DIFFERENT METHODS TO PRODUCE DISTRIBUTED SOIL DEPTH MAPS

DOCUMENTATION

by Livio Dreyer

Seminar geodata modeling

Institute of Geography

University of Bern

2020

Content

List of Figures

List of Tables

1. Introduction

Numerical models are widely used as predictive and analytical tools for hazardous surface processes (Zhang et al., 2016). A main challenge of hydrological and slope stability modelling is the availability and spatial representation of the of input data and parameters, as such models require detailed datasets about the spatial distribution of the input parameters (Van Westen et al., 2008). Soil depth, defined as depth to bedrock or the depth to the first marked layer with different hydrological properties, is among the most important factor controlling shallow landsliding (Catani et al., 2010). However, soil depth is one of most difficult to measure and least understood physical variable in slope stability modelling (Kuriakose, 2006). In catchment scale models, many researchers still apply straightforward and simplistic solutions by considering a constant soil depth value for the entire catchment, assigning one value for each geological or geomorphological unit or deriving a distributed soil depth map from a single topographic factor such as slope gradient, curvature or elevation (Kuriakose et al., 2009; Segoni et al., 2012; Tesfa et al., 2009). More complex methods such as process based methods (Catani et al., 2010; von Ruette et al., 2013) or geostatistical methods (Kuriakose et al., 2009; Szabó et al., 2019) are applied less frequently, as these approaches are time consuming to calibrate and are very site specific.

2. Objectives

Physically based hydrological and slope stability models often require soil depth as a spatially distributed input parameter. In order to generate distributed soil depth maps geostatistical approaches, a machine learning and an empirically based model are employed in this project and compared.

3. Data and methods

The Edisried-and Totenbühl catchment in the Canton of Obwalden, Switzerland, are used to predict soil depth in the corresponding catchment. The input datasets consists of a total of 194 soil depth measurements which are used as input date for point interpolation and model calibration. 140 measurements originate from the landslide inventory of WSL and 54 additional measurements have been conducted by the author (Fig.1). Auxiliary variables used in regression kriging, random forest classification, and the GIST model are derived from a digital elevation model (DEM) by Swisstopo. The DEM of an original resolution of 2m has been resampled by bilinear interpolation method in ESRI Arc GIS to 5 m resolution and the topographic covariates have been derived in SAGA GIS. The used geological map consists of the "Geologischer Atlas der Schweiz 1:25'000" by Swisstopo.

The following section will give a brief overview of the methods used to predict soil depth in the catchment. It is not intended to give a full and in depth description of the methods. References for each method are given and it is strongly recommended to consult theses references in order to understand the methods before applying the codes presented in this work. The presented codes are adopted from Hengl (2009); Rossiter, (2013) for ordinary/regression kriging, Malone et al. (2017) for the random forest model and the code for the GIST model has been constructed based on the work of (Catani et al., 2010).

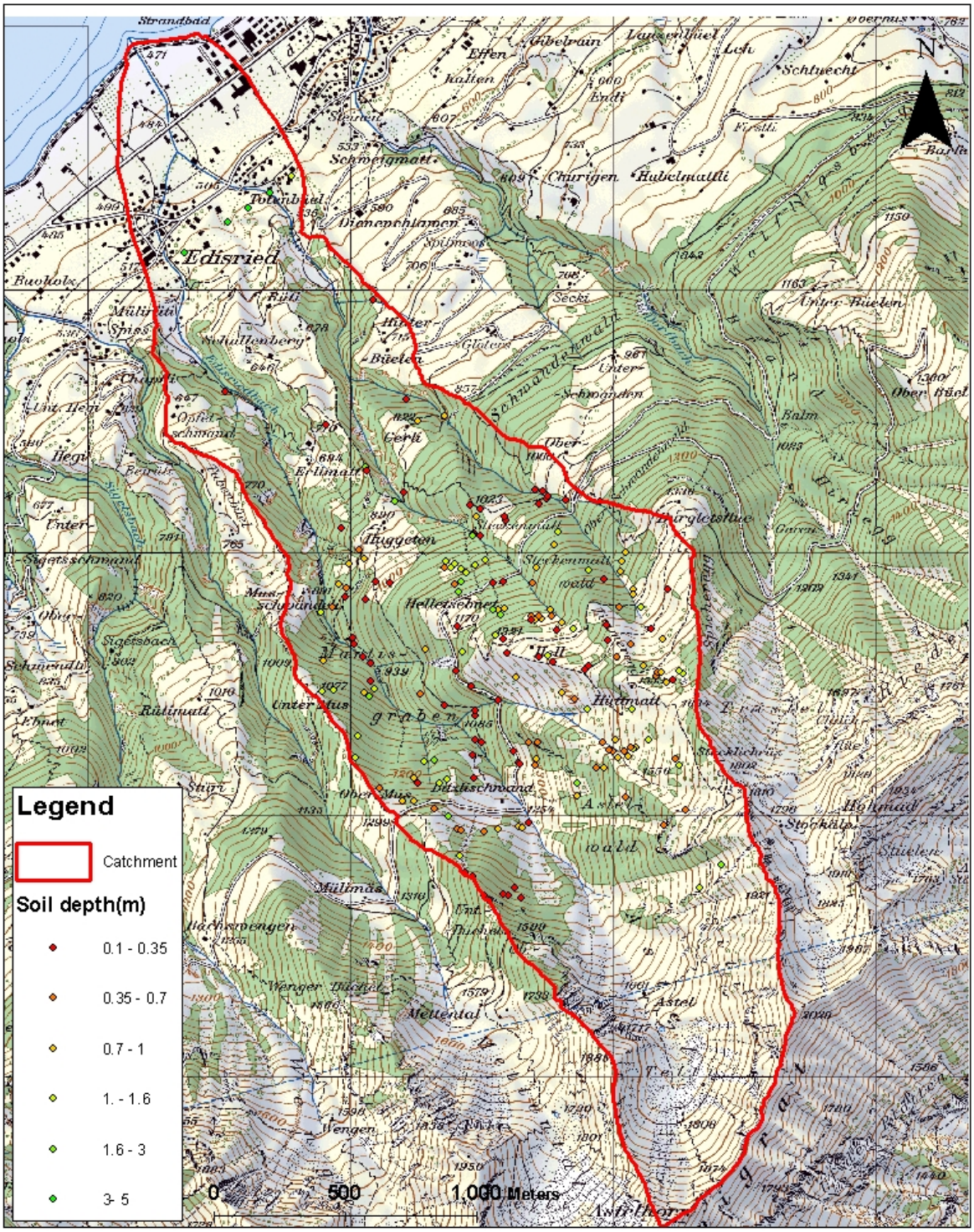

Fig. 1 Overview of catchment and soil depth measurements (dots)

3.2 Geostatistical methods

Geostatistical mapping is defined as the" analytical production of maps by using field observations, auxiliary information and a computer program that calculates values at locations of interest" (Hengl et al., 2007). Based on the theory of random fields, geostatistical methods take us of the spatial autocorrelation of a parameter in order to create a distributed map of the parameter of interest. Hengl, 2009 or Rossiter, 2013 provide a good introduction to geostatistics.

3.2.1 Ordinary Kriging

Ordinary kriging is the simplest form of kriging. In this approach the variable which has to be interpolated is assumed to be stationary. Thereby, the variable d is predicted by the ordinary kriging approach at the location Xi by:

$$d(Xi)=m+e\ (Xi)$$

where m is the regional mean and e(Xi) is a spatially random component which is estimated from the variogram (Rossiter, 2013).

3.2.2 Regression Kriging

Regression Kriging is a hybrid approach which considers both the local and long range data structure. Thereby, regression on auxiliary variables are used and are combined with simple kriging with a known mean to interpolate the residuals (Hengl et al., 2007).

3.3 Random Forest classification

Random forest is a nonparametric multivariate technique which belongs to the machine learning algorithms. The random forest technique is based on a combination of tree structured classifiers. For each node the best split is chosen of a random subset of the training data. Large numbers of trees are generated and the most popular class is chosen subsequently (Breiman, 2001). The theory behind random forests is comprehensively outlined in Breiman (2001).

3.4 Empirical-geomorphological based approach (GIST)

The geomorphologically Index Soil Thickness model (GIST) is based on three geomorphometric attributes and on geomorphologic and geological criteria. The model assumes that soil depth is linked to the relative position of the hillslope, curvature, and a defined threshold where mass movements may occur. These three factors are converted into the range of 0 to 1, where 1 indicate deep soils and 0 shallow soils. Each pixel is directly influenced by these factors, however the direct relationship of soil depth and the factors is not constant for the whole catchment. Therefore, the catchment can by subdivided into smaller land units and a specific relationship to the morphometric factors for each unit is assigned. Lastly, the soil depth estimates derived from topographic factors are calibrated for each lithological unit based on a few soil depth measurements. A full explanation of the method is found in Catani et al. (2010).

4. Code description

4.1 Ordinary and regression kriging

4.1.1 Model setup

The ordinary and regression kriging are executed in the R environment. The current code uses version 3.5.3. The following packages should be installed prior execution the code: gstat, maptools, rgdal, sp, raster, ithir.

4.1.2 Data preparation

Soil depth measurements should be prepared in a CSV file, where the first two columns consists of the x,y coordinates of the soil depth measurements and the third column is the actual soil depth measure (Fig.2). Regression kriging requires several distributed covariates for the soil depth prediction. The choice of covariates is site specific and depends on the correlation of the covariate and the soil depth and their availability. In the literature several covariates have been successfully used for predicting soil depth using regression kriging. The most common covariates reported in the literature are topographic derived variables such as slope, curvature, aspect, and distance from the stream network, topographic wetness index (TWI), land use data or geological/geomorphological land units. All the covariates raster maps should be in ASCII Grid format.

	POINT_X	POINT_Y	Soil_depth
1	2661437	1188923	1.20
2	2661417	1188938	1.20
3	2661395	1188958	1.10
4	2661360	1188968	1.20
5	2661379	1188939	1.20
6	2661357	1188940	0.65
7	2661417	1188889	1.80
8	2661395	1188877	0.40
9	2661409	1188849	2.00
10	2661570	1188760	2.00
11	2661559	1188782	0.80
12	2661541	1188780	2.00
13	2661581	1188786	1.00
14	2661547	1188683	1.50
15	2661470	1188950	1.30
16	2661773	1189031	0.80
17	2661787	1189078	1.10
18	2661902	1188934	1.10
19	2661963	1188897	0.10

Fig. 2 Soil depth table preparation

4.1.3 Running the model

Section 1: Load the required packages

Section 2: Read the depth file and make it a spatial data frame. Check if the soil depth measurements are normally distributed as this is a prerequisite of kriging. In this case (Fig.3) the data is not normally

distributed and the data needs to be log-transformed in order to be normally distributed.

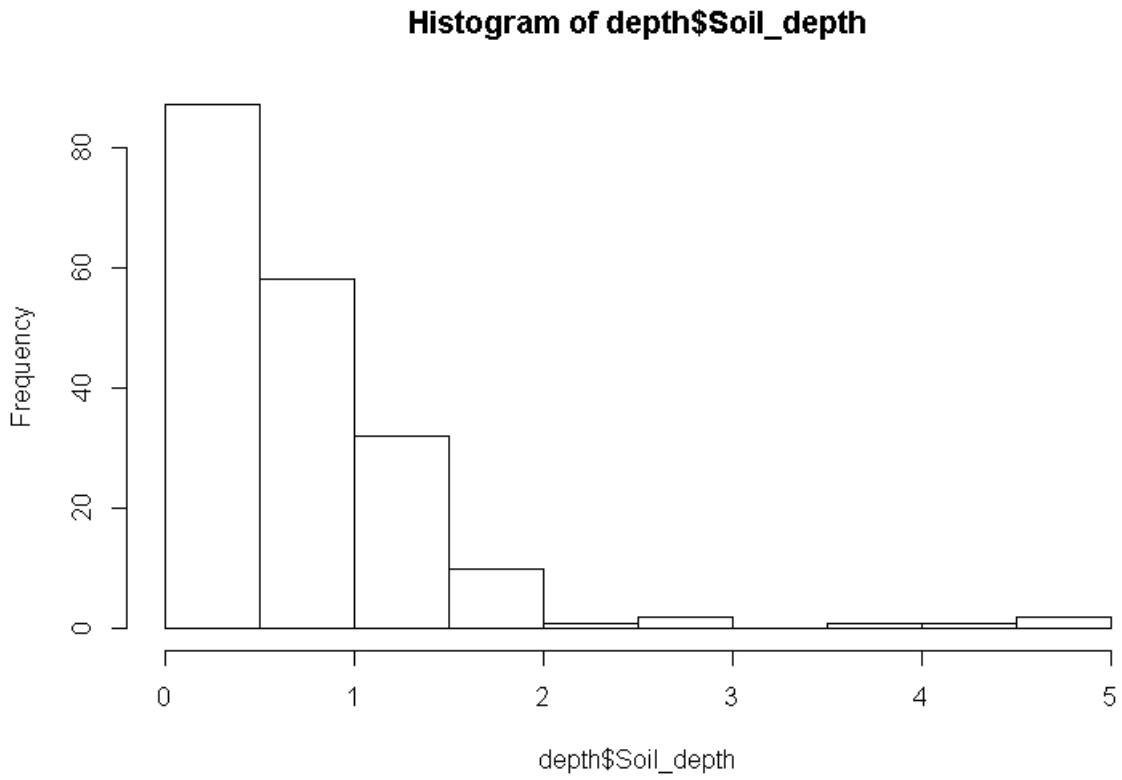

Fig. 3 Histogram of soil depth distribution

Section 3: Reads all the covariates raster maps. Categorical raster map such as a geological map where the raster values indicates a certain category and not a continues value such as slope are set to factor

Section 4: Values of covariate raster maps are copied at soil depth measurement locations and added to the soil depth file

Section 5: The soil depth file containing all soil depth values and covariates is converted into a data frame. Categorical values are set as factor. No data values in the data frame are identified and removed later. Then the data frame is converted to a spatial data frame and the ESPG projection code is added (in this case LV95).

Section 6: The linear multiple regression kriging model is fitted from the available covariates. The residuals of the fitted model are plotted to check if the residuals are normally distributed. If the residuals are not normally distributed regression kriging cannot be applied.

Section 7: This section serves to explore the correlation of the soil depth and the covariates.

Section 8: The significant covariates are chosen automatically by the step function and the covariates used for regression kriging are displayed

Section 9: The parameters nugget, range and sill are first fitted by eye to model the empirical variogram. Depending on the shape of the variogram one of the following variogram shapes should be chosen for “model” (Fig.4):

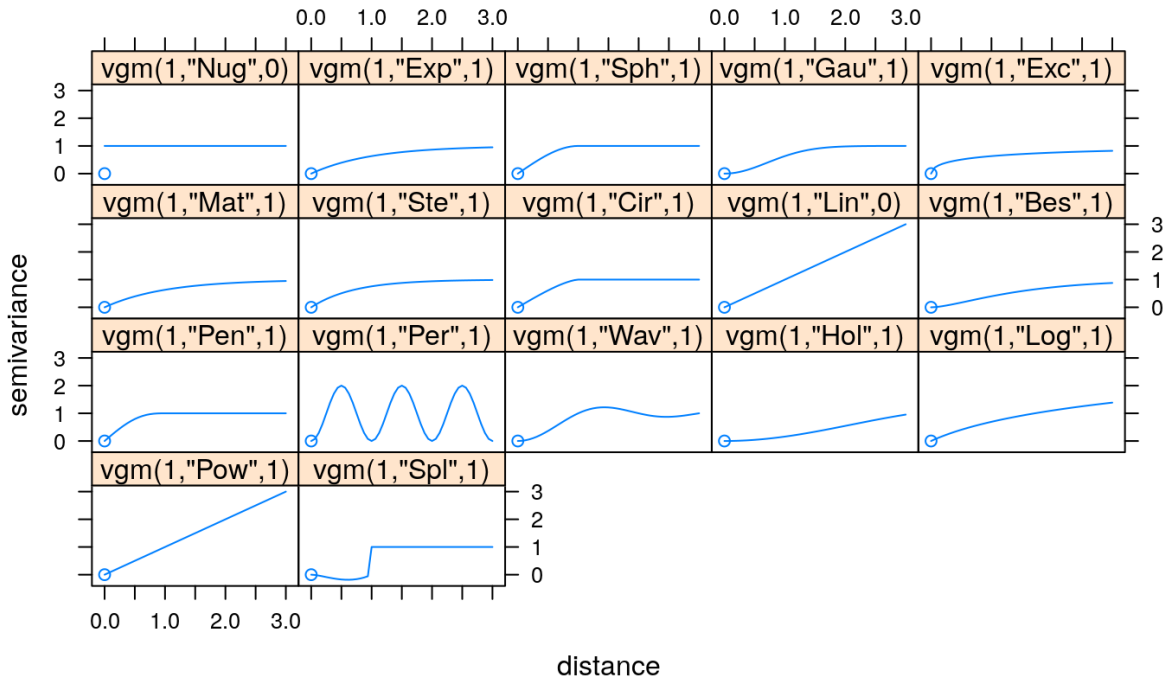

Fig. 4 Different variogram shapes available in the gstat package

Based on the empirical variogram the variogram is automatically fitted. In the final step of this section the ordinary kriging prediction is made

Section 10: The variogram for the residuals is equally modelled and fitted as outlined in section 9. Afterwards the regression kriging prediction is performed

Section 11: Due to the log-transformation of the soil depth measurements in section 2 the prediction of ordinary and regression kriging are back transformed

Section 12: Cross validation is performed and the RMSE is calculated in order to the asses the model performance

Section 13: The final model predictions of ordinary and-regression kriging are exported and written to the disk in the ASCII Grid format.

4.2 Random forest classification

4.2.1 Model setup

The random forest classification is executed in the R environment. The current code uses version 3.5.3. The following packages should be installed prior execution the code: ithir, raster, rgdal, sp and random forest

4.2.2 Data preparation

The data preparation for the random forest classification is the same as described in 4.1.2

4.2.3 Running the model

Section 1: Load the required packages

Section 2: Read the depth file and make it a spatial data frame

Section 3: Read all the covariates add save them in a raster stack

Section 4: Extract covariate values at depth location and save the values in a data frame.

Section 5: The data frame is checked for NA values are removed if found. Categorical variables are set as factor

Section 6: Builds the random forest model. In first step the data frame values are randomly shuffled and 70% of the data are used for training the model, the rest is used for validation purposes. After the model is build with the random forest package and the model summaries are displayed. An important output is the covariate importance, which can be displayed by varImpPlot (soildepth.RF) (Fig 5)

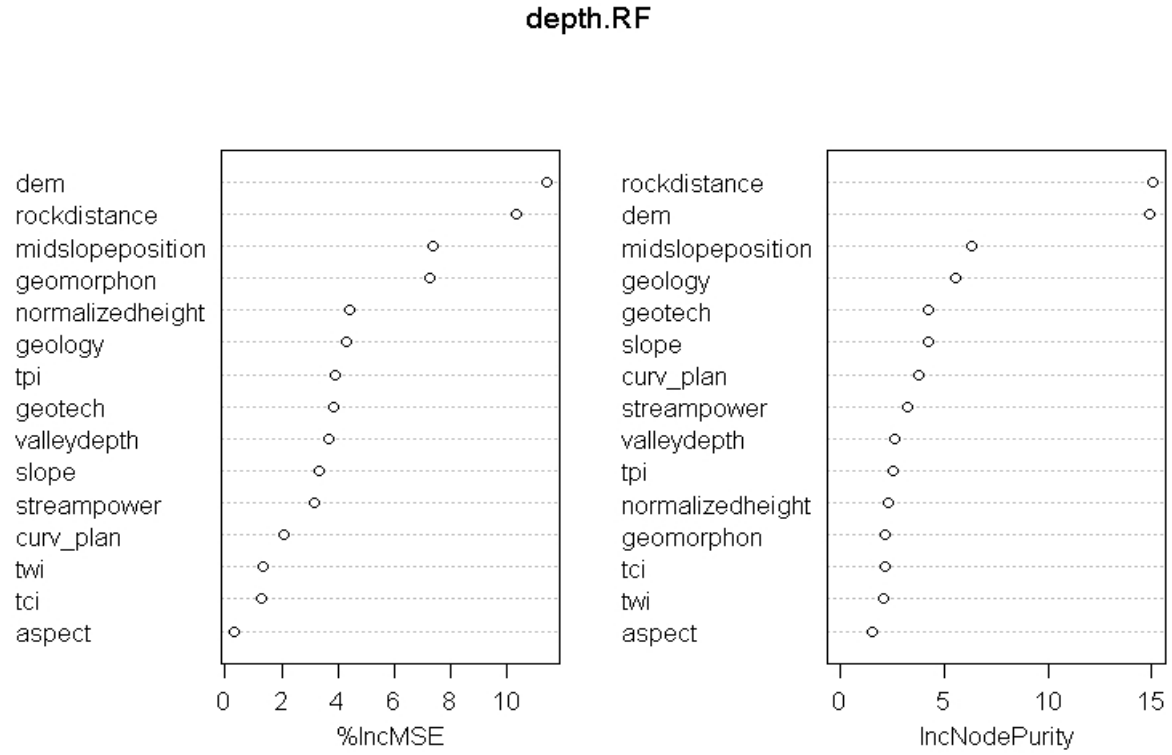

Fig. 5 Covariate importance

Section 7: Model accuracy is calculated both for the calibration dataset and the validation dataset which has been prepared in section 6

Section 8: The random forest model is applied spatially and the results are written to the disk and plotted.

4.3 GIST model

4.3.1 Model setup

The GIST model is executed in the python environment. The current code uses the Python version 3.6.9. The following packages should be installed prior running the code: nympy and pcraster

4.3.2 Data preparation

The model requires the following raster maps: slope map, curvature map, topographic position index (TPI), geological map, and a depth raster. Slope, curvature and TPI can be calculated easily in SAGA GIS from a digital elevation model. The soildepth-TPI trend can be obtained by the following way: The TPI values are extracted at the soil depth locations (e.g by the " Extract values by Point" tool of ArcGIS) and are exported as. dbf file in order to open the table in Excel. In Excel the soil depth and TPI should

be normalized between 0-1 by the following function: TPIi=xi−min(x)max(x)−min(x). Afterwards, the soil depth values are plotted against the TPI values in order to fit a trend line. In this example a fourth order polynomial function is best suited to describe the DTB-TPI relationship.

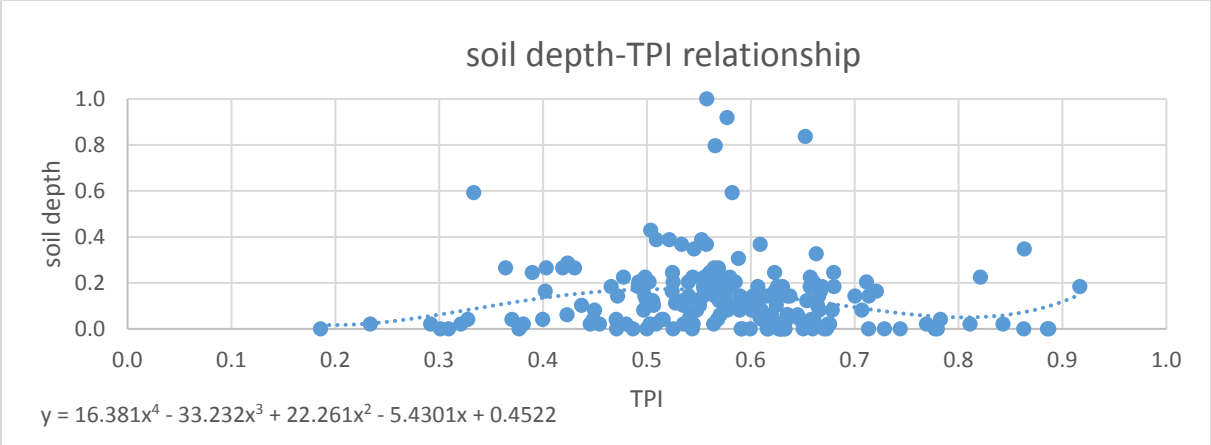

Fig. 6 Soil depth-TPI relationship

The depth raster should consist of a raster, where the measured soil depth is represented by one single pixel value. All raster cells which do not consist off a soil depth measurements should be set to "NoData" (Fig.7). All input raster maps should be in the .map format, which can be done with GDAL.

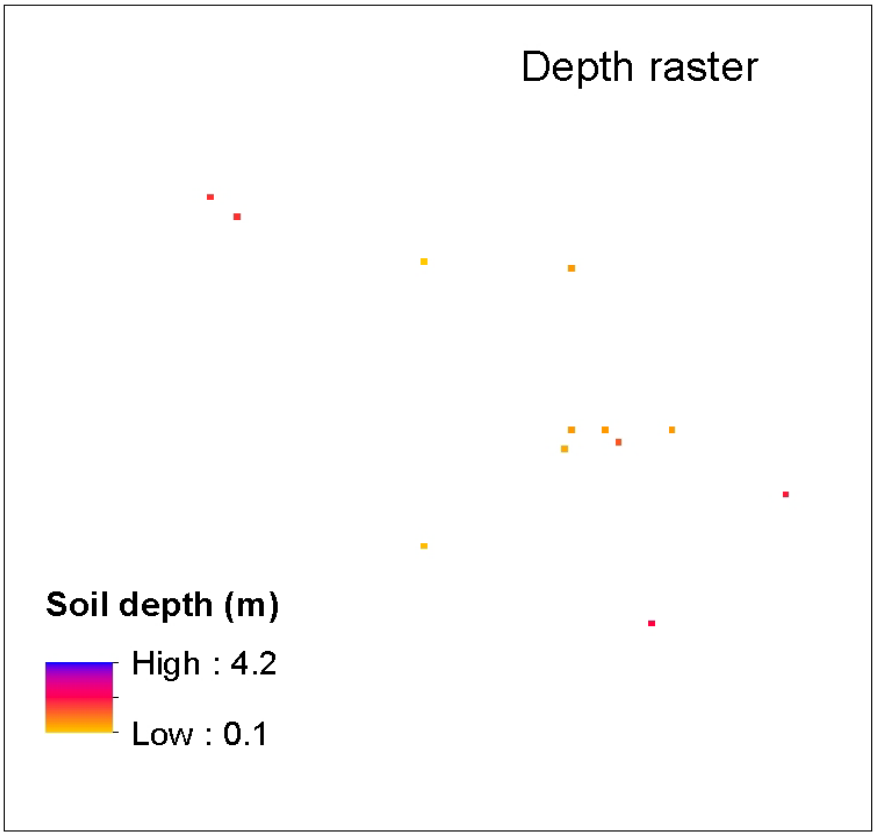

Fig. 7 Preparation of soil depth raster map

The threshold angle map should represent the mean slope angle for each geological unit where mass movements are triggered. The value usually lies between the internal friction angle of the slope

material and the Roering theoretical slope angle. In this example the threshold angle has been derived from Hamberger (2008).

4.3.3 Running the model

Section 1: Load the required modules and set working directory where the input data is stored

Section 2: All input maps are imported

Section3: The relevant input raster maps are normalized between the range of 0-1. For P the soil depth-TPI function should be inserted, where y equals P and x the TPI raster. Depending on the defined threshold, the mass movement raster (M) where the local slope value exceeds the threshold is created.

Section 4: Defines the definitions for the model. The definition "prepcal" prepares the dataset for the calibration of the geological units. The definition "cal" calibrates the soil depth for each geological unit. In this function the model loops trough a range of soil depth values for each geological unit. For each soil depth value the final soil depth is calculated by:

$$d = P*C*(1 + \tan\theta_{th})^{-1}$$

where d is the final soil depth (m), P=soil depth -TPI relationship, C= curvature and $(1 + \tan\theta_{th})^{-1}$ accounts only for pixels with a slope value higher than the defined mass movement threshold. The RMSE value for each initial soil depth is calculated and finally the soil depth which returns the lowest RMSE value is chosen for each geological unit. The calculation of the RMSE is defined in the "rmse" function

Section 5: The soil depth for each geological unit is automatically calibrated and for each unit the soil depth which returns the lowest RMSE is assigned

Section 7: The final soil depth is calculated and is written to the disk

Section 8: The RMSE is calculated for the final soil depth map

5. Results

Using the methods described earlier four different soil depth raster maps have been created in the Edisried-and Totenbüehl catchment (Fig.8). Accuracy assessment is displayed in Table 1. The ordinary kriging methods has the worst performance with a RMSE of 0.92. Due to the high spatial variability in short range of soil depth measurements, this method suffers from a nugget effect. Spatial autocorrelation is only found in close distance from the soil depth measurement points, after soil depth is quickly predicted as the regional mean of the dataset. Regression kriging performs slightly better due the correlation of the residuals of certain covariates. The random forest and GIST model have almost the same model performance with a RMSE of 0.63 and 0.64 respectively. The GIST model is characterised by a strong change of soil depth along geological boundaries, whereas the effect of geological units is much smaller in the random forest model.

Model	OK	RK	RF	GIST
RMSE	0.92	0.84	0.63	0.64

Tab. 1 RMSE for each soil depth prediction (OK=ordinary kriging, RK=regression kriging, RF=random forest)

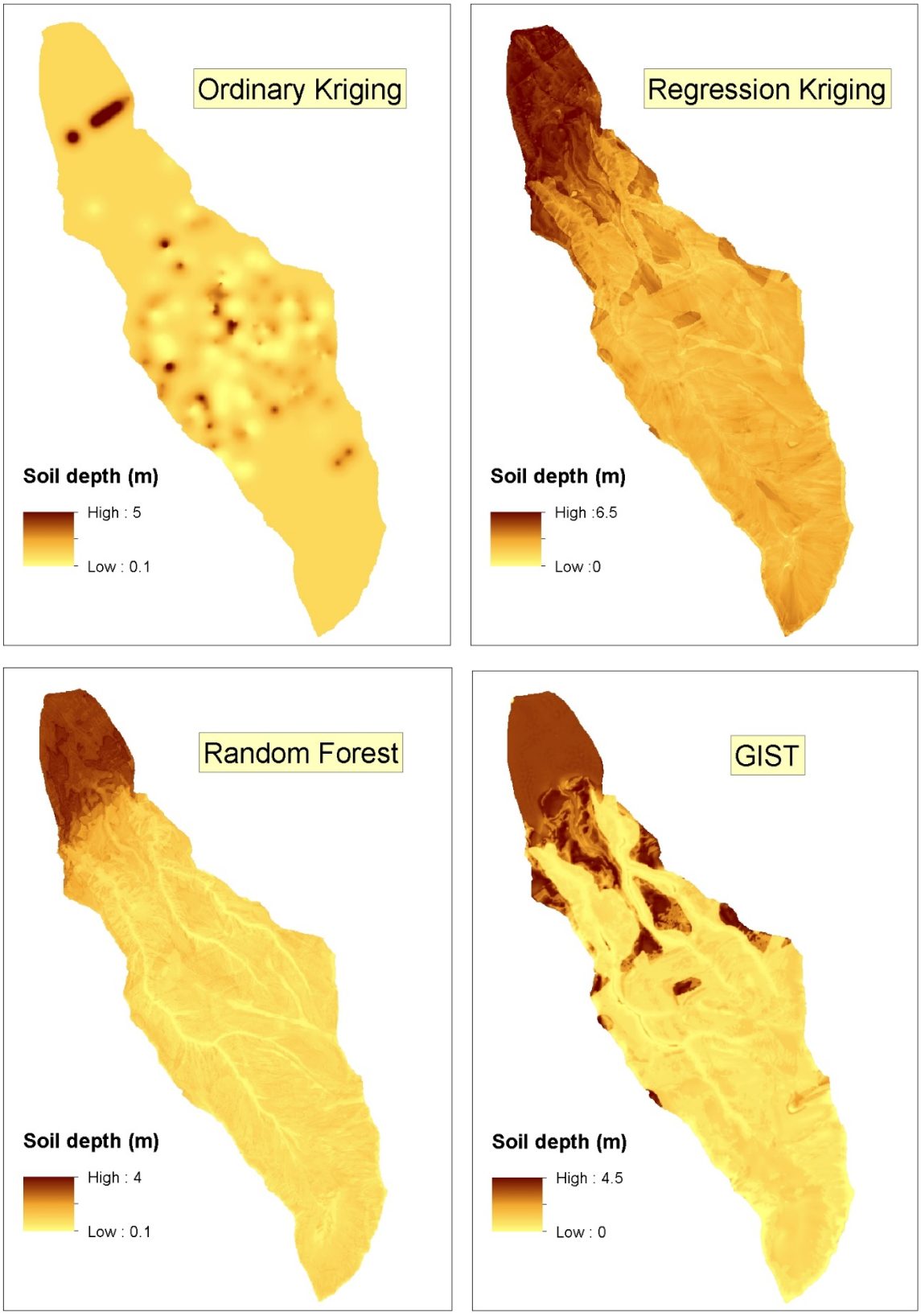

Fig. 8 Distributed soil depht predictions using ordinary, regression kriging, random forest and the GIST model

In a next step the performance and influence of the different soil depth models will be further evaluated in the coupled hydrological-slope stability model OpenLISEM hazard.

6. References

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Catani, F., Segoni, S., Falorni, G., 2010. An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale. Water Resour. Res. 46.

Hamberger, M., 2008. Rutschungserkennung mit Klassifikationssystemen am Beispiel Sachseln/Schweiz.

Hengl, T., 2009. A practical guide to geostatistical mapping.

Hengl, T., Heuvelink, G.B., Rossiter, D.G., 2007. About regression-kriging: From equations to case studies. Comput. Geosci. 33, 1301–1315.

Kuriakose, S.L., 2006. Effect of vegetation on debris flow initiation. The Netherlands: International Institute For Geo-Information Science

Kuriakose, S.L., Devkota, S., Rossiter, D.G., Jetten, V.G., 2009. Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the Western Ghats of Kerala, India. Catena 79, 27–38.

Malone, B.P., Minasny, B., McBratney, A.B., 2017. Using R for digital soil mapping. Springer.

Rossiter, D.G., 2013. An introduction to geostatistics with R/gstat.

Segoni, S., Rossi, G., Catani, F., 2012. Improving basin scale shallow landslide modelling using reliable soil thickness maps. Nat. Hazards 61, 85–101.

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., Pásztor, L., 2019. Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. Hydrol. Earth Syst. Sci. 23, 2615–2635.

Tesfa, T.K., Tarboton, D.G., Chandler, D.G., McNamara, J.P., 2009. Modeling soil depth from topographic and land cover attributes. Water Resour. Res. 45.

Van Westen, C.J., Castellanos, E., Kuriakose, S.L., 2008. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. Eng. Geol. 102, 112–131.

von Ruette, J., Lehmann, P., Or, D., 2013. Rainfall-triggered shallow landslides at catchment scale: Threshold mechanics-based modeling for abruptness and localization: MODELING SHALLOW LANDSLIDES AT CATCHMENT SCALE. Water Resour. Res. 49, 6266–6285. https://doi.org/10.1002/wrcr.20418

Zhang, K., Xue, X., Hong, Y., Gourley, J.J., Lu, N., Wan, Z., Hong, Z., Wooten, R., 2016. iCRESTRIGRS: a coupled modeling system for cascading flood–landslide disaster forecasting. Hydrol. Earth Syst. Sci. 20, 5035–5048.