and included observations when assimilating the temperature with the Buddy Check algorithm to help flag suspect observations (D = 200 km). The color of each point indicates how many times a location where an observation is available was taken into consideration or not. The map on the left indicates the how many times and which locations were not included in the assimilation (suspect observations). The map on the right, on the other hand, indicates how many times and which locations were included in the assimilation.

The results of the screening of the difference in the correlation coefficient using the Buddy Check algorithm show that there is no change in the quality of the assimilation (Figure 12, Figure 16 and Figure 17). The differences are equal to zero for the assimilation of all three variables

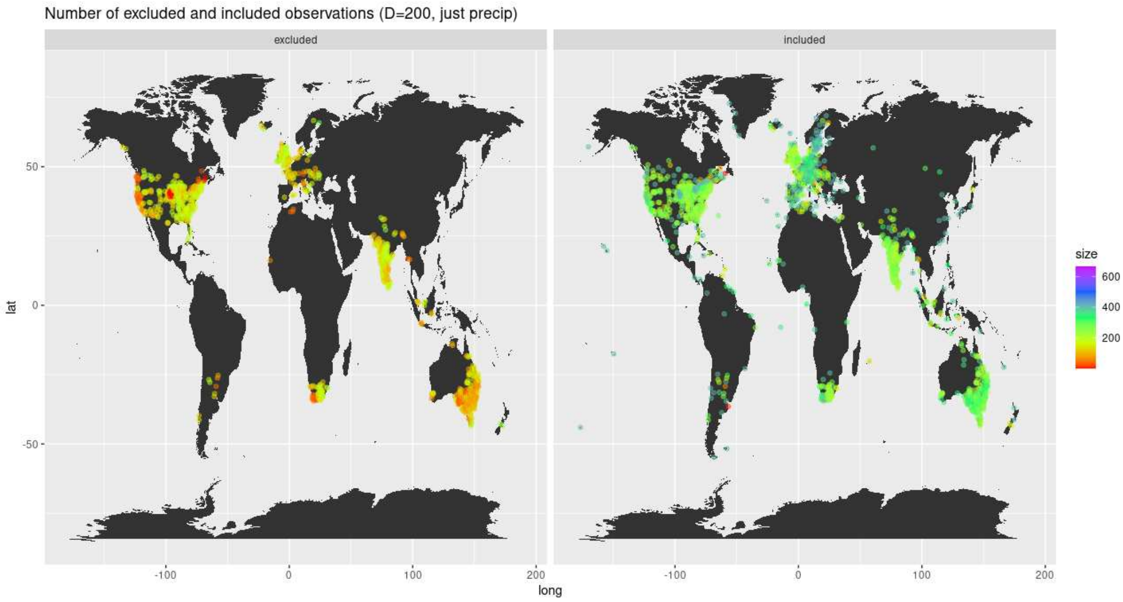

Figure 14: Cumulative number of excluded and included observations when assimilating the precipitation with the Buddy Check algorithm to help flag suspect observations (D = 200 km). The color of each point indicates how many times a location, where an observation is available, was taken into consideration or not. The map on the left indicates the how many times and which locations were not included in the assimilation (suspect observations). The map on the right, on the other hand, indicates how many times and which locations were included in the assimilation.

(temperature, precipitation, sea-level-pressure).

The difference in the correlation skill score shows a difference when the radius D is increased (Figure 18). The test performed, as stated in the introduction, is with a radius of 500 km. The difference in correlation skill score is different than zero this time. There is a big positive area in southern Asia and a negative area in the Southwestern region of North America.

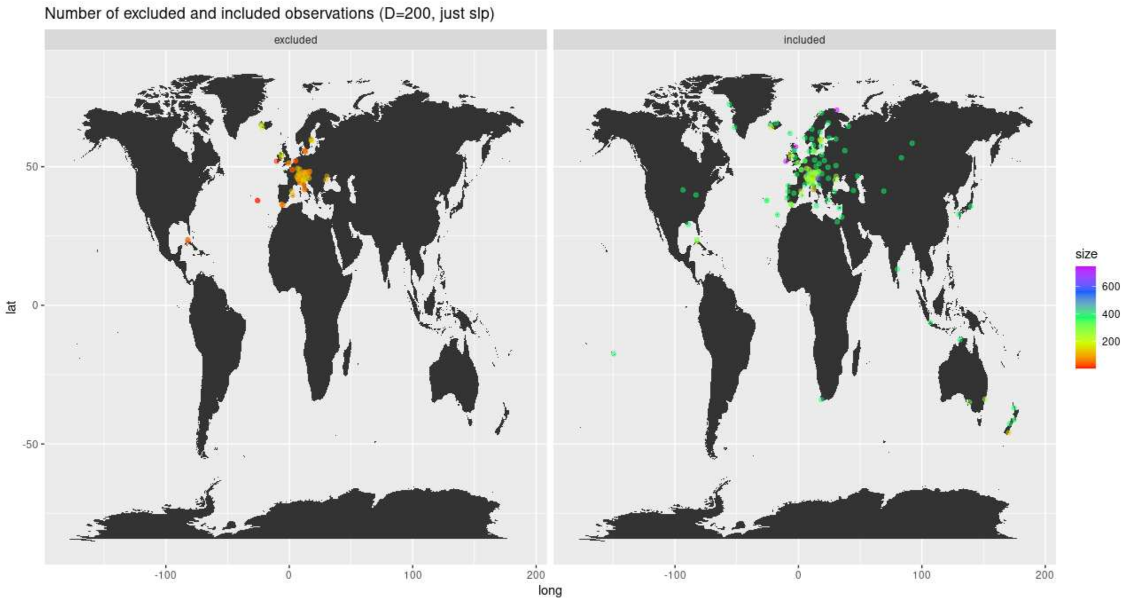

Figure 15: Cumulative number of excluded and included observations when assimilating the sea-level-pressure with the Buddy Check algorithm to help flag suspect observations (D = 200 km). The color of each point indicates how many times a location where an observation is available was taken into consideration or not. The map on the left indicates the how many times and which locations were not included in the assimilation (suspect observations). The map on the right, on the other hand, indicates how many times and which locations were included in the assimilation.

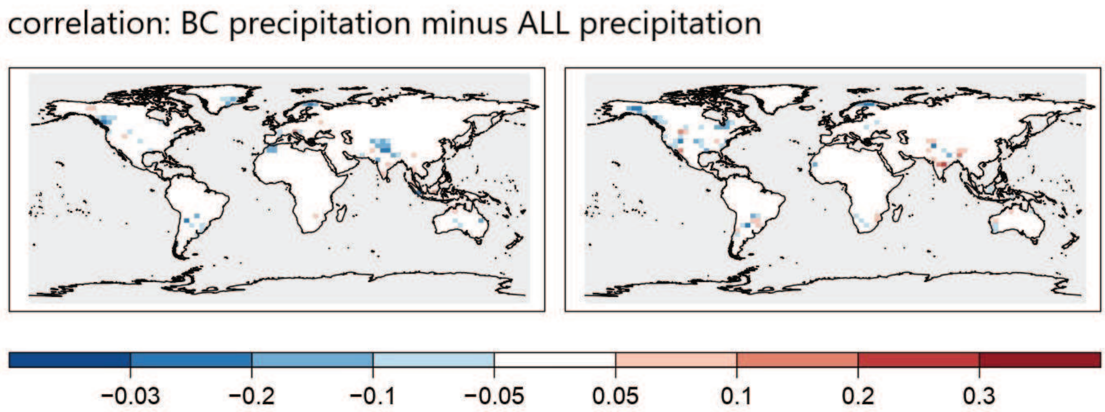

Figure 16: correlation coefficient difference between the assimilation with the Buddy Check algorithm of all the observations in the grid boxes of the precipitation (BC_D200_JUSTprecip_all_Sabrina_51-81) and the original assimilation of just the precipitation (NOBC_JUSTprecip_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

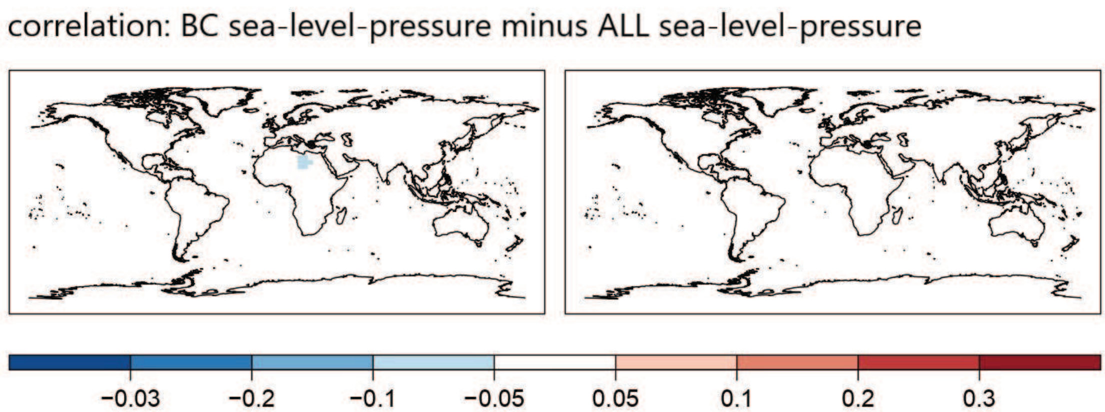

Figure 17: correlation coefficient difference between the assimilation with the Buddy Check algorithm of all the observations in the grid boxes of the sea-level-pressure (BC_D200_JUSTslp_all_Sabrina_51-81) and the original assimilation of just the sea-level-pressure (NOBC_JUSTslp_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

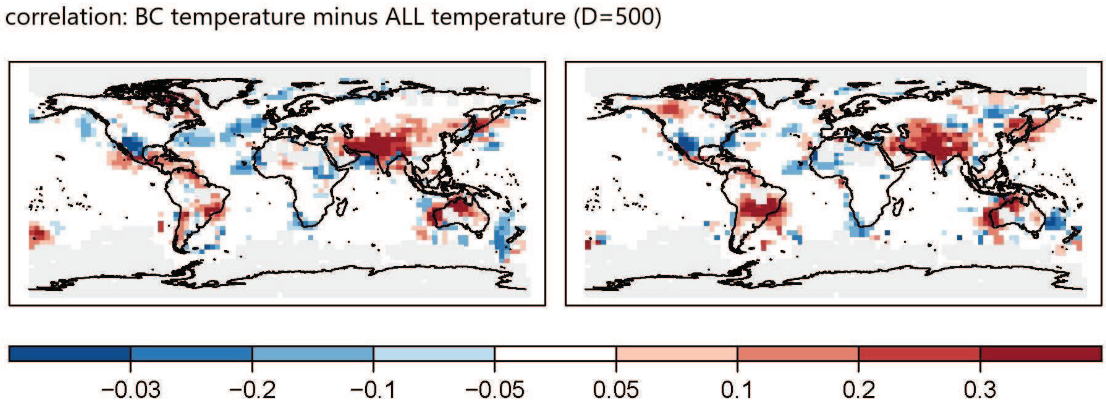

Figure 18: correlation coefficient difference between the assimilation with the Buddy Check algorithm of all the observations in the grid boxes of the temperature (BC_D500_JUSTtemp_all_Sabrina_51-81) using a radius of D = 500km and the original assimilation of just the temperature (NOBC_JUSTtemp_all_Sabrina_51-81). The panel on the left shows the correlation coefficient in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

3.4 Mean Squared Error skill score

The other skill score used to compare and assess the results of the data screening methods is the difference of the mean squared error skill score.

As before, the comparison will be done with a difference between the experiments and the original setup.

3.4.1 Average

Just like in the differences in correlation coefficient the first experiment is the difference between the assimilation of the average of the observation inside a grid box and the original setup.

The results of this difference when looking at the figures produced, when the variable assimilated is the temperature or the sea-level-pressure, are extremely low values (Figure 19 and Figure 21). This result suggests that assimilating the temperature observations and the sea-level-pressure observations by averaging the values in the same grid box does not improve the assimilation quality.

The results of the assimilation of the precipitation with this setting, just like for the other experiments, yields different results compared with the other two variables (Figure 20). The difference with this setting has positive results over Europe, Central-Eastern United States of America and parts of India for both seasons April-September and October-March. This suggests that in these regions the quality of the assimilation improved by averaging the observations in the same grid box. The range of this assimilation is -355.19 to 20.24. As this range suggests, however, there are also regions where the quality of the assimilation worsened by assimilationg the average of the observations in the same gidbox. Namely the regions of the Caucasus and Afghanistan.

3.4.2 First

The outcome of the difference between the random assimilation of an observation inside the grid boxes and the original assimilation shows, as before, similar results for the figures of the skill scores.

The results of the difference between the mean squared error skill score are negligible for the assimilation of the temperature and for the assimilation of sea-level-pressure (Figure 22 and Figure 24).

When this setting is used for the assimilation of the precipitation the difference is significant (-49.36 to 25.85). Similar to the results of the assimilation of the average of the observations in the same grid box (Chapter 3.4.1), the assimilation improved in quality over the United States (mostly in the months between April and September) and Southern Asia and Europe. With these settings the assimilation has a worse quality, according to the MSE skill score, over a large region near Mongolia in winter (Figure 23).

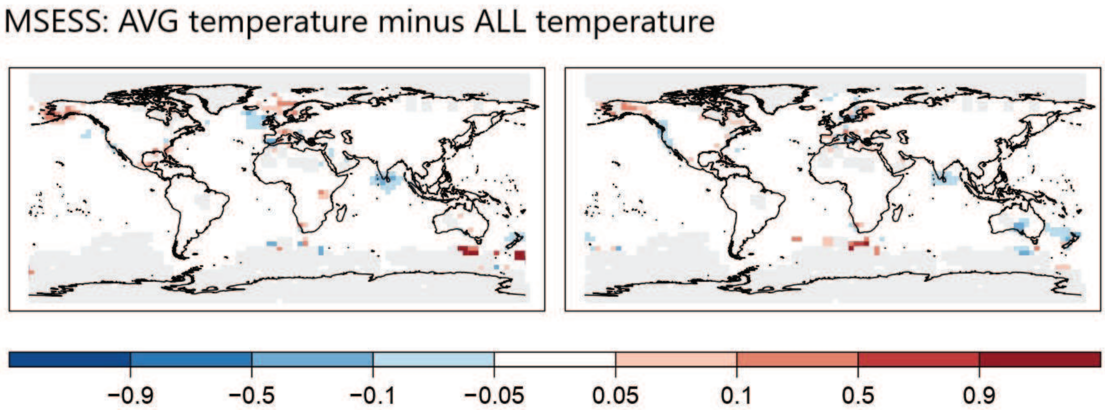

Figure 19: RE skill score (MSESS) difference between the assimilation of the average of all the observations in the grid boxes of the temperature(NOBC_JUSTtemp_avg_Sabrina_51-81) and the original assimilation of just the temperature (NOBC_JUSTtemp_all_Sabrina_51-81). The panel on the left shows the MSE skill in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

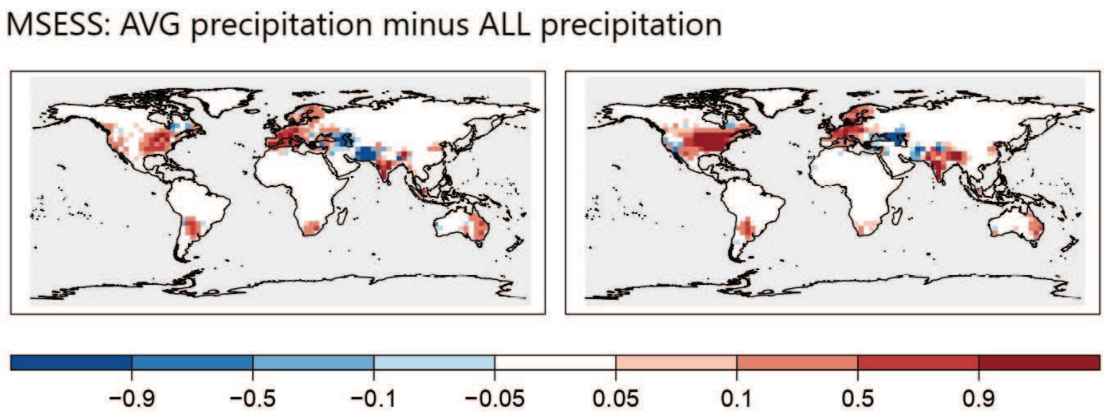

Figure 20: RE skill score (MSESS) difference between the assimilation of the average of all the observations in the grid boxes of the precipitation (NOBC_JUSTprecip_avg_Sabrina_51-81) and the original assimilation of just the precipitation (NOBC_JUSTprecip_all_Sabrina_51-81). The panel on the left shows the MSE skill score in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

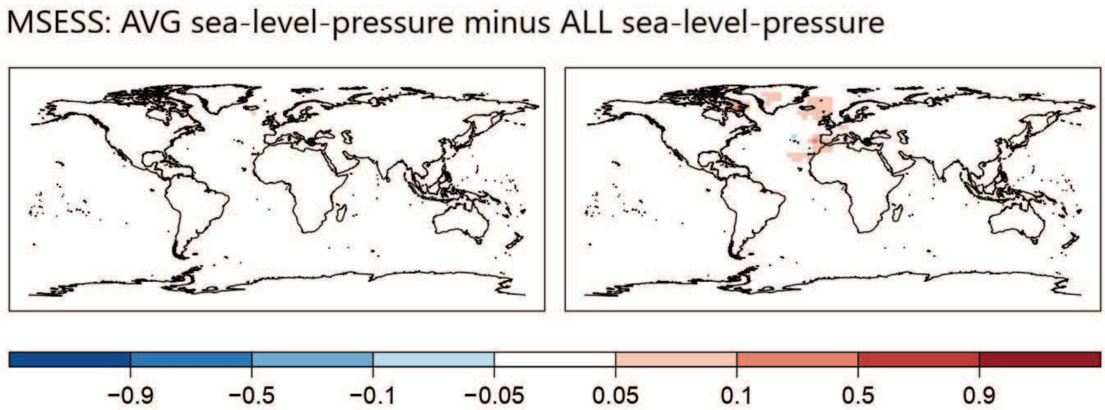

Figure 21: RE skill score (MSESS) difference between the assimilation of the average of all the observations in the grid boxes of the sea-level-pressure (NOBC_JUSTslp_avg_Sabrina_51-81) and the original assimilation of just the sea-level-pressure (NOBC_JUSTslp_all_Sabrina_51-81). The panel on the left shows the MSE skill score in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

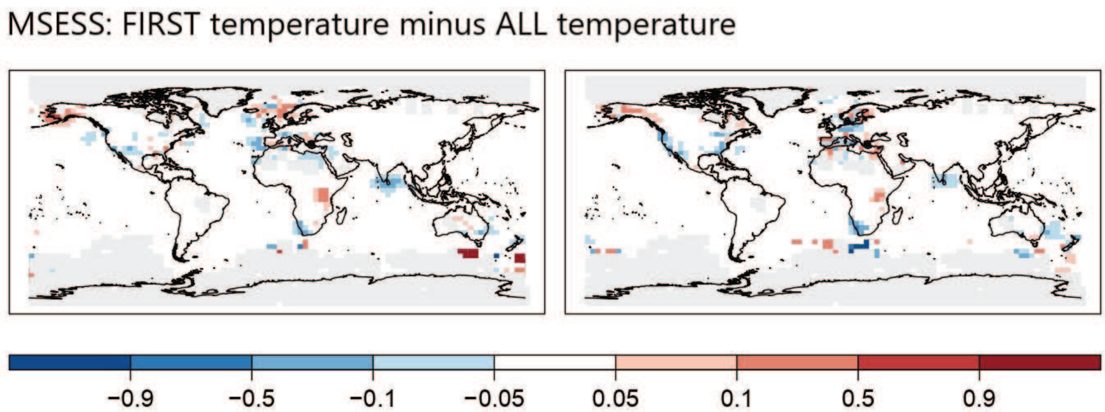

Figure 22: MSE skill score difference between the random assimilation of a temperature observation in a grid box if there were more observations in a grid box (NOBC_JUSTtemp_first_Sabrina_51-81) and the original assimilation of just the temperature (NOBC_JUSTtemp_all_Sabrina_51-81). The panel on the left shows the MSE skill score in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

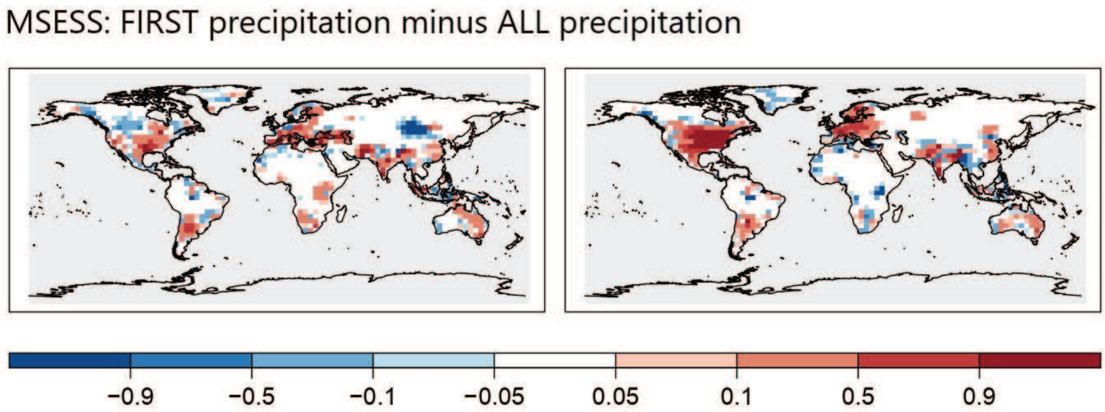

Figure 23: RE skill score (MSESS) difference between the random assimilation of a precipitation observation in a grid box if there were more observations in a grid box (NOBC_JUSTprecip_first_Sabrina_51-81) and the original assimilation of just the precipitation (NOBC_JUSTprecip_all_Sabrina_51-81). The panel on the left shows the MSE skill score in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

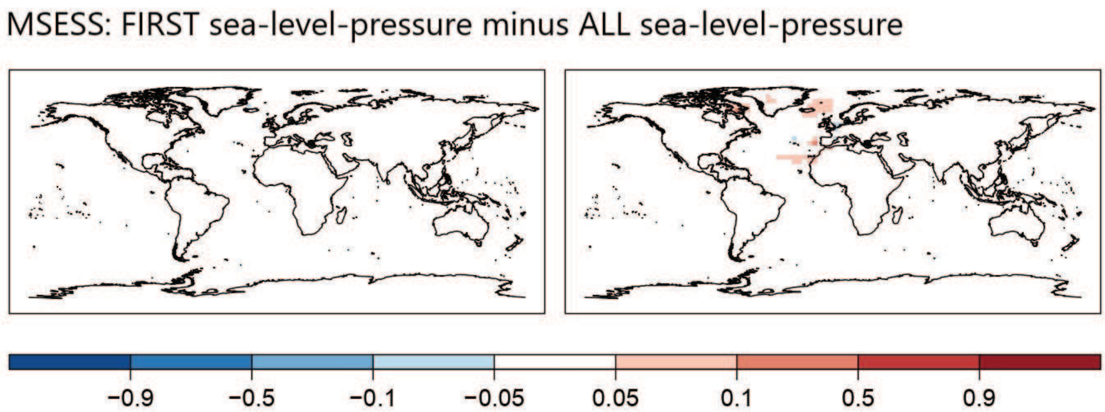

Figure 24: RE skill score (MSESS) difference between the random assimilation of a sea-level-pressure observation in a grid box if there were more observations in a grid box (NOBC_JUSTslp_first_Sabrina_51-81) and the original assimilation of just the sea-level-pressure (NOBC_JUSTslp_all_Sabrina_51-81). The panel on the left shows the MSE skill score in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

3.4.3 Buddy Check

Finally, the results of the difference in MSE skill score between the assimilation of the three variables taken in consideration (temperature, precipitation, sea-level-pressure) with the screening of the observations with the algorithm of the buddy check and the original setup show similar results to all the other experiments.

The difference in the MSE skill score between the assimilation of the temperature and the sea-level-pressure with the algorithm of the BC and the original setup show no differences in the figures (Figure 25 and Figure 27). The performance of the Buddy Check algorithm, when the precipitation was assimilated, shows positive results in the United States, Europe and South Asia (Figure 26). There are no negative differences in the $MSESS$ for this experiment (up to -0.51).

When the radius is increased to D = 500 km for the assimilation of temperature, there are again more differences in the MSE skill score. Just as in the correlation skill score, the assimilation improves over Southeast Asia and worsens in the Southwestern area of North America.

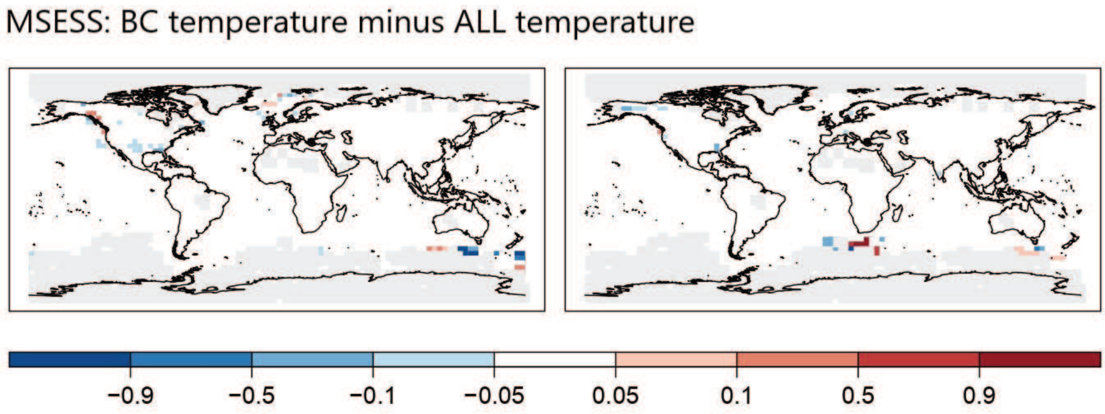

Figure 25: RE skill score (MSESS) difference between the assimilation with the Buddy Check algorithm of all the observations in the grid boxes of the temperature (BC_D200_JUSTtemp_all_Sabrina_51-81) and the original assimilation of just the temperature (NOBC_JUSTtemp_all_Sabrina_51-81). The panel on the left shows the MSE skill in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

3.5 Buddy Check and average

In order to have another measure of comparison for the Buddy Check it is interesting to research how it performs when the compared assimilation is not the original assimilation but the assimilation with the averaging of the observation in the same grid box. The same method as with the original setup is used. The differences in the correlation coefficient and the difference in the mean squared error is calculated.

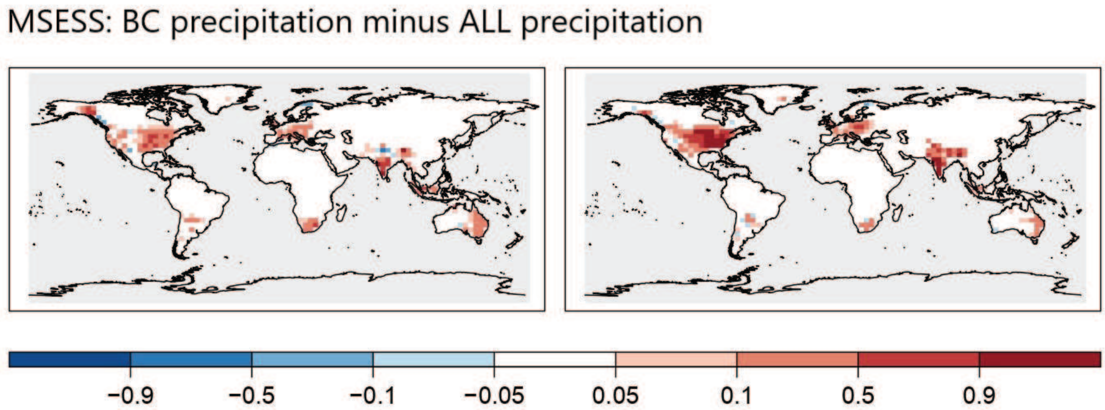

Figure 26: RE skill score (MSESS) difference between the assimilation with the Buddy Check algorithm of all the observations in the grid boxes of the precipitation (BC_D200_JUSTprecip_all_Sabrina_51-81) and the original assimilation of just the precipitation (NOBC_JUSTprecip_all_Sabrina_51-81). The panel on the left shows the MSE skill in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

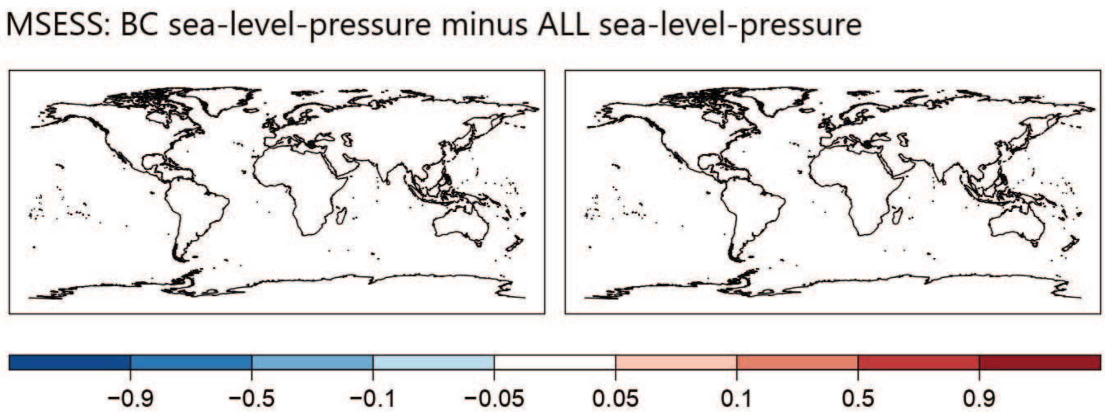

Figure 27: RE skill score (MSESS) difference between the assimilation with the Buddy Check algorithm of all the observations in the grid boxes of the sea-level-pressure (BC_D200_JUSTslp_all_Sabrina_51-81) and the original assimilation of just the sea-level-pressure (NOBC_JUSTslp_all_Sabrina_51-81). The panel on the left shows the MSE skill in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

3.5.1 Correlation coefficient AVG - BC

The figure (Figure 29) show that the difference lays in the same regions as the differences in the skill scores for the precipitation.

3.5.2 Mean squared error AVG - BC

The range of the difference in mean squared error between the assimilation of the temperature with Buddy Check algorithm and the assimilation of the temperature by averaging the observations in the same grid box is -19.72 to 355.19. However what interests us the most are the regions where the skill scores got worse or better, which are shown by the figures of the differences in the skill scores. These show for the season October-March an improvement in the skill around Afghanistan and Caucasus, while it shows a worsening of the skill in Europe and the United States for both seasons (Figure 28).

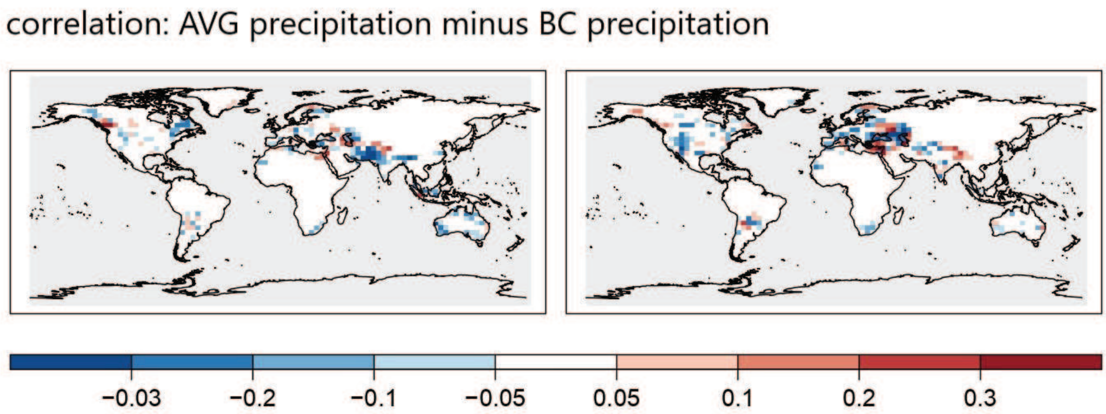

Figure 28: correlation coefficient difference between the assimilation of just precipitation with the Buddy Check algorithm (BC_D200_JUSTprecip_all_Sabrina_51-81) and the assimilation of all the observations averaging the observations in the same grid box for the precipitation (NOBC_JUSTprecip_avg_Sabrina_51-81). The panel on the left shows the MSE skill in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

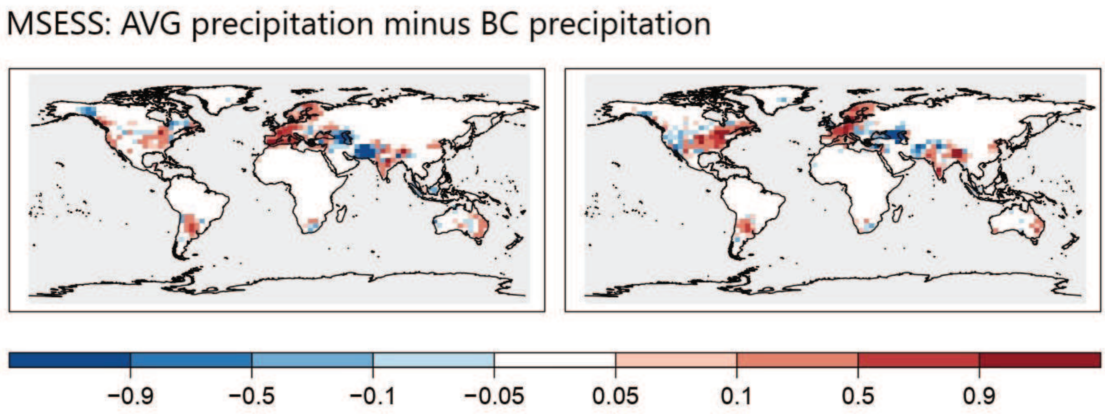

Figure 29: RE skill score (MSESS) difference between the assimilation of just precipitation with the Buddy Check algorithm (BC_D200_JUSTprecip_all_Sabrina_51-81) and the assimilation of all the observations averaging the observations in the same grid box for the precipitation (NOBC_JUSTprecip_avg_Sabrina_51-81). The panel on the left shows the MSE skill in the season October-March (boreal winter) while the panel on the right shows the skill for April-September (boreal summer).

3.5.3 Comparison

The results of the assimilations shows that the only assimilation setup that has different results is the assimilation of the precipitation. We are particularly interested in the performance of the Buddy Check algorithm. To understand better what is happening with the assimilation of the precipitation and the other variables using the Buddy Check algorithm, further investigations on the results of the assimilation using the Buddy Check algorithm were made.

The comparison of the skill scores of the precipitation show regions where the difference between the original setup and the assimilation with the average is clearly positive or negative. To examine the assimilation of the precipitation with the average, histograms and time series of interesting locations were looked into.

The locations that are taken into consideration are Afghanistan and Caucasus, which are regions where the difference in the MSE skill score between the assimilation using the average of the grid cells with more than one observation and the original setup are negative. Michigan and Nice are taken into consideration as examples of areas where this difference is positive.

Time series of the anomalies of the analysis are used to inspect how the temporal progress may have influenced the these results. The time series of Afghanistan and Nice, which have negative and positive differences in MSE skill score, were investigated in their the winter season because we had most differences in that season. The time series of the anomalies of the average, compared to the anomalies of the original setup, show interesting results. In the cases of Afghanistan and Caucasus, When the difference in the MSESS was negative, the time series always had a spike which is probably what worsened the assimilation.

Namely both the time series of the average of the Caucasus (Figure 32) and Afghanistan (Figure 33) show similarities in the amplitude (especially for Caucasus). However both have bad bias. The time series of the average of Nice (Figure 34) as well as the time series of the average of Michigan (Figure 35), in the other hand, show very good variability. A conclusion and discussion of these results will be drawn in the next chapter (Chapter 4.6).

To check whether we can trust the CRU data set I checked, in the climate explorer (WMO accessed: 09.10.2021) for ten stations around the location of Nice and Afghanistan. The plotted seasonal anomalies of the precipitation from the Global Historical Climatology Network daily (GHCNd), plotted for the whole year not just for summer or winter as the time series were plotted. I looked for anomalies in the observations of seasonal precipitation for the years 1951-1981. Thanks to the plots of the time series we can see that in the station of Caucasus (Figure 30) we have a positive anomaly around the year 1964 and around 1975 for the EKF400, in the GHCNd there were seasonal precipitation anomalies around the years 1968 and 1958, the anomalies are of much smaller magnitude than the EKF400 shows. For the station of the Afghanistan (Figure 31), in the other hand, the climate explorer shows that just before 1970 and just before 1975 there was a peak in the precipitation anomaly. Just as showed by the time series of the EKF400 but of much smaller magnitude.

Another interesting result that can be seen in the time series of the anomaly of the precipitation of Nice (Figure 34) is that in the more recent years there is no uncertainty in the EKF400. The reconstruction completely agrees with the reference (CRU).

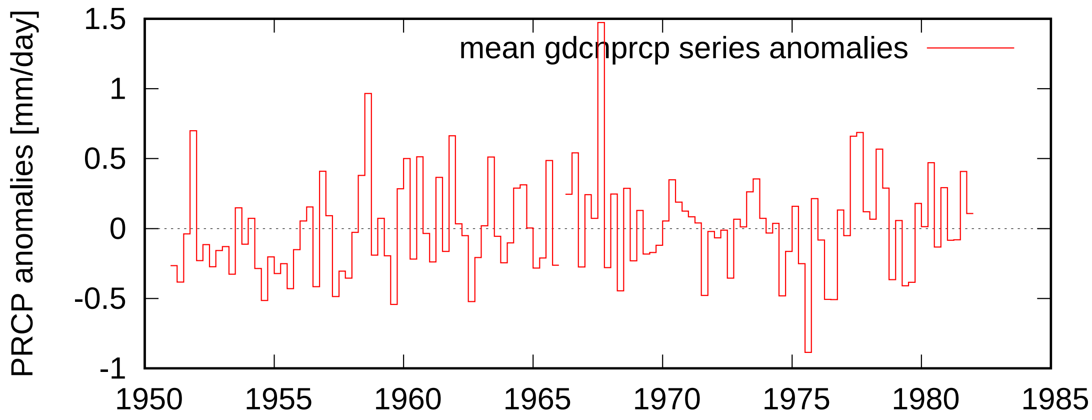

Figure 30: Time series of seasonal anomaly of the Global Historical Climatology Network daily for ten stations located around Caucasus for the whole year in the period 1951-1981.

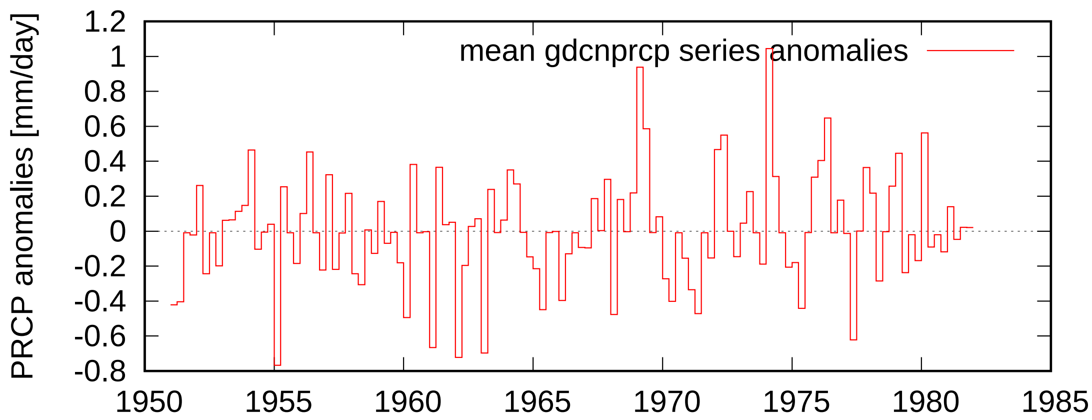

Figure 31: Time series of seasonal anomaly of the Global Historical Climatology Network daily for ten stations located around Afghanistan for the whole year in the period 1950-1980.

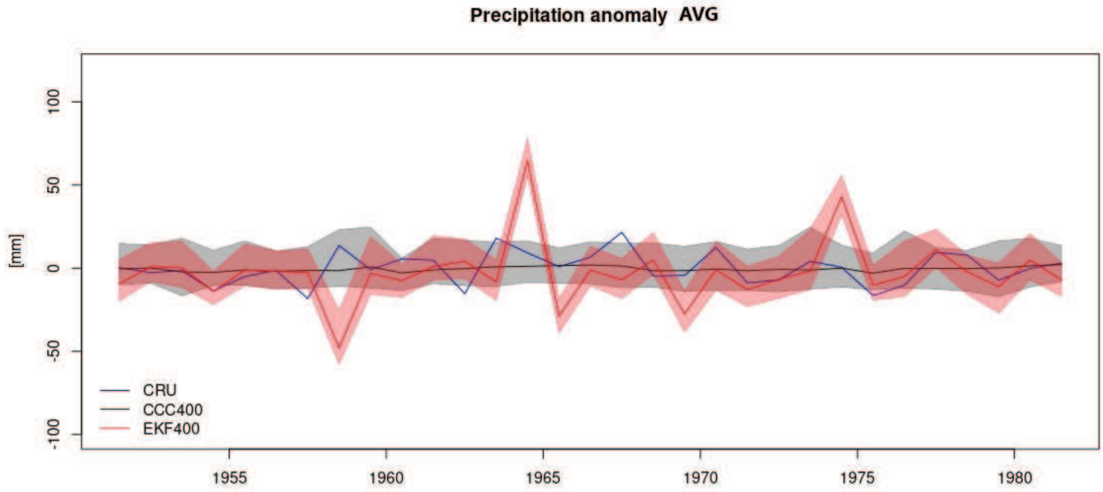

Figure 32: Time series of Caucasus in the summer season for the assimilation of the precipitation using the average of the observations in the same grid box.

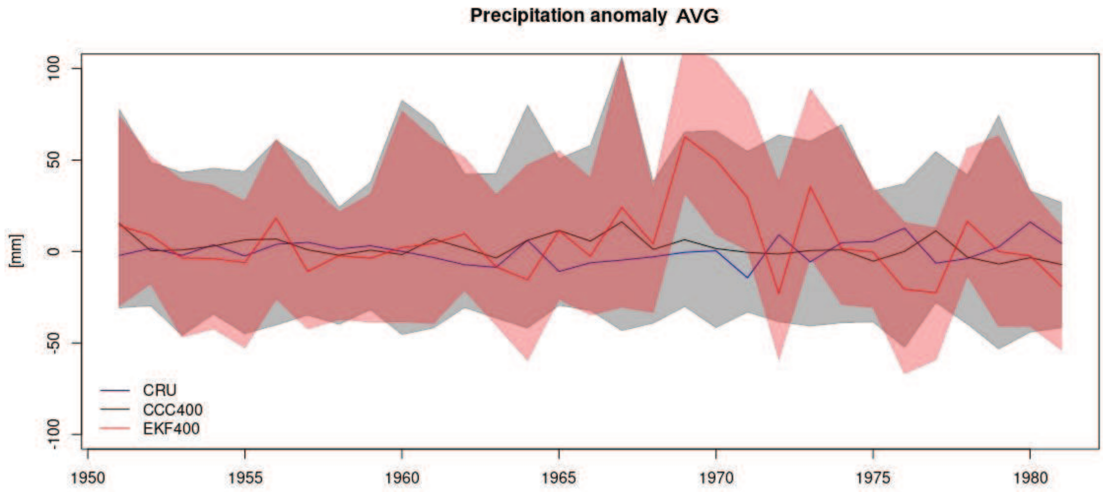

Figure 33: Time series of Afghanistan in the winter season for the assimilation of the precipitation using the average of the observations in the same grid box.

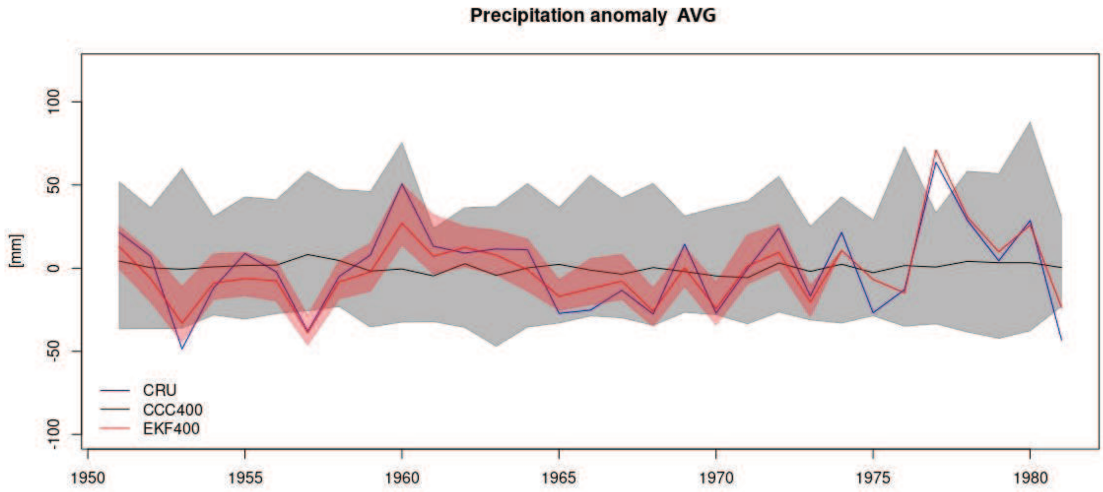

Figure 34: Time series of Nice in the winter season for the assimilation of the precipitation using the average of the observations in the same grid box.

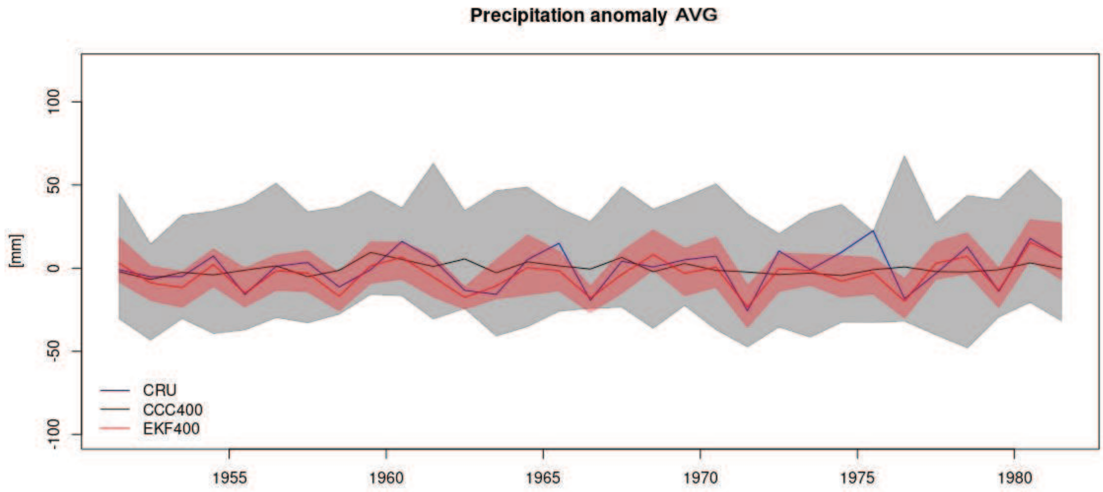

Figure 35: Time series of Michigan in the summer season for the assimilation of the precipitation using the average of the observations in the same grid box.

By analysing the time series of the anomalies it became clear the the mean of the anomalies is not zero. For this reason the next step in understanding this analysis was that of looking at the distribution of the data. For this purpose, histograms of the data sets for the same locations that were examined in the time series, were built.

The histograms show, unsurprisingly, that the data is skewed. The histograms for the assimilation of AVG of the precipitation, in the regions where the difference in the MSESS is positive, are left-skewed (Figure 36 and Figure 37). Since the assimilation of the AVG is left-skewed, it means that there are very few large data values and therefore has a right tail (Wilks 2011). The histograms of the assimilation of the AVG of the precipitation, for the regions where the difference in the MSESS is negative, show similar results to those of the regions where this difference is positive (Figure 38 and Figure 39).

Since, thanks to the the time series, the median of the anomalies seemed to be negative also when assimilating the precipitation with the original setup, histograms for the same locations but for the assimilation of precipitation with the original setup were built. These histograms are left-skewed for the regions were the difference in MSESS is positive (Figure 40 and Figure 41), with negative values up to -50 mm for Nice and up to -40 mm for Michigan. The other two investigated locations where the difference in MSESS is negative, Caucasus and Afghanistan, have left-skewed histograms as well like the other two locations. These last two locations had less negative values (up to -6 mm in Caucasus and up to -20 mm in Afghanistan) in comparison to the histograms of the locations where the difference in MSESS was positive (Figure 42 and Figure 43).

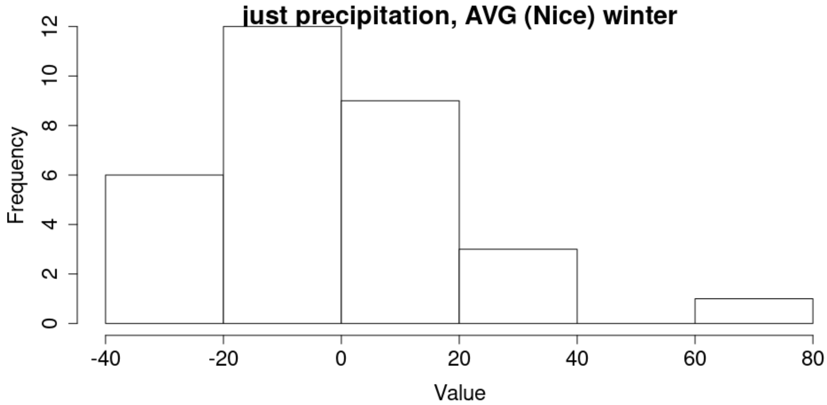

Figure 36: Histograms for the distribution of the assimilation of the station of Nice in the winter season using the average of the precipitation observations in the same grid box.

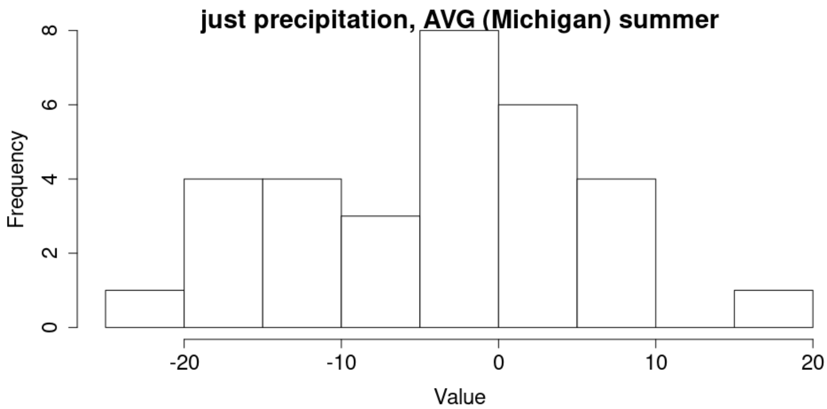

Figure 37: Histograms for the distribution of the assimilation of the station of Michigan in the summer season using the average of the precipitation observations in the same grid box.

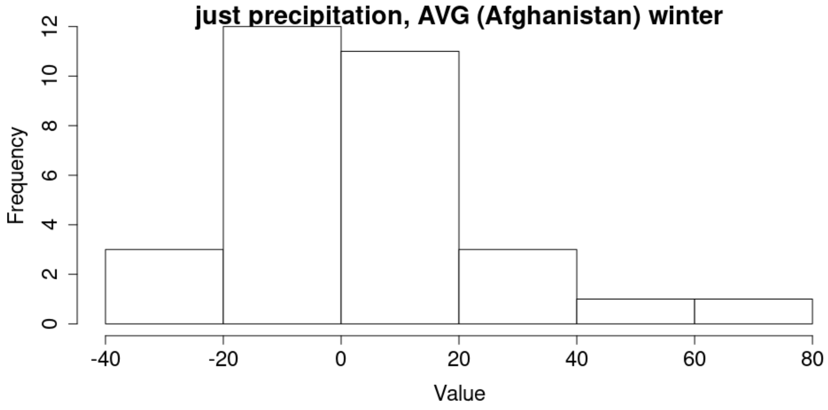

Figure 38: Histograms for the distribution of the assimilation of the station of Afghanistan in the winter season using the average of the precipitation observations in the same grid box.

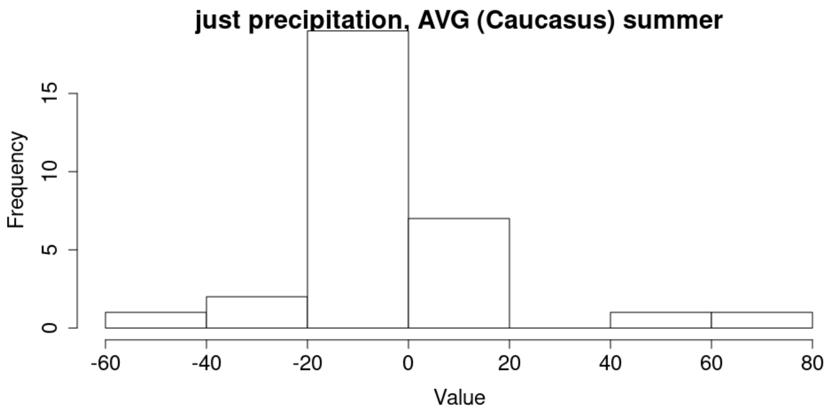

Figure 39: Histograms for the distribution of the assimilation of the station of Caucasus in the summer season using the average of the precipitation observations in the same grid box.

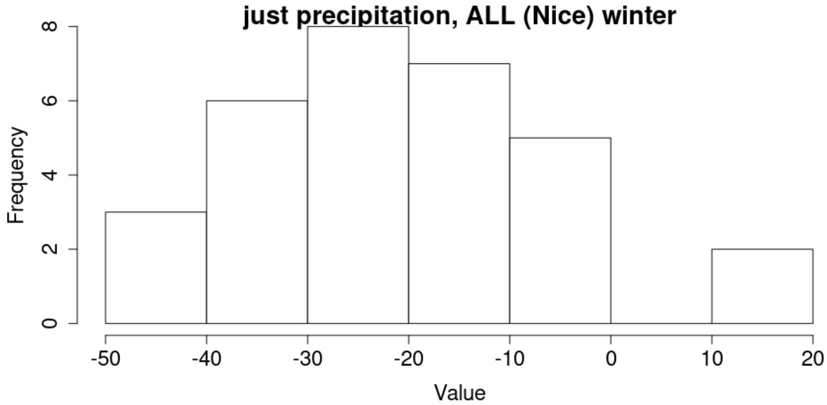

Figure 40: Histograms for the distribution of the assimilation of the station of Nice in the winter season using all the precipitation observations in the same grid box.

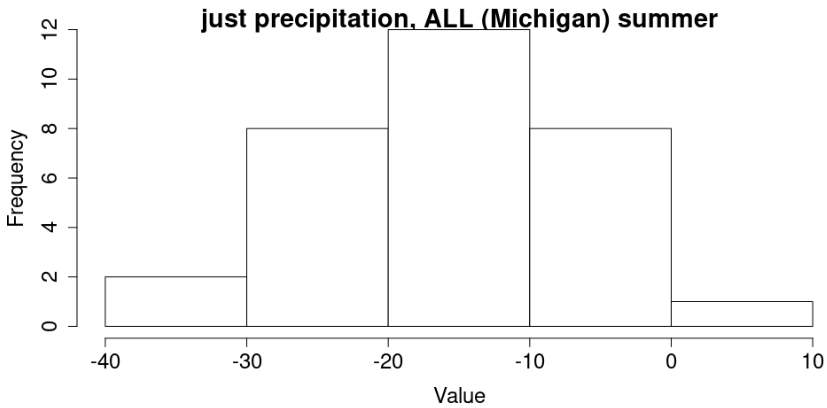

Figure 41: Histograms for the distribution of the assimilation of the station of Michigan in the summer season using all the precipitation observations in the same grid box.

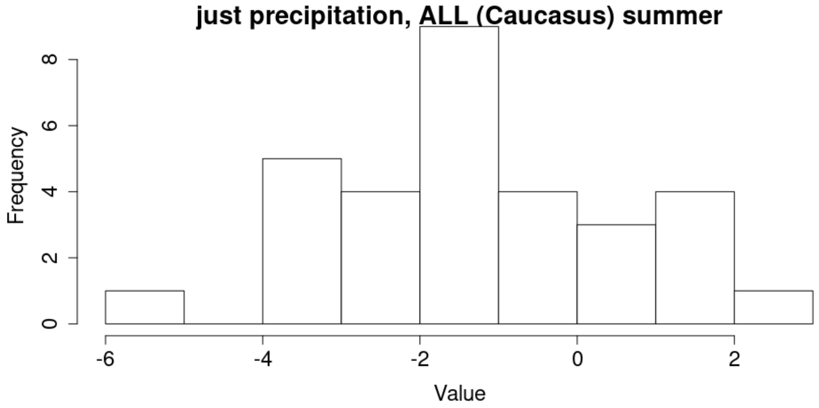

Figure 42: Histograms for the distribution of the assimilation of the station of Caucasus in the summer season using all the precipitation observations in the same grid box.

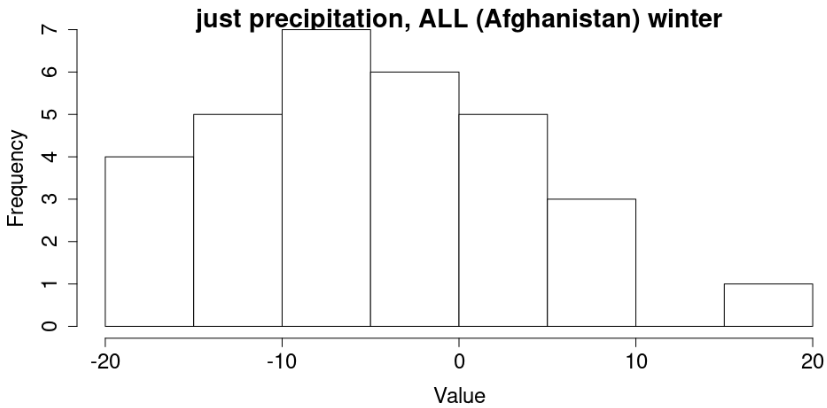

Figure 43: Histograms for the distribution of the assimilation of the station of Afghanistan in the winter season using all the precipitation observations in the same grid box.

As mentioned above, the variable which show the biggest differences, when calculating the difference in skill scores, is the precipitation. It is interesting to see that the maps for the included and the excluded observations, when assimilating the precipitation with the Buddy Check algorithm, are different compared to the other variables. These maps for the assimilation of precipitation, compared to the maps of the included and excluded observations of the assimilation of the temperature and the sea-level-pressure, show more diversity in the locations that have been included and excluded. The maximum times an observation was included, though, is greater for the assimilation of the sea-level-pressure (744 times for the same location). The other variables show a maximum of number of times the same location was included of 665 times for the precipitation and 372 times for the temperature.

Another map that was produced to try and visualize what is happening with the Buddy Check are the maps of the included and excluded observations divided by decades. Since the assimilations of all variables have been executed for the years 1951-1981 6 maps have been created. For each decade a maps was produced (51-61, 61-71 and 71-81) and each map of decade has been divided in included observations and excluded observations.

The reason behind the creation of these maps is to see if there is a difference between the decades in the locations that are assimilated. The expectation is that the further back in time we go the less observations get assimilated.

Figure 44, Figure 45 and Figure 46 show the maps of the decades divided by the included and excluded maps. The maps don't show any difference in the locations of the excluded or included observations in the course of time.

Lastly, a list of the locations that were most excluded in the experiments of the Buddy Check is produced so as to see if there is a specific location that may be corrupt that gets excluded a lot. The summary of the five most excluded locations is given in Table 4, Table 5 and Table 6. These tables represent the excluded locations for the assimilation of temperature, precipitation and sea-level pressure using the Buddy Check algorithm.

Table 4: Summary of the excluded stations for the assimilation of temperature using the Buddy Check algorithm

n. of Excluded observations	lon	lat
230	-82.45	27.95
202	-123.88	46.16
199	-132.39	56.47
188	-79.08	38.17
187	-82.08	29.16

Table 5: Summary of the excluded stations for the assimilation of precipitation using the Buddy Check algorithm

n. of Excluded observations	lon	lat
284	24.13	65.82
279	-7.88	53.08
273	-8.9	52.7
256	171	-42.7
246	78	27.2

Table 6: Summary of the excluded stations for the assimilation of sea-level-pressure using the Buddy Check algorithm

n. of Excluded observations	lon	lat
273	-22.73	65.08
225	12.95	47.01
190	-6.7	54.4
186	7.6	47.6
181	18.1	59.4

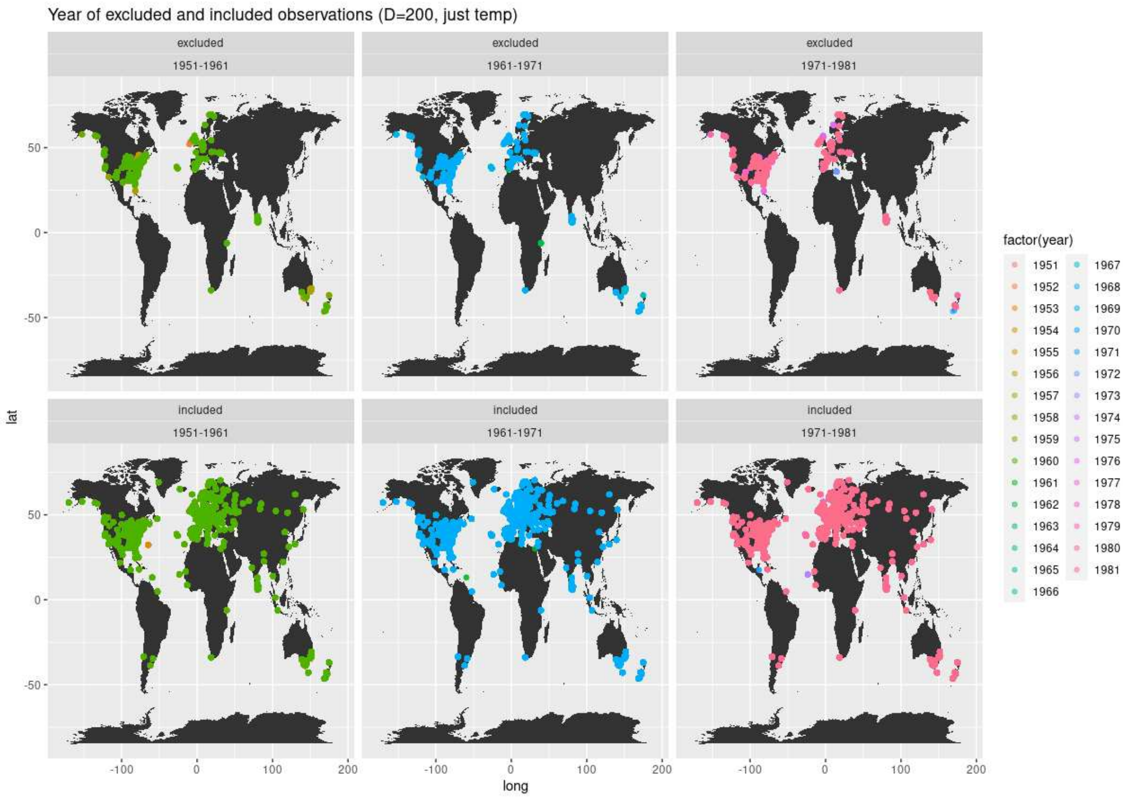

Figure 44: Location of excluded and included observations when assimilating the temperature with the Buddy Check algorithm to help flag suspect observations divided in decades. The columns represent the decade: the first column 1951-1960, the second 1961-1970 and the last column 1971-1981. The color of each point indicates how many times a location where an observation is available was taken into consideration or not. The map on the left indicates the how many times and which locations were not included in the assimilation (suspect observations). The map on the right, on the other hand, indicates how many times and which locations were included in the assimilation.

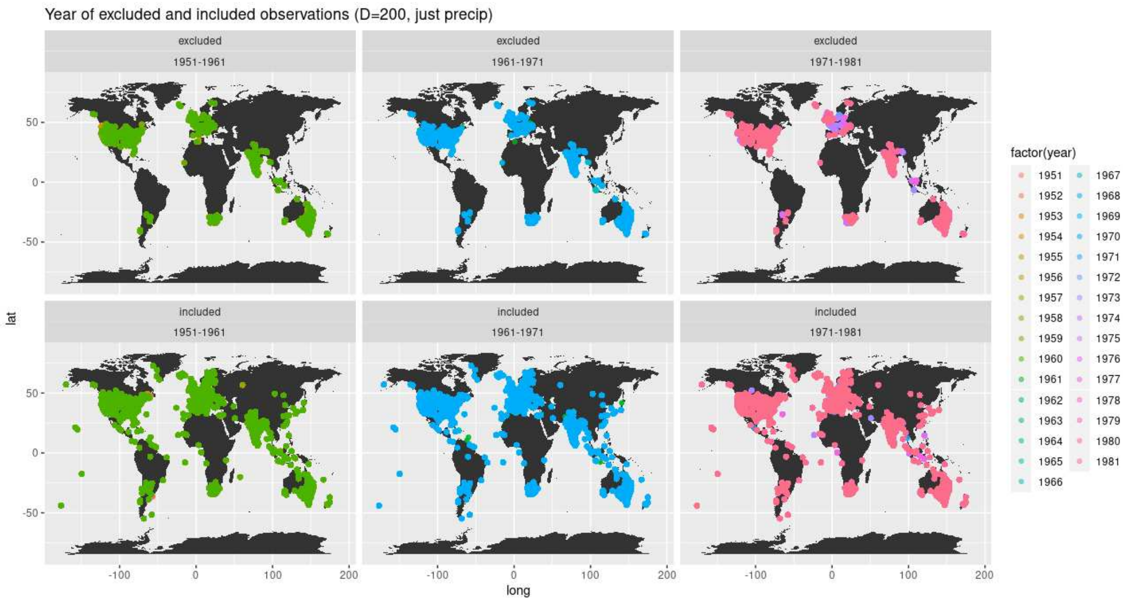

Figure 45: Location of excluded and included observations when assimilating the precipitation with the Buddy Check algorithm to help flag suspect observations divided in decades. The columns represent the decade: the first column 1951-1960, the second 1961-1970 and the last column 1971-1981. The color of each point indicates how many times a location where an observation is available was taken into consideration or not. The map on the left indicates the how many times and which locations were not included in the assimilation (suspect observations). The map on the right, on the other hand, indicates how many times and which locations were included in the assimilation.

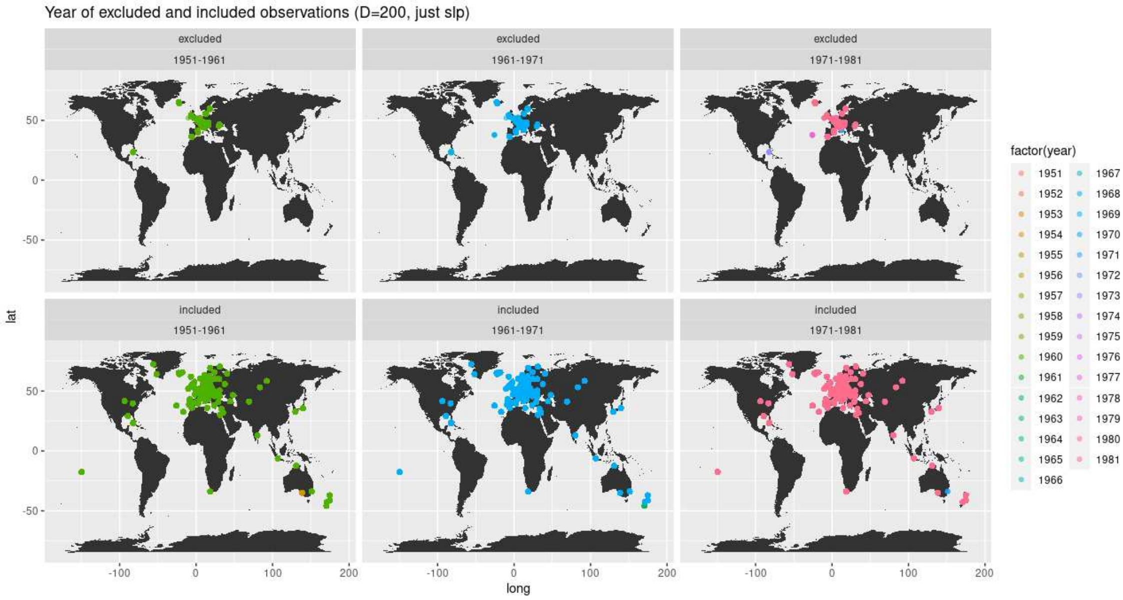

Figure 46: Location of excluded and included observations when assimilating the sea-level-pressure with the Buddy Check algorithm to help flag suspect observations divided in decades. The columns represent the decade: the first column 1951-1960, the second 1961-1970 and the last column 1971-1981. The color of each point indicates how many times a location where an observation is available was taken into consideration or not. The map on the left indicates the how many times and which locations were not included in the assimilation (suspect observations). The map on the right, on the other hand, indicates how many times and which locations were included in the assimilation.

4 Discussion and Conclusion

In the search for an optimal screening method to produce better analyses of the climate and the weather, I tested the ability of different data assimilation methods to reproduce the climate.

In the next sections a discussion on the results is presented in Chapter 3 and a conclusion will be drawn in Chapter 4.6.

4.1 Discussion

Valler, Brugnara, et al. 2020 used the EKF400 to make different tests on assimilating the precipitation and stated that it is a challenging task to assimilate precipitation of the past because of the uncertainty. Since the quality of the observation of the precipitation may be worse than that of the temperature it may be possible that we have positive results when assimilating the precipitation because the quality of the assimilation of all the observations available (original setup) of the EKF400 for the assimilation of the precipitation is lower compared to that of the assimilation of the temperature.

In the next section a discussion on the results of the three basic methods to assimilate data is given. A description of these tests is presented more precisely in the Methods section and in the Results sections.

4.2 Differences in skill scores

The results of the difference in the skill scores show that the variable which had the most differences between the skill scores of the tests and the original (assimilation of all observations) skill scores is the precipitation. The results of averaging the observations in the same grid box for the assimilation of the precipitation show that the range of the difference in correlation skill score has more negative differences than positive differences. Negative differences in the the correlation coefficient indicate that the assimilation of the precipitation using the average of the observations in the same grid box is lower than the correlation coefficient of the original setup for the precipitation. This suggest that, according to the correlation coefficient, the original setup for the assimilation of the precipitation has weaker correlation with the reference series.

When the differences in the correlation coefficient are plotted as maps, in the other hand, they reveal also small patches of positive differences in small regions of Central-South-Western Asia. The regions of negative difference still prevail over Central-South-Western Asia and Europe, but they are very small differences.

To have another measure to assess the averaging of the observations in one grid box we look at the maps of the $MSESS$. They indicate a majority of positive patches (Europe, USA and India). In this case positive differences signify that the mean squared error skill score is bigger in the new experiment than in the original setup (assimilation of all observations). This difference between the correlation coefficient and the $MSESS$ comes to place, as can be seen from the peaks in the time series, because the correlation coefficient ignores the biases and punishes errors in amplitude.

Just as in the maps for the $MSESS$, for the averaging of the observations in the same grid box when assimilating the precipitation, the other experiments for the precipitation (first, BC) also show positive differences in the $MSESS$. The maps for the difference in the $MSESS$ for assimilation of the precipitation using a random observation per grid box or the Buddy Check both show a majority of positive differences above all over Europe, the USA and India. Meaning that the mean squared skill score error when doing these assimilations is greater than in the original setup (assimilation of all the observations).

The attempt in reducing the signal-to-noise ratio for these experiments, which was the goal of averaging the data in the same grid box, seemed to work and produced an analysis which has more skill in the MSESS for precipitation than the original setup. Since, thanks to the time series, we have seen that the positive differences come to place because we had bias in some locations, the positive differences in these locations have to be neglected.

In addition, we have proven that the monthly precipitation data is not normally distributed. As a consequence, as stated in the Introduction (Chapter 1), the results are not optimal and negative differences in the MSESS could come to place because of this.

The difference in correlation coefficient of the precipitation for the random assimilation of an observation in the grid boxes shows a majority of negative differences in the United States, Europe, India and parts of Australia. There are also small patches of positive differences but the negative differences in correlation coefficients overwhelm the positive patches. There is a difference in the MSESS for the same experiment (assimilating a random observation per gridcell) just in the assimilation of precipitation. Here the difference is overwhelmingly positive, there are some negative regions but they are minimal.

Again, the attempt in improving the skill by using a random observation for every grid box that contains more than one observation has positive differences in the MSESS but this may come to place because of the bias.

In addition to the aforementioned discussion on the results, the assimilation of the precipitation using the average or a random observation per grid box show better results than the original setup probably because fewer observations are assimilated. With fewer observations the bias is also smaller implying better results in the skill scores.

The last basic test was the use of the Buddy Check to eliminated suspect observations from the assimilation. There were no changes in the correlation coefficient of the EKF400 when using the Buddy Check algorithm to eliminate suspect observations in respect to using the original setup. When looking at the differences in the MSESS, in the other hand, a positive difference can be seen when the precipitation is assimilated. Once more, the positive difference are in some of the investigated locations and could arise because the MSESS punishes the bias.

An important remark to make is that temperature is more homogeneous over larger regions than precipitation. An experiment with a larger radius for the "buddies" was used for this reason. The results showed positive skill just some regions over Asia. Over Europe and the United States of America the skill did not improve.

The use of the Buddy Check to try and produce an optimal data screening method for the EKF400 did not show particularly promising results by looking at the differences in the skill scores.

To further understand the behaviour of the Buddy Check and the other assimilations results were presented and they will be discussed in the next chapters.

4.3 Discussion on the comparison

Time series were created to understand the positive and negative results of the skill score in specific regions. In Chapter 4.2 it has already been commented that the correlation ignores the bias and the MSESS gives a lot of weight to outliers. The time series show that the regions where we have differences (positive and negative) in the MSESS are also regions where outliers are visible. With the help of histograms an important result could be discovered. The histograms of the precipitation in the locations investigated show that the precipitation distribution in right-skewed. This is a clear indication that the precipitation data is not normally distributed and thus the results show more negative anomalies and the EnKF will not show optimal results.

Box-plots also showed that the differences between all the data sets of the experiments are really small. There is a slight difference in the assimilation between the seasons, indeed the winter season has always a little higher skill than the summer season. Valler, Franke, Brugnara, et al. 2020 found, likewise, that the EKF400 performs better for the reconstruction of precipitation over Central and North America in the winter season (October-to-March). Other studies analysed by Bhend et al. 2012 also found that in winter the skill is higher than in summer. Bhend argues that this may be the case because data assimilation favors large-scale processes and large-scale circulation heavily influences the weather in the northern latitudes, where most of the observations are located. In summer, in the other hand, the weather is influenced by local processes and data assimilation is less successful. Other studies that show similar results are listed by Bhend et al. 2012: Rutherford et al. 2004 found optimal results reconstructions using tree rings in the cold season and Griesser et al. 2010 produced the best reconstructions for geopotential height in the Northern Hemisphere in the winter season.

Lastly, thanks to the maps of the included and excluded observations in the assimilation of the Buddy Check we observe that locations where the majority of the observations are assimilated are also the locations where we have negative or positive results in the differences of the skill scores.

When looking at the top five most excluded locations in all the experiments with the Buddy Check algorithm there isn't a location which is repeated between the experiments or a location which is excluded far more times than others. It can be noted, however, that the experiments when using the Buddy Check to assimilate the precipitation or the sea-level-pressure, exclude more observations than the experiments for the temperature. One reason behind this result could be the total number of observations available. If there are less temperature observations it could happen that we have less excluded observations as well as less included observations. The assimilation of the precipitation using the Buddy Check is the assimilation which has the most observations: 395923 included observations and 116787 excluded observations (total = 512710). The assimilation of temperature, in the other hand, has 166588 observations included and 22699 observations excluded (total = 139287). The assimilation of sea-level-pressure has even lower observations: 43755 observations included and 5993 observations excluded (total = 49748). Precipitation has more or less 3.6 times more observations than the temperature and far more than sea-level-pressure. The number of observations may also be the reason as to why the precipitation shows more differences in the experiments than the temperature (or the sea-level-pressure).

The reason behind this difference in quality between the use of the Buddy Check with the temperature or the precipitation may be that the quality of the assimilation of the temperature is already very high that the Buddy Check discards less suspect observations.

Thanks to the comparison of the anomalies in precipitation between the CRU data set and the data set GHCNd we found that the CRU data set may not be trusted for the very dry regions taken into consideration. This may have had an influence on negative results.

4.4 Duration of assimilation

We have seen in the previous sections that the results of the assimilation of the data sets with the Buddy Check algorithm do not improve the assimilation quality. The Buddy Check, as it was implemented in this research, will probably not be used because of this. Nonetheless it is interesting to see the how much time it took for the assimilation with the Buddy Check algorithm (or all the other data screening tests) to finish compared to how much time it took for the original assimilation to finish. Table 7 summarises the time it took for the assimilations to assimilate one year.

Table 7: Summary of approximate time it takes for these assimilations to assimilate one year

Description	Temperature	Precipitation	Sea-level-pressure
ALL	8 min	18 min	2 min
AVG	3 min	6 min	2 min
FIRST	11 min	9 min	2 min
BC	19 min	65 min	4 min

As predicted, the assimilations with the implementation of the Buddy Check algorithm take a longer time per year. The increase in time of computation is particularly interesting for the variable of the precipitation. For this variable the assimilation took more or less 3.6 times longer than the assimilation with the original setup.

4.5 Order of assimilation

Another important step in the understanding of how the EKF400 works, is to understand if the order of assimilation influences the results of the Buddy Check. The expected results are that the observations that get assimilated first are also the ones that are included more because the other observations have to agree with the first one to get assimilated.

Small clusters of observations in a 200 km radius are examined according to how soon they get assimilated and how often these observations are flagged as corrupt and thus not assimilated. The results indicate that the observations of temperature that get assimilated first may also be the ones that are included the most in the assimilation. This result is valid just for the cluster investigated for the assimilation of the temperature with the Buddy Check though. For the assimilation of the other variables with the Buddy Check the results are the opposite in some cluster and the same as in the assimilation of the temperature in other clusters.

The conclusion is that further research on the influence in the assimilation of the order of the assimilation of the observations when using the Buddy Check algorithm has to be done. We cannot draw any conclusions from the small set of clusters of observations investigated.

4.6 Conclusion

The title of this thesis, was well as the research question, is to find an optimal data screening method for a climate field reconstruction based on data assimilation.

Several experiments of data screening were assessed with different variables (temperature, precipitation and sea-level-pressure) using the correlation coefficient and the mean squared error skill score. The results suggest that, thanks to the skill scores used, most experiments didn't show a significant difference from the original setup. This is probably because the quality of the assimilation is already very good. Nevertheless when the precipitation was assimilated a difference could be seen, mostly in the difference of the $MSESS$.

The quality of the assimilations has improved, for some experiments, with these data screening methods. The assimilation of the precipitation using a random observation per grid box show a more negative correlation coefficient mostly over Europe and the United States, which is where the most observations are assimilated. The $MSESS$ for this setup show that the analysis has more skill than the original setup. As discussed before the skill is probably better in the assimilation of a random observation in a grid box and in the averaging of the observations in a grid box because less observations are assimilated in comparison to the original and thus less bias is introduced. The histograms also show that the data is not Gaussian. The consequence is unrealistic results for these experiments.

The difference in skill scores between the Buddy Check algorithm and the average result in patchy maps. There is no difference in the correlation coefficient over Europe and the United States, which is where the most observations are located. The difference in $MSESS$ show negative results over Europe and in parts of the United States. For these reasons we can say that the Buddy Check algorithm performs worse than the average of the grid boxes, at least in interesting areas for the reconstruction such as Europe and the United States.

4.7 Outlook

A good representation of the climate is essential in order to better understand the underlying processes of our climate. This good representation must use the best data assimilation method. The result might probably be a better assessment of the natural range of variation in extreme-event statistics. Additionally, it would provide an understanding on how variations in the ENSO and other climate variables affect those statistics.

Some of the analysis presented, which wants to cover data which extends back to many years in the past, have problematic results. An additional problem which we encountered, as stated before, is that the data of the precipitation is not normally distributed. The Kalman filter can be sensitive to assumptions (Brönnimann, Franke, et al. 2013). If a data screening, such as the one described by Dee, Rukhovets, et al. 2001 (Buddy Check), is going to be implemented in the EKF400 further analysis and tests must be done.

There are many possibilities to improve the analysis. Firstly, the assimilation of more than one variable would be an interesting experiment to do. The interaction of the variables or the inclusion of proxies could be analysed more in depth. Additionally, better analysis on other distances for the buddies must also be researched. This research focused only on a radius of 200 km with one experiment of a radius of 500 km for the temperature showing not very promising results. Finally, the order of the assimilation and its influence in the analysis has to be better understood to have the ability to better interpret the results.

5 Acknowledgements

I would like to express my special thanks to my supervisor Dr. Jörg Franke who helped me a lot and explained, with a lot of patience, many concepts, even more than once. I would also like to thank my family and my boyfriend who always supported me along my studies. I am glad I was given the opportunity to learn a lot on data assimilation and reconstruction methods.

6 References

Bhend, J. et al. (2012). “An ensemble-based approach to climate reconstructions”. In: *Clim. Past* 8, pp. 963–976. DOI: https://doi.org/10.5194/cp-8-963-2012.

Brönnimann, S., J. Franke, et al. (2013). “Transient state estimation in paleoclimatology using data assimilation”. In: *PAGES news* 21.2, pp. 74–75.

Brönnimann, S., C. Rohr, et al. (2020). “A Collection of Early Swiss Meteorological Series”. In: S. (Ed.) Swiss Early Instrumental Meteorological Series. DOI: https://doi.org/10.4480/GB2020.G96.01..

Collins, W. G. and L. S. Gandin (1990). “Comprehensive hydrostatic quality control at the National Meteorological Center”. In: *Mon. Weather Rev.* 118, pp. 2752–2767. DOI: https://doi.org/10.1175/1520-0493(1990)118<2752:CHQCAT>2.0.CO;2.

Compo, G. P. et al. (2011). “The Twentieth Century Reanalysis Project”. In: *Q. J. R. Meteorol. Soc.* 137, pp. 1–28. DOI: https://doi.org/10.1002/qj.776.

Dee, D. P., L. Rukhovets, et al. (2001). “An adaptive buddy check for observational quality control”. In: *Q. J. R. Meteorol. Soc.* 127, pp. 2451–2471. DOI: https://doi.org/10.1002/qj.49712757714.

Dee, D. P., S. M. Uppala, et al. (2011). “The ERA-Interim reanalysis: configuration and performance ofthe data assimilation system”. In: *Q. J. R. Meteorol. Soc* 137, pp. 553–597. DOI: https://doi.org/10.1002/qj.828.

Evensen, Geir (2003). “The Ensemble Kalman Filter: theoretical formulation and practical implementation”. In: *JOcean Dynamics* 53, pp. 343–367. DOI: https://doi.org/10.1007/s10236-003-0036-9.

Franke, J., S. Brönnimann, et al. (2017). “A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations”. In: *Scientific Data* 4, pp. 1–19. DOI: https://doi.org/10.1038/sdata.2017.76.

Franke, J., V. Valler, et al. (2020). “A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations”. In: *Clim. Past* 16, pp. 1061–1074. DOI: https://doi.org/10.5194/cp-16-1061-2020.

Griesser, T. et al. (2010). “Reconstruction of Global Monthly Upper-Level Temperature and Geopotential Height Fields Back to 1880”. In: *Journal of Climate* 23.11, pp. 5590–5609. DOI: https://doi.org/10.1175/2010JCLI3056.1.

Harris, I. et al. (2014). “Updated high-resolution grids of monthly climaticobservations – the CRU TS3.10 Dataset”. In: *International Journal Of Climatology* 34, pp. 623–642. DOI: https://doi.org/10.1002/joc.3711.

Hersbach, H. et al. (2020). “The ERA5 global reanalysis”. In: *Q. J. R. Meteorol. Soc.* 146, pp. 1999–2049. DOI: https://doi.org/10.1002/qj.3803.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Dover Publications, p. 376. ISBN: 9780080960906.

Jolliffe, T. and D. B. Stephenson (2011). *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*. 2nd ed. John Wiley Sons, Incorporated, p. 304. ISBN: 9781119960003.

Jones, P. D. et al. (2009). “High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects”. In: *The Holocene* 1.19, pp. 3–49. DOI: https://doi.org/10.1177/0959683608098952.

Kobayashi, S. et al. (2015). “The JRA-55 Reanalysis: General Specifications and Basic Characteristics”. In: *Journal of the Meteorological Society of Japan* 93.1, pp. 5–48. DOI: https://doi.org/10.2151/jmsj.2015-001.

Liu, Z. Q. and F. Rabier (2002). “The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study”. In: *Q. J. R. Meteorol. Soc* 128, pp. 1367–1386. DOI: https://doi.org/10.1256/003590002320373337.

Monmonier, M. (1999). *Air Apparent: How Meteorologists Learned to Map, Predict, and Dramatize Weather*. University of Chicago Press: Chicago. ISBN: 0226534227.

Murphy, A. H. (1988). “Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient”. In: *Monthly Weather Review* 116.12, pp. 2417–2424. DOI: https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Osborn, T. J. and P. D. Jones (2014). “The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google Earth”. In: *Earth Syst. Sci. Data* 6, pp. 61–68. DOI: https://doi.org/10.5194/essd-6-61-2014.

Rutherford, S. et al. (2004). “Proxy-Based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Method, Predictor Network, Target Season, and Target Domain”. In: *Journal of Climate* 8.13, pp. 2308–2329. DOI: https://doi.org/10.1175/JCLI3351.1.

Slivinski, L. C. et al. (2021). “An evaluation of the performance of the 20th Century Reanalysis version 3”. In: *Journal of Climate* 34.4, pp. 1417–1438. DOI: https://doi.org/10.1175/JCLI-D-20-0505.1.

Tingley, M. P. et al. (2012). “Piecing together the past: statistical insights into paleoclimatic reconstructions”. In: *Quaternary Science Reviews* 35, pp. 1–22. DOI: https://doi.org/10.1016/j.quascirev.2012.01.012.

Valler, V., Y. Brugnara, et al. (2020). “Assimilating monthly precipitation data in a paleoclimate data assimilation framework”. In: *Clim. Past* 16.4, pp. 1309–1323. DOI: https://doi.org/10.5194/cp-2019-137.

Valler, V., J. Franke, and S. Brönnimann (2019). “Impact of different estimations of the backgrounderror covariance matrix on climate recontructions based on data assimilation”. In: *Clim. Past* 15, pp. 1427–1441. DOI: https://doi.org/10.519/cp-15-1427-2019.

Valler, V., J. Franke, Y. Brugnara, et al. (2020). “An updated global atmospheric paleo reanalysis covering the last 400 years”. In: *Geoscience Data Journal* X, pp. XX–XX. DOI: https://doi.org/10.1002/gdj3.121.

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International geophysics series. Academic Press, p. 704. ISBN: 9780123850225.

WMO (accessed: 09.10.2021). *Climate explorer*. URL: https://climexp.knmi.nl/start.cgi.

Declaration of consent

on the basis of Article 30 of the RSL Phil.-nat. 18

Name/First Name: Sabrina Thaler Minke

Registration Number: 2012-056-883

Study program: Climate Sciences, Climate and Earth System Science

Bachelor ☐ Master ☑ Dissertation ☐

Title of the thesis: Development of an optimal data screening method for a climate field reconstruction based on data assimilation

Supervisor: Dr. Jörg Franke
Co-Supervisor: Prof. Dr. Stefan Brönnimann

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of 5 September, 1996 is authorized to revoke the title awarded on the basis of this thesis.
For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with future theses submitted by others.

Basel, 12.10.2021

Place/Date

Signature