Interpolation of temperature data for improved weather forecasts

FRIDA CRONQVIST
Interpolation of temperature data for improved weather forecasts

FRIDA CRONQVIST

Degree Projects in Optimization and Systems Theory (30 ECTS credits)
Degree Programme in Applied and Computational Mathematics (120 credits)
KTH Royal Institute of Technology year 2018
Supervisors at SMHI: Mikael Magnusson
Supervisor at KTH: Johan Karlsson
Examiner at KTH: Johan Karlsson
Abstract

To create weather forecasts at the Swedish Meteorological and Hydrological Institute (SMHI), it is needed to interpolate the temperature from the observation stations into a grid with 2.5 km between each point. Since the observed temperatures are known and without error, they should be held fix even after the interpolation. Usually, there is a strong relationship between temperature and elevation which needs to be accounted for in the interpolation method. In this study, regression kriging was investigated and compared to inverse distance weighted interpolation (IDW). Different choices to be made within regression kriging were investigated to optimize the method and the results were evaluated using leave-one-out cross validation for many different sets of data. The result was that regression kriging always had a smaller mean square error than IDW, as long as there were no violation of the required assumptions. If there were, regression kriging can lead to large errors and can therefore not be used. To avoid violating the assumptions, the regression part of regression kriging needs to be accurate enough. This might require more information than only the latitude and longitude coordinates and the elevation, which is what was known in this study.
Sammanfattning

Interpolation av temperaturdata för förbättrade väderprognoser

För att producera väderprognoser på Sveriges Meteorologiska och Hydrologiska Institut (SMHI) krävs det att temperaturerna från mätstationerna interpoleras till ett rutnät med 2.5 km mellan varje punkt. Eftersom de observerade temperaturerna är kända och utan fel ska de hållas fixa även efter interpolationen. Vanligtvis finns det ett starkt samband mellan temperatur och höjd över havet som också ska tas i beaktande i interpolationen. I den här studien undersöktes interpolationsmetoden regressionskriging och jämfördes med inverse distance weighted interpolation (IDW). Olika valmöjligheter inom regressionskrigingen undersöktes för att optimera metoden och resultaten från många olika datamängder evaluerades med hjälp av korsvalidering. Resultatet blev att regressionskriging alltid gav ett mindre genomsnittligt kvadratfel än IDW, så länge antagandena som krävs för metoden var uppfyllda. År de inte det kan regressionskriging leda till stora fel och kan därför inte användas. För att uppfylla dessa antaganden måste regressionsdelen i regressionskriging vara tillräckligt precis, vilket kan kräva mer information än enbart latitud- och longitudkoordinater samt höjddata, vilket var den enda kända informationen i denna studie.
Acknowledgements

I would like to start by thanking my supervisor at SMHI, Mikael Magnusson, for all the support and the interesting discussions we have had during the work with this thesis. Also a special thanks to Andreas Carlsson for explaining the meteorological background behind the temperature behaviour, about the interpolation method used today by SMHI and for helping me reading and writing data in the correct format to be able to compare my interpolation method to the one used today. I really appreciate that you have taken the time!

Finally I would like to thank my supervisor at KTH, Johan Karlsson, for very useful discussions and comments about my report during the entire process. Your input has been extremely helpful!
## Contents

1 **Introduction** .......................................................... 6  
   1.1 Problem description ................................................. 6  
   1.2 The interpolation method used today .......................... 6  
   1.3 Data and study area ................................................ 7  
   1.4 Difficulties .......................................................... 8  
   1.5 Choice of methodology ............................................. 9  
   1.6 Outline .............................................................. 9  

2 **Meteorological theory** .............................................. 10  
   2.1 Understanding temperature ....................................... 10

3 **Mathematical theory** .................................................. 13  
   3.1 Inverse distance weighted interpolation ........................ 13  
   3.2 Kriging .............................................................. 13  
       3.2.1 Simple kriging ............................................... 14  
       3.2.2 Regression kriging ......................................... 15  
       3.2.3 Semivariogram ............................................... 17  
       3.2.4 Binning ...................................................... 20  
       3.2.5 Model fit ..................................................... 22  
       3.2.6 Weighted least squares for model fitting ............... 25  
       3.2.7 Weighted least squares for estimating the regression coefficients .................................................. 27  
   3.3 Combined methods .................................................. 27

4 **Method** ............................................................... 28  
   4.1 Kriging .............................................................. 28  
       4.1.1 Linear regression analysis ................................... 30  
   4.2 Inverse distance weighted interpolation ....................... 31  
   4.3 Combined methods .................................................. 31  
   4.4 Leave-one-out cross-validation ................................... 31  
   4.5 Data sets used for evaluation ..................................... 32  
   4.6 Different parts of Sweden ......................................... 33

5 **Results** ............................................................... 35  
   5.1 Kriging .............................................................. 35  
       5.1.1 Different parts of Sweden .................................... 36  
       5.1.2 Variogram model .............................................. 39  
       5.1.3 Kriging neighbourhood ....................................... 41  
       5.1.4 Auxiliary variables .......................................... 42  
       5.1.5 Regression coefficients ..................................... 42  
       5.1.6 Weight formula ............................................... 43  
       5.1.7 Tests using all available observation stations ........... 44  
   5.2 Inverse distance weighted interpolation ........................ 49  
   5.3 Combined methods .................................................. 50
6 Discussion

6.1 Investigation of the data ............................................. 51
6.2 The regression part .................................................. 52
6.3 Instability of regression kriging .................................. 54
6.4 Comparison between IDW and regression kriging .......... 55
6.5 Tests using all available observation stations ............... 56
  6.5.1 Data from the 26th of March 2018 ......................... 56
  6.5.2 Data from the 20th of April 2018 ......................... 57
6.6 Methods for estimating the covariance matrix ............... 58
6.7 Semivariogram ....................................................... 59
6.8 Limitations and further work ................................... 60

7 References ................................................................. 62

Appendix A Figures ........................................................ 65

Appendix B Generalized least squares ............................... 86

Appendix C Distance between points on earth ................... 87

Appendix D Derivation of the bearing angle ....................... 89
1 Introduction

Humans have been trying to predict the weather for millennia and the thermometer and the barometer were invented already in the 17th century. In Sweden, the oldest official observation journals are from 1722 [1]. The Swedish Meteorological and Hydrological Institute (SMHI) was formed in 1919 [2] but has ancestors from 1873. Since then, the produced weather forecasts has undergone major improvements and today the improvements continue on a daily basis. Weather forecasts are used not only by the general population, but also within for example the transportation sector, agriculture and the national defence. Local forecasts are even used to control the heating in some buildings.

To create weather forecasts today, the data has to be interpolated from the observation stations to a tight grid. One step in the process of improving the forecasts is to create a more accurate interpolation. The aim of this thesis is to investigate the interpolation opportunities to try to find a more fitting interpolation method.

1.1 Problem description

To produce weather forecasts at SMHI, temperature data obtained from observation stations needs to be interpolated to a grid with 2.5 km between each point. This grid is then used in a meteorological model to create forecasts. Today the interpolation method used is an approximative method and the temperatures in the points of observation are not held fix. The method assumes that there are some errors in the observations and smooths the interpolated surface by allowing the value to differ slightly from the actual observed value. The method used today also does not take into account differences in elevation when interpolating. More details about this method is presented in Section 1.2. Since there usually is a relationship between temperature and elevation, including elevation in the method should significantly improve the interpolation.

The purpose of this study is to investigate and evaluate spatial interpolation methods, taking elevation into account and also to hold the temperature in the observation points fix. The optimization problem becomes a problem in three dimensions where both the horizontal and the vertical distance between the observation stations and the interpolation point are to be considered. Two methods were implemented and investigated. Those were inverse distance weighted interpolation (IDW) and regression kriging. Regression kriging is an advanced method with many opportunities and to be able to fairly evaluate its full potential, a careful investigation was performed where the different opportunities was varied one at a time. The theory about the methods is presented in Section 3.

1.2 The interpolation method used today

Today, the method used is a version of bilinear interpolation. An interpolated "background field", consisting of second-order curves connecting the observations and the target point, that is okay but not good enough is used from the start and then corrected depending on Kalman filtered observations. If, for example the background field has underestimated the temperature, it is corrected upwards in that point. As mentioned above the method is approximative, meaning that it assumes there are some errors in the observation points and the value is allowed to differ slightly from the actual observed value, which is not desired. The other problem noted with this method is that the correction causes too steep gradients which leads to over-/undershooting and too varying
temperatures between close locations. Especially hot spring days cause problems since the water has a much cooler temperature than the land surfaces and the method cannot capture the steep temperature gradients close to the coasts.

On the other hand, in some weather conditions the method used today works well. Typical such examples are cloudy days when the temperature is evenly distributed over the whole area and there are no steep gradients. An example of the temperature interpolated with the method used today is showed in Figure 1. Here, the data is from the 26th of March and the method works quite well.

Figure 1: The temperature interpolated with the method used today. The red lines are temperature lines and the colour also represents the temperature in the region.

1.3 Data and study area

The area of interest to which the interpolation method is to be performed is the whole of Sweden. The landscape in Sweden is very varying, including mountains, coasts, inland and lakes and the temperature is varying in both time and space.

Temperature data is available from many years back in time and also with several readings per day. The number of observation stations in Sweden measuring temperature is about 200 for most of the years. At most, about 400 stations were active around year 1950. Especially in the mountainous part of Sweden, the number of observation stations
is sparse. For practical reasons, the stations are usually located in the valleys. For the Swedish observation stations, accurate elevation measures in metres above the sea level are available. The stations available in Sweden today is showed in Figure 2. Here the colours represent the temperature the 10th of February 2018. The coordinates for the observation stations are given in latitude and longitude and transformed by basically ”moving the equator” to the middle of Sweden. This is to make the coordinate system as rectangular as possible.

Figure 2: Available observation stations in Sweden today. The temperature at each station is showed in a colour scale and measured 10th February.

Temperature data from observation stations in Norway, Denmark, Finland and the Baltic states are also available but here, the elevation needs to be estimated from a map which makes it less accurate than for the Swedish observation stations. For this reason, and because it is easier obtained, data from Swedish stations were used in the process of developing the interpolation method. When testing and comparing the developed interpolation method to the one used today, all observation stations from the countries named above were used.

1.4 Difficulties

The temperature is a very complex variable that does not always behaves linearly. The relationship between temperature and for example elevation can also be different in different parts of Sweden. This is due to factors such as latitude, inversion, wind direction and wind speed, the impact from the sea and special types of micro climate that affect the temperature. This is explained in more detail in Section 2.1. In this study, only the elevation and location of the points are known. In some cases this might not be enough to explain the temperature’s behaviour.

Another difficulty is that the density of observation stations is sparse, especially in the mountainous region, and the stations are mostly located in the valleys. This makes it
hard (if not impossible) to accurately model the temperature during weather conditions when it varies nonlinearly with elevation and behaves differently in different locations.

What makes it hard to find the most suitable interpolation method is that the temperature behaviour also changes over time. What is found to be optimal for one specific day and time may not be the best choice for another. It is impossible to minimize the errors for every time, day, month and year since this would require an infinite amount of testing. What is possible within the scope of this study is to run cross-validation tests for different times of the year to try to capture as many different weather situations as possible. It is left to further studies to investigate if it would be more appropriate to use different models or methods for different situations and perhaps for different times of the year.

1.5 Choice of methodology

In the beginning of the study, many reports were read to find conclusions regarding which methods worked best for spatial interpolation in similar studies. The performance of the interpolation methods differed depending on the area of interest and the spatial variable considered, but in most of the studies regression kriging or some other kriging method was the one showing the best performance [36][17][21][12][32][20][25]. Therefore regression kriging was chosen to be deeply investigated in this study. Kriging is a common method for spatial interpolation and the mathematical theory behind it is presented in Section 3.2.2. The process of how regression kriging was implemented and used is explained in Section 4.1. Since regression kriging is a complex method and to have time to fairly evaluate all its possibilities, this was the only advanced method included. The alternative would have been to investigate several interpolation methods briefly, but the result of such a study would not be reliable since the methods would not be optimized for this particular purpose. The simpler method inverse distance weighted interpolation (IDW) was also selected to be used as a comparison to regression kriging and the theory behind it is presented in Section 3.1.

1.6 Outline

The structure of the thesis is as follows: In Section 2 the meteorological theory is explained which will later be used to analyse the results and help to understand the temperature behaviour in different weather situations. Section 3 presents the mathematical theory behind the chosen interpolation methods regression kriging and inverse distance weighted interpolation. In the end of the section, the theory behind a combination of these two methods is given. The methodology used in this study is presented in Section 4. First, there are some general explanations about the process of the study which is followed by detailed information about how the methods were implemented and investigated. In the end of the section, it is explained how the results were evaluated. The results then follows in Section 5. Since there were many different choices to be made to optimize the regression kriging method, tests were run where one such choice was varied at a time. All the results from these tests are presented in Section 5.1. The results from inverse distance weighted interpolation and a combination of the methods are shown in Section 5.2 and 5.3 respectively. Section 6 contains a deep analysis of the results and also includes a discussion about the chosen method and the choices to be made within regression kriging. In the end of the section, limitations and further work are discussed.
2 Meteorological theory

To be able to analyse the results, some basic meteorological understanding is required. The temperature is a complex variable that does not always behave linearly. For that reason, some theory about the temperature is presented in this section.

2.1 Understanding temperature

There are several factors that affect the temperature at a specific location and a specific time. Therefore it is hard to mathematically model the temperature only by considering the available numbers. Some meteorological knowledge is required to understand the complex behaviour of the temperature and to understand why the mathematical model gives larger errors at some locations and in some situations than in others. The factors that cause temperature variations are called controls of temperature \[22\]. One important such factor is the difference in incoming solar radiation. Since the sun angle and the length of the days depend on latitude, the temperature is usually lower in polar regions. Large temperature variations during a day can also be caused by the variations in the sun angle. High latitudes experience less variation in sun angle and will therefore have less temperature variation during the day. The incoming rays from the sun at a specific location varies, of course, also with the season. Latitude is maybe the most intuitive control of temperature, but there are many more.

Another control of temperature is different heating of land and water. Depending on the surface of the earth, a varying amount of solar energy is absorbed and reflected. This causes the temperature in the air above to vary as well. The difference is largest between land and water surfaces since land heats and cools much faster than water. This causes the temperature above land surfaces to vary more than the temperature over water. One reason why it takes more time for the temperature to rise and fall in water is because water is highly mobile. Turbulence in the water moves the heat through a larger mass of water and thus, a greater volume of water needs to be heated for the temperature to rise. Water is also more transparent than land and therefore allows the solar radiation to penetrate to a greater depth than in a land surface. On land surfaces, the heat instead remains in the top layer. Hence, a thick layer of water needs to be heated to affect the temperature, but on land, only a thinner layer is heated which leads to a much higher temperature. This process is also true for cooling; water cools much slower than land surfaces. The cooled surface water cools and sinks, giving place to warmer water rising from below. Therefore a large amount of water needs to be cooled for the temperature at the surface to drop significantly. There are also other contributing factors to the slow heating and cooling of water compared to land. One is that the specific heat (i.e. the amount of heat needed to increase the temperature of 1 gram water by 1 degree Celsius) is three times greater for water than for land. Therefore water requires more heat to increase the temperature compared to land. Another factor is the evaporation, which is a kind of vaporization that occurs at the surface of a body as the liquid changes to gaseous phase, is greater from water surfaces than from land surfaces.

With this in mind, it is reasonable to conclude that locations close to water typically have more evenly distributed temperature during the year, i.e. warmer winters and colder summers, compared to locations far inland. Some geographical locations are also more affected than others by the proximity to water. Such locations are for example windward coastal locations that will experience a more evenly distributed temperature over the year.
than leeward coastal locations.

The control of temperature mostly considered in this report is the elevation. In average the temperature drops by 6.5 degrees per kilometre in the atmospheric layer closest to earth. This is due to the decrease in atmospheric pressure which allows the air to expand. The expanding process requires energy which causes the temperature to drop. But the temperature-elevation relationship is not as simple as that. For example the absorption and re-radiation of solar energy by the ground surface can make it warmer than expected at higher elevations. Also the lower density at higher elevations makes the atmosphere absorb and reflect less solar radiation and that makes the intensity of insolation increase with increasing elevation. As a consequence, the temperature rises faster during the day and also decreases faster during the night.

The phenomenon where temperature increases with elevation in a defined layer in the atmosphere is called inversion. This phenomenon can be caused by the wind. If warm air is coming in perpendicular to a valley, it will not mix with the colder air down in the valley. Cold air have a higher density than warm air and the temperature will then increase with elevation up to approximately the top of the surrounding mountains, and then start to decrease again. In some cases it can be several such layers where the temperature increases and decreases more than one time with increasing elevation. If the wind is parallel to the valley, the relationship between elevation and temperature will typically be more linear. In Sweden the valleys are in general located in northwest-southeast direction and thus, the temperature-elevation relationship in Sweden is usually more complicated when the wind is coming from northeast or southwest.

Yet another factor that have a large impact on the temperature is the extent of cloud cover. Clouds have a high albedo and therefore reflects back much of the sunlight to space. That is why we experience darker days when the cloud cover is thick. A decrease in incoming solar radiation decreases the temperature during the day. At night, on the other hand, the clouds absorb the outgoing terrestrial radiation and re-radiates it back to earth. As a consequence, the nights are warmer if the cloud cover is thick. Thus, a thick cloud cover makes the temperature less varying during the day and night. But also surfaces covered by snow and ice have high albedos. That means that the temperature a winter day will be lower if the ground is covered by snow than if it is free ground that could absorb the heat and make the temperature rise.

Typically the lowest temperature is found about sunrise and then the temperature rises until about 2-5 p.m. After that it decreases again until the next sunrise. It can be surprising that the temperature peaks in the afternoon because the intensity of sunlight reaches its maximum about noon. The reason why the temperature continues to rise during the afternoon is because of the outgoing terrestrial energy from earth. When there is no incoming solar energy, the atmosphere and the surface of earth cool since the heat radiates away while no solar energy is replacing the lost heat. After sunrise the incoming solar energy increases again causing the minimum temperature to occur just before sunrise. The delay of the maximum temperature compared to the maximal incoming solar radiation is called the lag of the maximum. After the peak in insolation about noon, the intensity of incoming solar radiation still exceeds the outgoing radiation from earth’s surface, causing the temperature to continue to rise for a few hours. The temperature starts to fall when the rate of terrestrial energy loss exceeds the input of solar energy. It is also the case that the rate at which the earth radiates heat to the atmosphere is not the same as the rate at which the atmosphere radiates heat to space. In general, for a few hours after the peak in insolation, more heat is given to the atmosphere from
the surface of earth than what is emitted by the atmosphere to space. This causes the increase in temperature until the afternoon.

The length of the lag of the maximum depends on the location and the climate. On a clear day in a dry region the absorbed amount of solar energy will generally be high which makes the maximum temperature occur late. A shorter lag time will be found in humid regions.

All these controls of temperature cooperates and makes the behaviour of the temperature very complex. With that in mind, it is understandable that the accuracy in a mathematical model trying to capture behaviour of the temperature will vary depending on the weather conditions.
3 Mathematical theory

The main focus in this study is to investigate interpolation methods to improve the method used today for interpolation of temperature data. Two methods were implemented and tested, inverse distance interpolation (IDW) and regression kriging. A combination of the two methods was also tested. IDW is a much simpler method than regression kriging and is mostly used as a comparison to the advanced regression kriging. To evaluate the results, leave-one-out cross validation was used. In this section, the theory behind IDW, regression kriging and a combination of the methods is presented.

3.1 Inverse distance weighted interpolation

The idea behind inverse distance weighted interpolation (IDW) is that the observation points close to the target point are considered more important and are thus given larger influence on the target value than those far away. The interpolating function is

\[
\hat{z}(s_0) = \begin{cases} 
\frac{\sum_{i=1}^{N} w(s_i)z(s_i)}{\sum_{i=1}^{N} w(s_i)}, & \text{if } d(s_0, s_i) \neq 0 \text{ for all } i \\
z(s_i), & \text{if } d(s_0, s_i) = 0 \text{ for some } i 
\end{cases}
\]  

(1)

where

\[
w(s_i) = \frac{1}{d(s_0, s_i)^p}
\]

(2)

is the definition of the weighting function defined by Shepard’s original formula \[^{33}\], \(s_0\) is an unknown point to be interpolated, \(s_i\) for \(i = 1, ..., N\) are the known observation points, \(d(s_0, s_i)\) is the distance between \(s_0\) and \(s_i\) and \(p\) is the so-called power parameter which is a positive real number to be chosen by the user.

IDW is an exact interpolation method, i.e. values in known points remain the same after interpolation. This is of course preferable if it is known that there are no errors in the observations. Also, in IDW, the maximum and minimum values can only occur in the observation points and it is not possible to extrapolate outside the range of the observed values. This method does not take the elevation into consideration and is in this study mostly used as a comparison to be able to determine the performance of the more advanced method implemented.

3.2 Kriging

Kriging, named after the mining engineer Krige from South Africa \[^{34}\] who introduced the procedure, is a common method for interpolating spatial data, i.e. data that can be mapped in a geographic coordinate system. Kriging applies the Best Linear Unbiased Estimation (BLUE) \[^{17}\]. It is a minimization problem where the constraint is that the method must be unbiased, and the mean squared error of the residuals is minimized (best estimate).

There are several variants of kriging that have different characteristics. In this study, regression kriging was the method used. Regression kriging consists of two parts, one regression part and one simple kriging part. To perform regression kriging, the first thing to do is to compute the regression coefficients in the regression part. Usually this is done by using generalized least squares (GLS). GLS requires a known covariance matrix which is estimated from a so-called semivariogram, which is introduced in Section 3.2.2 and...
explained in detail in Section 3.2.3. The residuals from the regression part together with the estimated covariance structure is then used in the simple kriging part to compute an interpolated residual in the target point. The final interpolated value is then obtained by adding the two parts together. Thus, the interpolated surface is a function of the data from the closest observations, but under the constraint that the data satisfies some chosen model for the spatial variability, the semivariogram. The details are explained in the following sections.

3.2.1 Simple kriging

Simple kriging is an interpolation method where the idea is to give a weight to every observed point and then compute the target value from the mean plus a linear combination of the weights times the difference between the mean and the value in the observation points \[26\], i.e.

\[
\hat{z}(s_0) = \mu + \sum_{i=1}^{N} w(s_i)(z(s_i) - \mu)
\]  \hspace{1cm} (3)

where \(z(s_i)\) for \(i = 1, ..., N\) are observed random variables at location \(s_i\), \(w(s_i)\) are weights to be computed and \(\mu\) is the mean which is assumed to be known and stationary, \(E[z(s_i)] = \mu\) for all \(i\).

The goal is to determine the weights that minimize the variance of the estimator

\[
\text{Var}(\hat{z}(s_0) - z(s_0)) = \begin{bmatrix} 1 & -w^T \end{bmatrix} \hat{C} \begin{bmatrix} 1 \\ -w \end{bmatrix} = \begin{bmatrix} C(0) & b \\ b^T & C \end{bmatrix} \begin{bmatrix} 1 \\ -w \end{bmatrix} = (4)
\]

\[
= C(0) + w^T C w - 2b^T w
\]

where \(C(h)\) is the covariance matrix that depends on the distance vector \(h\) and \(b\) is the vector of covariances between \(z(s_0)\) and \(\begin{bmatrix} z(s_1) & z(s_2) & \ldots & z(s_N) \end{bmatrix}\). The covariances are assumed to be translation-invariant, i.e.

\[
C(h) = C(s_i - s_j) = \text{Cov}(z(s_i), z(s_j)), \text{ for all } s_i, s_j \text{ with } s_i - s_j = h.
\]

This, together with the mentioned assumption that \(E[z(s_i)] = \mu\) for all \(i\), is called the assumption of second order stationarity.

The constraint in the minimization problem is that the kriging method must be unbiased, which is automatically satisfied in simple kriging since

\[
E[\hat{z}(s_0) - E[z(s_0)]] = E[\hat{z}(s_0) - \mu] = \mu + \sum_{i=1}^{N} w(s_i)(z(s_i) - \mu) - \mu = \sum_{i=1}^{N} w(s_i)E[z(s_i)] - \mu = \sum_{i=1}^{N} w(s_i)(E[z(s_i)] - \mu) = 0.
\]  \hspace{1cm} (5)

Hence, the minimization problem is

\[
\min_w C(0) + w^T C w - 2b^T w
\]  \hspace{1cm} (6)
without any constraint. Solving this by for example taking the partial derivatives with respect to the weights and setting them to zero gives the solution

$$
\begin{bmatrix}
w(s_1) \\
\vdots \\
w(s_N)
\end{bmatrix} = \begin{bmatrix}
C(s_1 - s_1) & \ldots & C(s_1 - s_N) \\
\vdots & \ddots & \vdots \\
C(s_N - s_1) & \ldots & C(s_N - s_N)
\end{bmatrix}^{-1} \begin{bmatrix}
C(s_0 - s_1) \\
\vdots \\
C(s_0 - s_N)
\end{bmatrix}.
$$

(7)

One advantage with the kriging method is that it is an exact interpolation method. By exact we mean that the estimation/prediction in the point is the same as the observed value in that point, i.e. \( \hat{z}(s_0) = z(s_i) \) if \( s_0 = s_i \). This holds since the vector \( c = [C(s_0 - s_1) \ldots C(s_0 - s_N)]^T \) equals column \( i \) in the matrix of covariances above, i.e. \( [C(s_1 - s_i) \ldots C(s_N - s_i)]^T \). From this it follows that the weight vector \( [w(s_1), \ldots, w(s_i) \ldots w(s_N)]^T = [0, \ldots, 1, \ldots, 0]^T \), i.e. \( w(s_i) = 1 \) and \( w(s_j) = 0 \) for \( j \in \{1, \ldots, N\}, j \neq i \) is a (unique) solution. Hence, \( \hat{z}(s_0) = \mu + \sum_{i=1}^{N} w(s_i)(z(s_i) - \mu) = z(s_i) \) if \( s_0 = s_i \).

### 3.2.2 Regression kriging

Regression kriging is a method that combines simple kriging with mean zero and linear regression. The regression part is used to model the mean function, which is assumed to depend on the location \( s_0 \), and then simple kriging is applied to the residuals from the regression part.

The target value is thus modelled as

$$
z(s_0) = \mu(s_0) + e(s_0).
$$

(8)

where \( \mu(s_0) \) is the deterministic mean function to be estimated by linear regression and \( e(s_0) \) is the error function which is estimated by simple kriging as mentioned above. The regression part is made by creating a model of the relationship between the target value and the auxiliary variable(s) at the observed points. If using multiple linear regression, the model is [15]

$$
\hat{\mu}(s_0) = \sum_{j=0}^{p} \hat{\beta}_j \cdot q_j(s_0), \quad q_0(s_0) \equiv 1
$$

(9)

where \( q_j(s_0), \) for \( j = 1, \ldots, p \) are the values of the \( p \) auxiliary variable(s) at the target point \( s_0 \) and \( \hat{\beta}_j \) are the estimated regression coefficients. The estimated intercept is \( \hat{\beta}_0 \).

Now, using regression kriging we combine the simple kriging described above with mean zero and the linear multiple regression. This gives the estimated target value [7]

$$
\hat{z}(s_0) = \hat{\mu}(s_0) + \hat{e}(s_0) = \sum_{j=0}^{p} \hat{\beta}_j \cdot q_j(s_0) + \sum_{i=0}^{N} w(s_i) \cdot e(s_i)
$$

(10)

where \( \hat{\mu}(s_0) \) is the fitted deterministic part (the trend component) and \( \hat{e}(s_0) \) is the interpolated residual describing the small-scale variability.

The estimated regression coefficients, \( \hat{\beta}_j \), are commonly computed using generalized least squares (GLS) [15] and are then defined as

$$
\hat{\beta}_{GLS} = (X^T C^{-1} X)^{-1} X^T C^{-1} z
$$

(11)
where $C$ is the covariance matrix of the residuals from the regression part $e(s) = z(s) - \hat{\mu}(s)$,

$$C = \begin{bmatrix}
C(e(s_1), e(s_1)) & \cdots & C(e(s_1), e(s_N)) \\
\vdots & \ddots & \vdots \\
C(e(s_N), e(s_1)) & \cdots & C(e(s_N), e(s_N))
\end{bmatrix}. \quad (12)$$

The matrix $X$ has dimension $(N \times p + 1)$ and contains the auxiliary values and a first column of ones

$$X = \begin{bmatrix}
1 & q_1(s_1) & \cdots & q_p(s_1) \\
1 & q_1(s_2) & \cdots & q_p(s_2) \\
\vdots & \vdots & & \vdots \\
1 & q_1(s_N) & \cdots & q_p(s_N)
\end{bmatrix}. \quad (13)$$

and $z$ is the vector of observed values of the target variable. To compute $\hat{\beta}_{GLS}$, an iterative process is performed. Usually, more than one iteration is not necessary because the difference after several iterations is small and therefore will not affect the final prediction. The iteration process is as follows:

The first step is to derive a linear model of the target value as a function of the auxiliary variable(s). Using ordinary least squares (OLS) this is done by solving

$$z = X\beta_{OLS} + e \quad (14)$$

for $\beta_{OLS}$ \[9\], which leads to

$$\hat{\beta}_{OLS} = (X^TX)^{-1}X^Tz. \quad (15)$$

When the OLS regression coefficients are calculated, the next step is to derive the OLS residuals at all observation points $s_i$ for $i = 1, \ldots, N$. This is done by calculating

$$e_{OLS}(s) = z(s) - X\hat{\beta}_{OLS}. \quad (16)$$

Now we want to compute the covariance matrix for the OLS residuals that is needed for computing the GLS coefficients. To estimate the covariances, usually the semivariances which for the OLS residuals are defined as

$$\gamma_{OLS}(h) = \frac{1}{2}E[(e_{OLS}(s_i) - e_{OLS}(s_j))^2] \text{ for all } s_i, s_j \text{ with } s_i - s_j = h \quad (17)$$

are estimated instead via a so-called semivariogram and then the relationship

$$C(h) = C(0) - \gamma(h). \quad (18)$$

is used to compute the covariances. This will be further explained in Section \[3.2.3\].

Now the GLS coefficients can be estimated from the equation

$$\hat{\beta}_{GLS} = (X^TC_{OLS}^{-1}X)^{-1}X^TC_{OLS}^{-1}z \quad (19)$$

where $C_{OLS}$ is the covariance matrix of the OLS residuals and $X$ and $z$ are defined above (Equation \[13\] and \[14\]). See the derivation in Appendix \[3\]. There should not be a large difference between $\hat{\beta}_{GLS}$ and $\hat{\beta}_{OLS}$ if there is no significant clustering of the observation points. Once the GLS residual coefficients are determined, the GLS residuals can be derived in the same way as the OLS residuals, i.e.

$$e_{GLS}(s) = z(s) - X\hat{\beta}_{GLS}. \quad (20)$$
If needed, more than one iteration can be performed, and in that case new regression
coefficients, $\hat{\beta}_{\text{GLS}}$, are estimated by creating a semivariogram from the previously com-
puted GLS residuals, instead of the OLS residuals, and the procedure is repeated. An
alternative way of computing the regression coefficients and residuals is to use weighted
least squares. This is explained in Section 3.2.7.

Both OLS and GLS assume that the expected value of the error term is zero but the
difference lies in the variance. In OLS it is assumed that $\text{Var}(e) = \sigma^2 I$, where $I$ is the
identity matrix, i.e. the covariances/correlations between the errors are zero. GLS, on
the other hand, assumes the variance $\text{Var}(e) = \sigma^2 C$, where $C$ is the covariance matrix.
By estimating the covariance matrix via a model derived from the semivariogram we
expect the GLS coefficients to more accurately reflect the true spatial variation among
the errors. This is the reason why an extra effort from estimating the GLS coefficients
is usually necessary. In the covariance matrix, this dependence between the errors shows
as non-zero elements in the off-diagonal. The estimated covariance structure is later also
used in the simple kriging part of regression kriging.

In the end, the covariance structure for the final residuals, obtained after one or
several iterations, needs to be computed in order to use simple kriging and calculate
the kriging weights as described in Section 3.2.1. Again, this is done by determining
the autocorrelation function from the semivariogram for the final residuals. Now, we
are ready to calculate the interpolated residual at the target point. The weights are
calculated from Equation (7) and then the interpolated residual is determined from

$$
\hat{e}(s_0) = \sum_{i=0}^{N} w(s_i) e(s_i)
$$

which corresponds to the simple kriging formula in Equation 3 with mean zero. If the
total number of points is large it is possible to set a restriction so that $N$ is the $N$ closest
points to the target $s_0$. This is called the kriging neighbourhood.

The last step in regression kriging is to calculate the target value by inserting $\hat{\mu}(s_0)$
and $\hat{e}(s_0)$ in Equation (10). In matrix notations this equation can be written as

$$
\hat{z}(s_0) = q^T \cdot \hat{\beta}_{\text{GLS}} + w^T \cdot (z - X \hat{\beta}_{\text{GLS}})
$$

where $\hat{z}(s_0)$ is the predicted target value at the point $s_0$, $q$ is the column vector of $p+1$
predictors at the target point (containing the predictors and in addition number one in
the first row) and $w$ is the vector of $n$ kriging weights used to interpolate the residuals
at the point $s_0$.

### 3.2.3 Semivariogram

The dependence between the target values at different locations is called autocorrelation.
Estimating the spatial autocorrelation is called spatial modelling or structural analysis.
The first step of spatial modelling is to create a so-called semivariogram plot. A semivari-
ogram is a plot of estimations of the theoretical variogram, which will soon be defined,
against the distance between the points. Then a model is fitted to the semivariogram
which represents the spatial autocorrelation and from which the covariance structure can
be estimated.

The theoretical variogram is defined as

$$
\gamma(h) = \frac{1}{2} \text{E}[(z(s_i) - z(s_j))^2] = \frac{1}{2} \text{Var}(z(s_i) - z(s_j)) \text{ for all } s_i, s_j \text{ with } s_i - s_j = h
$$
where $z(s_i)$ and $z(s_j)$ are the target values in the points $s_i$ and $s_j$ respectively and $h \in \mathbb{R}^3$ is the distance vector between them. This equality holds because of the assumption $E[z(s_i)] = \mu$, for all $i$, mentioned above, which leads to that $E[(z(s_i) - z(s_j))] = 0$, for all $s_i, s_j$. The function $\gamma(h)$ can be interpreted as the "dissimilarity between point values as a function of distance". There is some confusion regarding the terminology for the function $\gamma$; some authors use the word variogram while others use semivariogram (or semivariance) and defines the semivariogram $\gamma(h)$ as half the variogram $2\gamma(h)$. It is also common to use the same name for both. Here, from now on, both words will be used for $\gamma(h)$ to avoid confusion. The relationship between the covariances and the semivariances is derived as follows:

Using the assumption that the covariances are translation-invariant, i.e. $C(h) = C(s_i - s_j) = \text{Cov}(z(s_i), z(s_j))$, for all $s_i, s_j$ with $s_i - s_j = h$ together with the variance addition theorem we have that

$$2\gamma(h) = \text{Var}(z(s_i) - z(s_j)) =$$

$$= \text{Var}(z(s_i)) + \text{Var}(z(s_j)) - 2\text{Cov}(z(s_i), z(s_j)) =$$

$$= 2\text{Var}(z(s_i)) - 2\text{Cov}(z(s_i), z(s_j)).$$

Let $\sigma^2 = \text{Var}(z(s_i)) = C(0)$ and $C(h) = \text{Cov}(z(s_i), z(s_j))$. Hence, the relationship

$$C(h) = \sigma^2 - \gamma(h)$$

is obtained.

The most common estimator of the semivariance showed in Equation (23) is called the method of moments estimator and is defined as

$$\gamma_c(t) = \frac{1}{2n(t)} \sum_{i,j} \|z(s_i) - z(s_j)\|^2$$

where $n(t)$ is the number of paired data that is separated by the distance $t - \epsilon \leq \|s_i - s_j\| \leq t + \epsilon$. Thus, the estimator is rotational symmetric. As we can see, the estimator of the semivariance is the average squared target value between all points at a certain distance divided by two.

After estimating the semivariances according to Equation (26), the semivariogram plot is created by plotting the distances $\|h\|$ on the x-axis and the estimation of the semivariances $\gamma_c(\|h\|)$ calculated for all pairs of points on the y-axis. If there are many points, which there usually are, it is possible to group them into lag bins and plot the mean of the groups instead of the individual points. This is explained in detail in Section 3.2.4. Plotting the semivariance for all pairs of points against the distance is called the semivariogram cloud and can be used in the early analysis of the data. An example of a variogram cloud is showed in Figure 3.
Figure 3: Example of a semivariogram cloud, i.e. where the semivariances of all pairs of points are plotted against the distance.

When the pairs are grouped, the variogram plot will look somewhat like Figure 4. The slope is decreasing as the distance $\|h\|$ increases and at a certain distance the curve levels out. At this distance the samples can be considered as (spatially) independent of each other. We define the range, $a$, as the certain distance to where the model levels out [6]. The nugget, $c_0$, is the (small) distance from the origin to where the curve starts on the y-axis. The sill, $c_0 + c_1$, is defined the same way as the range but on the y-axis. The sill is identical to the variance $\sigma^2 = \text{Var}(z(s_0))$ and thus, the relationship in Equation 25 can be expressed as

$$C(h) = c_0 + c_1 - \gamma(h).$$

(27)
When the semivariogram is plotted, the next step is to fit a model to the points. This model influences the predictions and it is therefore critical to find a model as precise as possible in order to obtain good predictions. How to find such an adequate variogram model function is discussed in Section 3.2.5.

3.2.4 Binning

The number of pairs for N point observations is \( N(N - 1)/2 \) and obviously the number of pairs increases very fast with the number of points [14]. Plotting the distance between all pairs versus their corresponding semivariances is, as mentioned, called the variogram cloud. A variogram cloud can be helpful in the early analysis of the data but is usually hard to describe visually. Therefore it is often necessary to group the pairs, where the distance and direction between the points can be considered and are more or less similar between the groups. This procedure is named binning [28]. The distance between points in the same group (bin) are averaged and plotting such averaged distances against their corresponding semivariances is called a standard experimental or sample variogram.

The distances between the pairs in the groups are called lags. For example, we can group the pairs in one group at which the samples are 100 m apart, one group at which they are 200 m apart, one 300 m apart and so on. Then the lag distance or lag size (distance between lags) is 100 m and the number of lags is the number of such groups. Thus, the maximum distance at which the variogram is calculated is given by the number of lags times the lag distance. This is called the variogram coverage.

It is necessary to choose a proper lag size in the binning process. Too large lag sizes will make the short-range autocorrelation "disappear" and too small lag sizes will result in too few samples in each bin which gives each single pair too large influence and the averages will not be representative [19]. There is no best way of determining the size of the lags but a rule of thumb is to restrict the lag size times the number of lags (i.e. the variogram coverage) to be about half to three-fourth times the maximal distance among all the pairs. This is because at about this distance, which is called the distance of reliability, the number of points usually decreases fast, which means the reliability of the plotted points decreases. Also, the very distant pairs have a risk of being located in very different geographical areas and their semivariances are thus not representative. This rule of thumb can of course be modified depending on the range of the fitted semivariogram.
model. If the range is small relative to the extent of the empirical semivariogram, the lag size can be decreased, and vice versa. It is also recommended to have at least about 30 pairs in each bin. Ideally there would be equally many points in each bin.

The simplest way to group the pairs is to use the omnidirectional variogram. In this case no attention is paid to the direction the points are related to each other. Only the distance between the pairs is taken into consideration. Pairs with similar distance are grouped together in the same bin and plotted in the semivariogram. The distance is calculated as explained in Appendix C.

Sometimes there is a reason to believe that there is some dependency even between the direction and the elevation angle between the pairs. Then there are several options to consider in the binning procedure. The direction in which the pairs are related to each other is determined by the so-called true bearing, here simply referred to as bearing \[\beta\]. This is the angle measured clockwise from the north line. If the bearing is calculated from point A to B, it will not give the same answer as if calculated from B to A. This angle should always be in the range \([0, 180^\circ]\), and it is therefore needed to choose the direction (A to B or B to A) such that the bearing \(\beta \in [0, 180^\circ]\). The formula for the bearing angle is derived in Appendix D.

When the distance and direction for every pair of observations are calculated, the pairs can be grouped into lag bins. For each bin, the average distance must be computed and plotted against the corresponding semivariance. To group the pairs, imagine that we move every distance vector such that they all start from the origin and point towards the right hand half-plane. The bearing angle \(\beta\) is then the angle from the y-axis to the vector. There are several ways to group these vectors into bins. One way is to place a grid over this imagined coordinate system, where each cell in the grid forms a bin \[19\]. This is shown in Figure 5. Another way of binning is to use a radial sector pattern and place it on the coordinate system of distance vectors, as shown in Figure 6. Here the pairs are grouped depending on first the angle and then the length of the distance vector. We can also compute the elevation angle and use it in the binning. The elevation angle
Figure 6: One way of binning where the part of each sector that is located within a certain distance interval from the origin forms a bin.

is defined by

\[
elevation \ angle = \arctan\left(\frac{\Delta_{\text{elev}}}{\|h\|}\right)
\]

(28)

where \(\Delta_{\text{elev}}\) is the elevation difference between the points and \(\|h\|\) is the separating distance. This can be used in a three dimensional counterpart to both the 2D grid and sector method. Again consider the pairs moved to the right hand half-plane, but now also add the elevation angle so all pairs are represented by a vector in the second quarter of the 3D coordinate system. Now the pairs can be grouped three-dimensionally in the same way as explained above.

### 3.2.5 Model fit

As mentioned above, it is important to find a good model for the variogram in order to obtain reasonable kriging weights and good predictions in the end. The model fitting procedure consists of two parts. One part is to select an appropriate model and the other is to determine the parameters needed in the model. The requirement is that the variogram measure \(\hat{\gamma}(h)\) must be positive definite for all distances \[13\]. This makes sure that the kriging equations can be solved and that the kriging variance is positive.

Any positive definite model function can be used but some well-known models are listed below \[4\][27]:

- The spherical model:

\[
\hat{\gamma}(h) = \begin{cases} 
  c_0 + c_1 \left( \frac{3\|h\|}{2a} - \frac{1}{2} \left( \frac{\|h\|}{a} \right)^3 \right), & 0 < \|h\| < a \\
  c_0 + c_1, & \|h\| \geq a \\
  0, & \|h\| = 0
\end{cases}
\]

(29)

where \(c_0\) is the nugget, \(c_0 + c_1\) is the sill and \(a\) is the range.
• The exponential model:

\[ \hat{\gamma}(h) = \begin{cases} 
    c_0 + c_1(1 - e^{-\|h\|/a}), & \|h\| > 0 \\
    0, & \|h\| = 0.
\end{cases} \]  

(30)

• The Gaussian model:

\[ \hat{\gamma}(h) = \begin{cases} 
    c_0 + c_1(1 - e^{-\|h\|^2/a^2}), & \|h\| > 0 \\
    0, & \|h\| = 0.
\end{cases} \]  

(31)

• The power model:

\[ \hat{\gamma}(h) = \begin{cases} 
    c_0 + c_1\|h\|^a, & \|h\| > 0 \\
    0, & \|h\| = 0.
\end{cases} \]  

(32)

This model has no sill and the parameter \(a\) is dimensionless \([29]\), typically \(0 < a < 2\). For \(a = 1\) we obtain a straight line.

• The Sine Hole Effect model:

\[ \hat{\gamma}(h) = \begin{cases} 
    c_0 + c_1\left(1 - \frac{\sin(\pi\|h\|/a)}{\pi\|h\|/a}\right), & \|h\| > 0 \\
    0, & \|h\| = 0.
\end{cases} \]  

(33)

In this model the semivariance increases with increasing distance and after reaching the sill it oscillates around the sill value. The amplitude of the model then decreases with increasing distance.

• The Matérn model:

\[ \hat{\gamma}(h) = \begin{cases} 
    c_0 + c_1\left(1 - \frac{1}{2^\nu\Gamma(\nu)}\left(\frac{\|h\|}{a}\right)\nu K_\nu\left(\frac{\|h\|}{a}\right)\right), & \|h\| > 0 \\
    0, & \|h\| = 0.
\end{cases} \]  

(34)

where \(\nu\) is the smoothness parameter \(\in [0, \infty]\), \(\Gamma(\cdot)\) is the gamma function and \(K_\nu(\cdot)\) is the modified Bessel function of the second kind. The parameter \(\nu\) is critical. Larger \(\nu\) gives a smoother function. When \(\nu = 0.5\) the Matérn model is the same as the exponential model \([30]\).

It is also possible to combine the models into a so-called nested model. This can be useful if for example the empirical semivariogram first increases, thereafter flattens out and then starts to increase again at larger distances. A nested model should be used if it fits the empirical variogram more accurately and thus leads to better estimations, but should not be used when there is a simpler model with equally good performance. The nested model is created by summing simpler models \([10]\):

\[ \hat{\gamma}(h) = \sum_{i=1}^{n} |\lambda_i|\hat{\gamma}_i(h) \]  

(35)

where \(n\) is the number of models involved and \(\lambda_i\) for \(i = 1, ..., n\) are weights defined for each individual model. In fact, the models presented above are all nested models where \(\gamma_1 = c_0\) is the nugget effect. The combination with nugget effect and another well known
model is so common that they are often combined into a nested model but are still called the name of the well known model. In the case of the nested model, the sill is the sum of the nugget $c_0$ and the $c_1$-coefficients from the models involved. The covariance from Equation (27) therefore becomes

$$C(h) = c_0 + \sum_{i=1}^{n} c_{1,i} - \gamma(h).$$  \quad (36)

In some cases, it can be reasonable to consider fitting a model in more than one dimension. For example the multivariate (N-dimensional) Gaussian model becomes

$$\hat{\gamma}(x) = c_0 + c_1 (1 - e^{-1/2(x-\mu)^T C^{-1} (x-\mu)})$$  \quad (37)

where $x = [x_1, x_2, ..., x_N]$ is a vector-valued random variable, $\mu \in \mathbb{R}^N$ is the mean and $C \in \mathbb{S}_+^N$ is the covariance matrix. In two dimensions this can be written explicitly as

$$\hat{\gamma}(x_1, x_2) = c_0 + c_1 (1 - \exp[-(C_{11}(x_1-\mu_1)^2 + 2C_{12}(x_1-\mu_1)(x_2-\mu_2) + C_{22}(x_2-\mu_2)^2)])$$  \quad (38)

where $c_0$ is the offset, $c_1$ is the amplitude and the coefficients $C_{11}, C_{12}$ and $C_{22}$ can be interpreted as

$$\begin{bmatrix}
C_{11} &=& \cos^2(\theta) & \sin^2(\theta) \\
C_{12} &=& \frac{\sin(2\theta)}{2\sigma_{x_1}} & \frac{\sin(2\theta)}{2\sigma_{x_2}} \\
C_{22} &=& \frac{\cos^2(\theta)}{2\sigma_{x_2}} & \frac{\cos^2(\theta)}{2\sigma_{x_2}}
\end{bmatrix}$$  \quad (39)

where $\sigma_{x_1}$ and $\sigma_{x_2}$ are the spreads in $x_1$- and $x_2$-direction and $\theta$ is the clockwise rotation of the Gaussian blob. The matrix

$$\begin{bmatrix}
C_{11} & C_{12} \\
C_{12} & C_{22}
\end{bmatrix}
$$

is the covariance matrix. Using a two dimensional model gives more flexibility for the model to fit the data and taking into consideration the two important variables $x_1$ and $x_2$ that affect the target value. By also grouping (binning) the data points using these factors it is possible to create a grid of points with similar $x_1$- and $x_2$-values and fit the model to these.

As in the case with the one-dimensional model, it is also possible to combine different models (or the same model with different parameters) into a nested model. This is done in the same way as in the 1D case. In 2D, Equation (35) becomes

$$\hat{\gamma}(x_1, x_2) = \sum_{i=1}^{n} |\lambda_i| \hat{\gamma}_i(x_1, x_2).$$  \quad (40)

When the model is selected, the parameters in it need to be determined. Historically, it was common to estimate the nugget, sill and range parameters just by eye or by ordinary least squares. There are, however, improved ways of doing this. One is to estimate these parameters by using weighted least squares. In that way more weight can be given to for example points that are close in distance, have a small semivariance (i.e. close to the origin in the semivariogram plot), bins containing many points etc.

---

1Recall that $\mathbb{S}_+^N = \{A \in \mathbb{R}^{N \times N} : A = A^T$ and $x^T Ax > 0, \forall x \in \mathbb{R}^N$ such that $x \neq 0\}$, i.e. $\mathbb{S}_+^N$ is the space of symmetric positive definite matrices of size $N \times N$. 24
The general formulation of the minimization problem is \[ \min_{\theta} |\gamma - \hat{\gamma}(\theta)|^T \Omega |\gamma - \hat{\gamma}(\theta)| \] (41)

where \( \gamma = [\gamma(h_i), l = 1, \ldots, L]^T \) is the vector of sample semivariances for all \( L \) lags, \( \hat{\gamma}(\theta) = [\hat{\gamma}(h_i; \theta), l = 1, \ldots, L]^T \) is the corresponding model values calculated using the parameters \( \theta \) and \( \Omega = [w_{ll}, l, l' = 1, \ldots, L] \) is the matrix of weights. Depending on \( \Omega \) the minimization problem becomes the same as ordinary least squares (OLS), weighted least squares (WLS) or generalized least squares (GLS). Next we examine these three cases:

- If \( \Omega = I \) (identity matrix) this is the same as minimizing OLS. OLS assumes that the residuals \( \gamma - \hat{\gamma}(\theta) \) are independent and with the same uncertainty. If this is not true, OLS is not a good choice.
- If \( \Omega \) has non-zero and different diagonal elements and zeros everywhere else, this becomes WLS and the optimization problem is then usually solved using iterative procedures. This is explained in detail in Section 3.2.6. This is the most common method to use for the model fit.
- If \( \Omega = C^{-1} \) i.e. the inverse of the covariance matrix of the sample semivariogram, calculated from the chosen semivariogram model, this becomes the problem of minimizing GLS. Using GLS is slightly more efficient than WLS but it is also more complicated because it requires calculating the variances of the estimates of the sample variogram at each lag and also the covariances between them. Therefore WLS is recommended.

3.2.6 Weighted least squares for model fitting

The idea of using weighted least squares in the case of fitting the semivariogram model is to minimize the squared distance between the empirical variogram (the exact points) and the model variogram, i.e.

\[
\min_{\theta} \sum_{l=1}^{L} w_l \left( \gamma(h_i) - \hat{\gamma}(h_i; \theta) \right)^2
\] (42)

where \( \gamma(h_i) \) is the sample semivariance at lag distance \( ||h_i|| \) and \( \hat{\gamma}(h_i; \theta) \) is the model semivariance at lag distance \( ||h_i|| \) for the parameters \( \theta \). The weights \( w_l \) are usually chosen such that more weight is given to the points close to the origin in the semivariogram plot. Note that this is the same as Equation (41) in the WLS case.

There is a number of possible ways of determine the weights \( w_l \). Some common options are \[24\]

- all squared residuals receive the same weight:
  \[ w_l = 1 \quad \text{for all } l \]
- more weight is assigned to large lag bins:
  \[ w_l = n(||h_i||) \]

where \( n(||h_i||) \) is the number of pairs in the bin separated by the lag distance \( ||h_i|| \).
- small distances $\|h_i\|$ receive more weight (since they have lower $\hat{\gamma}(h_i; \theta)$), i.e. the points close to the origin are more important:

$$w_i = \frac{1}{\hat{\gamma}(h_i; \theta)^2}$$

- points in large bins with small semivariances are more important:

$$w_i = \frac{n(\|h_i\|)}{\hat{\gamma}(h_i; \theta)^2}$$

- more weight is assigned to squared residuals at small distances and with more sample pairs in the bin:

$$w_i = \frac{n(\|h_i\|)}{\|h_i\|^2}$$

The optimization problem is non-linear and is therefore solved using some numerical method.

Using weighted least squares there is an iterative procedure that can be performed as follows: Let $\|h_0\|$ be the smallest distance among all pairs and $\|h_{end}\|$ the largest distance. Then the starting guess $\theta_0$ can either be approximated by looking at the graph or by the following formulas [18]:

- For all models:

$$c_{0,\text{initial}} = \max(0, \gamma(h_0) - \frac{\|h_0\|}{\|h_1\| - \|h_0\|}(\gamma(h_1) - \gamma(h_0))) \quad (43)$$

- For all models except the power model:

$$\begin{cases} c_{1,\text{initial}} = \frac{\gamma(h_{end-2}) + \gamma(h_{end-1}) + \gamma(h_{end})}{3} - c_{0,\text{initial}} \\ a_{\text{initial}} = \frac{\|h_{end}\|}{2} \end{cases} \quad (44)$$

- For the power model:

$$\begin{cases} c_{1,\text{initial}} = \frac{(\gamma(h_{end-2}) + \gamma(h_{end-1}) + \gamma(h_{end}))}{3} - c_{0,\text{initial}} \\ a_{\text{initial}} = 1.0 \end{cases} \quad (45)$$

It is critical to find initial guesses that are good enough. If the initial sum of squares is too high, the convergence may fail. This can also happen if the wrong model is chosen such that the experimental semivariogram and the model have shapes that are not similar enough.

The second step in the iterative procedure is to compute the weights from the chosen one of the formulas above. Depending on which weight formula that is chosen, $\hat{\gamma}(h_i; \theta_0)$ may first need to be computed. Then some minimizing method is used to minimize Equation (42) to obtain new parameters $\theta = c_0, c_1, a$. Inserting these new parameters in the chosen equation for the weights gives the updated weights which then can be used in the minimization of Equation (42) etc. The iteration procedure continues until some convergence criterion is satisfied. Of course, if the chosen weight formula does not depend on the parameters $\theta$, there is no need to iterate more than once, since the weights do not need to be updated.
3.2.7 Weighted least squares for estimating the regression coefficients

An alternative to using generalized least squares to compute the regression coefficients is to use weighted least squares (WLS).

The ordinary least squares approach assumes that the standard deviation of the error term is constant over all values of the predictor. If this is not the case and the errors can be assumed to be uncorrelated, weighted linear regression is one way to optimize the parameter estimation. In this case, there is no need to create a semivariogram as explained in Section 3.2.3 to estimate the spatial correlation and instead the covariance matrix of the errors is assumed to be of the form

\[
C = \sigma^2 \begin{bmatrix}
1/w_1 & \ldots & 0 \\
0 & 1/w_2 & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 1/w_n
\end{bmatrix}
\]  

where \( w_i \) for \( i = 1, \ldots, n \) are the weights. Let \( W = C^{-1} \). Then the normal equations are

\[
(X'WX)\hat{\beta}_{WLS} = X'Wz
\]  

and thus the weighted least squares estimator is

\[
\hat{\beta}_{WLS} = (X'WX)^{-1}X'Wz
\]  

and the residuals

\[
e_{WLS}(s) = z(s) - X\hat{\beta}_{WLS}.
\]

The drawback with weighted least squares method is that the weights must be known, which they usually are not, and in that case they need to be estimated. To do this, residual analysis can be performed. In some cases, the error variance is a function of one or several of the regressors. If, for example, \( \text{Var}(e_i) = \sigma^2 x_{ij} \), the weights are \( 1/x_{ij} \). Sometimes it is also necessary to perform several iterations where the weights are estimated and recomputed after evaluation.

Note that even if using weighted least squares for estimating the regression coefficients, it is still needed to use a semivariogram to estimate the covariance structure required for computing the kriging weights.

3.3 Combined methods

To increase the overall performance or decrease the data capacity needed, it is possible to combine some of the above mentioned methods. One such example is to first use the regression part of regression kriging (i.e. multiple linear regression) to compute the residuals and then use IDW to compute the interpolated residual instead of kriging. In that way the method will also take the auxiliary variables from the regression part into consideration, and not only the distance as in the regular IDW. The regression part will "detrend" the data before using the IDW method.
4 Method

As mentioned in Section 1.5, regression kriging was chosen to be deeply investigated in this study. How regression kriging was implemented is explained in Section 4.1. Inverse distance weighted interpolation was also selected to be used as a comparison to regression kriging and the implementation of it is explained in Section 4.2. The results were evaluated by leave-one-out cross validation, which is explained in Section 4.4, for different areas and using different data sets. Section 4.5 explains in detail which data sets were used.

4.1 Kriging

Regression kriging was used with different parameters and models to find the combination that minimizes the mean square error obtained from leave-one-out cross validation. When using the regression kriging method, the first thing to do was to investigate the data. The measured temperature was plotted in a colour scale in each observation point to get an overview. The temperature was also plotted against elevation, longitude and latitude to examine these relationships and to decide which auxiliary variables to use. When the auxiliary variables were decided, the OLS coefficients and residuals from the regression part could be computed. The residuals were then used to analyse the goodness of the regression part, see Section 4.1.1. The GLS coefficients and residuals were then estimated by creating a semivariogram and fitting a model to the data. Several iterations were performed to make sure the coefficients had converged. The estimated covariance structure was then used in simple kriging to obtain an interpolated residual in the target point. The last step was to add the regression part and the kriging part to obtain the final interpolated temperature.

To determine which settings to use within the regression kriging method, one setting at a time was changed and the result was evaluated using leave-one-out cross validation for different data sets, see Section 4.4 and 4.5. The choices to be made to find the optimal regression kriging method were:

- Which auxiliary variables to include in the regression part. Even though only the latitude and longitude coordinates and the elevation were known it was possible to include for example the variables squared, some combination of the variables or some transformation of them. Several combinations were investigated to find the optimal choice. For one data set from each month, tests were run using latitude, longitude and elevation as auxiliary variables and also using only latitude and elevation. The results are showed in Section 5.1.4. During the developing process even some combinations, transformations and squared and cubic variables were tested but were rejected without further investigation.

- How to estimate the regression coefficients. One option was to use generalized least squares and estimate the covariance matrix by fitting a semivariogram model and another option was to use weighted least squares and let the weights depend on for example difference in latitude and/or elevation. Both options were tested and when using weighted least squares, the weights were assumed to depend equally on latitude and elevation. The results can be seen in Section 5.1.5.

- Which model to fit to the semivariogram. There is an infinite number of models to choose from since every positive definite model can be used and they can also be
combined into nested models. The models mentioned in Section 3.2.5 were implemented and tested for data to be able to draw some conclusions on which models fit the available data best. In addition, a nested model consisting of the spherical and the power model and also two two-dimensional models was implemented and tested. Those were the Gaussian model and a nested model consisting of the sum of two Gaussian models. The result is displayed in Section 5.1.2.

- How to choose the initial guesses for the nugget, sill and range parameters in the chosen semivariogram model. Here, it was assumed that the automatically generated initial guesses displayed in Equation (43) to (45) in Section 3.2.6 were good enough for this purpose. When the tests were run, the coefficients were iterated 20 times to make sure they had converged. For the two-dimensional models, the initial guesses were estimated by hand.

- How to group (bin) the pairs. Several different alternatives were tested, including the omnidirectional method, the grid method, the sector method and to separate the pairs depending on their vertical distance and their horizontal distance. In the two-dimensional case the points were separated by vertical distance and horizontal distance. There were also some choices to be made within the binning, such as the size of the bins, how distant bins to consider etc. Since the direction of the distance vector was not considered in the estimator of the semivariances, and since one of the purposes of the regression part in regression kriging was to remove the directional dependence, only the omnidirectional semivariogram was used. The grid method and the sector method for binning were implemented but it was directly noted that they were not as accurate as the omnidirectional semivariogram with the current settings. The binning was performed in order to obtain at least 30 pairs in each bin and having approximately equally many pairs in each bin. In most cases, the number of pairs in each bin was about 100-200.

- How to choose the weights in weighted least squares when fitting the semivariogram model. Some common choices are showed in Section 3.2.6. All of the most common choices were tested and also weighting by only considering the elevation difference between the points, in this case \( w_l = \frac{1}{\Delta \text{elev}_l} \) and \( w_l = \frac{n(\|h_l\|)}{(\Delta \text{elev}_l)^2} \), where \( \Delta \text{elev}_l \) is the mean elevation difference between the pairs in bin \( l \). The results are showed in Section 5.1.6. For the test using all the available observation stations from Sweden, Norway, Denmark, Finland and the Baltic states the weight formula \( w_l = \frac{n(\|h_l\|)}{\hat{\gamma}(h_l; \theta)^2} \) was used.

Another choice to be made is how large to chose the tolerance when iterating to find the parameters in the semivariogram model. If the weight model is not containing the parameters, only one iteration is needed, but otherwise a tolerance has to be chosen. Here, the tolerance \( 10^{-2} \) was used.

- How to choose the kriging neighbourhood. Sometimes it might be optimal to restrict the kriging part of regression kriging to only consider the closest points in horizontal distance and/or vertical distance. If so, the rest of the points are given zero weight. Here, tests were run where all points were included in the kriging neighbourhood, where only the 30 closest points in Euclidean distance were included and when the kriging neighbourhood was chosen to be the 30 closest points to the target in horizontal distance and among those, the 15 closest points in vertical distance were chosen and used. The results are showed in Section 5.1.3.
• How to measure the distance between the points. The performance can differ slightly depending on if the distance is measured as the great circle distance or the Euclidean distance. Here, the difference in result should be minor as we are only considering a relatively small area. A larger area is of course more affected by the choice of distance measurement. To generalize the method, both the great circle distance and the Euclidean distance were implemented and tested but since there were very small differences, the Euclidean distance was the method used.

4.1.1 Linear regression analysis

Linear multiple regression is as mentioned in Section 3.2.2 a part of the regression kriging technique. A linear regression model is fitted first via ordinary least squares as explained above and then the covariance matrix is estimated from the OLS residuals to obtain the generalized least squares coefficients and residuals. If needed the computed GLS residuals are used to repeat the procedure in several iterations and to obtain more accurate regression coefficients.

To investigate the behaviour of the data and the model fit, it is common to plot the residuals against the fitted values. Preferably the externally studentized residuals are used [9], defined as

\[ t_i = \frac{e_i}{\sqrt{S^2_{(i)}(1 - h_{ii})}}, \quad i = 1, 2, \ldots, n \]  

(50)

where

\[ e_i = z_i - \hat{z}_i, \quad i = 1, 2, \ldots, n \]  

(51)

where \( z_i \) is an observation and \( \hat{z}_i \) is the corresponding fitted value, \( h_{ii} \) is the diagonal elements of the hat matrix \( H = X(X'X)^{-1}X' \) and

\[ S^2_{(i)} = \frac{(n - p)MS_{Res} - e_i^2/(1 - h_{ii})}{n - p - 1} \]  

(52)

where

\[ MS_{Res} = \frac{SS_{Res}}{n - p} = \frac{\sum_{i=1}^{n} e_i^2}{n - p} \]  

(53)

and \( p \) is the number of parameters.

If the regression model fits the data accurately, the points in the plot should lie in a horizontal band with mean zero. If not, there are some patterns to look for, see Figure 7. If the pattern is shaped like a megaphone, the so-called outward-opening funnel pattern, the variance increases as the target variable increases. Of course, an inward-opening funnel pattern is also possible. This problem of heteroscedasticity, i.e. nonconstant error variance, can be dealt with by transforming the regressor or some of the response variables or by using weighted least squares. If the plot shows a curved pattern, it is an indicator of nonlinearity in the data. To increase performance in that case, other or more regressor variables can be included in the model or some transformation can be used to capture the nonlinear behaviour.
To plot the residuals against each regressor can also give more information about the data and the model. Those plots can show the same patterns as plotting the residuals against the fitted values. This can give an indication of which regressor variable that may need transformation etc.

### 4.2 Inverse distance weighted interpolation

Inverse distance weighted interpolation (IDW) was used as a comparison to regression kriging. Since IDW is a much simpler method, it should not perform better than regression kriging, unless something is wrong with the regression kriging. IDW was easily implemented by first calculate the distance from the target point to each observation point and from that compute the weights according to Equation (2). The weights were inserted in Equation (1) and the target value computed. For determining the power parameter \( p \) needed in IDW, data from 1\textsuperscript{st} January 2018 was used and the mean square errors for \( p = 2, 3, 4 \) and 5 were calculated. The value of \( p \) that gave the smallest MSE was then used in the rest of the computations.

### 4.3 Combined methods

Tests were also run where the combination of the regression part of regression kriging was used together with the inverse distance weighted interpolation method. The regression coefficients were determined by generalized least squares where the covariance matrix was estimated from a semivariogram using the spherical model. Then the IDW method was used on the residuals from the regression part.

### 4.4 Leave-one-out cross-validation

When the interpolation is performed, a method for estimating the error is needed to evaluate the result and to compare the different interpolation methods. Resampling methods are methods that estimate this error by repeatedly drawing samples from the available training set of data and refitting the model of interest, or perform interpolation, using each sample. Then the error can be estimated by using the known information about the remaining data, not used in the interpolation. The obvious drawback using such methods is that the interpolation has to be performed several times which makes it computationally heavy.
Leave-one-out cross validation (LOOCV) is a resampling method that uses a validation set and a training set. The available \( N \) data points are divided into a validation set that only contains one data point and a training set that contains all the remaining points. Then the interpolation is performed on the \( N-1 \) points in the training set and a prediction is performed in the last point, the one in the validation set. Let \( \hat{z}_i \) for \( i = 1, ..., N \) be the predicted value in point \( i \) using all points but point \( i \) as the training set. Since this point was not included in the fitting process but its true value, \( z_i \), is known, the mean square error \( MSE_i = (z_i - \hat{z}_i)^2 \) can be calculated. This procedure can now be performed for all points \( i = 1, ..., N \) and the LOOCV estimate for the test mean square error is calculated from

\[
CV_{(N)} = \frac{1}{N} \sum_{i=1}^{N} (z_i - \hat{z}_i)^2,
\]

i.e. the mean of the squared errors when all points are used as validation sets, one at a time.

One advantage with the LOOCV method is that almost all points are used in the fitting procedure. Therefore the error is not overestimated as much as it would have been if fewer points were used in the fitting and a larger validation set used. The second advantage is that no randomness is included in the process. All points are used as validation sets which means the same result will be obtained every time the error estimation is done.

4.5 Data sets used for evaluation

The results were evaluated by leave-one-out cross validation done for several sets of data to ensure that many different weather conditions were covered. It was assumed that the results of the interpolation method would differ between different seasons of the year. Usually the relationship between temperature and elevation is more stable during the summer and is more varying during the winter. It is also expected that a clear day will be harder to interpolate since the temperature varies more during such conditions, see Section 2.1. Therefore, to evaluate the different choices presented in Section 4.1, temperature data from the 10th each month in the last year was used. One such choice at a time was varied during the developing process and the results are presented in Section 5. The results were used to arrive at the different parameters, factors and models included in regression kriging. In addition, Sweden was divided into three groups representing different meteorological characteristics, see Section 4.6. The results for regression kriging and IDW in each such group were evaluated and compared to further investigate the performance of the methods. Many tests were also run during the developing process of the method to conclude which ideas might increase the performance and thus are worth examining more or which to reject directly.

When a final method with all its parameters and possibilities was determined, the performance was also evaluated in a "real-world-situation" and compared to the interpolation method used today at SMHI. To do this, the regression kriging method was used to interpolate the temperature data to a grid with 2.5 km between each grid point using all the observation stations from Sweden, Norway, Denmark, Finland and the Baltic states. This comparison test was run for data from the 26th of March 2018. At this date the temperatures were distributed reasonable evenly over the area which is the kind of weather situation where the method used today usually works well. To evaluate the regression kriging method in "hard" weather conditions, it was also tested for data from
20\textsuperscript{th} April 2018 where also all available stations were used. This day the temperature was +27.2 degrees at a maximum over the mainland but the water was still cold. Thus, this is a typical example of a weather situation where the method used today performs poorly because of the steep temperature gradients.

The Swedish stations available are showed in Figure 8. These are the observation stations used for all tests except for the comparison tests using data from 26\textsuperscript{th} March 2018 and 20\textsuperscript{th} April 2018.

![Observed data](image)

Figure 8: The observed temperature in all the Swedish observation stations for the 10\textsuperscript{th} of February.

4.6 Different parts of Sweden

To evaluate the results further, Sweden was divided into three groups, see Figure 9. This was because the nature in Sweden varies much and it is interesting to divide the area into groups that are more homogeneous seen to the meteorological aspects. In that way, the interpolation methods can be compared and evaluated for one group at a time and conclusions can be drawn about how well the methods capture the temperate behaviour in regions with different characteristics. Note that the latitudes and longitudes are in the rotated coordinate system, where the equator approximately lies in the middle of Sweden.
Figure 9: The observation stations were grouped into three groups with different meteorological characteristics. Here the colours represents the temperature in each observation and the data is from 10th February 2018.

The first group was chosen to be the mountainous north-western part of Sweden. In the rotated coordinate system this part was estimated to be where the latitude was greater than 1 and also greater than the line $\text{lat} = 7/3 \cdot \text{lon} + 1.5$. This group contains most of the mountain range in Sweden which makes the interpolation hardest since the temperature often shows a complex behaviour in this region. The north-eastern group does not contain as much mountains and instead the temperature can be influenced by the coastal effects. The south group contains mostly flatland and is probably the most homogeneous group seen to the meteorological effects. With this grouping, the number of observation stations in the north-western group was about 50, the north-eastern group about 40 and the south group about 90. The results were compared to that of IDW to identify the strengths and weaknesses of regression kriging in comparison to the simpler method IDW.
5 Results

The accuracy of the interpolation methods were determined by leave-one-out cross validation. To cover many different weather situations this was done for at least one day each month in the last year. The choices to be done in regression kriging were varied one at a time to determine the optimal settings. For these tests, the temperature data from all available Swedish observation stations were used for interpolation.

Also, two tests were run where all available data from observations in Sweden, Norway, Denmark, Finland and the Baltic states was used. The first test was from 26th March 2018 when the temperatures were distributed reasonably evenly, i.e. the kind of weather condition when the method used today works well. The result from regression kriging was compared to the result from the method used today. The second test was from 20th April 2018 when it was very high temperatures over the land surfaces but still cold temperatures over the water which is a weather condition when the method used today usually performs poorly. The results are presented in the following sections.

5.1 Kriging

The data from all observation stations in Sweden, Norway, Denmark, Finland and the Baltic states from 26th March 2018 that was used for the comparison test was also used in a leave-one-out cross validation test. The result from this test was that the cross validation mean square error was 1.1095 degrees. Here the auxiliary variables used was longitude, latitude and elevation and the regression coefficients were estimated by generalized least squares. The semivariogram model used was the spherical model and the fit was done to the omnidirectional semivariogram. The weight formula in weighted least squares used for determining the model parameters was \( w_l = \frac{n(\|h_l\|)}{\gamma(h_l; \theta)^2} \). In the omnidirectional semivariogram, the points considered were the ones that lie at a maximum of 0.5 times the maximum distance among all pairs and the minimum number of pairs in each bin was set to 30. The number of bins was chosen to be 100, which gave about 30-300 pairs in each bin. No restriction was set on the kriging neighbourhood and the Euclidean measure was used to compute the distance between the locations. The result from interpolating to a grid using these settings is showed in Section 5.1.7.

With the same choices of parameters, methods and models as used in the comparison test, the cross validation result for the 10th of each month was as shown in Table 1. Here, only the available Swedish observation stations were used for interpolation.
<table>
<thead>
<tr>
<th>Month</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>11.7912</td>
</tr>
<tr>
<td>February</td>
<td>0.5859</td>
</tr>
<tr>
<td>Mars</td>
<td>0.9659</td>
</tr>
<tr>
<td>April</td>
<td>1.3494</td>
</tr>
<tr>
<td>May</td>
<td>1.1323</td>
</tr>
<tr>
<td>June</td>
<td>3.2539</td>
</tr>
<tr>
<td>July</td>
<td>1.8006</td>
</tr>
<tr>
<td>August</td>
<td>0.9604</td>
</tr>
<tr>
<td>September</td>
<td>0.5847</td>
</tr>
<tr>
<td>October</td>
<td>0.3041</td>
</tr>
<tr>
<td>November</td>
<td>convergence issues</td>
</tr>
<tr>
<td>December</td>
<td>3.2831</td>
</tr>
</tbody>
</table>

Table 1: Results from the leave-one-out cross validation tests performed for each month using the same choices for the parameters, models and methods as used in the comparison test.

For the November data, the regression coefficients did not converge when iterating according to what is explained in Section 3.2.2 using GLS. This issue will be further discussed in Section 6.

5.1.1 Different parts of Sweden

Sweden was divided into three groups according to Figure 9 and the mean square error was calculated for each group each month. All settings to be done in regression kriging were the same as in the comparison test and also the same as in Table 1. It is interesting to compare the regression kriging method with the inverse distance weighted interpolation method for the three different groups to get an overview of how the performance of the methods differ. The two methods were both tested for each month and the results are showed in Table 2 - 4.

In addition, plots of the difference between the observed temperature and the interpolated temperature in each observation station is showed in Figure 25 - 30 in Appendix A. Here it is not the mean square error that is plotted, but the "real" error. For example if the interpolated temperature was 2 degrees too cold, the colour in the plot is the one representing -2. This is to gain some knowledge about the performance in different parts of Sweden and to get an overview of over- or underestimated temperatures in the different points. Note that the scale is different between the plots, this was necessary since the errors vary between each month and if the same scale was used for all, we would not be able to see the differences between the errors for the months with only small errors.
<table>
<thead>
<tr>
<th>Month</th>
<th>MSE RK</th>
<th>MSE IDW</th>
<th>Improvement [percent]</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>31.6335</td>
<td>49.9921</td>
<td>36.7</td>
</tr>
<tr>
<td>February</td>
<td>0.8657</td>
<td>3.5661</td>
<td>75.7</td>
</tr>
<tr>
<td>Mars</td>
<td>1.7427</td>
<td>3.9291</td>
<td>55.6</td>
</tr>
<tr>
<td>April</td>
<td>0.8469</td>
<td>4.2911</td>
<td>80.3</td>
</tr>
<tr>
<td>May</td>
<td>0.8177</td>
<td>5.3374</td>
<td>84.7</td>
</tr>
<tr>
<td>June</td>
<td>2.1469</td>
<td>8.3926</td>
<td>74.4</td>
</tr>
<tr>
<td>July</td>
<td>2.1053</td>
<td>5.5060</td>
<td>61.8</td>
</tr>
<tr>
<td>August</td>
<td>0.6566</td>
<td>4.5510</td>
<td>85.6</td>
</tr>
<tr>
<td>September</td>
<td>0.8903</td>
<td>4.0572</td>
<td>78.1</td>
</tr>
<tr>
<td>October</td>
<td>0.2273</td>
<td>4.4680</td>
<td>94.9</td>
</tr>
<tr>
<td>November</td>
<td></td>
<td></td>
<td>convergence issues</td>
</tr>
<tr>
<td>December</td>
<td>10.3197</td>
<td>12.2028</td>
<td>15.4</td>
</tr>
</tbody>
</table>

Mean improvement: 67.6

Table 2: Results from the leave-one-out cross validation tests performed for the mountainous north-western part of Sweden using the same choices for the parameters, models and methods as used in the comparison test. The improvement column is showing how many percent lower the MSE is for regression kriging than for IDW.

<table>
<thead>
<tr>
<th>Month</th>
<th>MSE RK</th>
<th>MSE IDW</th>
<th>Improvement [percent]</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>7.6366</td>
<td>14.5144</td>
<td>47.4</td>
</tr>
<tr>
<td>February</td>
<td>1.0477</td>
<td>1.3137</td>
<td>20.3</td>
</tr>
<tr>
<td>Mars</td>
<td>1.1595</td>
<td>2.2273</td>
<td>47.9</td>
</tr>
<tr>
<td>April</td>
<td>0.6834</td>
<td>0.7159</td>
<td>4.5</td>
</tr>
<tr>
<td>May</td>
<td>1.9401</td>
<td>2.2489</td>
<td>13.7</td>
</tr>
<tr>
<td>June</td>
<td>6.1914</td>
<td>10.9771</td>
<td>43.6</td>
</tr>
<tr>
<td>July</td>
<td>2.4477</td>
<td>2.8748</td>
<td>14.9</td>
</tr>
<tr>
<td>August</td>
<td>0.6959</td>
<td>1.6825</td>
<td>58.6</td>
</tr>
<tr>
<td>September</td>
<td>0.3526</td>
<td>1.1497</td>
<td>69.3</td>
</tr>
<tr>
<td>October</td>
<td>0.1328</td>
<td>0.7261</td>
<td>81.7</td>
</tr>
<tr>
<td>November</td>
<td></td>
<td></td>
<td>convergence issues</td>
</tr>
<tr>
<td>December</td>
<td>0.4369</td>
<td>1.1559</td>
<td>62.2</td>
</tr>
</tbody>
</table>

Mean improvement: 42.2

Table 3: Results from the leave-one-out cross validation tests performed for the northeastern part of Sweden using the same choices for the parameters, models and methods as used in the comparison test. The improvement column is showing how many percent lower the MSE is for regression kriging than for IDW.
<table>
<thead>
<tr>
<th>Month</th>
<th>MSE RK</th>
<th>MSE IDW</th>
<th>Improvement [percent]</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>3.2088</td>
<td>6.4803</td>
<td>50.5</td>
</tr>
<tr>
<td>February</td>
<td>0.2551</td>
<td>0.6898</td>
<td>63.0</td>
</tr>
<tr>
<td>Mars</td>
<td>0.5041</td>
<td>1.0444</td>
<td>51.7</td>
</tr>
<tr>
<td>April</td>
<td>1.8423</td>
<td>2.4059</td>
<td>23.4</td>
</tr>
<tr>
<td>May</td>
<td>1.0010</td>
<td>1.4166</td>
<td>29.3</td>
</tr>
<tr>
<td>June</td>
<td>2.7754</td>
<td>3.2393</td>
<td>14.3</td>
</tr>
<tr>
<td>July</td>
<td>1.4206</td>
<td>2.1534</td>
<td>34.0</td>
</tr>
<tr>
<td>August</td>
<td>1.2135</td>
<td>1.6535</td>
<td>26.6</td>
</tr>
<tr>
<td>September</td>
<td>0.5116</td>
<td>0.9837</td>
<td>48.0</td>
</tr>
<tr>
<td>October</td>
<td>0.4030</td>
<td>0.9886</td>
<td>59.2</td>
</tr>
<tr>
<td>November</td>
<td>convergence issues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>December</td>
<td>0.7727</td>
<td>2.0571</td>
<td>62.4</td>
</tr>
</tbody>
</table>

Mean improvement: 42.0

Table 4: Results from the leave-one-out cross validation tests performed for the south part of Sweden using the same choices for the parameters, models and methods as used in the comparison test. The improvement column is showing how many percent lower the MSE is for regression kriging than for IDW.
5.1.2 Variogram model

Many different variogram models were implemented and tested. Some of them could be rejected directly since the shape of the model did not fit at all to the data which caused convergence problems.

Test results based on January and October data are showed in Table 5 as examples. As seen in Table 1, the January data gave the largest cross validation mean square error while the October data gave the smallest mean square error. Therefore these months are used here as examples. The semivariogram plots with the different models fitted to the binned pairs of points are showed in Figure 31 - 32 in Appendix A. The results of the leave-one-out cross validation tests for the different models are showed in Table 5. For the Matérn model, the parameter $\nu$ was chosen to $\nu = 0.7$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Month</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spherical</td>
<td>January</td>
<td>11.7912</td>
</tr>
<tr>
<td>Spherical</td>
<td>October</td>
<td>0.3041</td>
</tr>
<tr>
<td>Exponential</td>
<td>January</td>
<td>11.9500</td>
</tr>
<tr>
<td>Exponential</td>
<td>October</td>
<td>0.3001 convergence issues</td>
</tr>
<tr>
<td>Gaussian</td>
<td>January</td>
<td>12.4414</td>
</tr>
<tr>
<td>Gaussian</td>
<td>October</td>
<td>0.3570</td>
</tr>
<tr>
<td>Matérn</td>
<td>January</td>
<td>12.0605</td>
</tr>
<tr>
<td>Matérn</td>
<td>October</td>
<td>0.3001</td>
</tr>
<tr>
<td>Nested, spherical + power</td>
<td>January</td>
<td>12.8552</td>
</tr>
<tr>
<td>Nested, spherical + power</td>
<td>October</td>
<td>0.4972</td>
</tr>
</tbody>
</table>

Table 5: Results from the leave-one-out cross validation tests performed for January and October data using different semivariogram models.

For the October data, it was noticeable that the regression coefficients did not converge for the exponential model as they did for for example the spherical model. The coefficients jumped back and forth while iterating and the norm of the difference between the coefficients was as much as 5, even for some of the late iterations. The power model and the sine hole effect model were not used since the form of the models differed too much from the shape of the binned residuals for these two sets of data. There was also some convergence issues for the nested model for the October data. After 20 iterations, the regression coefficients seemed to have converged and the norm of the difference between the 20th and the 19th iteration was 0.02. To compare, the corresponding difference for the other models, except for the exponential model for the October data, was about $10^{-7}$.

Two two-dimensional models were also implemented and tested. An example of a scatterplot and the corresponding fitted surface from 10th January is showed in Figure 10 and 11. Here, the 2D Gaussian model was used for fitting. The other model tested was the nested model consisting of the sum of two 2D Gaussian models. This model was more flexible but also had a larger risk of overfitting. The largest issue using the 2D models was that the starting guesses required for fitting needed to be accurately estimated. Since the formulas for the automatically computed starting guesses for the 1D semivariogram models, used for estimating the nugget, sill and range parameters ($c_0$, $c_1$ and $a$) does not apply in the 2D case, the starting guesses needed to be estimated by hand. Of course, there were also more parameters to estimate than in the 1D models. The results of the
tests of the 2D models was that is was too difficult to find accurate enough starting guesses and therefore also hard to get the regression coefficients to converge while using these models. Since the benefits of using 2D models were not large enough to motivate using them despite the difficulties they implied, they were not used in further investigations.

Figure 10: Scatterplot showing all binned pairs, grouped based on horizontal and vertical distance. The data is from 10th January 2018.
Figure 11: The fitted surface computed using the 2D Gaussian model. The data is from 10th January 2018.

5.1.3 Kriging neighbourhood

To evaluate how to optimally choose the kriging neighbourhood, again the data from the 12 different months was used. Here, only the results from January and October are presented. All other choices to be made in regression kriging is the same as used in the comparison test. The kriging neighbourhood was set to

- use all points.
- use the 30 closest points in Euclidean distance.
- use the 30 closest points in horizontal distance and among them, pick the 15 closest points in vertical distance.

The result of the tests for the January and October data is showed in Table 6.

<table>
<thead>
<tr>
<th>Kriging neighbourhood</th>
<th>Month</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>All points</td>
<td>January</td>
<td>11.7912</td>
</tr>
<tr>
<td>All points</td>
<td>October</td>
<td>0.3041</td>
</tr>
<tr>
<td>30 closest Euclidean distance</td>
<td>January</td>
<td>12.2151</td>
</tr>
<tr>
<td>30 closest Euclidean distance</td>
<td>October</td>
<td>0.3114</td>
</tr>
<tr>
<td>30 closest horizontal dist, 15 closest vertical dist</td>
<td>January</td>
<td>11.8819</td>
</tr>
<tr>
<td>30 closest horizontal dist, 15 closest vertical dist</td>
<td>October</td>
<td>0.3484</td>
</tr>
</tbody>
</table>

Table 6: Results from the leave-one-out cross validation tests performed for January and October data using different kriging neighbourhoods.
5.1.4 Auxiliary variables

The correlation between temperature and elevation, latitude and longitude for the different months is showed in Figure 22 - 24 in Appendix A. Note that the latitudes and longitudes are in the rotated coordinate system. By looking at the plots one can note that there is usually no linear relationship between temperature and longitude. Therefore tests were run with only latitude and elevation as auxiliary variables. The results are displayed in Table 7. All other choices to be made in the regression kriging method were the same as in the comparison test.

<table>
<thead>
<tr>
<th>Month</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>11.6902</td>
</tr>
<tr>
<td>February</td>
<td>0.5882</td>
</tr>
<tr>
<td>Mars</td>
<td>1.0394</td>
</tr>
<tr>
<td>April</td>
<td>1.3774</td>
</tr>
<tr>
<td>May</td>
<td>1.1393</td>
</tr>
<tr>
<td>June</td>
<td>3.3831</td>
</tr>
<tr>
<td>July</td>
<td>1.8141</td>
</tr>
<tr>
<td>August</td>
<td>0.9725</td>
</tr>
<tr>
<td>September</td>
<td>0.5858</td>
</tr>
<tr>
<td>October</td>
<td>1.7213</td>
</tr>
<tr>
<td>November</td>
<td>2.6323</td>
</tr>
<tr>
<td>December</td>
<td>3.0197</td>
</tr>
</tbody>
</table>

Table 7: Results from the leave-one-out cross validation tests performed for each month using only latitude and elevation as auxiliary variables.

For the December data, the regression coefficients had some convergence issues but were still used for the interpolation.

5.1.5 Regression coefficients

The regression coefficients can be estimated by ordinary least squares (OLS), generalized least squares (GLS) or by weighted least squares (WLS). When using WLS, the variances were assumed to depend equally on latitude and elevation difference. The weights were thus estimated to \( w_i = 0.5w_i,\Delta{\text{lev}} + 0.5w_i,\Delta{\text{lat}} \). Note that the semivariogram is still needed to estimate the kriging weights, even if WLS is used for determining the regression part.

With all other choices the same as for the comparison test, the result using WLS is presented in Table 8.
<table>
<thead>
<tr>
<th>Month</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>11.7439</td>
</tr>
<tr>
<td>February</td>
<td>0.6739</td>
</tr>
<tr>
<td>Mars</td>
<td>1.1839</td>
</tr>
<tr>
<td>April</td>
<td>1.5495</td>
</tr>
<tr>
<td>May</td>
<td>1.1668</td>
</tr>
<tr>
<td>June</td>
<td>3.5038</td>
</tr>
<tr>
<td>July</td>
<td>2.4101</td>
</tr>
<tr>
<td>August</td>
<td>1.0798</td>
</tr>
<tr>
<td>September</td>
<td>0.6025</td>
</tr>
<tr>
<td>October</td>
<td>0.3694</td>
</tr>
<tr>
<td>November</td>
<td>3.0273</td>
</tr>
<tr>
<td>December</td>
<td>4.6829</td>
</tr>
</tbody>
</table>

Table 8: Results from the leave-one-out cross validation tests using weighted least squares for determining the regression coefficients.

With all other choices except for the auxiliary variables the same as used in the comparison test, but with only the auxiliary variables latitude and elevation, the result using WLS was as displayed in Table 9.

<table>
<thead>
<tr>
<th>Month</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>14.4578</td>
</tr>
<tr>
<td>February</td>
<td>0.6686</td>
</tr>
<tr>
<td>Mars</td>
<td>1.3923</td>
</tr>
<tr>
<td>April</td>
<td>1.6855</td>
</tr>
<tr>
<td>May</td>
<td>1.2252</td>
</tr>
<tr>
<td>June</td>
<td>3.3181</td>
</tr>
<tr>
<td>July</td>
<td>2.2827</td>
</tr>
<tr>
<td>August</td>
<td>1.1984</td>
</tr>
<tr>
<td>September</td>
<td>0.6016</td>
</tr>
<tr>
<td>October</td>
<td>0.4770</td>
</tr>
<tr>
<td>November</td>
<td>3.1702</td>
</tr>
<tr>
<td>December</td>
<td>5.8037</td>
</tr>
</tbody>
</table>

Table 9: Results from the leave-one-out cross validation tests using weighted least squares for determining the regression coefficients. The auxiliary variables used was only latitude and elevation.

### 5.1.6 Weight formula

When fitting the semivariogram model, weighted least squares was used as explained in Section 3.2.6. Different weight formulas were tested and the results are showed in Table 10.
Table 10: Results from the leave-one-out cross validation tests performed for January and October data using different weight formulas.

<table>
<thead>
<tr>
<th>Weight formula, ( w_i )</th>
<th>Month</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>January</td>
<td>11.7804</td>
</tr>
<tr>
<td>1</td>
<td>October</td>
<td>0.3012</td>
</tr>
<tr>
<td>( n(|h_i|) )</td>
<td>January</td>
<td>11.7420</td>
</tr>
<tr>
<td>( n(|h_i|) )</td>
<td>October</td>
<td>0.4483</td>
</tr>
<tr>
<td>( 1/\gamma(h_i; \theta)^2 )</td>
<td>January</td>
<td>11.7928</td>
</tr>
<tr>
<td>( 1/\gamma(h_i; \theta)^2 )</td>
<td>October</td>
<td>0.3058</td>
</tr>
<tr>
<td>( n(|h_i|)/\gamma(h_i; \theta)^2 )</td>
<td>January</td>
<td>11.7912</td>
</tr>
<tr>
<td>( n(|h_i|)/\gamma(h_i; \theta)^2 )</td>
<td>October</td>
<td>0.3041</td>
</tr>
<tr>
<td>( n(|h_i|)/|h_i|^2 )</td>
<td>January</td>
<td>11.7515</td>
</tr>
<tr>
<td>( n(|h_i|)/|h_i|^2 )</td>
<td>October</td>
<td>0.3100</td>
</tr>
<tr>
<td>( 1/\Delta\text{elev}_i )</td>
<td>January</td>
<td>11.7885</td>
</tr>
<tr>
<td>( 1/\Delta\text{elev}_i )</td>
<td>October</td>
<td>0.3054</td>
</tr>
<tr>
<td>( n(|h_i|)/(\Delta\text{elev}_i)^2 )</td>
<td>January</td>
<td>11.7589</td>
</tr>
<tr>
<td>( n(|h_i|)/(\Delta\text{elev}_i)^2 )</td>
<td>October</td>
<td>0.3109</td>
</tr>
</tbody>
</table>

5.1.7 Tests using all available observation stations

Data from 26\(^{\text{th}}\) March 2018 was used in a test to compare the method used today with the regression kriging method. The temperature was interpolated to a grid with 2.5 km between each grid point. The area was on latitude -2 to 12 in the rotated coordinate system and on longitude -5 to 8. This was to cover the whole area where there are observation stations available from this particular day. The south part of Sweden had no observation stations available in the post process model and could therefore not be used for interpolation. Outside this region the temperature was just set to zero. Since there are no observation stations in the water, the evaluation of the method is restricted to the land surfaces. Since regression kriging is not developed for extrapolation, the extrapolated temperatures in the water is just ignored when comparing the methods.

The interpolated data were plotted and is showed in Figure 12a. Zoomed in plots are showed in Figure 34 - 36 in Appendix A. To compare, the corresponding temperatures from the method used today is showed in Figure 12b and the corresponding zoomed in plots are showed in Figure 37 - 39 in Appendix A. The red lines show the temperature, which is constant along the lines and the grey lines are lakes, rivers etc. The colours represent the temperature and the scale is showed to the right of the figures. The temperature is measured in degrees Celsius. The black numbers are the temperature in the observation stations, located at the black dots. In addition, the topography grid used for interpolation is showed in Figure 40 and 41 in Appendix A, just to compare if the temperature follows the elevation pattern.
(a) The interpolated temperature using regression kriging.

(b) The temperature interpolated with the method used today.

Figure 12: The temperature interpolated to a grid with 2.5 km between each point.

This particular day, the temperature was more or less evenly distributed over the area with no large gradients or differences. In such conditions, the method used today usually performs well and the problems with over-/undershooting and that the interpolated temperature close to the observation stations is not close enough to the observed temperature does not occur much. Anyway, the resulting plots will be analysed in detail in Section 6.5.

The other data set from all available observation stations was from the 20th of April 2018. The temperature was +27.2 degrees at a maximum but the temperature in the water was still cold. This is a typical example of a “hard weather condition” to model and is thus an interesting case. This is the kind of situations when the interpolation method used today usually does not perform very well. A first investigation of the data was done and the temperature in the observation stations were plotted, see Figure 13. Note the very wide span of temperatures included in the area. The temperature was also plotted against elevation, latitude and longitude to conclude which auxiliary variables to use. The results are showed in Figure 14a - 14c. Note that there is no clear relationship between temperature and longitude.

When only elevation and latitude were used as auxiliary variables the residual plots became as showed in Figure 15a - 15c, i.e. where the externally studentized residuals were plotted against the fitted values and against each regressor. Here, the regression coefficients were determined using ordinary least squares.
Figure 13: The temperature in the observation stations for the 20th of April 2018.

(a) Correlation between temperature and elevation for data from the 20th of April 2018.
(b) Correlation between temperature and latitude for data from the 20th of April 2018.
(c) Correlation between temperature and longitude for data from the 20th of April 2018.

Figure 14: Correlations between temperature and elevation, latitude and longitude for data from the 20th of April 2018.
(a) Externally studentized residuals from OLS plotted against the fitted values for data from the 20\textsuperscript{th} of April 2018. 

(b) Externally studentized residuals from OLS plotted against elevation for data from the 20\textsuperscript{th} of April 2018. 

(c) Externally studentized residuals from OLS plotted against latitude for data from the 20\textsuperscript{th} of April 2018.

Figure 15: Externally studentized residuals from OLS plotted against the fitted values, elevation and latitude for data from the 20\textsuperscript{th} of April 2018.

When instead latitude, elevation and elevation squared were used as auxiliary variables, the corresponding plot showing residuals plotted against elevation is showed in Figure 16.

Figure 16: Externally studentized residuals from OLS plotted against elevation for data from the 20\textsuperscript{th} of April 2018. Here the auxiliary variables were latitude, elevation and elevation squared.

The OLS residuals from the regression part were plotted in a colour scale with longitude on the x-axis and latitude on the y-axis. The result is showed in Figure 17.
Different semivariogram models were tried to find out which one fit the empirical semivariogram best. The spherical model fitted to the semivariogram from the OLS residuals is showed in Figure 18. Unfortunately, none of the models made the regression coefficients converge. Different weight formulas and different semivariogram binning were also tried but with so success. Tests were run anyway using the regression coefficients obtained from OLS and WLS but the result was not satisfactory since the estimated covariance matrix needed for computing the kriging weights was close to singular which caused large errors for some locations in the cross validation. The resulting interpolated temperatures were thus concluded not reliable.
Figure 18: The spherical variogram model (green line) fitted to OLS residuals from data from the 20th of April 2018. Here the auxiliary variables were latitude, elevation and elevation squared.

5.2 Inverse distance weighted interpolation

Data from 1st January 2018 was used for finding the optimal value for the power parameter \( p \). Tests was run for \( p = 2, 3, 4, 5 \). The result is showed in Table 11.

<table>
<thead>
<tr>
<th>( p )</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>7.7838</td>
</tr>
<tr>
<td>3</td>
<td>6.2536</td>
</tr>
<tr>
<td>4</td>
<td>6.2137</td>
</tr>
<tr>
<td>5</td>
<td>6.3551</td>
</tr>
</tbody>
</table>

Table 11: Results for data from January 1st 2018 for different values of the power parameter \( p \).

Since \( p = 4 \) gave the smallest mean square error, that choice of the power parameter was used for all interpolations using the IDW method.

When performing a leave-one-out cross validation test with the data from all observation stations available in Sweden, Norway, Denmark, Finland and the Baltic states from 26th March 2018, the result was that the mean square error was 3.7157 degrees for the IDW method. Performing the cross validation test for the 10th of each month using only the Swedish observation stations, the result for IDW was as displayed in Table 12.
Month | Mean Squared Error
--- | ---
January  | 19.8652
February  | 1.6011
Mars  | 2.0536
April  | 2.5908
May  | 2.6389
June  | 6.0887
July  | 3.2021
August  | 2.4605
September  | 1.8517
October  | 1.8531
November  | 2.6762
December  | 4.6184

Table 12: Results from the leave-one-out cross validation tests performed for each month using IDW.

### 5.3 Combined methods

As explained in Section 4.3, tests were also run where the regression part of regression kriging was used together with IDW instead of the simple kriging part. The semivariogram model used to estimate the GLS coefficients for the regression part was the spherical model and then IDW was used on the residuals from the regression part. The result for January and October is showed in Table 13.

<table>
<thead>
<tr>
<th>Month</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>14.7151</td>
</tr>
<tr>
<td>October</td>
<td>0.5042</td>
</tr>
</tbody>
</table>

Table 13: Results from the leave-one-out cross validation tests performed for January and October data using the regression part of regression kriging and then inverse distance weighted interpolation on the residuals. The auxiliary variables used was latitude, longitude and elevation.
6 Discussion

In this section, the results from the study will be analysed. Even the method, including some of the choices done in regression kriging, will be discussed and a comparison between regression kriging and IDW as well as between regression kriging and the method used today at SMHI will be done. In the end, the limitations and further work are discussed.

6.1 Investigation of the data

To improve the interpolation method, some investigation of the data was done. The temperature was plotted in a colour scale with longitude and latitude on the x- and y-axis respectively, see an example in Figure 8 to get an overview of the data. Also, the temperature was plotted against the variables latitude, longitude and elevation to investigate these relationships, see Figure 22 - 24 in Appendix A. Typically the temperature decreased with increasing latitude and with increasing elevation and in many cases the relationship looked more or less linear. Those cases were of course easiest to interpolate and usually gave the most accurate results. The data from January, June and December shows very nonlinear relationships between both temperature and elevation and between temperature and latitude. As noted in the result section, these months also gave the largest mean squared errors, except for November where there was convergence problems.

In some cases it was noted that the temperature decreased with elevation in two different ways, see an example of this in Figure 19 and in Figure 22a. Here it looks like there is an approximate linear relationship for some of the points at high elevation, located in the mountainous region in the north-western part of Sweden, and another linear relationship for points at lower elevation. This is an example where the temperature-elevation relationship looks different in different parts of the country which is hard to model mathematically. The complex behaviour of the temperature can partly be explained by inversion as written in the meteorological theory section.
Figure 19: Correlation between temperature and elevation for data from the 12\textsuperscript{th} of March. Note that the data shows at least two different relationships.

The plots of temperature vs longitude usually showed no linear relationship. If that is the case, one must consider removing longitude from the auxiliary variables in the regression part. However, when the same tests were run with only latitude and elevation as auxiliary variables, the performance was usually approximately the same as when longitude was included. Here comes the difficulty with this large amount of data that is varying over time. It becomes a balance to choose the option that works best for most of the cases. Perhaps including longitude can make a significant improvement for the cases where the temperature and longitude shows a somewhat linear relationship but the decrease in performance for the cases where there are no relationship is not significant. Then the optimal choice is probably to include longitude.

6.2 The regression part

When investigating the regression part more closely, the residuals were plotted against both the fitted values and against each regressor. This is a way to test and find patterns in the data, to determine how to estimate the covariance matrix and to decide which auxiliary variables to use. In some cases, i.e. for some data sets, it was noted that the residuals plotted against the fitted values showed some nonlinear behaviour and also that there was some heteroscedasticity. After plotting the residuals against latitude it was concluded that the nonlinear behaviour was usually probably due to some nonlinearity in the temperature-latitude relationship. A series of tests were then run to investigate that nonlinear relationship and find out how to get rid of the heteroscedasticity seen in the plots. One way of correcting for heteroscedasticity is to transform either one or several of the regressors or transform the target variable. Different transformations were tried but
since none of them significantly improved the result and since the heteroscedasticity did not exist in many of the data sets investigated, it was concluded that a transformation was not needed. A transformation that gives good results for some particular data set might even make the method perform worse for another data set.

The nonlinearity and heteroscedasticity can also be corrected for by adding auxiliary variables to the regression part. For example adding squared terms or even cubic terms might be helpful. One example of this is the data from 12\textsuperscript{th} March 2018. Figure 20a shows the calculated externally studentized residuals plotted against the fitted temperature values for this date. Here the auxiliary variables were latitude, longitude and elevation. Note the nonlinear shape of the plot and also the heteroscedasticity. The variances are much larger for lower temperatures. Plotting the residuals against latitude shows almost the same pattern, see Figure 20b, indicating that the variance is a function of latitude in this case and also that the nonlinearity comes from the relationship with latitude. The corresponding plots for longitude and elevation are shown in Figure 20c and 20d. Here, there is no clear pattern that can explain the nonlinearity.

![Residuals](image)

(a) Externally studentized residuals plotted against the fitted values from the regression part.

(b) Externally studentized residuals plotted against latitude.

(c) Externally studentized residuals plotted against longitude.

(d) Externally studentized residuals plotted against elevation.

Figure 20: The residuals plotted against the fitted values and each auxiliary variable. The auxiliary variables were latitude, longitude, elevation.
When adding latitude squared and latitude cubed to the auxiliary variables, the corresponding plot to Figure 20a is showed in Figure 21. Now the nonlinear relationship is not as clear as it was in the previous figure.

![Externally studentized residuals plotted against the fitted values from the regression part. The auxiliary variables were latitude, longitude, elevation, latitude squared and latitude cubed.](image)

Figure 21: Externally studentized residuals plotted against the fitted values from the regression part. The auxiliary variables were latitude, longitude, elevation, latitude squared and latitude cubed.

In cases like this, one would expect the interpolation to perform better when the auxiliary variables latitude squared and latitude cubed are included. However, in other cases when there is no nonlinear relationship between temperature and latitude, these variables should probably not be added. It is also the case that the kriging part of regression kriging might take care of some amount of nonlinearity and it is therefore hard to predict the result of the interpolation just by looking at the initial data.

It might also be the case that more auxiliary variables are needed, except for latitude, longitude and elevation. When the temperature behaviour cannot be explained by only these three variables, it can be necessary to include for example the distance to the coast, the wind speed, the wind direction or some other variables that affect the temperature.

### 6.3 Instability of regression kriging

It is noticeable from the result section that it is critical that the regression coefficients converge. A badly chosen semivariogram model or auxiliary variables can lead to convergence problems. For example the sine hole effect model did not converge for the data from January and October and was therefore rejected directly, but it might still be a good choice for some other set of data. Herein lies the problem of the large amount of data that the same model should be fitted to for the interpolation to work. When there are
convergence issues, the errors in the interpolated temperature can turn out to be huge. To
guard oneself against such huge errors, it is of course possible to set a restriction that for
example the difference between the interpolated value from regression kriging and some
other simpler method is not allowed to exceed some pre-defined value. Another, possibly
better solution, is to make a check if the regression coefficients have converged or not and
use for example inverse distance weighted interpolation if they did not converge. One
could also use OLS or WLS if the coefficients did not converge when using GLS, since this
should at least improve the chances of obtaining a stable result. The covariance structure
still needs to be estimated in order to perform the kriging part and for some weather
situations even this turned up to be very hard. For example the data from the 20th of
April 2018 was such a case. It was noted that when the estimated covariance matrix was
close to singular, solving the matrix equation for the kriging weights was impossible and
did not yield reliable results. It is hence essential to find an accurate enough semivari-
ogram model that gives a positive definite covariance matrix. The reason why there was
convergence problems for this particular date is further discussed in Section 6.5.2. For
the covariance matrix to be positive definite, the assumption of second order stationarity
must be satisfied. For the semivariogram models to fit the data, the assumption that
close points are more related to each other than distant points also cannot be violated. If
this is not the case, i.e. the regression part does not capture the temperature behaviour
good enough, there will be convergence problems.

The complexity of some of the data sets that might cause convergence problems is
also a reason for not using a too complex semivariogram model. Even if a nested or a
two-dimensional model can fit very well to some particular data set, it is likely not to
fit well to another and there is a large risk of overfitting. This was noted in the tests
for the nested model. Here, there was some convergence issues for the October data and
even if the resulting mean squared error was small it is likely that convergence problems
will occur when using other data sets and there is no guarantee that it will not be large
errors.

Another way of solving this problem would have been to develop another way of
automatically finding the initial guesses that are more suited for this particular problem
(which might be difficult due to the variety of data) or selecting them manually. Of course
the choices and parameters also affect one another. This means that for example some
of the models that did not converge for the chosen set of parameters might converge for
other choices. An example of this is that the spherical model gave a good result for the
December data when the auxiliary variables were latitude, longitude and elevation but had
some converge issues when the auxiliary variables were only latitude and elevation. For
the November data the reverse was true. By selecting for example another weight formula
when calculating the parameters in the variogram model the result would probably be
different. This makes it extremely hard to fairly evaluate all choices to find the optimal
ones. To perform a strictly statistical test all combinations should be included, which is
of course not possible since there are infinitely many variogram models possible.

6.4 Comparison between IDW and regression kriging

When comparing the results from the tests run for each month for regression kriging
and inverse distance weighted interpolation, it was noted that the regression kriging had
smaller mean squared error for all months except November, where there was conver-
gence problems when latitude, longitude and elevation were used as auxiliary variables.
Since IDW does not take elevation into consideration, it should not perform as well as regression kriging in the mountainous region. The exception may be if the temperature behaviour is very complex and impossible for regression kriging to capture. As noted in Table 2 in Section 5.1.1, the improvement of using regression kriging instead of IDW was overall largest in the mountainous region of Sweden, just as expected. When comparing the percentage showing how much lower the mean square error was for each month for regression kriging compared to IDW one can note that the percentage was highest for the mountainous region for all months except January and December. When studying the correlation plots between temperature and elevation, Figure 22 - 24 in Appendix A one can note that for January and December it looks like the temperature decreases with elevation in two different ways. A somewhat similar behaviour can be seen in the corresponding correlation plot for the March data, and maybe also for June. For March, the performance improvement was about the same for all three regions of Sweden, but for June, the improvement was much larger for the mountainous region. The conclusion from this is that when the relationship between temperature and elevation is linear, the regression kriging method is exceptionally good even though it always outperforms IDW as long as the regression coefficients converges. When the temperature-elevation relationship is more complicated and varies between different parts of Sweden, the improvement of using regression kriging is not that large.

What is common for most of the error plots in Figure 25 - 30 in Appendix A is that it looks like the regression kriging can decrease the largest errors that appears in the inverse distance weighted interpolation method. For example in the interpolations from April, Figure 26 and 27, there are some large errors in the south-western part of Norrland for the IDW method that are almost completely gone when using regression kriging. Another example is the interpolation from March. IDW gives large errors in the south part of Sweden, estimating a too cold temperature. Regression kriging, estimates a little bit too cold but is far better than IDW. On the other hand, for the interpolations from January for example, the errors are large at about the same locations for both methods.

6.5 Tests using all available observation stations

6.5.1 Data from the 26th of March 2018

What is noticeable when looking at the plots in Figure 34 - 39 in Appendix A is that the temperature varies more in the plots showing the interpolated temperature from the method used today than when regression kriging was used. This is expected since SMHI have experienced problems with over-/undershooting. This is due to the (too smooth) background field in their method is forced to approach the kalman filtered observation points that is higher or lower than the temperatures in the background field, which causes over-/undershooting in neighbouring points.

The two main goals of this study was to include elevation in the interpolation method to improve the performance especially in the mountainous region, and to fix the temperature in the observation points. The temperature was fixed in the observation points by choosing to investigate the exact interpolation method regression kriging. What is left to analyse is if the interpolated temperature a small distance from the observation stations also is close to the observed temperature. Since the temperature is interpolated to a grid covering the whole area where there are observation stations available, it is possible to closely analyse if the method manages to keep the temperature close to the observed one at nearby points. By looking at the plots, we can conclude that the temperature is close
to the observed at nearby points for both methods, but of course it varies with location. No significant difference between the two methods can be identified.

For this particular day, it was hard to identify the problem with over-/undershooting on a local scale. What can be seen in Figure 39, showing the south part of the area, is that the temperature in the Baltic sea is -4 degrees when interpolated with the method used today but the observation stations close to the coast have temperatures around -1 to -2 degrees. The observed temperatures in the mountains are cooler and perhaps forcing the interpolation method to approach -1 close to the coast made it turn to -4 in the sea. The regression kriging method shows -2 even in the Baltic sea, which is probably a better estimation.

When it comes to the elevation analysis, the first thing to note is that the temperature behaviour is much more complex in the mountainous region. One can note that when the elevation increases fast, the temperature gradients are steeper, which is seen as more red temperature lines lies close to each other in the plots (Figure 34 - 39 in Appendix A). The temperature is colder at higher elevations and when comparing to the topography in Figure 40 and 41 in Appendix A it looks like the temperature follows the same pattern as the elevation. This is also true for the method used today at SMHI and it is hard to determine which method that best captures the temperature-elevation relationship.

This particular day the temperatures were more or less evenly distributed over the area. In such weather conditions the method used today works well and is thus hard to compete with. By only a subjective judgement, both methods works reasonably well for this set of data.

6.5.2 Data from the 20th of April 2018

As mentioned above, the weather condition this day was especially hard to model mathematically. This is due to the very high temperatures over the land surfaces in the south part of the region but cold temperatures in the water, causing steep temperature gradients. Unfortunately none of the models and other choices that were tried could make the regression coefficients converge for this data set. When performing the cross validation anyway, using the regression coefficients from OLS and estimating the covariance structure by the semivariogram showed in Figure 18, the covariance matrix became close to singular which led to some very large kriging weights. This caused huge errors in the interpolated temperature for some locations in the cross validation, and the conclusion was that the regression kriging method can not be used when the coefficients do not converge since the resulting interpolated temperatures then are not reliable.

When plotting the residuals against the fitted values, Figure 15a, some heteroscedasticity was noted but no nonlinear behaviour. The heteroscedasticity also appears a little when the residuals were plotted against latitude. The relationship between the residuals and elevation on the other hand, shows some nonlinear behaviour. This nonlinear behaviour then disappeared when elevation squared was added to the auxiliary variables, see Figure 16.

It was noted that the residuals from the regression part varied much, even between close locations. This can be seen in Figure 17 where the colours represent the OLS residuals to which the kriging part is to be applied. This can cause problems for the interpolation and can be the main explanation why there were convergence issues. The reason is that the large variances between residuals at close locations violates the assumption of second order stationarity, i.e. that the covariances depend on distance and
not on location. As seen in the plot, the maximum and minimum values of the residuals are located close to each other, which does not agree with the assumption needed for the semivariogram that close points are more related to each other than distant points. The large variances at close locations can of course inflate the semivariances at short lags and perhaps the range will become too short.

When studying the empirical semivariogram, Figure 18, one can note that there are only few points with small semivariances at short distances. The number of grouped data in these bins is also sparse. This agrees with the pattern noted above, where close points have a very wide spread of values. This can be a reason why there was convergence problems when using the weight formula $w_l = \frac{n(\|h_l\|)}{\hat{\gamma}(h_l; \theta)^2}$ for fitting the semivariogram model. The small number of points in the important groups with small semivariance and short distance makes these points receive too small weight. When using for example OLS (i.e. $w_l = 1$ for all $l$) the spherical model at least was possible to fit to the data but still not converged when performing the iterations. For this reason, another way of performing the binning was tried where the data points with small semivariances were grouped into several smaller bins and were given larger weight. However, this did not lead to convergence. Perhaps the convergence problems would be smaller if there were more data points from the beginning that had small semivariances at short distances.

Because of the large spread of the residuals from the regression part at close locations, the computed nugget, sill and range parameters needed for the kriging part were probably not accurate enough. The fitted model did not represent the true behaviour of the data. Even if the variogram parameters were instead for example estimated by hand, the variation in the area with large variances at close locations would be smoothed and some of the information about the temperature behaviour would be lost. The problem probably instead lies in the regression part of regression kriging. An accurate regression part should make the variance of the residuals small at close locations and then there would be no problem in the kriging part. In this case, note that the residuals have a large negative value close to the coasts in the south of the region and a large positive value in the mainlands of Sweden and Denmark. This pattern should be possible to capture if the distance to the coast could be used as an auxiliary variable! If only latitude, longitude and elevation are used, there is no chance to model the coastal effects which are dominating in this particular data set. Unfortunately, this leads to convergence problems and instability of the interpolation method. Of course no tests were run including the distance to the coast since that information was not known in this study and the suggestion that including it would improve the method could not be verified.

6.6 Methods for estimating the covariance matrix

The estimation of the covariance matrix is a critical part of regression kriging. Using weighted least squares there is no way of estimating the dependence between the errors from the regression part which are therefore assumed independent. Intuitively this is a drawback since we expect the temperature to be correlated for points at close locations and therefore we also expect the errors to be somewhat similar for such points. However, in the mountainous region in the north-western part of Sweden the variances between the errors can be large, as seen when plotting the residuals against the fitted values. An example of such a plot is showed in Figure 20a. Thus, in the mountainous region, estimated relationships among errors close to each other are likely to be wrong and it is possible that it is more correct to assume local independence here. The other
advantage with using weighted least squares instead of estimating the covariance matrix with generalized least squares using a semivariogram is that it is easier to control which points are important and which factors that should affect the weights (variances). Since temperature is such a complex variable it can be an advantage to be able to do this more or less manually. It is reasonable to assume that the variances depend on the distance in latitude and elevation to the target point. Close points in elevation and in latitude should be given a larger weight when determining the regression coefficients. But how much weight should be given to each? Here, it was assumed that the elevation distance and the latitude distance is of equal significance and the weight given to each point is just the average of the weights from elevation distance and latitude distance. This can of course be tested more carefully but is left for further studies.

On the other hand, usually the temperature shows some linear behaviour and close points are somewhat related, at least globally over the whole area. This can be seen in the semivariogram. If the residuals were independent, the semivariogram would be just a horizontal line, which corresponds to independence. Since we can see that the semivariances increase with distance, and thus correlation decreases with distance, it is needed to incorporate a spatial structure. Thus, in many situations it may be more correct to assume dependence between the error terms and the semivariogram can help finding that dependence. In such cases GLS should be used. Again this becomes a judgement of what works best in most situations.

If possible, it would be interesting to investigate if using WLS in the mountainous region where the variances between the errors for some cases are large and using GLS in the rest of the area could improve the result. This would however require a large amount of observation stations and is outside the scope of this study.

6.7 Semivariogram

In this study, the method of moments estimator of the semivariances, showed in Equation [26] was used. It was assumed that the semivariances depend only on the norm of the distance vector $\|h\|$, and not on the direction. When using the regression kriging method, one purpose of the regression part is to remove the directional dependence. If this is succeeded, the semivariances should not depend on the direction of the distance vector. Therefore, as seen in the result section, the omnidirectional semivariogram was used for computations. One can argue that using a semivariogram in several dimensions can capture the relationship between temperature and elevation and between different directions, and can therefore be used instead of the regression part. This is left to further investigations but it is probably hard to find a model that is not too complex, which may cause convergence problems, that still fits all the different sets of data. Also, this would require more observation points in all different directions used in the multi-dimensional model.

When it comes to determine the most suited semivariogram model the difficulty with the large amount of different weather situations is palpable. In some situations it is possible to use a quite complex model, for example a nested model or the sine hole effect model, to accurately fit the data points. This may lead to a good interpolation of the temperature for that particular measurement but can lead to overfitting and large errors for another. In some cases it was observed that some of the closest points in distance had a very different temperature and thus a large semivariance. This could cause a complex semivariogram model to ”make a turn” upward close to the origin to fit those
large semivariances. The close points with large semivariance may not be representative for other points and if the model fits too well to those it can lead to large errors.

Also two two-dimensional models were implemented, separating the horizontal distance and the vertical distance. The conclusion from testing those models was that the number of points available were too few to obtain a 2D-model that is accurate enough to represent the spatial variability, and also that there were some problems with the convergence, mostly because the starting guesses became hard to estimate. Thus, in this study it was concluded that a somewhat simple model, such as the spherical model, is most "safe" to use since it does not have as large risk of overfitting. If more observation stations were available one may consider using a more complex model.

6.8 Limitations and further work

Since the temperature is such a complex variable there are many aspects to consider that can potentially improve the performance of the regression kriging method. More tests need to be run on different sets of data to more accurately determine which factors and models that are optimal in most of the cases. It is also possible that different settings should be used in different weather situations. If so, a deeper investigation needs to be made to detect some pattern for which weather situations a certain model, method, parameters etc. are most suitable.

It is also possible that a simpler interpolation method is enough for some weather conditions and can be used to speed up the computations for such cases. If one could identify which weather factors that are relevant to determine which method to use, this could potentially decrease the computational power needed for interpolation, and thus be of interest. Also, it might be worth to investigate if using different semivariogram models for different weather conditions can improve the result. It is probably hard to find the pattern for when to use which model, based on only the initial data, but if a large enough amount of data is investigated it might be possible for example using machine learning. If succeeded, perhaps a more complicated model could be used in some particular situations. For example the sine hole effect model seemed to fit the data very well in some cases but did not converge in other. If it was possible to identify when to use this model, the result could probably be improved.

Further work also involves investigating the possibility to use different settings in different parts of Sweden. Since the nature in Sweden is very varying with mountains, lakes, coasts etc. it is reasonable to assume that the temperature will show different behaviours in different locations. What makes this hard is the sparsity of observation stations. With the temperature observed at more locations the identified relationships would also be more reliable. For example dividing Sweden into parts where the temperature-elevation relationship is modelled differently at different parts is probably not possible today without any more observation stations. Too few stations would mean that every single station would be given a too large influence over the model which is then likely to be overfitted. As mentioned above, it would also be interesting to investigate if using WLS in some parts of Sweden and GLS in others when estimating the covariance matrix would improve the results. But again, this would require more observation stations.

Even more information from each observation station that can help explain the temperatures behaviour is needed for the regression kriging method to reach its full potential. As seen when examining the data from 20th April 2018, adding the distance to the coast to the auxiliary variables could potentially increase the performance of the method.
Another possible improvement that needs further investigation is the analysis of potential outliers. In some situations, the temperature can be more or less independent between two neighbouring stations. It requires meteorological knowledge and probably a study of a large amount of data to conclude if or when some specific observations should not be allowed to affect the interpolated temperature at nearby points, or maybe be given a less weight than it otherwise would have according to kriging.

It would also be interesting to examine the time factor. It is possible to incorporate the time dependence in regression kriging even though it is outside the scope of this study. If the time factor is included, even other time-dependent explanatory variables such as wind speed and wind direction could be taken into account. The wind factor is also potentially interesting even in the case of interpolation only in space, and not in time. Since inversion depends partly on the wind speed and wind direction it could be used as an auxiliary variable in the regression part that might improve the interpolation. At least locally the relationship between elevation and temperature can be affected by the wind. However, the wind speed and wind direction is today not known in all of the grid points, to which it is needed to interpolate, which makes it impossible to use this as an explanatory variable. One would need to estimate the wind based on meteorological models and knowledge which in turn would introduce another source of error in the interpolation method.
7 References

References


A Figures
Figure 22: Correlations between temperature and elevation, latitude and longitude for the data from January to April.
Figure 23: Correlations between temperature and elevation, latitude and longitude for the data from May to August.
Figure 24: Correlations between temperature and elevation, latitude and longitude for the data from September to December.
Figure 25: Error plots for the data from January and February.

(a) Errors for the January data using regression kriging.

(b) Errors for the January data using IDW.

(c) Errors for the February data using regression kriging.

(d) Errors for the February data using IDW.
(a) Errors for the March data using regression kriging.

(b) Errors for the March data using IDW.

(c) Errors for the April data using regression kriging.

(d) Errors for the April data using IDW.

Figure 26: Error plots for the data from March and April.
(a) Errors for the May data using regression kriging.

(b) Errors for the May data using IDW.

(c) Errors for the June data using regression kriging.

(d) Errors for the June data using IDW.

Figure 27: Error plots for the data from May and June.
(a) Errors for the July data using regression kriging.

(b) Errors for the July data using IDW.

(c) Errors for the August data using regression kriging.

(d) Errors for the August data using IDW.

Figure 28: Error plots for the data from July and August.
(a) Errors for the September data using regression kriging.

(b) Errors for the September data using IDW.

(c) Errors for the October data using regression kriging.

(d) Errors for the October data using IDW.

Figure 29: Error plots for the data from September and October.
(a) Errors for the November data using regression kriging.

(b) Errors for the November data using IDW.

(c) Errors for the December data using regression kriging.

(d) Errors for the December data using IDW.

Figure 30: Error plots for the data from November and December.
(a) Spherical model fitted to the residuals from January data.
(b) Spherical model fitted to the residuals from October data.
(c) Exponential model fitted to the residuals from January data.
(d) Exponential model fitted to the residuals from October data.

Figure 31: Different models fitted to the residuals from January and October data.
Figure 32: Different models fitted to the residuals from January and October data.

(a) Gaussian model fitted to the residuals from the January data.
(b) Gaussian model fitted to the residuals from the October data.
(c) Matérn model fitted to the residuals from the January data.
(d) Matérn model fitted to the residuals from the October data.
(a) Nested (spherical + power) model fitted to the residuals from the January data. (b) Nested (spherical + power) model fitted to the residuals from the October data.

Figure 33: Different models fitted to the residuals from January and October data.
Figure 34: The interpolated temperature using regression kriging. Here, the north part of the interpolated area is shown.
Figure 35: The interpolated temperature using regression kriging. Here, the middle part of the interpolated area is shown.
Figure 36: The interpolated temperature using regression kriging. Here, the south part of the interpolated area is shown.
Figure 37: The temperature interpolated with the method used today. Here, the north part of the interpolated area is shown.
Figure 38: The temperature interpolated with the method used today. Here, the middle part of the interpolated area is shown.
Figure 39: The temperature interpolated with the method used today. Here, the south part of the interpolated area is shown.
Figure 40: The topography used for interpolation. Here, the north part of Sweden is shown.
Figure 41: The topography used for interpolation. Here, the south part of the interpolated area is shown.
B Generalized least squares

The ordinary least squares (OLS) with regression model \( y = X\beta + e \) (55)

assumes that the errors have mean zero and constant variance, i.e.

\[
\begin{cases}
  E[e] = 0 \\
  \text{Var}(e) = \sigma^2 I
\end{cases}
\] (56)

This model cannot be applied when the variance is not constant and instead \( \text{Var}(e) = \sigma^2 V \), where \( \sigma^2 V = C \) is the covariance matrix. If this is the case the model can be transformed to one that satisfies the assumptions in Equation (56). This is done as follows:

By multiplying Equation (55) by \( K^{-1} \), which is the inverse of the square root of \( V \), i.e. \( K'K = KK = V \), we obtain

\[
K^{-1}y = K^{-1}X\beta + K^{-1}e.
\] (57)

The nonsingular symmetric matrix \( K \) exists since \( V \) is nonsingular and positive definite.

Define

\[
\begin{align*}
  z &= K^{-1}y \\
  B &= K^{-1}X \\
  g &= K^{-1}e
\end{align*}
\] (58)

Hence, Equation (57) becomes

\[
z = Be + g
\] (59)

which can be shown satisfies the assumptions in Equation (56). The expectation becomes

\[
E[g] = K^{-1}E[e] = 0
\] (60)

and the variance

\[
\text{Var}(g) = \{[g - E[g]]'[g - E[g]]\}' = E[gg'] = E[K^{-1}ee'K^{-1}] = K^{-1}E[ee']K^{-1} = \sigma^2 K^{-1}VK^{-1} = \sigma^2 K^{-1}KKK^{-1} = \sigma^2 I.
\] (61)

Thus, the OLS assumptions are satisfied for this new transformation and we can apply ordinary least squares. The function to be minimized is

\[
S(\beta) = gg' = e'V^{-1}e = (y - X\beta)'V^{-1}(y - X\beta)
\] (62)

which leads to the normal equations

\[
(X'V^{-1}X)\hat{\beta} = X'V^{-1}y
\] (63)

with the solution

\[
\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y.
\] (64)

The solution \( \hat{\beta} \) in Equation (64) is the generalized least-squares estimator of \( \beta \).

One can show that \( \hat{\beta} \) is the best linear unbiased estimator of \( \beta \). The covariance matrix is

\[
\text{Var}(\hat{\beta}) = \sigma^2 (B'B)^{-1} = \sigma^2 (X'V^{-1}X)^{-1}.
\] (65)
C Distance between points on earth

To be able to perform the interpolation, the distance between the known observation points and the target points as well as the distance between the observation points themselves need to be known. Since the Earth is more or less spherical, there are different ways to measure the distance. One is to consider the great-circle distance (or orthodromic distance) between the points. This is the shortest possible distance measured along the surface of Earth \[31\]. This distance is calculated using what is called the Haversine formula and the deduction is explained below.

The first step in the derivation is to compute the Euclidean distance, i.e. the straight line burrowing through the Earth. Recall the spherical coordinates

\[
\begin{align*}
  x &= R \cos(\theta) \cos(\phi) \\
  y &= R \cos(\theta) \sin(\phi) \\
  z &= R \sin(\theta)
\end{align*}
\]

where, in our case, \( R \approx 6378 \) km is the approximate radius of the Earth.

Now consider a coordinate system where the origin is at the centre of Earth, the positive z-axis is passing through the North Pole and the x-axis is passing through \( 0^\circ \) longitude (the Prime Median). Now \(-90^\circ \leq \theta \leq 90^\circ\) is the latitude and \(0 \leq \phi \leq 360^\circ\) is the longitude measure. The negative values of \( \theta \) corresponds to the southern hemisphere and the positive values corresponds to the northern hemisphere.

Now we want to calculate the Euclidean distance between two given points, \( P_1 \) and \( P_2 \). Let \( P_1 = P_1(x_1, y_1, z_1) \) and \( P_2 = P_2(x_2, y_2, z_2) \), where \( x, y, z \) are determined by the spherical coordinates defined in Equation (66). The Euclidean distance is given by the Pythagorean theorem

\[
d^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 = \tag{67}
\]

\[
R^2 \left( (\cos \theta_1 \cos \phi_1 - \cos \theta_2 \cos \phi_2)^2 + (\cos \theta_1 \sin \phi_1 - \cos \theta_2 \sin \phi_2)^2 + (\sin \theta_1 - \sin \theta_2)^2 \right) = ... = R^2 (2 - 2 \cos \theta_1 \cos \theta_2 \cos(\phi_1 - \phi_2) - 2 \sin \theta_1 \sin \theta_2).
\]

Now consider Figure 42 where the searched distance is the arc length \( D \). Note that

\[
\sin(\alpha/2) = \frac{d}{2R} \Rightarrow \cos(\alpha/2) = \sqrt{1 - \sin^2(\alpha/2)} = \sqrt{1 - \left(\frac{d}{2R}\right)^2}. \text{ From the trigonometric}
\]
relation "double angle" we obtain

\[
\sin(\alpha) = 2 \sin(\alpha/2) \cos(\alpha/2) = \frac{d}{R} \sqrt{1 - \left(\frac{d}{2R}\right)^2} = \frac{d}{2R^2} \sqrt{4R^2 - d^2}.
\]

This leads to

\[
D = R\alpha = R \sin^{-1} \left( \frac{d}{2R^2} \sqrt{4R^2 - d^2} \right). \tag{68}
\]

Usually the distance is expressed in terms of the haversine function which is defined by

\[
\text{haversine}(A) = \sin^2(A/2) = \frac{1 - \cos(A)}{2} \Rightarrow \cos(A) = 1 - 2\text{haversine}(A). \tag{69}
\]

Inserting this in Equation (67) yields

\[
d^2 = \ldots = R^2 (4\text{haversine}(\theta_1 - \theta_2) + 4 \cos \theta_1 \cos \theta_2 \text{haversine}(\phi_1 - \phi_2) \tag{70}
\]

\[
\Rightarrow \left( \frac{d}{2R} \right)^2 = \text{haversine}(\theta_1 - \theta_2) + \cos \theta_1 \cos \theta_2 \text{haversine}(\phi_1 - \phi_2).
\]

Thus, the Euclidean distance penetrating the Earth can be written as

\[
d = 2R \sqrt{\text{haversine}(\theta_1 - \theta_2) + \cos \theta_1 \cos \theta_2 \text{haversine}(\phi_1 - \phi_2)} = \tag{71}
\]

\[
2R \sqrt{\sin^2 \left( \frac{\theta_1 - \theta_2}{2} \right) + \cos \theta_1 \cos \theta_2 \sin^2 \left( \frac{\phi_1 - \phi_2}{2} \right)}.
\]

But since \( (\frac{d}{2R})^2 = \sin^2(\alpha/2) = \text{haversine}(\alpha) \) we can combine this with Equation (68) and obtain the formula for the distance \( D \) as follows:

\[
D = R\alpha = 2R \sin^{-1}(\sqrt{\text{haversine}(\alpha)}). \tag{72}
\]

Of course it is also possible to include the elevation in the distance calculation. Let the elevation in point \( P_1 \) be defined by \( \text{elev}_1 \) and in point \( P_2 \) by \( \text{elev}_2 \). In spherical coordinates we have

\[
\begin{aligned}
x_i &= (R + \text{elev}_i) \cos(\theta_i) \cos(\phi_i) \\
y_i &= (R + \text{elev}_i) \cos(\theta_i) \sin(\phi_i) \\
z_i &= (R + \text{elev}_i) \sin(\theta_i)
\end{aligned} \tag{73}
\]

for \( i = 1, 2 \). The euclidean distance is now obtained by plugging in \((x_1, y_1, z_1)\) and \((x_2, y_2, z_2)\) in Equation (67) and taking the square root. The rest of the derivation of the Haversine formula is the same.
D Derivation of the bearing angle

Define three unit vectors, $e_N$ that goes against the north pole, $e_A$ that goes against the initial point A, and $e_B$ that goes against the final point B. Imagine two planes, one that goes through $e_N$ and $e_A$ and another that goes through $e_A$ and $e_B$. The bearing is the angle between those two planes [23], which is the same as the angle between the normals perpendicular to the two planes, i.e $e_N \times e_A$ and $e_A \times e_B$.

Rotate the coordinate system such that the point A has latitude $\theta_A$ and longitude 0 and point B has latitude $\theta_B$ and longitude $\Delta\phi$ which is the difference between the true longitudes of A and B. Recall the spherical coordinates

\[
\begin{align*}
  x &= R \cos(\theta) \cos(\phi) \\
  y &= R \cos(\theta) \sin(\phi) \\
  z &= R \sin(\theta)
\end{align*}
\]

This gives

\[
\begin{align*}
  e_N &= (0, 0, 1) \\
  e_A &= (\cos(\theta_A), 0, \sin(\theta_A)) \\
  e_B &= (\cos(\theta_B) \cos(\Delta\phi), \cos(\theta_B) \sin(\Delta\phi), \sin(\theta_B))
\end{align*}
\]

and the cross products becomes

\[
\begin{align*}
  e_N \times e_A &= (0, \cos(\theta_A), 0) \\
  e_B \times e_A &= (\sin(\theta_A) \cos(\theta_B) \sin(\Delta\phi), \cos(\theta_A) \sin(\theta_B) - \sin(\theta_A) \cos(\theta_B) \cos(\Delta\phi), -\cos(\theta_A) \cos(\theta_B) \sin(\Delta\phi))
\end{align*}
\]

We now want to compute the angle between those two vectors. We know that the first one, $e_N \times e_A$, is parallel to the y-axis which makes the computations simpler. We can compute the component of $e_B \times e_A$ in the x-z-plane (which is the square root of the sum of squares of the x- and z-components) divided by the y-component as the tangent of the angle between $e_B \times e_A$ and the y-axis. This gives

\[
\tan(\beta) = \frac{\sqrt{(\sin(\theta_A) \cos(\theta_B) \sin(\Delta\phi))^2 + (-\cos(\theta_A) \cos(\theta_B) \sin(\Delta\phi))^2}}}{\cos(\theta_A) \sin(\theta_B) - \sin(\theta_A) \cos(\theta_B) \cos(\Delta\phi)} = \frac{\sqrt{\cos(\theta_B)^2 \sin(\Delta\phi)^2 + \sin(\theta_A)^2 \cos(\theta_A)^2}}{\cos(\theta_A) \sin(\theta_B) - \sin(\theta_A) \cos(\theta_B) \cos(\Delta\phi)} = \frac{\cos(\theta_B) \sin(\Delta\phi)}{\cos(\theta_A) \sin(\theta_B) - \sin(\theta_A) \cos(\theta_B) \cos(\Delta\phi)}
\]

Taking arctan() on both sides gives the bearing $\beta$. The problem using arctan() is that then we obtain an angle in $[-\pi/2, \pi/2]$ and cannot distinguish between opposite directions. Therefore we instead use the function arctan2() defined as

\[
\text{arctan2}(y, x) = \begin{cases} 
  \arctan(y/x) & x > 0 \\
  \arctan(y/x) + \pi & x < 0 \text{ and } y \geq 0 \\
  \arctan(y/x) - \pi & x < 0 \text{ and } y < 0 \\
  \pi/2 & x = 0 \text{ and } y > 0 \\
  -\pi/2 & x = 0 \text{ and } y < 0 \\
  \text{undefined} & x = 0 \text{ and } y = 0
\end{cases}
\]

89
i.e. it returns $\arctan(y/x)$ but with an adjustment depending on the sign of $x$ and $y$. Then the angle $\beta$ becomes the angle of $(x, y)$ in polar coordinates.

The last two steps in this derivation is to use the modulus function $\mod(\ldots, 2\pi)$ to convert the angle from the range $(-\pi, \pi]$ to $(\pi, 2\pi)$ and finally convert the final answer to degrees instead of radians. Thus, the final formula for the bearing $\beta$ is

$$\beta = \mod\left(\arctan\!\left(\frac{\cos(\theta_B) \sin(\Delta \phi)}{\cos(\theta_A) \sin(\theta_B) - \sin(\theta_A) \cos(\theta_B) \cos(\Delta \phi)}\right), 2\pi\right) \cdot \frac{180}{\pi}.$$  

(79)