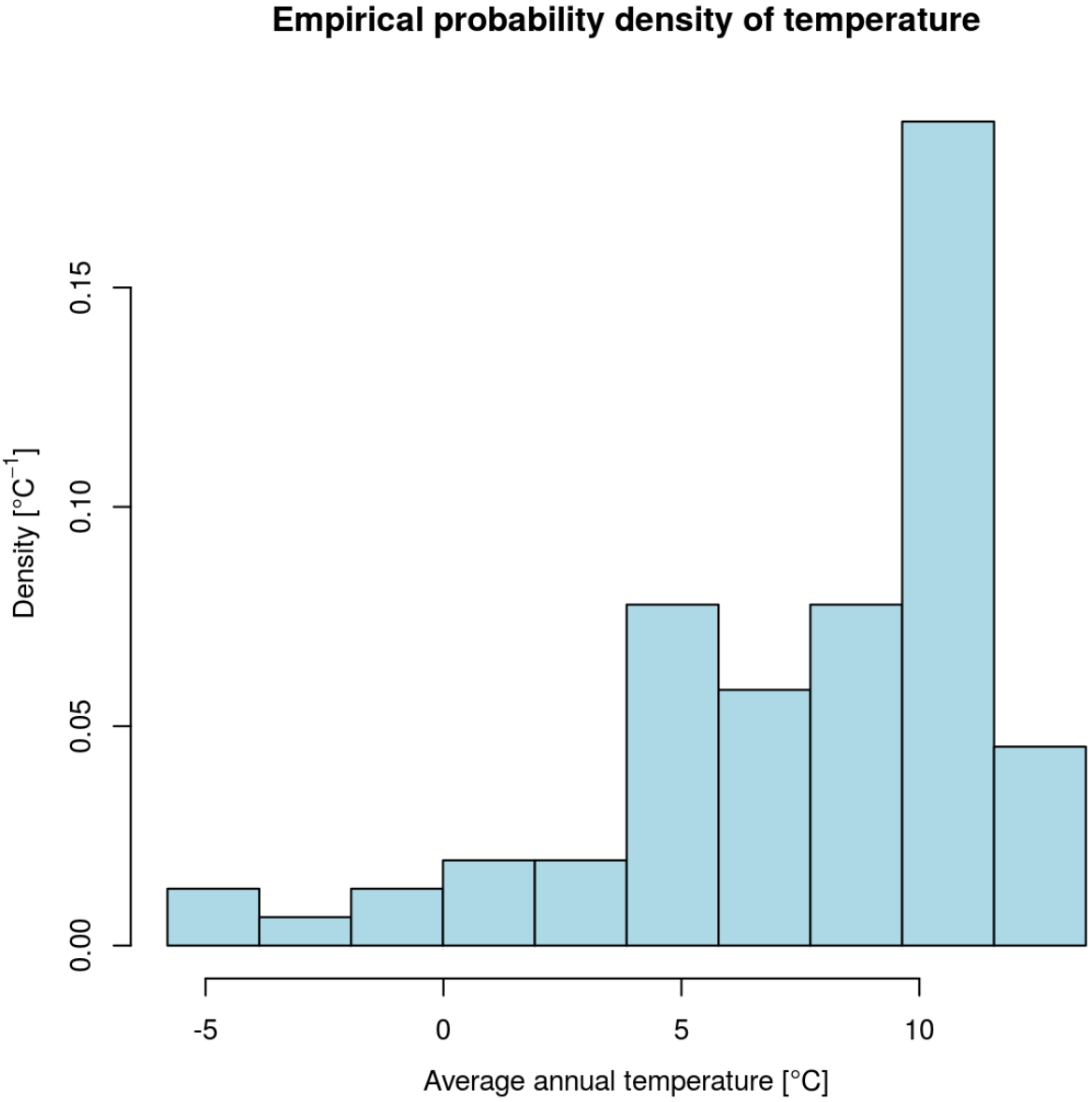
Exercise 5 - Universal kriging - Solutions

Office hours: Friday 09:00-12:00

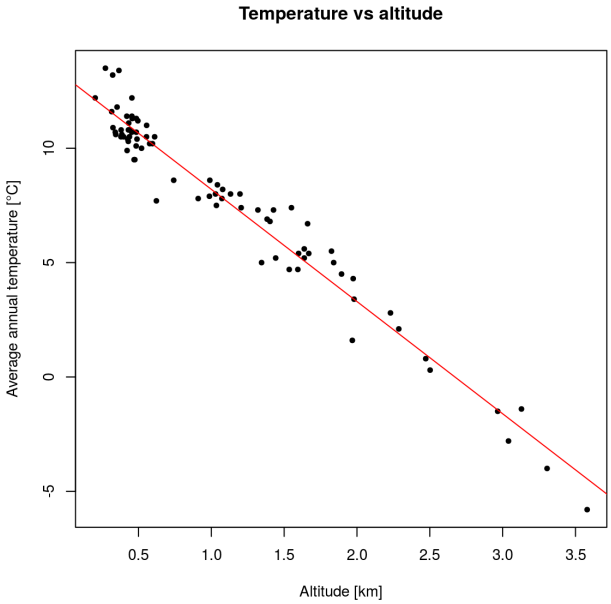
A. Berne alexis.berne@epfl.ch
G. Ghiggi gionata.ghiggi@epfl.ch
M. Guidicelli matteo.guidicelli@epfl.ch

1. *Read the prediction dataset and plot the empirical probability density function (pdf) of the temperature values. Is normality a good assumption?*



Normality is not a good assumption. The distribution is negatively skewed. The mean is smaller than the median. The classical sample variogram (e.g., Matheron) might not be adequate.

2. *Plot the average annual temperature with respect to the altitude. Is there a trend?*

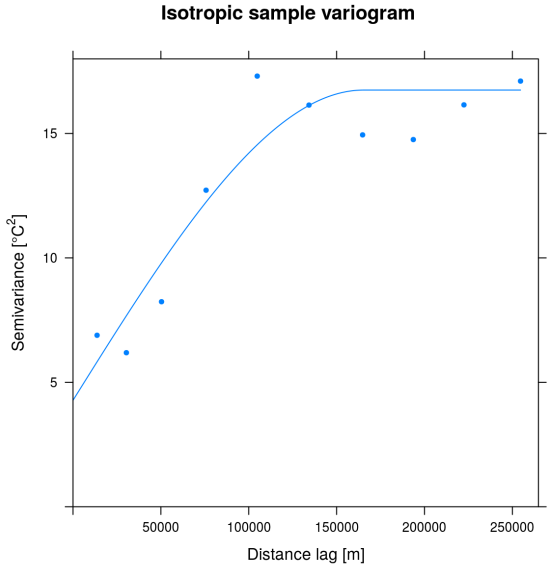


Yes, there is a strong negative linear trend between the temperature and the altitude, as shown by the strong negative correlation coefficient between these variables (-0.978).

3. *Use the function "lm()" to fit a linear regression model for the temperature vs the altitude. What are the coefficients (intercept and slope) of the fitted model and the residual standard error?*

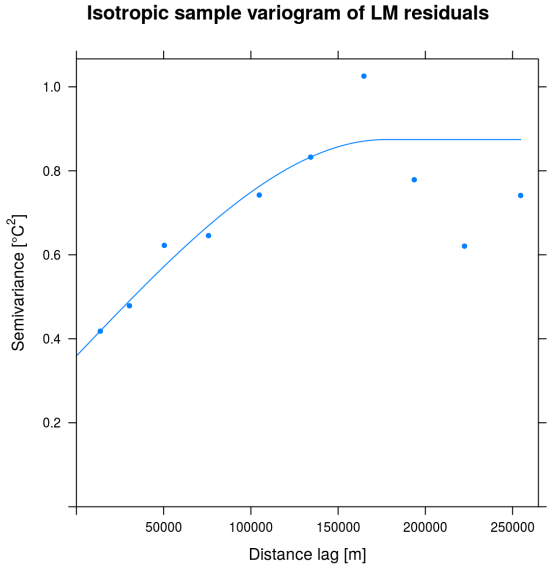
   The fitted intercept is 13.109 °C, the slope is -4.906 °C km^{-1} and the residual standard error 0.866 (°C)^2.

4. *Compute the isotropic sample variogram of the temperature values using Cressie's robust estimator and the boundaries provided in the code. Fit a spherical model on it. Show the sample variogram and the fitted model on the same graph. What are the values of the fitted nugget, sill and range?*



   The fitted nugget is 3.972 $(°C)^2$, the sill 16.755 $(°C)^2$ and the range 161.9 km.

5. *Compute the isotropic sample variogram of the temperature values using Cressie's robust estimator and a linear trend with respect to the altitude. Fit a spherical variogram model on it and provide the fitted values of the nugget, the sill and the range. What can you say about the new values of the nugget and the sill?*



The fitted nugget is 0.370 $(°C)^2$, the sill 0.873 $(°C)^2$ and the range 180.2 km. Both the nugget and the sill are significantly smaller than before. The sill represents the variance of the residuals after subtracting the linear trend w.r.t. the altitude. It is very close to the residual standard error computed in 3.

6. *Interpolate the temperature values at the locations defined in the validation dataset.*

   (a) *using ordinary kriging (OK) and the variogram defined in 4.*

   (b) *using the linear model (LM) defined in 3.*

   (c) *using universal kriging (UK) and the variogram defined in 5.*

   Compare the interpolated values with the temperature measurements of the validation dataset. Compute the bias and the root mean squared error (rmse) of the predicted values for each method. Which method performs best? And which performs worst? Explain why!

	bias [°C]	rmse [°C]
OK	-0.540	3.083
LM	0.289	0.767
UK	0.191	0.695

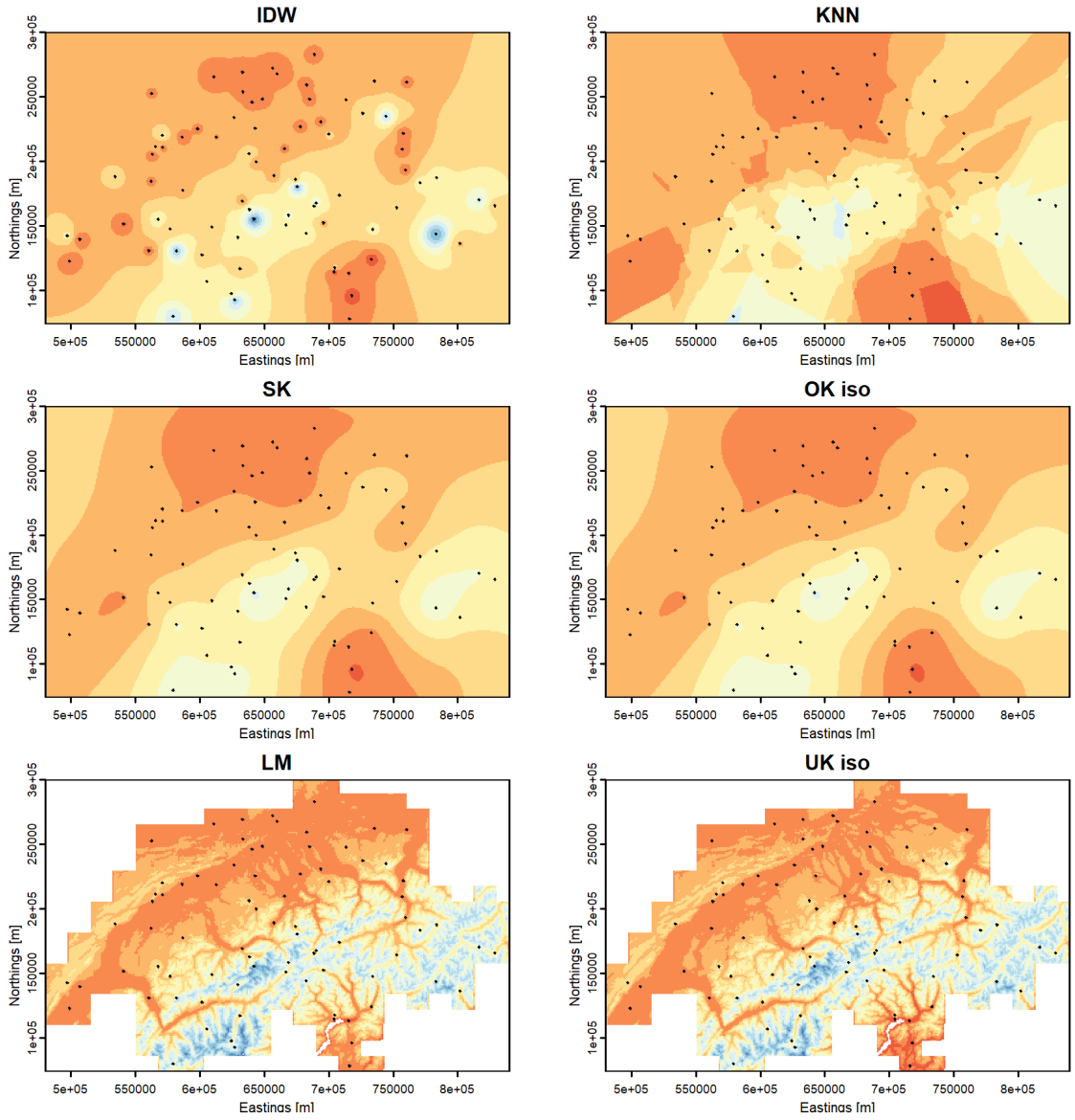
   The method that performs best is universal kriging (UK). It takes into account the linear relationship between the temperature and the altitude and the spatial structure of the residual errors. It clearly has the smallest values of bias and rmse. The method that performs worst is ordinary kriging (OK) because it does not take into account the strong linear trend of the temperature with respect to the altitude and assumes a constant mean over the entire domain.

7. *Read the provided 1-km digital elevation model (DEM) for Switzerland and interpolate the temperature values at the locations given by the DEM.*

   (a) *using ordinary kriging (OK)*

   (b) *using universal kriging (UK)*

   Plot the maps with the interpolated temperature values over Switzerland. What are the main differences between the considered interpolation methods?



Ordinary kriging does not produce realistic results because it assumes that the average temperature is constant at all altitudes. It therefore overestimates the temperatures in the Alps and underestimates them at low altitudes and in the valleys. Universal kriging better takes into account this temperature gradient with the altitude and therefore produces much more realistic results.