Lund University

Spacial Statistics with Image Analysis

Home assignment 1: Classic methods for Gaussian fields

Author:
David Larsson
tfy15dla@student.lu.se

Author:
Sofie Hellmark
mat15she@student.lu.se

27 november 2018

1 Introduction

Ordinary least squares, OLS, and Kriging are two methods that can be used to estimate unknown values, here rainfall, from known observations at some locations. From the known data points a model for how the rain depends on elevation, x-distance and y-distance can be estimated. Using Ordinary Least Squares we first have to predetermine the form of the model, then estimate the parameters. Since a spacial dependence in the errors from using OLS was found, this is not the best method. Kriging was used to model the errors and make a new estimation of the field. All the rain data was transformed with the square root before processing the data. For this problem a total of hundred locations with the measured rain was supplied. Of these 90 was used to make the model and the ten remaining was saved to be compared to as validation of the model. The unknown location that the model was applied to consisted of a total of 4440 points with given elevation, x- and y-coordinates.

2 Ordinary Least Squares

With Ordinary Least Squares the goal is to minimize the residual, $\boldsymbol{e}$, between a regression model and the data. The regression model describes the dependence between the dependent data point, here rain y_i, and a linear combination of independent data, here elevation x_1^i, x-distance x_2^i and y-distance x_3^i. Using OLS first the form of the model has to be predetermined, then the parameters $\boldsymbol{\beta}$ can be estimated. The regression model is given by: $y_i = f(x_1^i, x_2^i, x_3^i, \beta) + e_i \quad i = 1, ..., n$, where n is the number data points. For all points we write: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$. Using OLS $\boldsymbol{\beta}$ is estimated as: [2]

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}) \tag{1}$$

A first check of how and if the rainfall was dependent on the independent variables was to plot the observed rainfall against the covariates in the known points, see figure 1. X-distance seems to have some linear dependence, elevation and y-distance dependence is hard to distinguish in the figures.

Different models were tested, for each model tested $\hat{\boldsymbol{\beta}}$ was estimated using Least Squares and $\hat{\mathbf{Y}}_{valid}$ was calculated using our validation points. To decide which parameters to use the norm of the residual $||\hat{\mathbf{Y}}_{valid} - \mathbf{Y}_{valid}||_2$ was checked for different combinations. For example a linear dependence on all variables: $\mathbf{X} = \begin{bmatrix}\mathbf{1} & \boldsymbol{x_1} & \boldsymbol{x_2} & \boldsymbol{x_3}\end{bmatrix}$ was tested as well as a polynomial dependence on the elevation and no dependence on y-distance: $\mathbf{X} = \begin{bmatrix}\mathbf{1} & \boldsymbol{x_1}^2 & \boldsymbol{x_2}\end{bmatrix}$. The last one was one of the best models found, and this $\mathbf{X}$ is the one referred to in the rest of the text. Adding a dependence on y-distance did not make much difference, corresponding parameter was very small, and as can seen in figure 1, y-distance do not seem have a strong correlation with the rainfall.

Using this $\mathbf{X}$, x-coordinates and elevation for the known points, and equation 1 $\hat{\boldsymbol{\beta}}$ was estimated to: $\hat{\boldsymbol{\beta}} = \begin{bmatrix}1.6323 & -0.0477 & -0.0016\end{bmatrix}^T$. The rainfall in all the unknown points was estimated using the created model: $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. The reconstructed field is shown in figure 2.

From the plot the known values marked by the red circles does not actually have the same value as the estimates close to them which is bad. From a good model we should at least expect it to be able to estimate the values very close to our known values with good precision.

2.0.1 Variance of Least Squares

The variance of $\hat{\boldsymbol{\beta}}$ and the variance of the residuals is shown in the equation below[1].

$$
\begin{aligned}
V(\hat{\boldsymbol{\beta}}|\sigma_\epsilon^2) &= \sigma_\epsilon^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
\hat{\sigma_\epsilon^2} &= \frac{||\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}||_2^2}{n-p}
\end{aligned}
$$

This then provides an expression for the variance for each estimated expected value $\hat{\mu}_0 = \boldsymbol{x}_0\hat{\beta}$. Furthermore the variance for the prediction $y_0 = \mu_0 + \epsilon$.

$$
\begin{aligned}
V(\hat{\mu}_0) &= \boldsymbol{x}_0 V(\hat{\beta})\boldsymbol{x}_0^T \\
V(\hat{y}_0) &= \sigma_\epsilon^2 + \boldsymbol{x}_0 V(\hat{\beta})\boldsymbol{x}_0^T = \sigma_\epsilon^2 + V(\hat{\mu}_0)
\end{aligned}
$$

Using these expressions to compute the variance for the ordinary least squares estimate for ten validation points the estimates were compared with real values with the addition of the 95% confidence interval provided by taking $\hat{y}_0 \pm 1.96\sqrt{V(\hat{\mu}_0) + \sigma_\epsilon^2}$. The result from plotting theses is shown in figure 5a where the validation values are within the prediction interval of the estimated values showing that all though it is not completely accurate the real values are within the uncertainty. A plot of the variance for the estimated values using Ordinary Least Squares are shown in figure 6a. The standard deviation seems to be quiet high even for the estimations close or even on the know values which again confirms that the model is not that good since when we already know the value for a location the model should not have a high standard deviation at that location.

2.1 Universal Kriging

To study the spacial dependence in the errors a 95% confidence interval was created using bootstrap with 100 permutations. The confidence interval is plotted together with the non-parametric covariance function for the residuals from the Least Squares method in figure 3. To get a better estimation of $\hat{\boldsymbol{\beta}}$ it is re-estimated using the function function *covest_ml*. This gave $\hat{\boldsymbol{\beta}} = \begin{bmatrix}1.5546 & -0.0313 & -0.0016\end{bmatrix}^T$. The same function uses maximum likelihood and provides the parameters of a matern covariance function: $\kappa = 0.0719,\ \sigma^2 = 0.1696,\ \nu = 1.1117,$ and $\sigma_\epsilon^2 = 9.361 \cdot 10^{-17} \approx 0$. The fact the the variance of the nugget σ_ϵ^2 was so close to zero is questionable but since the model seemed to be good enough this was ignored. The parametric covariance estimate is given using κ, σ^2 and ν and the matern covariance function, also seen in figure 3. In the figure we see that there is an spacial dependence for short distances. The parametric covariance function follows the non-parametric and is much smoother, the matern covariance function fitted best. One can argue that there is a slightly negative correlation between 50-100 km, this can not be modeled with a matern covariance function. Since we can see that there is a spacial dependence - and because of this it is a good idea to model the residuals using Universal Kriging.

This is done by creating the covariance matrix, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_M + \boldsymbol{I}\sigma_\epsilon^2$, for all grid points, including the known points.[1] Using the estimated parameters κ, σ^2 and ν, $\boldsymbol{\Sigma}_M$ can be created with the matern covariance function and the distance matrix for all points. Separating the covariance matrix, $\mathbf{\Sigma}$, into $\mathbf{\Sigma}_{kk}$, $\mathbf{\Sigma}_{uu}$ and $\mathbf{\Sigma}_{uk}$ reconstruction of the unknowns can be done using Gauss-Markov theorem, giving us that the best linear unbiased predictions see equation 2[2]. The reconstructed field together with the known points is shown in figure 4.

$$\hat{\mathbf{Y}}_u = \mathbf{X}_u\hat{\boldsymbol{\beta}} + \mathbf{\Sigma}_{uk}\mathbf{\Sigma}_{kk}^{-1}(\mathbf{Y}_k - \mathbf{X}_k\hat{\boldsymbol{\beta}}) \tag{2}$$

$\mathbf{\Sigma}_{kk}$ is the covariance matrix for the known values, $\mathbf{\Sigma}_{uu}$ is the covariance matrix for the unknown values and the $\mathbf{\Sigma}_{ku}$ and $\mathbf{\Sigma}_{uk}$ are the cross-covariance matrices for the known and unknown respectively the unknown and the known values. $\boldsymbol{X}_u$ is the chosen data points for the grid-points that should be estimated $\hat{\boldsymbol{Y}}_u$ and the known data and rain values are $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$.

2.1.1 Variance of Universal Kriging

The conditional variance for the unknown values ,$\boldsymbol{Y}_u$, can be calculating with the following equation which takes the estimation of the parameter β into account [1]

$$V(\boldsymbol{Y_u}|\boldsymbol{Y_k},\boldsymbol{\theta}) = \mathbf{\Sigma}_{uu} - \mathbf{\Sigma}_{uk}\mathbf{\Sigma}_{kk}^{-1}\mathbf{\Sigma}_{ku} + (\boldsymbol{X}_u^T - \boldsymbol{X}_k^T\mathbf{\Sigma}_{kk}^{-1}\boldsymbol{X}_k)^{-1}(\boldsymbol{X}_u^T - \boldsymbol{X}_k^T\mathbf{\Sigma}_{kk}^{-1}\mathbf{\Sigma}_{ku})$$

The variance for the estimated values can then be found on the diagonal of the covariance matrix computed in the equation above. The square root of the variance is the standard deviation and these are shown in figure 6b. Comparing this plot with the standard deviation for the ordinary least squares shown in figure 6a the deviations are lower for Kriging case than for ordinary least squares. This is due to that in the Kriging case we refined our estimate of the parameters, $\hat{\beta}$ and also modeled the spatial dependence for the residuals which then provided more accurate estimates. The standard deviation seems to be zero close to the known points and becomes larger where an unknown point is far away from a known point. This is an improvement from the ordinary least squares and it is also more realistic that the estimate close should have lower standard deviation and especially the estimates on the known point should be close to zero which it is for the Kriging model. We see that is is especially large close to the edges where known points do not surround an unknown field which also agrees with what you would expect.

In the case of using Universal Kriging to model our spatial dependence not only is the standard deviation smaller but the estimates for the validation points shown in figure 5 are clearly more accurate then the estimates made from using ordinary least squares shown in figure 5a. The estimates which are displayed as the red line clearly is closer the the real values when applying Kriging versus using ordinary least squares. The real values are still within the prediction interval for the Kriging case as it is for the ordinary least squares as well, however the interval is more narrow for the Kriging estimates which show a lower uncertainty for the predictions.

3 Appendix A

3.1 Theory Questions

3.1.1

The predictions are linear in the observation, i.e that $\hat{\boldsymbol{Y}}_u = \boldsymbol{\lambda}^T \boldsymbol{Y}_k$ for some $\boldsymbol{\lambda}$. Where $\hat{\boldsymbol{Y}}_u$ is the estimate of the unknown precipitation of rain and $\boldsymbol{Y}_k$ is the know values:

The estimated vector is given from

$$\hat{\boldsymbol{Y}}_u = \mathbf{1}_u \hat{\beta} + \boldsymbol{\Sigma}_{uk} \boldsymbol{\Sigma}_{kk}^{-1} (\boldsymbol{Y}_k - \mathbf{1}_k \hat{\beta}) \tag{3}$$

and since $\hat{\beta}$ is given by

$$\hat{\beta} = (\mathbf{1}_k^T \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{1}_k)^{-1} \mathbf{1}_k^T \boldsymbol{\Sigma}_{kk}^{-1} \boldsymbol{Y}_k \tag{4}$$

Assuming a known covariance-matrix

$$\begin{bmatrix} \boldsymbol{Y}_k \\ \boldsymbol{Y}_k \end{bmatrix} \in \boldsymbol{N}(\begin{bmatrix} \mathbf{1}_k \beta \\ \mathbf{1}_u \beta \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{kk} & \boldsymbol{\Sigma}_{ku} \\ \boldsymbol{\Sigma}_{uk} & \boldsymbol{\Sigma}_{uu} \end{bmatrix}) \tag{5}$$

Evaluating equation 4 $\hat{\beta}$ shows that the first parenthesis in the expression just becomes a scalar, α and then the equation is reduced to $\hat{\beta} = \alpha \mathbf{1}_k^T \boldsymbol{\Sigma}_{kk}^{-1} \boldsymbol{Y}_k$. By doing the vector-matrix multiplication with $\mathbf{1}_k^T \boldsymbol{\Sigma}_{kk}^{-1}$ it then becomes a row-vector. This row-vector is then multiplied by the column-vector $\boldsymbol{Y}_k$ and this then become a linear combination of the entries of $\boldsymbol{Y}_k$ and can be written as $\sum_{i=0}^{n} a_i Y_i$, n being the length of $\boldsymbol{Y}_k$. Putting this result of $\hat{\beta}$ into equation 3 the whole equation can be seen as a linear combination of $\boldsymbol{Y}_k$ and therefore be written as $\boldsymbol{\lambda}^T \boldsymbol{Y}_k = \sum_{i=0}^{n} \lambda_i Y_i$. Where $\boldsymbol{\lambda}^T$ is given by equation 6

$$\boldsymbol{\lambda}^T = (\mathbf{1}_k^T \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{1}_k)^{-1} \mathbf{1}_k \boldsymbol{\Sigma}_{kk}^{-1} + \boldsymbol{\Sigma}_{uk} \boldsymbol{\Sigma}_{kk}^{-1} (\boldsymbol{I} - (\mathbf{1}_k^T \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{1}_k)^{-1}) \mathbf{1}_k^T \boldsymbol{\Sigma}_{kk}^{-1}) \tag{6}$$

3.1.2

Show that the predictions are unbiased, i.e. that $\boldsymbol{E}(\hat{\boldsymbol{Y}}_u) = \boldsymbol{E}(\boldsymbol{Y}_u)$.

Firstly using the result from evaluating $\hat{\beta}$ in question 1 is used to calculate

$$\boldsymbol{E}(\hat{\beta}) = \boldsymbol{E}(\alpha \mathbf{1}_k \boldsymbol{\Sigma}_{kk}^{-1} \boldsymbol{Y}_k) = \alpha \mathbf{1}_k \boldsymbol{\Sigma}_{kk}^{-1} \boldsymbol{E}(\boldsymbol{Y}_k) = \alpha \mathbf{1}_k \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{1}_k \beta$$

The last equality is given from the expected value in equation 5. And inserting the expression for the scalar alpha then gives

$$\boldsymbol{E}(\hat{\beta}) = \alpha \mathbf{1}_k \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{1}_k \hat{\beta} = (\mathbf{1}_k \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{1}_k)^{-1} (\mathbf{1}_k \boldsymbol{\Sigma}_{kk}^{-1} \mathbf{1}_k) \beta = \beta \tag{7}$$

Using this and the expression in equation 3 the expected value, $\boldsymbol{E}(\hat{\boldsymbol{Y}}_u)$ can then be shown to be unbiased:

$$\boldsymbol{E}(\hat{\boldsymbol{Y}}_u) = \mathbf{1}_u \boldsymbol{E}(\hat{\beta}) + \boldsymbol{\Sigma}_{uk}\boldsymbol{\Sigma}_{kk}^{-1}(\boldsymbol{E}(\boldsymbol{Y}_k) - \mathbf{1}_k \boldsymbol{E}(\hat{\beta})) = \mathbf{1}_u\beta + \boldsymbol{\Sigma}_{uk}\boldsymbol{\Sigma}_{kk}^{-1}(\mathbf{1}_k\beta - \mathbf{1}_k\beta) = \mathbf{1}_u\beta = \boldsymbol{E}(\boldsymbol{Y}_u)$$

3.1.3

Show that the predictor is the minimum variance by comparing it to another unbiased predictor $(\boldsymbol{\lambda} + \boldsymbol{\nu})^T \boldsymbol{Y}_k$ by proving showing the following to be true. Expressed in words this means that the prediction errors is uncorrelated from the observations and therefore the observations can add no further information

$$\boldsymbol{\nu}^T \boldsymbol{C}(\boldsymbol{Y}_k, \hat{\boldsymbol{Y}}_u - \boldsymbol{Y}_u) = 0$$

$$\begin{aligned}
&\boldsymbol{\nu}^T \boldsymbol{C}(\boldsymbol{Y}_k, \hat{\boldsymbol{Y}}_u - \boldsymbol{Y}_u) = \boldsymbol{C}(\boldsymbol{\nu}^T\boldsymbol{Y}_k, \hat{\boldsymbol{Y}}_u - \boldsymbol{Y}_u) = \boldsymbol{C}(\boldsymbol{\lambda}^T\boldsymbol{Y}_k - \boldsymbol{Y}_u, \boldsymbol{\nu}^T\boldsymbol{Y}_k) = \\
&\boldsymbol{C}(\boldsymbol{\lambda}^T\boldsymbol{Y}_k, \boldsymbol{\nu}^T\boldsymbol{Y}_k) - \boldsymbol{C}(\boldsymbol{Y}_u, \boldsymbol{\nu}^T\boldsymbol{Y}_k) = \boldsymbol{\lambda}^T\boldsymbol{C}(\boldsymbol{Y}_k, \boldsymbol{Y}_k)\boldsymbol{\nu} - \boldsymbol{C}(\boldsymbol{Y}_u, \boldsymbol{Y}_k)\boldsymbol{\nu} = \\
&(\boldsymbol{\lambda}^T\boldsymbol{\Sigma}_{kk} - \boldsymbol{\Sigma}_{uk})\boldsymbol{\nu}
\end{aligned}$$

To continue the calculation the result from showing the linearity in the first question is used and more specifically the expression for $\boldsymbol{\lambda}^T$ from equation 6. This together with that the prediction has to be unbiased, $\boldsymbol{E}(\hat{\boldsymbol{Y}}_u) = \boldsymbol{E}(\boldsymbol{Y}_u)$:

$$\boldsymbol{E}(\hat{\boldsymbol{Y}}_u) = \boldsymbol{E}((\boldsymbol{\lambda} + \boldsymbol{\nu})^T\boldsymbol{Y}_k) = \boldsymbol{E}(\boldsymbol{\lambda}^T\boldsymbol{Y}_k) + \boldsymbol{E}(\boldsymbol{\nu}^T\boldsymbol{Y}_k) = \boldsymbol{E}(\boldsymbol{Y}_u) + \boldsymbol{\nu}^T\mathbf{1}_k\beta$$

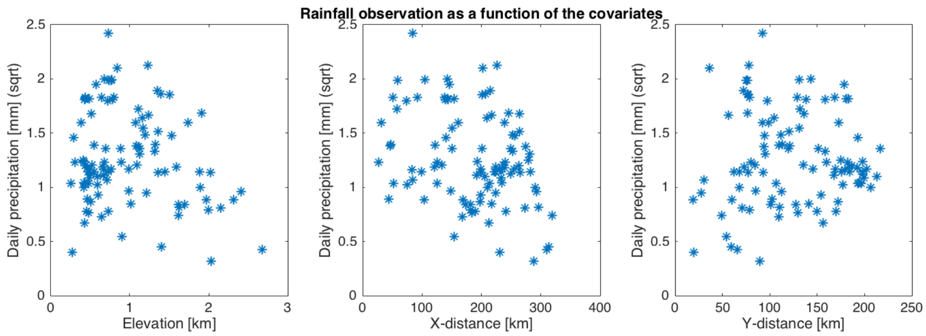
This then give the condition that $\boldsymbol{\nu}^T\mathbf{1}_k = 0$ which is equal to $\mathbf{1}_k^T\boldsymbol{\nu} = 0$.

$$\begin{aligned}
&(\boldsymbol{\lambda}^T\boldsymbol{\Sigma}_{kk} - \boldsymbol{\Sigma}_{uk})\boldsymbol{\nu} = \\
&\Big((\mathbf{1}_k^T\boldsymbol{\Sigma}_{kk}^{-1}\mathbf{1}_k)^{-1}\mathbf{1}_k\boldsymbol{\Sigma}_{kk}^{-1} + \boldsymbol{\Sigma}_{uk}\boldsymbol{\Sigma}_{kk}^{-1}(\boldsymbol{I} - (\mathbf{1}_k^T\boldsymbol{\Sigma}_{kk}^{-1}\mathbf{1}_k)^{-1})\mathbf{1}_k^T\boldsymbol{\Sigma}_{kk}^{-1}\Big)\boldsymbol{\Sigma}_{kk} - \boldsymbol{\Sigma}_{uk})\boldsymbol{\nu} = \\
&\boldsymbol{\Sigma}_{uk}\boldsymbol{\nu} - \boldsymbol{\Sigma}_{uk}\boldsymbol{\nu} = 0
\end{aligned}$$

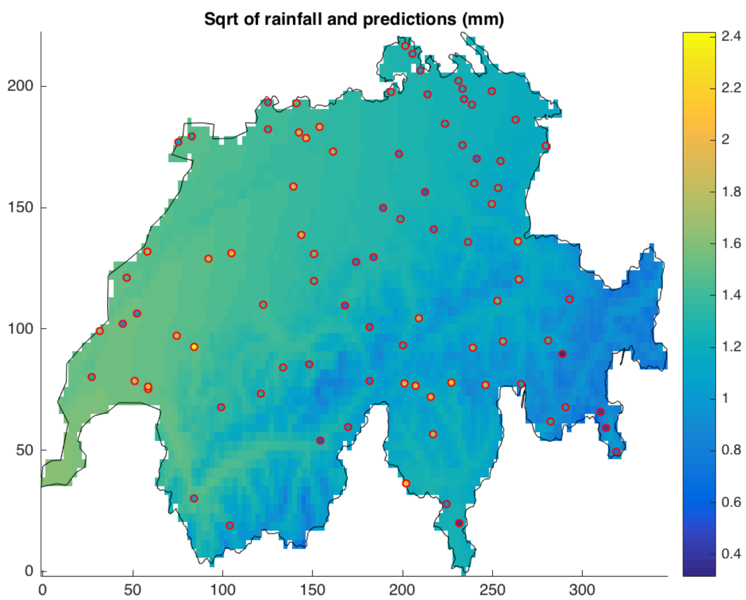
This then proves the initial statement that the prediction error is uncorrelated to the observations and therefore the observation can add no further information and we have the optimal predictions.

4 Appendix B

4.1 Ordinary least squares

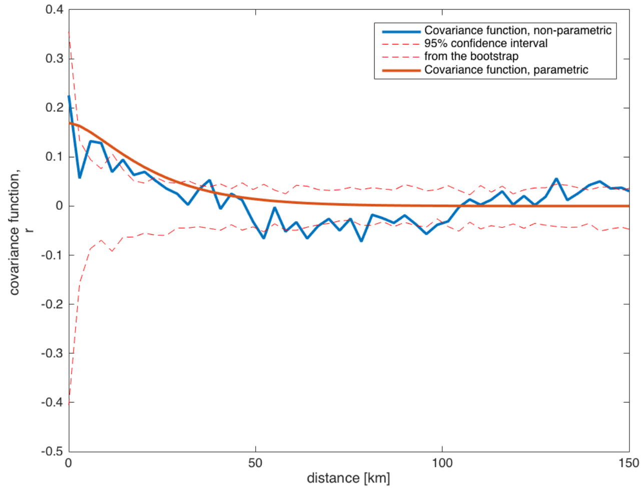


Figur 1: Observed rainfall, with a square root transform, as a function of elevation, x-distance and y-distance.

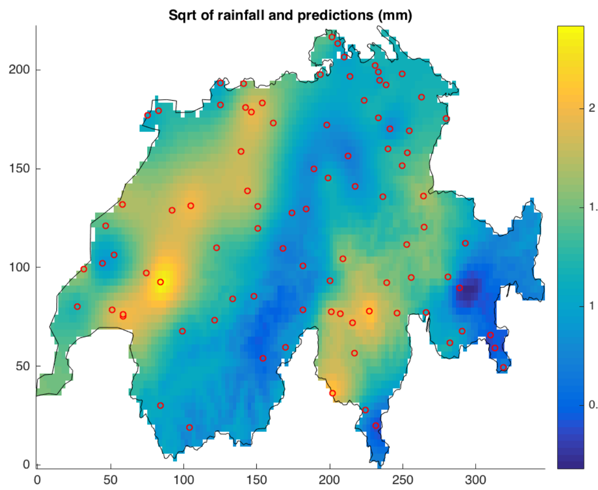


Figur 2: Prediction of the rainfall using ordinary least square. Note that the result is in square root of millimeter rain and shown by the colors where the color bar gives us the values. The circles contains the known true values.

4.2 Universal Kriging

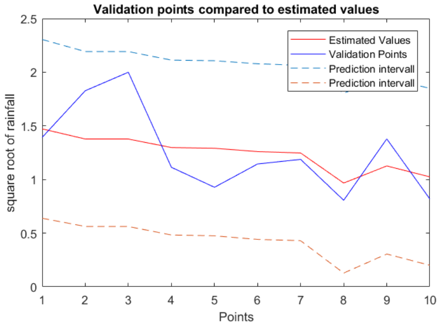


Figur 3: Parametric and non-parametric covariance estimations for the residuals, along with a 95% confidence interval created using bootstrap approach with 100 permutations. This shows a significant spacial dependence in the residuals since for some values the covariance is outside of the confidence interval.

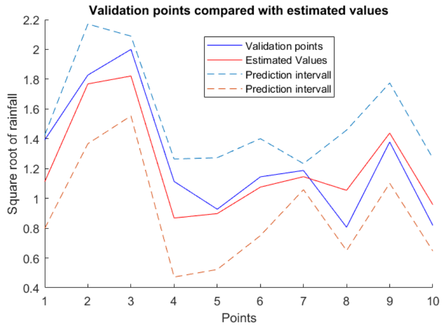


Figur 4: The predictions of rainfall using kriging. Note that the result is in square root of millimeter rain and shown by the colors where the color bar gives us the values. The circles contains the known true values.

4.3 Validation comparison



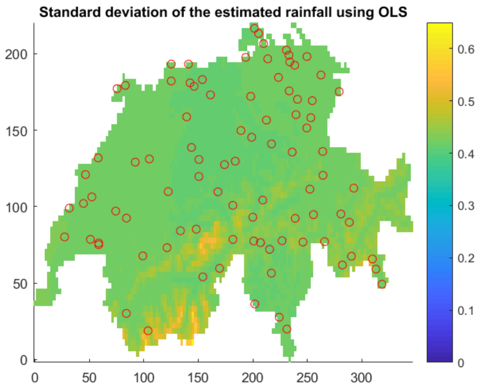
(a) Estimated values using least squares compared to the actual value of the 10 points.



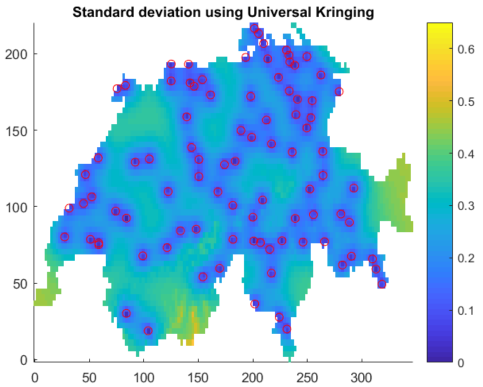
(b) The estimated values using universal Kriging compared to the actual 10 values validation points.

Figur 5: Comparison between the estimates for the validation points using ordinary least squares and then universal Kriging

4.4 Standard Deviation comparison



(a) The standard deviation for the estimates from using Ordinary Least Squares



(b) The standard deviations for the estimates from using universal Kriging.

Figur 6: Comparison between the standard deviation for the estimates from using Least Squares, left and Kriging, right. The color bar shows the standard deviation of the square root of precipitation rain in millimeters.

Referenser

[1] J. Lindström, Mathematical Sciences, LTH, *Spatial Statistics with Image Analysis Lecture 4*.

[2] J. Lindström, Mathematical Sciences, LTH, *Spatial Statistics with Image Analysis Lecture 3*