Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development Discussions
Open Access

The design, deployment and testing of Kriging models in GEOframe

Marialaura Bancheri[1], Francesco Serafin[1], Michele Bottazzi[1], Wuletawu Abera[3], Giuseppe Formetta[2], and Riccardo Rigon[1]

[1] Department of Civil, Environmental and Mechanical Engineering, University of Trento, Italy
[2]Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, UK
[3]International Center for Tropical Agriculture (CIAT), P.O.BOX 5689, Addis Ababa, Ethiopia

Correspondence to: Marialaura Bancheri (marialaura.bancheri@unitn.it)

Abstract. This work presents a package for the interpolation of climatological variables, such as temperature and precipitation, using Kriging techniques. The purposes of the study are (1) to present a geostatistical software easy to use and easy to plug-in in a hydrological model, (2) to show a practical example of an accurately designed software in the perspective of reproducible research, (3) to show the goodness of the software applications, in order to have a reliable alternative to other traditionally used tools. Ten types of theoretical semivariograms and four types of Kriging were implemented and gathered into Object Modelling System compliant components. The package provides real time optimization for semivariogram and kriging parameters. The software was tested using temperature and rainfall data retrieved from 97 meteorological stations in the Isarco River basin, Italy. For both variables, good interpolation results were obtained and then compared to the results from the R package, $gstat$.

1 Introduction

Meteorological forcing data such as rainfall, temperature, solar radiation and others are the dominant controlling factors of the hydrological cycle, energy balance and ecosystem processes (Ly et al., 2013). These data, besides being important by themselves, are the natural input to distributed and semi-distributed hydrological models, where their quality and precision affect the accuracy of results.

Several algorithms for the spatial interpolation of meteorological data are available in literature: Thiessen polygons (e.g., Thiessen, 1911; WMO, 1994), inverse distance methods (Ly et al., 2013), interpolation with splines (e.g., Hutchinson, 1995; Mitášová and Mitáš, 1993), Kriging (e.g., Matheron, 1981; Goovaerts, 1997) or other types of interpolation (e.g., Robeson, 1992; Li and Heap, 2011, and references therein). They were assessed by several authors, (Tabios and Salas, 1985; Jarvis and Stuart, 2001), anong others, concluding that Kriging are one of the best techniques for the interpolation of the spatial behavior of climatological variables. For monthly rainfall and storm totals, Tabios and Salas (1985) and Creutin and Obled (1982), shown that it is preferable to other rainfall interpolation methods. Goovaerts (2000), Lloyd (2005), Basistha et al. (2008), Ly et al. (2011), confirmed these results.

Kriging can be applied to a wide range of datasets (e.g., Stahl et al., 2006; Phillips et al., 1992), allowing the estimation of the variance of interpolated quantities (e.g., Li and Heap, 2011). Auxiliary variables can be used to improve the interpolation, such

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

as terrain-related parameters (e.g., relief, slope and aspect) as investigated in Attorre et al. (2007). Not surprisingly, Carrera-Hernández and Gaskin (2007) found that the use of elevation as a secondary variable improves temperature prediction.

However, the interpolation with Kriging could be computationally more demanding than other techniques. To overcome this problem, most of applications, which implement Kriging interpolators, either use long time series and long time-step, such as daily, (Verfaillie et al., 2006; Buytaert et al., 2006), monthly or yearly (Hevesi et al., 1992; Goovaerts, 2000; Boer et al., 2001; Todini, 2001), or short time series and shorter time steps (such as rainfall events) [e.g., Haberlandt (2007)]. Having tools which implement efficient computations would help to extend the interpolations to real time processes. Furthermore, between the geostatistical tools available to the scientific community, few are open-source (i.e., to our knowledge just SAGA GIS kriging (www.saga-gis.org), R gstat (www.cran.r-project.org) and the High Performance Geostatistics Library HPGL (www.github.com/hpgl), but none implements a quick way to plug-in to hydrological models, and to automatic calibration algorithms.

Based on these premises, this work has two objectives. One is to have an efficient and precise tool to make spatial estimations and interpolation of environmental quantities. The second is to use an implementing strategy that favors the usability of the software, its maintenance, its inspection, its extension and, perhaps, makes easier the scientific work. The latter goal falls under the contemporary efforts to promote practices of open science (e.g. https://www.fosteropenscience.eu/). However, in this paper, for maintaining the right focus, we will not discuss the open science aspects and philosophy openly, rather the Kriging software and its design.

The Spatial Interpolation Kriging package (GEOframe-SIK, from now simply SIK), which makes hourly (or sub-hourly, when it is reasonable) estimates of any spatially distributed environmental data, is presented. SIK is designed according to the Object Modeling System v.3 (OMS3) framework (David et al., 2013) to be compatible with the JGrass-NewAGE system, (Formetta et al., 2014). The package can be integrated with other JGrass-NewAGE components and connected with them at run-time to form a variety of Modelling Solutions (MS). In particular, in this work four components are deployed to obtain the optimization of the parameters of the theoretical semivariograms, to perform the Kriging interpolation and to automatically and easily perform a jackknife, to assess the error of estimates.

SIK inherits some previous code used, for instance, in (Formetta et al., 2014).In order to make the old code easily extensible, maintainable, SIK was completely refactored and a systematic use of Design Patterns (DP), (Gamma, 1994; Freeman et al., 2004) was introduced.

The present paper is organized as follows: first the theory of the Kriging is introduced in section 2; then the structure of the package and the informatics are presented in section 3. Section 4 describes the study area and the experimental setup. The results of the application of the SIK package on temperature and rainfall datasets are discussed in section 5. Finally, a comparison of results of the interpolation of the temperature dataset obtained with the R $gstat$ is presented in section 6.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development Discussions

2 A little of Kriging theory

Kriging is a group of geostatistical techniques used to interpolate the value of random fields based on spatial autocorrelation of measured data.

The measurements value $z(\boldsymbol{x}_\alpha)$ and the unknown value $z(\boldsymbol{x})$, where $\boldsymbol{x}$ is the location given according to a certain cartographic projection, are considered as particular realizations of random variables $Z(\boldsymbol{x}_\alpha)$ and $Z(\boldsymbol{x})$ (Goovaerts, 1997; Isaaks and Srivastava, 1989). Let the estimation of the (true) random variable $Z(\boldsymbol{x})$ be $Z^\lambda(\boldsymbol{x})$. It is obtained as a linear combination of the N random variables at surrounding points, denoted as $\boldsymbol{x}_\alpha$ with $\alpha = \{1, \cdots N\}$, as in Goovaerts (1999):

$$Z^\lambda(\boldsymbol{x}) - m(\boldsymbol{x}) = \sum_{\alpha=1}^{N} \lambda_\alpha(\boldsymbol{x}_\alpha)[Z(\boldsymbol{x}_\alpha) - m(\boldsymbol{x}_\alpha)] \tag{1}$$

$m(\boldsymbol{x})$ and $m(\boldsymbol{x}_\alpha)$ are the expected values of the random variables $Z(\boldsymbol{x})$ and $Z(\boldsymbol{x}_\alpha)$; $\lambda(\boldsymbol{x}_\alpha)$ at varying α is the N-ple of weights assigned to the random variable $Z(\boldsymbol{x}_\alpha)$ at measured sites. The superscript λ in $Z^\lambda(\boldsymbol{x})$ denotes that this new random variable is parameterized by the weights. These are chosen to satisfy the conditions of minimizing the error of variance of the estimator σ^2_λ, that is:

$$\underset{\lambda}{\operatorname{argmin}}\ \sigma^2_\lambda \equiv \underset{\lambda}{\operatorname{argmin}}\ Var\{Z^\lambda(\boldsymbol{x}) - Z(\boldsymbol{x})\} \tag{2}$$

under the constraint that the estimate is unbiased, i.e.

$$E\{Z^\lambda(\boldsymbol{x}) - Z(\boldsymbol{x})\} = 0 \tag{3}$$

The latter condition, implies that:

$$\sum_{\alpha=1}^{N} \lambda_\alpha(\boldsymbol{x}_\alpha) = 1 \tag{4}$$

As shown in various textbooks, e.g. Kitanidis (1997), the above conditions bring to a linear system whose unknown is the N-uple of weights, and the the system matrix depends on the semivariograms (defined below in a simplified case) among the couples of the known sites. In synthetic notation, this "Kriging equation" can be written as:

$$\Gamma\,\Lambda = B \tag{5}$$

where Γ is the matrix of the two point variograms (defined below), Λ is the N-ple of the unknown weights and B (the so called known term) is an N-ple containing the variograms among the ungauged site and the measured sites. Further information is required for (5) to be a solvable linear system. In fact, B is actually unknown at this stage.

When it is made the assumption of isotropy of the spatial statistics of the quantity analyzed, the semivariogram is given by, (e.g. Cressie and Cassie (1993)):

$$\gamma(h) := \frac{1}{2N_h} \sum_{i=1}^{N_h} (Z(\boldsymbol{x}) - Z(\boldsymbol{x}_i))^2 \tag{6}$$

where N_h denotes the number of observation points at location $\boldsymbol{x}_i$ at distance h apart from $\boldsymbol{x}$ for any h. When random variables are substituted by their available realizations (i.e. $z(\boldsymbol{x}_i)$, indicated with normal letters) an empirical semivariogram is obtained. In order to be extended to any distance, $\gamma(h)$ needs to be fitted to a theoretical semivariogram model, i.e. an assumed function form, as those detailed in Appendix A. The latter operation is also necessary to get B. In fact, when a theoretical semivariogram is selected, just the information about the location of the ungauged place is required to get its semivariogram with any of the measured places.

Eventually, once determined B, the system (5) can be solved. This procedure is clearly delineated in literature and explained for instance in Kitanidis (1997).

Three main variants of Kriging can be distinguished, (Goovaerts, 1997):

- Simple Kriging (SK), which considers the mean, $m(\boldsymbol{x})$, to be known and constant throughout the study area;
- Ordinary Kriging (OK) , which account for local fluctuations of the mean, limiting the stationarity to the local neighborhood. In this case, the mean in unknown.
- Kriging with a trend model (here Detrended Kriging, DK), which considers that the local mean $m(\boldsymbol{x}_{\boldsymbol{\alpha}})$ varies within the local neighborhood.

The trend can be, for example, a linear regression model between the investigated variables and a auxiliary variable, such as elevation or slope. According to the procedure shown in Goovaerts (1997), the DK is performed as follows: i) the trend is subtracted from the original data and the OK of the residuals performed, ii) the final interpolated values are the sum of the interpolated values and the previously estimated trend.

Variants of OK and DK are the local ordinary kriging (LOK) and local detrended Kriging (LDK). In this case the estimate is only influenced by the measurements belonging to a neighbor, which are usually defined either in a maximum searching radius or as a number of stations closer to the interpolation point. In LDK case, the trend is estimated locally too, and therefore it can take account for local trend variations.

The SIK package implements the OK and the DK, since local mean may vary significantly over the study area and the SK assumption about the mean could be too strict, (Goovaerts, 1997).

The workflow for solving an interpolation problem with a Kriging is then summarized in the following steps:

1 - get the data from gauges;

2 - build the empirical semivariogram;

3 - interpolate a theoretical model for the semivariogram;

4 - use the theoretical model for solving the Kriging system;

5 - produce maps or pointwise time-series of the quantity desired in any point of the domain;

6 - calculate estimation errors.

The last step underlines that we are not only interested in estimating a variable (temperature, rainfall intensity or other scalars) but also in evaluating the errors of our estimate. Kriging returns, besides the spatial variable estimate, a variance of the estimate. However, Goovaerts (1997) states that the standard deviation cannot be used as a direct measures of estimation precision, since the the Kriging variance is only a ranking index of data geometry (and size), not a measure of the local spread errors, Goovaerts (1997); Deutsch and Journel (1992).

Therefore, to estimate the errors produced by interpolations using Kriging techniques, we chose the Leave-One-Out (LOO) cross validation technique (Efron and Efron, 1982; Isaaks and Srivastava, 1989; Martin and Simpson, 2003; Aidoo et al., 2015). LOO cross validation consists of removing one data point at a time and performing the interpolation for the location of the removed point, using the remaining stations. The approach is repeated until every sample has been, in turn, removed and estimates are calculated for each point. This procedure is straightforward, but cumbersome to be produced manually. Therefore, a special module (component) was programmed and delivered to obtain it. LOO estimates errors just over the location where measures are available and, eventually, these errors can be interpolated themselves to obtain an error estimation in any point of the spatial domain.

3 Design of the SIK package

On the base of the analysis of the mathematical problems, and the use case delineated in the previous section, the design of the software was organized at four levels of granularity whose rational is explained below. To deploy SIK we used various contemporary tools that, for the reader convenience, are presented in A.

3.1 Overall design of the SIK components

The component-based environmental modeling framework Object Modelling System v3 (OMS3), (David et al., 2013), was chosen for the development of the SIK code. Components, here, mean self-contained building blocks, modules or units of code (e.g. Argent, 2004; Van Ittersum et al., 2008). Each component implements a single modeling concept, and the components can be joined together to obtain a Modeling Solution (MS), which can accomplish a complicate task. The OMS user does not need to have a extensive knowledge of OMS libraries, and, as its authors state in David et al. (2013): "there are no interfaces to implement, no classes to extend, no polymorphic methods to override and no framework-specific data types to use". Besides, when the workflow allows it, components are run in parallel without particular cares from the researcher who programs them (this property is often called “implicit parallelism”).

The advantage of constructing upon a modular software framework, besides minimizing couplings, is the production of code that is more flexible, easier to maintain and to be inspected by third parties. Multiple algorithms can be implemented within the same component or in various components, and inserted in MS as alternatives, thus opening the way to compare, inside the same chain of tools, different approaches to the same hydrological problem.

More details on OMS3 can be found in David et al. (2013), in Formetta et al. (2014) and in Bancheri (2017). It is clear that the adoption of such a framework as a basis for our programs has an impact on our software design.

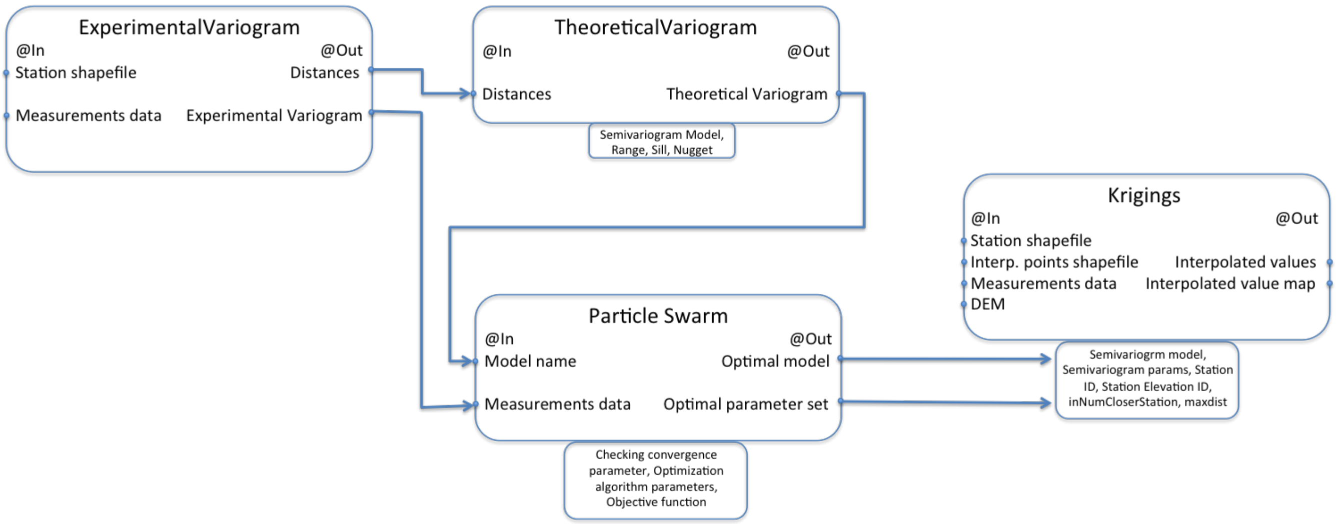

Figure 1. The MS optimizes model parameters connecting of the experimental variogram component to the theoretical variogram to the optimization tool.The output of the optimization are the sill, nugget and range, which, together with the type of model, are the input of SIK-K. Outputs of the MS can be time-series and maps of interpolated variables.

The initial implementation of Kriging, used in Formetta et al. (2014) and in Abera et al. (2017) was designed to group all the tasks into one single component.

In this case we thought useful to split it into four components: the first, SIK-EV, related to the production of experimental variogram from data; the second, SIK-TV, related to the selection and the parameter estimation of the theoretical variogram; the third, SIK-K, for the solution and mapping of the Kriging system; the fourth SIK-LOO, to manage the error estimation. In turn SIK-LOO does not work alone to produce its results, but can use the other three to generate the spatial estimates of errors.

Figure 1 exemplifies the use of the components in the OMS environment. It shows the MS obtained linking the SIK-ET, SIK-TV and to the optimization algorithm, a component embedded in OMS3. This MS optimizes the variogram parameters (sill, nugget and range) and feeds the SIK-K to produce the final desired results. The inputs of SIK-TV are the time series of the measured variables and the geometry, in shapefile format, with the spatial coordinates of the gauge stations. The outputs are the experimental variogram values and the distance vector, which feed the theoretical variogram component together with the name of the model chosen and the sill, nugget and range. Particle swarm is the component that actually optimizes the theoretical model parameters. Further inputs of the calibrator are the objective function to be optimized (in this case the RMSE) and other internal parameters, such as the number of iteration and the tolerance. Final outputs of the MS are either time series or map of interpolated values, which can be directly visualized in a GIS system.

Figure 2 show the implementation of the iterative procedure necessary to estimate errors of interpolation, called SK-LOO. Given n spatially distributed measures, $n-1$ measures are used for the interpolation, while the remaining one is used for comparison to produce the error estimate. The operation is repeated n times excluding each time a different gauged location,

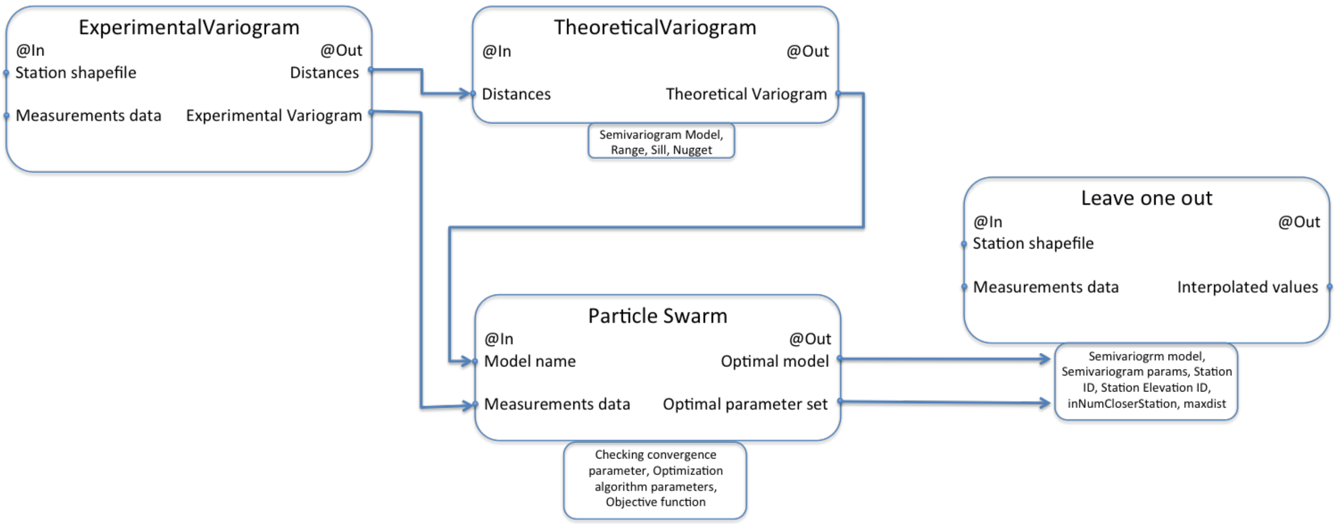

Figure 2. The MS, respect to the one presented in figure 1, connects the particle swarm to the automatic leave-one-out for the assessment of the model performances.

so to obtain a set of n estimated errors. Because our package is required to deal with time-varying fields, the operation is repeated for each time step, when the measures are available, and the site error is actually a temporal mean over a period that must be controlled by the user.

3.2 Internal classes design characteristics

Each of the four OMS3 components shown in the previous section can contain alternative solutions. For example, in the SIK-TV, the software design allows to have multiple theoretical variogram models, while in the SIK-K component, we want to implement, at least, the four types of Kriging listed in section 2.

In principle, we could have implemented a component for any type of Kriging and any type of variogram but this should have exploded the number of software modules to maintain. However, to "close the code to modification and keep it open to extensions" (Martin, 2002) and to maintain the code abstract enough to avoid the code disruption at any addition("program to a interfaces", e.g. Gamma (1994)), we adopted a systematic use of Design Patterns (DPs, Gamma (1994)). This was a further enhancement with respect to the previous version of the SIK package.

In general, DPs implement rules that allows, for instance, to separate code parts that are going to vary from those who are going to remain the same. The adoption of these DPs, once their rational is understood, makes the code more "readable" and maintainable. While largely known among programmers, DPs did not penetrate enough in the scientific community, which remained largely impervious to these techniques, and just a few examples of good practices can be found in scientific literature, (Gardner and Manduchi 2002; Rouson et al. 2011; Donatelli and Rizzoli 2008).

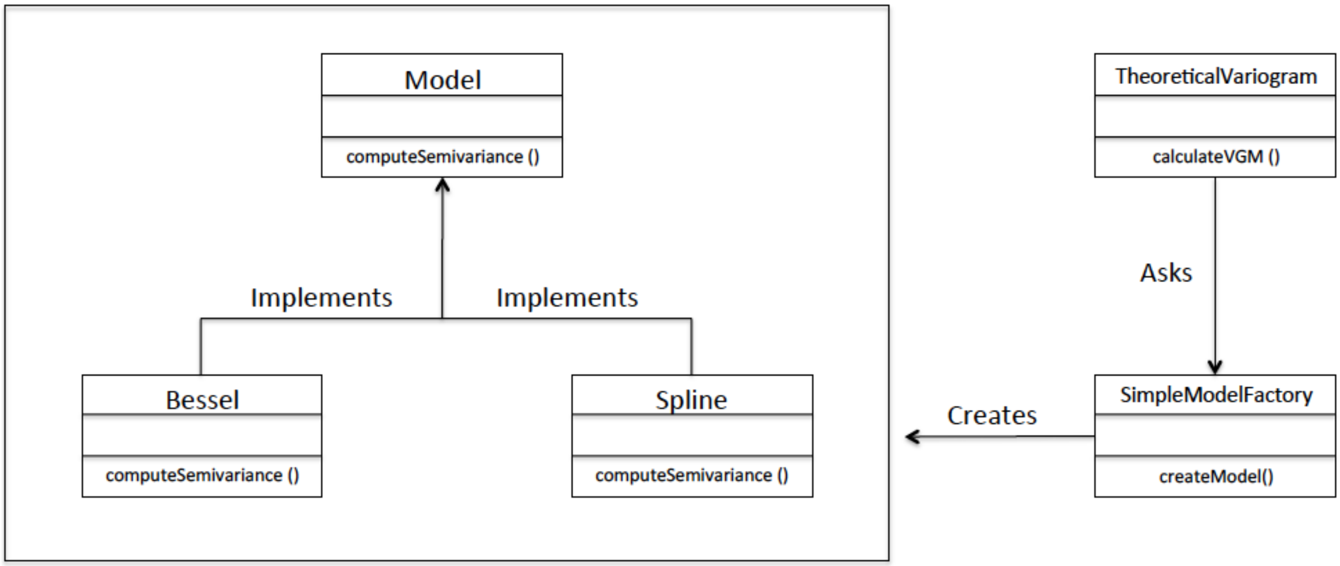

Figure 3. Implementation of the Java simple factory for the choice of the theoretical variogram model in the component SIK-TV.

The various theoretical semivariogram models or Kriging types, to be chosen at run-time, were encapsulated by using the Simple Factory class (Freeman et al., 2004). In this way, adding a new type of variogram or deleting an obsolete one to the SIK-TV is straightforward and requires few changes, confined just in a class.

Figure 3 shows the implementation of the simple factory for the choice of the theoretical semivariogram model. The concrete classes, Bessel and Spline, implement the same interface, Model. The SimpleModelFactory generates objects of a concrete class from a given information (a string containing the name of the chosen model). The component TheoreticalVariogram class uses the pattern to get the object of the concrete class.

The dependency inversion principle was also strictly respected in all the programming. The dependencies of the classes are not demanded to the concrete subclasses but only to the abstract classes and interfaces, (Ellis et al., 2007). So any changes in a concrete (sub)classes does not affect the overall structure of the program and remain limited to it.

4 Testing and simulations setup

4.1 Study area and data description

To test the performances of the modeling solutions presented in Figures 1 and 2, we used the SIK components to interpolate temperature and rainfall data from 97 stations, located in the Isarco River valley, Italy, shown in Figure 4 and detailed in D. Isarco River is a left tributary of the Adige River, in Trentino-Alto Adige Region, northern Italy.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development Discussions

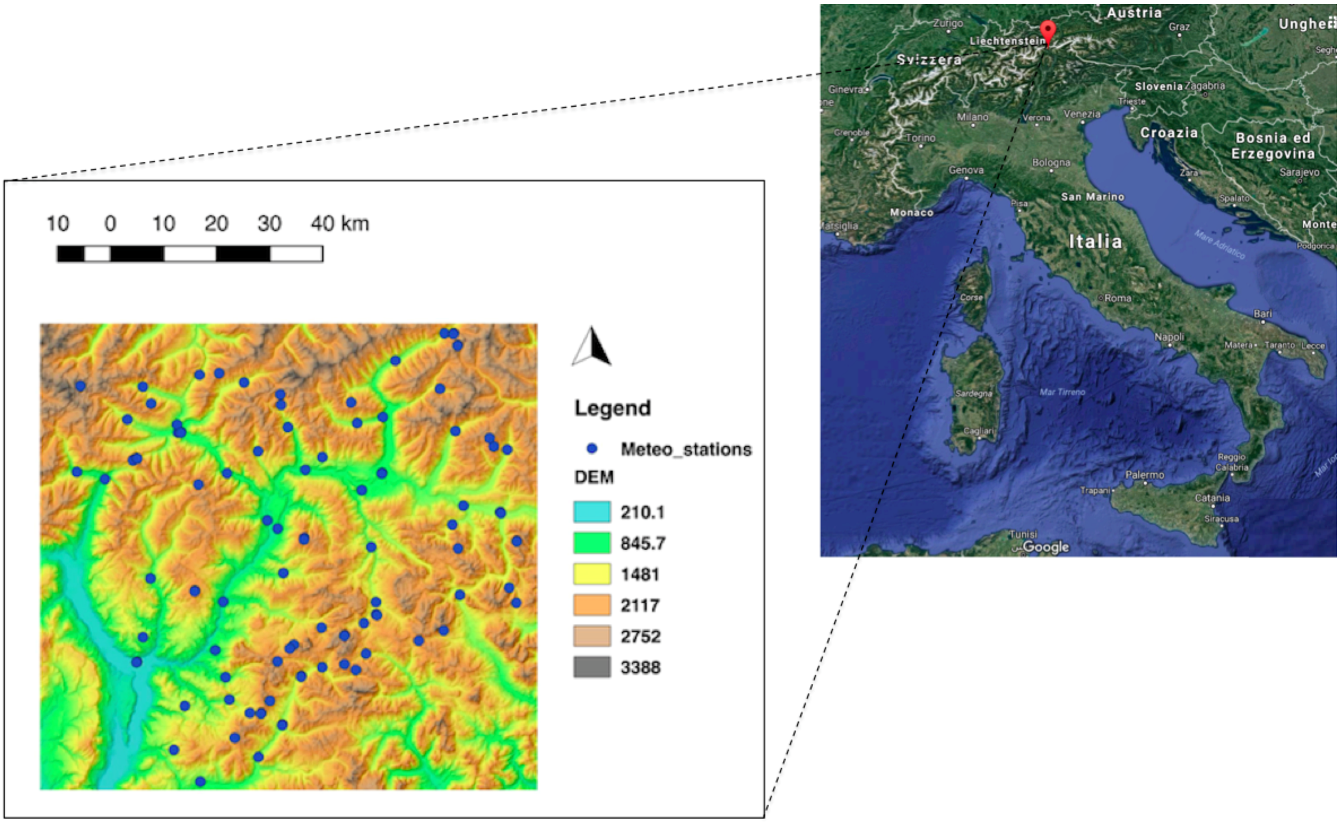

Figure 4. Study area: Isarco River Valley is situated in the North-Est part of Italy and it is one of the main valley in the Alto-Adige region.

The catchment area is around 4200 km^2, the river length is around 95 km and the altitude spans from 210 m a.s.l. to 3400 m a.s.l. Climate is typically alpine, characterized by dry winters, snow and glacier-melt in spring, and humid summers and autumns. Data used in the testing were provided by Provincia Autonoma di Bolzano, and collected into the Adige database (http://abouthydrology.blogspot.it/2016/09/the-adige-database-or-database-newage.html) during the CLIMAWARE and GLOBAQUA projects.

4.2 Setup

In the available dataset (2003-2013) we detected the year with the smallest number of missing data, which was the 2008 and we used it to test the SIK components.

A quality check was made, in order to eliminate the out-layers from the dataset. Moreover, the spatial distribution of the no-value was analyzed, in order to asses the number of bins of distances in which compute the semivariance. For each time step, we found that around the 10% of stations were not recording data. Therefore, since the mean number of active stations for each time step was 70-80, we decided to consider 8 bins. These choice was also supported by a visual inspection of the

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

experimental semivariance shape, which confirmed that using 8 bins, the number of stations involved were nor too low or too high.

In order to asses the goodness of SIK performances, two applications were performed:

- an interpolation of one year of hourly temperature data;
- an interpolation of a rainfall event, also at hourly time steps.

Firstly, the analysis of the semivariance was performed and experimental semivariograms were fitted using all the 10 theoretical models. The models that gave the best fitting where then used for the interpolation of the temperature and rainfall variables using the 4 types of Kriging. Thus, Kriging performances were assessed using the leave-one-out cross validation. Finally, results obtained from the interpolation of the temperature dataset were compared to the results obtained with R $gstat$, in order to assess the differences between the two packages, their easiness of use and their performances.

5 Simulations results

5.1 Application of SIK on temperature dataset

The first application of SIK components was made using the temperature dataset. The hourly experimental semivariograms were computed and then fitted using the 10 available theoretical models.

Figure 5 shows the results of the fitting of the experimental semivariogram for a single time step, 15th June 2008. The black dots represent the experimental semivariance, while each colored curve represents a different optimized theoretical model.

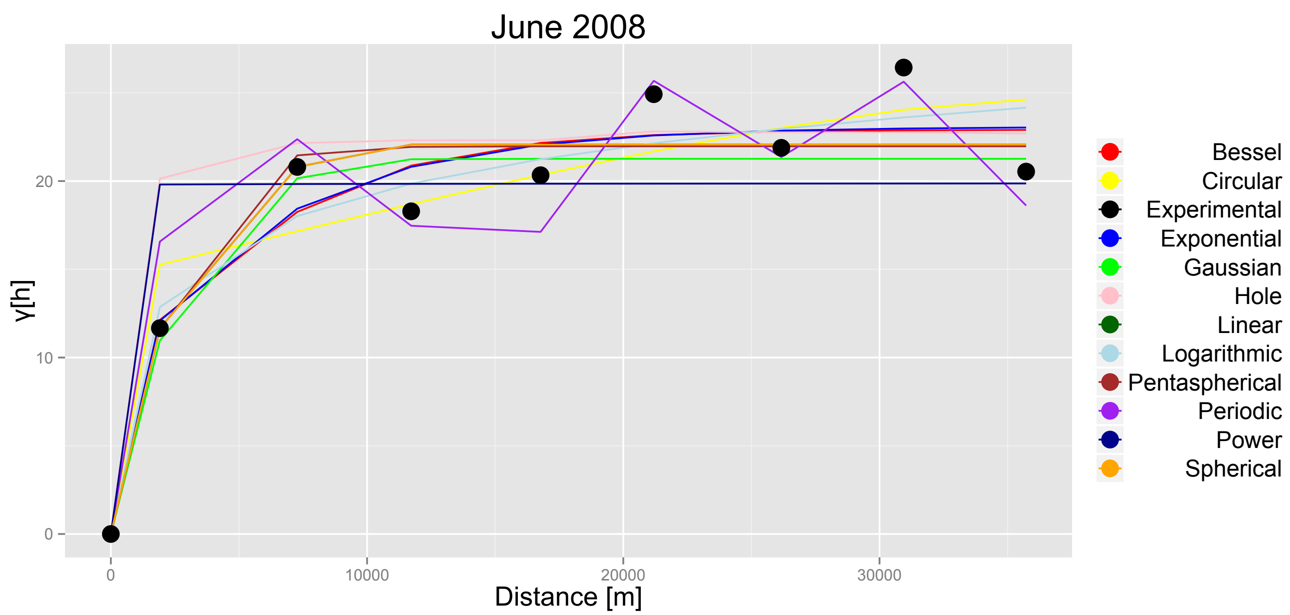

Figure 5. Fitting of the experimental semivariogram for the 15th June 2008 12:00 CET. The 10 theoretical semivariogram models were optimized using the PSO.

Table 1 reports the main indexes of goodness, NSE, RMSE, R^2 (the correlation coefficient) and PBIAS, computed between the experimental semivariogram and the 10 theoretical semivariogram models. All the previous quantities are defined in Appendix B. All the models gave satisfactory results and, therefore, we chose to use the best 5: Bessel, Exponential, Gaussian, Linear and Spherical for the interpolation of the temperature dataset.

In order to assess the goodness of the 4 typer of Kriging, OK, LOK, DK, LDK, we performed the leave-one-out cross validation using the optimized hourly values of sill, nugget and range.

Figures 6 show the results in terms of NSE between the measured and interpolated values of temperature using the four types of Kriging and the five semivariogram models. Each point represents the averaged monthly NSE over the 97 meteorological stations. The two local cases were performed using the ten closest stations to the interpolation point. In the both OK and LOK the performances are very poor, meaning that mean temperature would have been a better predictors than the interpolation.

In fact, a strong trend between temperature and elevation ($R^2 \sim 0.9$) was detected during the quality check phase (and was obviously known to exists). Therefore, interpolation results obtained using the DK and the LDK present, as expected, higher and optimal values of the goodness of fit index compared to the OK and LOK cases.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Semivariogram model	NSE	RMSE	R^2	PBIAS
Bessel	0.92	2.14	0.92	-0.20
Circular	0.88	2.59	0.88	0.0
Exponential	0.92	2.10	0.92	-3.80
Gaussian	0.90	2.39	0.91	0.35
Hole	0.77	3.61	0.81	7.90
Linear	0.91	2.28	0.91	0.0
Logarithmic	0.92	2.17	0.92	0.0
Pentaspherical	0.91	2.29	0.91	0.0
Periodic	0.90	2.18	0.92	0.0
Power	0.72	3.99	0.73	-3.70
Spherical	0.91	2.28	0.91	0.0

Table 1. Results in terms of goodness of fit indices of the fitting between the experimental and the 10 theoretical semivariograms. All the models shown a good agreement: Bessel model proved to be the best while the Power the worst.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development Discussions

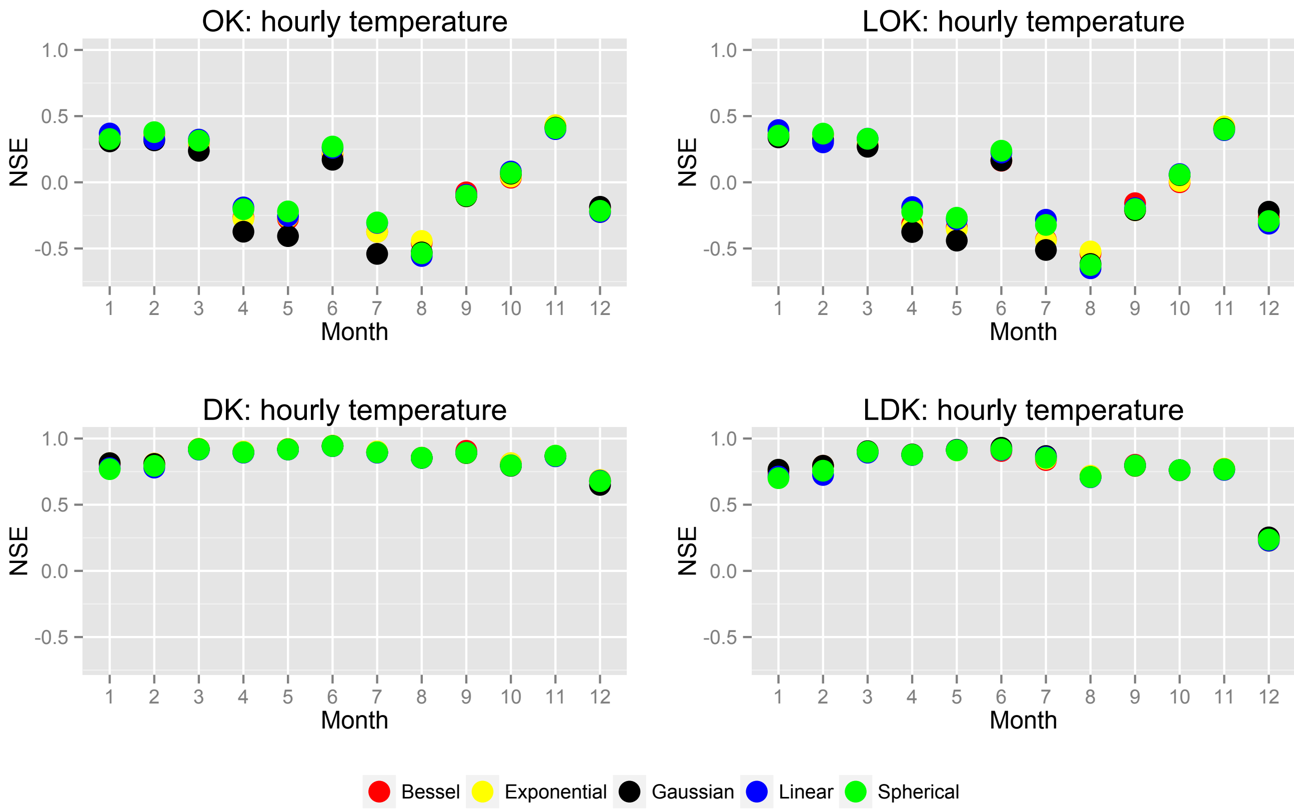

Figure 6. Monthly variation of the NSE index: each dot is the averaged NSE over the entire dataset. The theoretical semivariogram models shown are: Bessel (red dots), Exponential (yellow dots), Gaussian (black dots), Linear (blue dots), Spherical (green dots).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development Discussions

The spatialization of the temperature was also made for each pixel of the DEM (100 m resolution), applying the LDK and the exponential semivariogram model. 7 show the maps and the histograms obtained for two different dates in the 2008, one during winter (15th February 2008 12:00) and one in summer (15th June 2008 12:00).

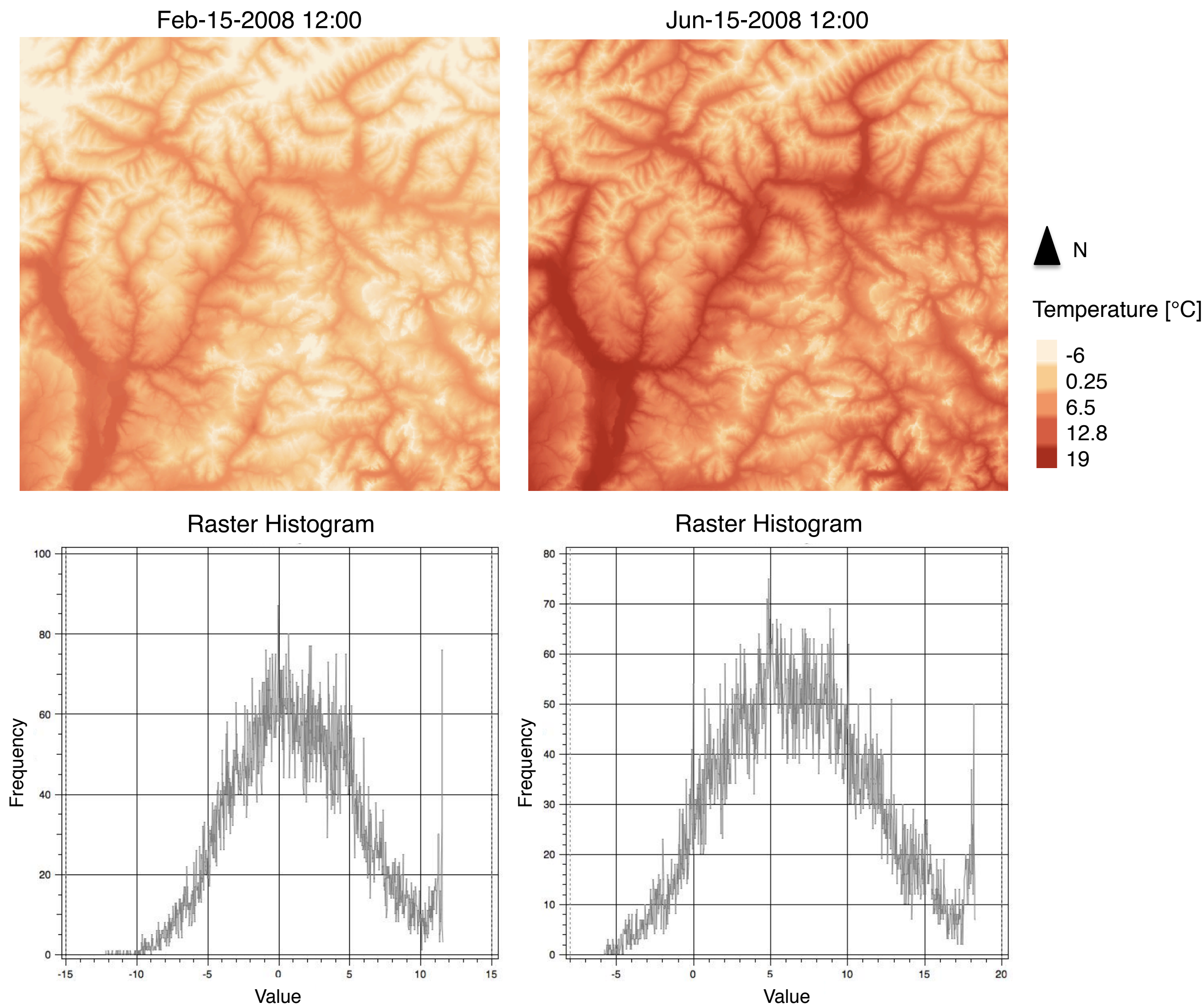

Figure 7. Maps of spatialized temperature for the 15th February 2008 and for the 15th June 2008. The two histograms, in the bottom plots, show the distributions of the temperature values for the two selected dates.

Figure 8 shows the bubble plots of the RMSE obtained between the measured hourly temperature in June 2008 and the interpolated with the OK and DK, overlapped to the DEM. The size of the bubble is representative of the magnitude of the error: bigger errors are obtained in the case of OK for the stations at higher elevation, which are corrected in the case of DK.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

The biggest error in the OK interpolation ($RMSE = 11.95°C$) is obtained for the station ID 90145 (Z=3399 alms) which is then reduced to $1.83°C$.

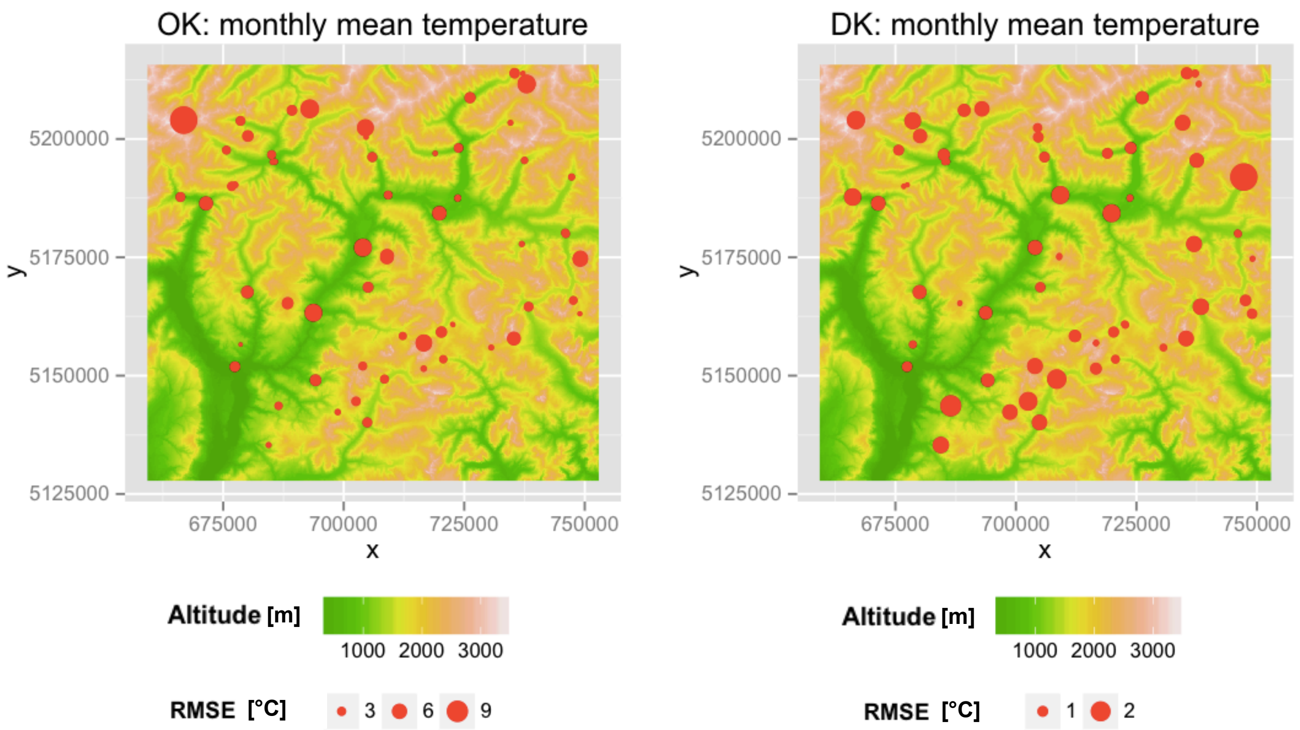

Figure 8. The two bubble plots show the RMSE obtained using the OK (box on the left) and DK (box on the right). The scales of the bubbles are not the same for the two plots for visualization reasons. In fact, being the errors in the DK case is much smaller than the OK case, a unique bubble scale didn't allow to appreciate the RMSE values in the DK case. Errors were estimated by means of the leave-one-out (LOO) procedure.

5.2 Application of SIK on rainfall dataset

The application on the rainfall dataset was made at event scale. We chose a rainfall event of 11 h between the 29th and 30th June 2008. The event was chosen since it is the longest and the most intense recorded from the highest number of stations for that year.

Figure 9 shows the boxplots of the 11 hourly semivariograms with 8 bins of lag distance, while red line represents the best theoretical semivariogram, which in this case was obtained using the Bessel model.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

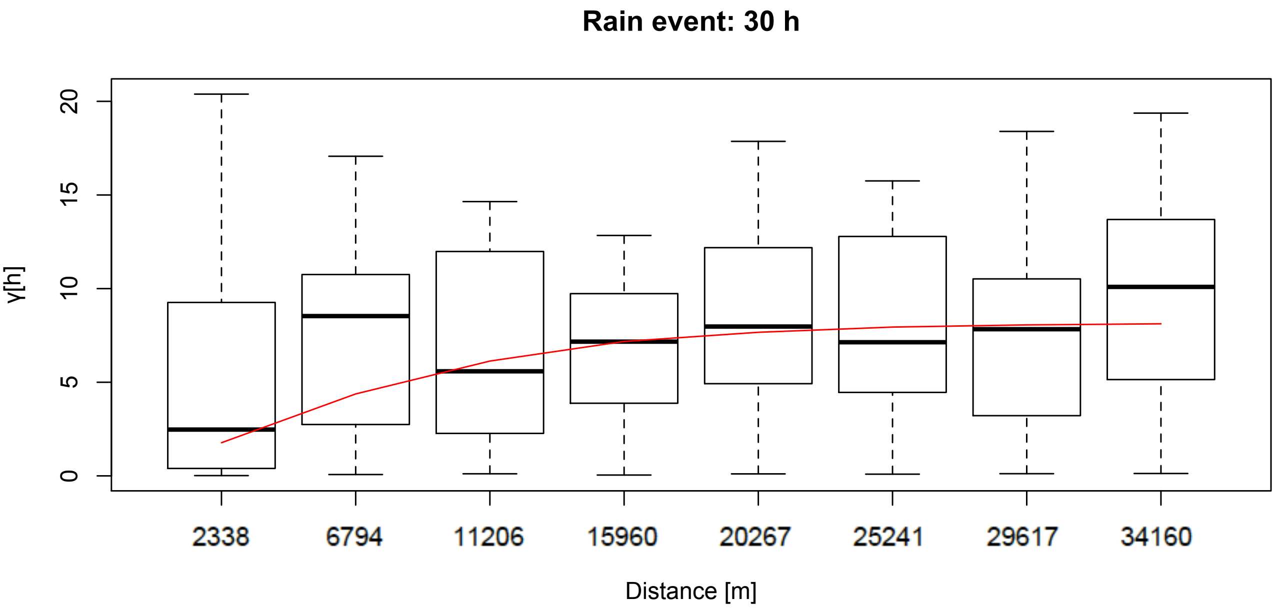

Figure 9. Boxplots of the semivariograms of the precipitation event of 29th and 30th June 2008: the horizontal line in the middle shows the median, the bottom and top end of the box show the 25th and 75th percentile, respectively, the whiskers (vertical line) shows the range of the data.

The optimized value of range, nugget and sill were used for the 4 types of Kriging interpolations. Figure 10 shows the comparison of the results obtained for two stations (ID 1152 and ID 1270), chosen at different elevation (953 m a.s.l. and 2100 m a.s.l., respectively).

Table 2 shows the indexes of goodness between the measured and the interpolated rainfall for the 4 types of Kriging and the two stations. In this case, no trend between the rain and elevation was detected. Therefore, the performance are overall good in the case of station ID 1152 and the best interpolator is the local ordinary kriging computed using the 5 closer stations. Results of the station ID 2170 are slightly worse in the case of OK, and worst in cases of LOK, DK, LDK, probably due to the highest elevation of the station. In this case the best interpolator is the LOK computed using the 5 closer stations.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

	ID 1152			ID 1270		
Kriging type	NSE	RMSE	R^2	NSE	RMSE	R^2
OK	0.69	1.64	0.77	0.60	1.48	0.64
LOK	0.73	1.53	0.74	0.31	1.88	0.45
DK	0.61	1.58	0.77	0.46	1.66	0.54
LDK	0.66	1.71	0.72	0.35	1.82	0.37

Table 2. Results in terms of goodness of fit indices of the fitting between the measured and interpolated rainfall values for two stations (ID 1152 and ID 1270).

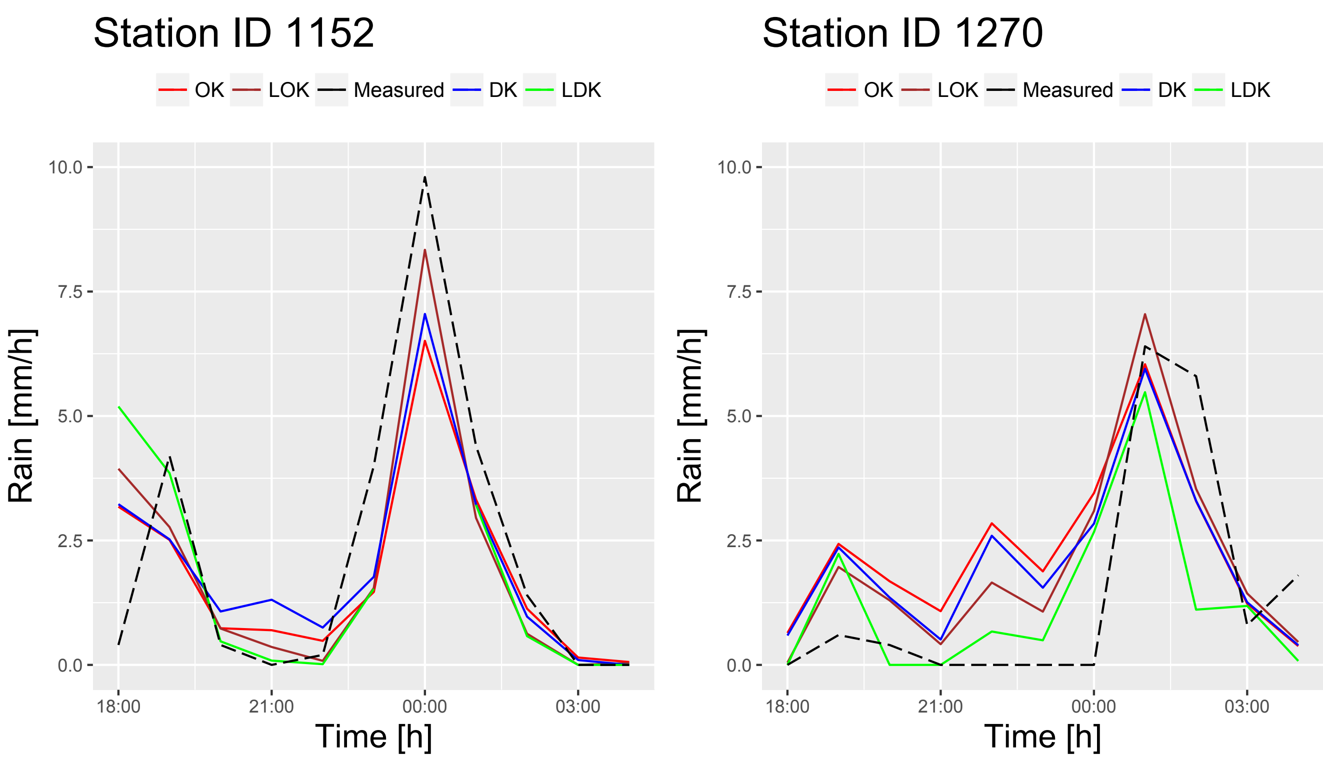

Figure 10. Comparison between the four types of Kriging (OK, red solid line, DK, blue solid line, LOK, brown solid line and LDK, green solid line) and the measured rainfall (black dashed line).

The spatial interpolation of the precipitation was also made for each pixel of the DEM (100 m resolution), applying the OK and the linear semivariogram model. Figure 11 show the results of the interpolation for the June 30th 2008 at 00:00. As it appears from the map, the rainfall intensities are higher in the river valley, with a value of 9.8 $[mm/h]$ measured at the station ID 1152.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

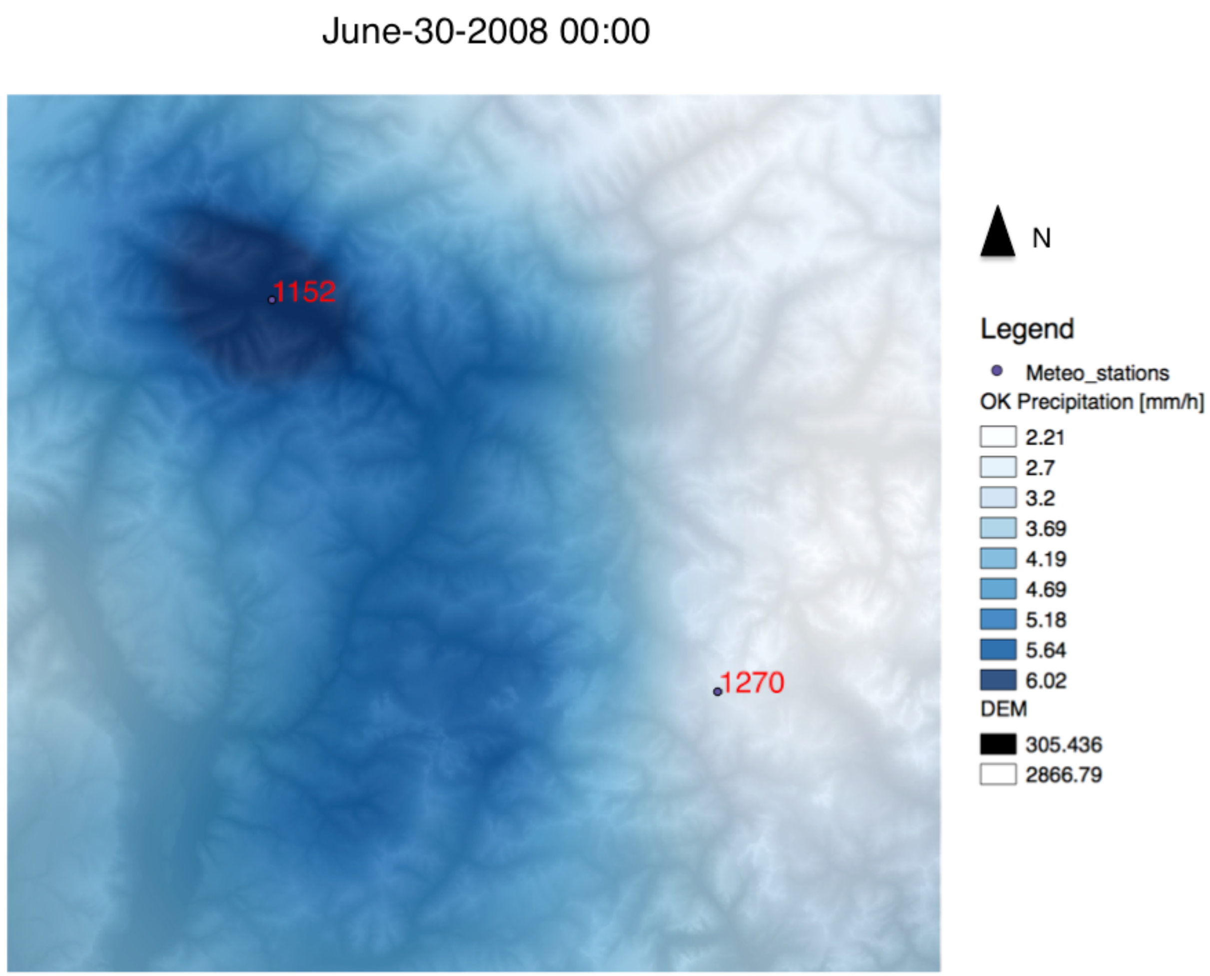

Figure 11. Spatial interpolation of the precipitation made for each pixel of the DEM (100 m resolution), applying the OK and the linear semivariogram model.

6 A qualitative comparison with R package gstat

A comparison with the R package *gstat* was made in order to highlight the differences and similarities of both softwares, and to justify the introduction of an alternative software.

gstat is developed in C with a part of the code in R language and must be executed using the R various environments. SIK is developed in Java (using Java 7 and older features) as a group of OMS components and can be executed inside the OMS console. Moreover the SIK components can be used as a stand-alone Java programs or embedded in other codes.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Moving to functional differences, *gstat* computes both omnidirectional and directional semivariongram, while SIK, so far, does not implement directional semivariograms. Moreover, *gstat* makes available four more theoretical semivariogram models respect to SIK: Matern, Matern with Stein's parameterizations, Wave and Legendre. However, adding the desired model to any to SIK-TV components is easy and straightforward, thanks to the design pattern implemented, as shown in figure 3.

Despite C language is usually considered faster than Java, the interpolation of a year of hourly data took way longer in *gstat* than SIK. The main reason is data management and the use of R *for-loops* for managing multiple time steps. These procedures using the R interpreter definitely slow down the computation, while the pure Kriging opertion could be fast but we could not measure it. Besides, *gstat* also requires implementation efforts from users-side. In our opinion, using SIK within the OMS environment is easier and more straightforward, at least for the tasks we describe in the paper.

Regarding the estimate that the two packages offers, they are usually different. Comparison were made either with the OK and the DK by utilizing the same temperature data used in section 4. Semivariograms were computed using the same number of bins and the same cutoff distance.

Figure 12 shows the results of the comparison in terms of NSE: the overall performances of both tools are good. However, SIK performs always better, and sometimes significantly better.

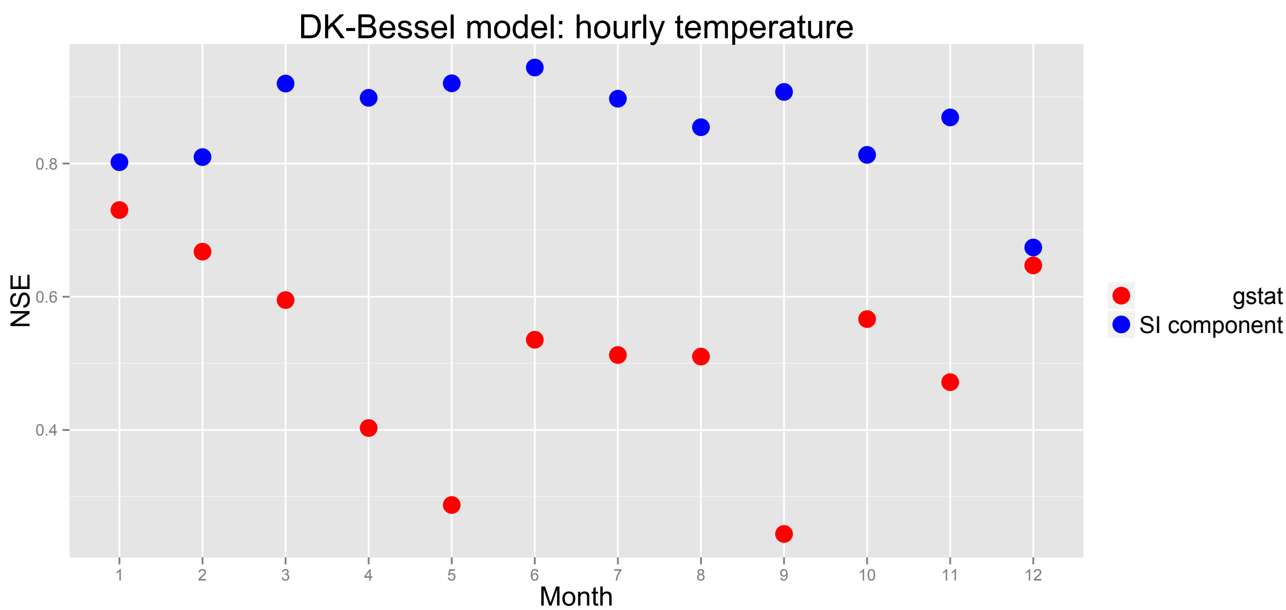

Figure 12. Comparison between *gstat* and SIK package in the interpolation performances using the DK.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development
Discussions

7 Conclusions

This paper presents a new modelling package for the spatial interpolation of environmental variables. It includes 10 theoretical semivariogram models and 4 types of Kriging interpolations.

Several characteristics make the SIK package a good competitor tool among the available in literature. From the user perspective:

- it can be used as a stand-alone;
- it can plugged-in the hydrological modelling system JGrass-NewAge;
- it can use with all the OMS compliant components, such as the calibration tools for the optimization of the parameters;
- a tool for the automatic estimation of errors is included;
- results are presented in data formats directly visualised by GIS;
- a variety of Modelling solutions can be obtained, according to user needs;
- it is faster than gstat in every-day use routine;

From the programmer perspective the implementation of design patterns makes the package easily maintainable and suitable for future improvements. All the elements are close to modification and open to extension. Further developments of the package such as the possibility of integrating new types of Kriging, implementing a different selection method of the gauge stations, the addition of non-linear relationships between the interpolated variable and an auxiliary variable, are easy and straightforward.

To test the performance of the SIK package, two applications are performed: the interpolation of one year of temperatures and the interpolation of a rainfall event. Data comes from a dataset of 97 stations located in the Isarco Valley in Italy. Both the interpolation of the temperature and the interpolation of the rainfall gave very good results, with an high agreement between the measured and the interpolated variables. The tests show also how it is possible to choose among ten variograms and four Kriging alternatives and to compare easily the outcomes. As expected, temperature showed a trend with elevation and only detrended Kriging performed appropriately. Single event rainfall, on the contrary do not show trend with elevation.

In comparison with $gstat$, the SIK package proved to be a good alternative, as regards both the easiness of use and the accuracy of the interpolation.

Appendix A: SIK deployment

The initial code was already available from a control version system under GPL v3 license, but the repository was owned by the initial Author, while we believed that a a-personal repository was better suited to host a collaborative work. Therefore, for SIK and its companion tools, the collective GEOframe organization repository was created under Github (www.github.com), thus using Git (www.git-scm.com) instead, for instance, of Mercurial, (www.mercurial-scm.org). The organization can be found at (www.github.com/geoframecomponents).

The original code did not included any building tool. These tools can be considered a contemporary evolution of the Unix "make" (e.g. www.gnu.org/software/make/) and take care of gathering the various libraries that concur to form the final executable file and link them to produce it. In our case, the choices possible for Java projects includes Apache Ant (www.ant.apache.org), Maven (www.maven.apache.org) and Gradle (www.gradle.org). All of them provide ways to solve the software dependencies; Maven and Gradle, in particular, can download and update the remote resources needed. Our favor, among the possibilities, was for Gradle for its more concise syntax depending on use of the Groovy language (www.groovy-lang.org) respect to the XML (www.w3.org/XML) used by Maven. Using building tools have also a practical effect to abstracting from the use of IDEs. In current Java market there exist at least three major IDEs for managing large projects: Netbeans (www.netbeans.org), Eclipse (www.eclipse.org) and IntelliJ (www.jetbrains.com). Some programmers feel more comfortable with one of them instead than another All of them support Gradle and Maven (and Ant) and can import seamlessly a Gradle or Maven (or Ant) project. These tools are widespread in programmers' community, but are very much less used by scientists, making increasingly difficult for them maintain their own code. At the same time, researchers, with these tools, can more easily master others' codes which otherwise could not, even if they are open source. Therefore we think that adopting a proper building tool, certainly not necessary to do good science, is useful to promote collaborative work and open science.

Another important step in the management of the code was the implementation of continuous integration system (www.jenkins.io). Travis tool (www./travis-ci.org) provides automatically to run the building system and run all the software tests (in our case JUnit tests), upon the "commit" of the last changes to the repository (in our case GitHub). Eventually, major codes commits are tagged with release numbers, under the GPL v3 license (www.gnu.org/licenses/gpl-3.0.en.html).

Since Github is a repository and not an archival system, we also decided to use Zenodo (www.zenodo.org) and provide our products with a Digital Object Identifier (DOI) (alternatives are, among others, Figshare (www.figshare.com) and Open Science Framework (www.osf.io)). The assignment of the DOI allows to retrieve exactly that code in the foreseen future also to researchers peers. This could be important to reconstruct which software version was used in a paper where relevant results were obtained, and besides makes life easier inside a research group, at least our.

Appendix B: List of symbols

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development Discussions
Open Access EGU

Variable	Description
$m(\boldsymbol{x})$	expected values of $Z(\boldsymbol{x})$
$m(\boldsymbol{x}_\alpha)$	expected values of $Z(\boldsymbol{x})_\alpha$
$\boldsymbol{x}$	coordinates of the point where the variable is estimate
$\boldsymbol{x}_\alpha$	coordinates of the α-th point where the measurement were taken
$z(\boldsymbol{x})$	particular realization of the random variable in $\boldsymbol{x}$
$z(\boldsymbol{x}_\alpha)$	particular realization of the random variable in $\boldsymbol{x}$
N	number of observation points
NSE	Nash-Sutcliffe efficiency
PBIAS	Percent bias
R^2	Coefficient of determination
RMSE	Root mean square error
$Z(\boldsymbol{x})$	random variable with values in $\boldsymbol{x}$
$Z^\lambda(\boldsymbol{x})$	true value of the random variable in $\boldsymbol{x}_\alpha$
$\gamma(h)$	semivariogram of $Z(\boldsymbol{x})$
$\lambda_\alpha(\boldsymbol{x}_\alpha)$	kriging weight assigned to the z variable evaluated in the α-th position $z(\boldsymbol{x}_\alpha)$
σ^2_λ	variance of $Z(\boldsymbol{x})$

Appendix C: List of semivariogram model implemented in SIK

Using n to represent the nugget, h to represent lag distance, r to represent range, and s to represent sill, the ten theoretical semivariogram models most frequently used in literature are:

– Bessel semivariogram

$$\gamma(h) = s \cdot \left(1 - frachr \cdot k1\left(\frac{h}{r}\right)\right) \tag{C1}$$

– Circular semivariogram

$$\begin{cases} \gamma(h) = n + s \cdot \left\{\frac{2}{\pi} \cdot \left[\frac{h}{r} \cdot \sqrt{1 - \left(\frac{h}{r}\right)}^2\right] + \arcsin\left(\frac{h}{r}\right)\right\} & \\ & h < r \\ \gamma(h) = n + s & h \geq r \end{cases} \tag{C2}$$

– Exponential semivariogram

$$\gamma(h) = n + s \cdot (1 - e^{-\frac{h}{r}}) \tag{C3}$$

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development Discussions

– Gaussian semivariogram

$$\gamma(h) = n + s \cdot [1 - e^{-(\frac{h}{r})^2}] \tag{C4}$$

– Hole semivariogram

$$\gamma(h) = n + s \cdot \left[1 - \frac{\sin(\frac{h}{r})}{\frac{h}{r}}\right] \tag{C5}$$

– Linear semivariogram

$$\begin{cases} \gamma(h) = n + s \cdot \frac{h}{r} & h < r \\ \gamma(h) = n + s & h \geq r \end{cases} \tag{C6}$$

– Logarithmic semivariogram

$$\gamma(h) = n + s \cdot \log\left(\frac{h}{r}\right) \tag{C7}$$

– Pentaspherical semivariogram

$$\begin{cases} \gamma(h) = n + s \cdot \left\{\frac{15}{8}\frac{h}{r} + \left(\frac{h}{r}\right)^3 \cdot \left[-\frac{5}{4} + \frac{3}{8}\left(\frac{h}{r}\right)^5\right]\right\} & h < r \\ \gamma(h) = n + s & h \geq r \end{cases} \tag{C8}$$

– Periodic semivariogram

$$\gamma(h) = n + s \cdot \left[1 - \cos\left(2\pi\frac{h}{r}\right)\right] \tag{C9}$$

– Power semivariogram

$$\gamma(h) = n + s \cdot h^r \tag{C10}$$

– Spherical semivariogram

$$\begin{cases} \gamma(h) = n + s \cdot \left[1.5 \cdot \frac{h}{r} - 0.5 \cdot \left(\frac{h}{r}\right)^3\right] & h < r \\ \gamma(h) = n + s & h \geq r \end{cases} \tag{C11}$$

Appendix D: Kriging dataset: the Isarco River Basin

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development
Discussions

Station ID	Elevation	X	Y
1008	254	677379	5151854
1010	560	703978	5177054
1025	1250	746077	5179955
1123	1205	704851	5139977
1131	821	723632	5187440
1137	2906	749059	5174631
1138	1990	677518	5190276
1139	2145	676739	5189931
1140	3105	737918	5211545
1142	2006	737151	5213768
1145	2985	716594	5156827
1146	2050	722698	5160726
1147	2260	688344	5165237
1152	943	685746	5195128
1153	2473	708956	5175007
1260	2777	692951	5206370
1262	1645	720648	5153469
1270	2100	730594	5155931
1274	1314	738357	5164560
1284	2142	716596	5151487
1311	1736	748986	5163080
1324	2265	747634	5165909
1326	2615	735292	5157843
1332	1750	700847	5142153

Station ID	Elevation	X	Y
1343	1385	708425	5149125
90072	1147	666135	5187742
90074	644	671444	5186398
90130	1330	689271	5206060
90133	1246	678592	5203835
90135	1960	680117	5200634
90138	948	685044	5196675
90140	1440	697647	5204620
90145	3399	666809	5203984
90147	1364	675668	5197644
90148	2145	676779	5190080
90149	1990	677414	5190356
90155	943	685228	5195148
90156	943	685786	5195277
90159	850	694445	5187499
90162	590	702029	5178611
90166	1219	745961	5180192
90168	1285	736938	5177815
90170	2340	737994	5173329
90172	1131	739010	5181375
90175	1412	747278	5191964
90176	2747	743952	5194145
90177	2152	744722	5192575
90182	1320	737507	5195474

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Geoscientific Model Development Discussions
Open Access
EGU

Station ID	Elevation	X	Y
90186	2006	737193	5213916
90187	3105	737960	5211693
90189	1450	735444	5213892
90192	1080	726201	5208726
90193	2155	717838	5200867
90196	1562	734591	5203425
90202	1141	718972	5196967
90203	870	723822	5198107
90211	828	723674	5187588
90216	1558	720262	5159216
90218	1428	722518	5163202
90220	2050	722543	5160892
90222	2985	716635	5156974
90225	1150	721627	5173593
90230	820	719843	5184315
90232	750	709213	5188155
90233	1349	712423	5190536
90234	2808	704501	5202375
90235	2050	704653	5200382
90236	1159	705920	5196174
90239	1410	700296	5191661
90266	490	693721	5163281
90267	840	692236	5154144
90269	1022	694161	5149040

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Station ID	Elevation	X	Y
90273	1616	698761	5142305
90275	1128	694888	5144827
90276	2125	695895	5137598
90287	2100	689061	5185368
90293	966	680012	5167662
90296	2260	688384	5165385
90298	1140	678615	5156585
90312	254	677473	5151945
90337	1470	686486	5143630
90354	1562	684427	5135344
90387	2906	749101	5174778
90458	1750	700886	5142300
90459	1205	704891	5140124
90467	1000	689420	5129334
90532	2040	702547	5144578
90533	2050	703949	5152028
90534	1385	708465	5149272
90631	1465	712386	5150914
90639	2376	718717	5150433
90651	1350	707040	5155154
90652	1050	700371	5133994

Appendix: Computer Code Availability

A OSF project with all the components necessaries to reproduce the results obtained on this paper has been created and is available at the following link: https://osf.io/24rgv. The interested researcher can find the entire OMS project, containing input data, output, sim files, jar files and R script used for the plots. Moreover, also the links to the source codes and to the documentation of the SIK components are available in the OSF project. In particular, for the present work, version 0.9.8 is the version of the codes of the GEOframe-SIK package that we used, available at the following link: https://github.com/geoframecomponents/Krigings/tree/v0.9.8.

Appendix: Author contribution

Marialaura Bancheri, Giuseppe Formetta and Francesco Serafin developed the model code integrated in the GEOframe-SIK package. Marialaura Bancheri, Francesco Serafin, Michele Bottazzi and Wuletawu Abera designed the experiments and performed the simulations. Marialaura Bancheri and Riccardo Rigon prepared the manuscript with contributions from all co-authors.

Appendix: Acknowledgments

The authors acknowledge Trento University project CLIMAWARE (http://abouthydrology.blogspot.it/search/ label/CLIMAWARE) and European Union FP7 Collaborative Project GLOBAQUA (Managing the effects of multiple stressors on aquatic ecosystems under water scarcity, grant no. 603629-ENV- 2013.6.2.1) that partially financed this research.

References

Abera, W., Formetta, G., Borga, M., and Rigon, R.: Estimating the water budget components and their variability in a pre-alpine basin with JGrass-NewAGE, Advances in Water Resources, 104, 37–54, 2017.

Aidoo, E. N., Mueller, U., Goovaerts, P., and Hyndes, G. A.: Evaluation of geostatistical estimators and their applicability to characterise the spatial patterns of recreational fishing catch rates, Fisheries research, 168, 20–32, 2015.

Argent, R. M.: An overview of model integration for environmental applications—components, frameworks and semantics, Environmental Modelling & Software, 19, 219–234, 2004.

Attorre, F., Alfo, M., De Sanctis, M., Francesconi, F., and Bruno, F.: Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale, International Journal of Climatology, 27, 1825–1843, 2007.

Bancheri, M.: A flexible approach to the estimation of water budgets and its connection to the travel time theory, Ph.D. thesis, University of Trento, 2017.

Basistha, A., Arya, D., and Goel, N.: Spatial distribution of rainfall in Indian himalayas–a case study of Uttarakhand region, Water Resources Management, 22, 1325–1346, 2008.

Boer, E. P., de Beurs, K. M., and Hartkamp, A. D.: Kriging and thin plate splines for mapping climate variables, International Journal of Applied Earth Observation and Geoinformation, 3, 146–154, 2001.

Buytaert, W., Celleri, R., Willems, P., Bièvre, B. D., and Wyseure, G.: Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes, Journal of Hydrology, 329, 413–421, 2006.

Carrera-Hernández, J. and Gaskin, S.: Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico, Journal of Hydrology, 336, 231–249, 2007.

Cressie, N. A. and Cassie, N. A.: Statistics for spatial data, vol. 900, Wiley New York, 1993.

Creutin, J. and Obled, C.: Objective analyses and mapping techniques for rainfall fields: an objective comparison, Water Resources Research, 18, 413–431, 1982.

David, O., Ascough Ii, J., Lloyd, W., Green, T., Rojas, K., Leavesley, G., and Ahuja, L.: A software engineering perspective on environmental modeling framework design: The Object Modeling System, Environmental Modelling & Software, 39, 201–213, 2013.

Deutsch, C. V. and Journel, A. G.: Geostatistical software library and user&s guide, vol. 1996, Oxford university press New York, 1992.

Donatelli, M. and Rizzoli, A.-E.: A design for framework-independent model components of biophysical systems, in: Proceedings of the iEMSs Fourth Biennial Meeting, Barcelona, Catalonia. International Congress on Environmental Modelling and Software iEMSs 2008., pp. 727–734, 2008.

Efron, B. and Efron, B.: The jackknife, the bootstrap and other resampling plans, vol. 38, SIAM, 1982.

Ellis, B., Stylos, J., and Myers, B.: The factory pattern in API design: A usability evaluation, in: Proceedings of the 29th international conference on Software Engineering, pp. 302–312, IEEE Computer Society, 2007.

Formetta, G., Antonello, A., Franceschi, S., David, O., and Rigon, R.: Hydrological modelling with components: A GIS-based open-source framework, Environmental Modelling & Software, 55, 190–200, 2014.

Freeman, E., Robson, E., Bates, B., and Sierra, K.: Head First Design Patterns: A Brain-Friendly Guide, " O'Reilly Media, Inc.", 2004.

Gamma, E.: Design patterns: elements of reusable object-oriented software, Pearson Education India, 1994.

Gardner, H. and Manduchi, G.: Design Patterns for e-Science, 2002.

Goovaerts, P.: Geostatistics for natural resources evaluation, Oxford university press, 1997.

Goovaerts, P.: Geostatistics in soil science: state-of-the-art and perspectives, Geoderma, 89, 1–45, 1999.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of hydrology, 228, 113–129, 2000.

Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, Journal of Hydrology, 332, 144–157, 2007.

Hevesi, J. A., Istok, J. D., and Flint, A. L.: Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis, Journal of applied meteorology, 31, 661–676, 1992.

Hutchinson, M.: Interpolating mean rainfall using thin plate smoothing splines, International journal of geographical information systems, 9, 385–403, 1995.

Isaaks, E. H. and Srivastava, R. M.: An introduction to applied geostatistics, vol. 561, Oxford university press New York, 1989.

Jarvis, C. H. and Stuart, N.: A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part I: The selection of "guiding" topographic and land cover variables, Journal of Applied Meteorology, 40, 1060–1074, 2001.

Kitanidis, P. K.: Introduction to geostatistics: applications in hydrogeology, Cambridge University Press, 1997.

Li, J. and Heap, A. D.: A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors, Ecological Informatics, 6, 228–241, 2011.

Lloyd, C.: Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain, Journal of Hydrology, 308, 128–150, 2005.

Ly, S., Charles, C., and Degre, A.: Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium, Hydrology & Earth System Sciences, 15, 2011.

Ly, S., Charles, C., and Degré, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review, 2013.

Martin, J. D. and Simpson, T. W.: A study on the use of kriging models to approximate deterministic computer models, in: Proceedings of DETC, vol. 3, pp. 2–6, 2003.

Martin, R. C.: Agile software development: principles, patterns, and practices, Prentice Hall, 2002.

Matheron, G.: Splines and kriging: their formal equivalence, Down-to-earth statistics: solutions looking for geological problems, 8, 77–95, 1981.

Mitášová, H. and Mitáš, L.: Interpolation by regularized spline with tension: I. Theory and implementation, Mathematical geology, 25, 641–655, 1993.

Phillips, D. L., Dolph, J., and Marks, D.: A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain, Agricultural and Forest Meteorology, 58, 119–141, 1992.

Robeson, S. M.: Spatial interpolation, network bias, and terrestrial air temperature variability, 1992.

Rouson, D., Xia, J., and Xu, X.: Scientific software design: the object-oriented way, Cambridge University Press, 2011.

Stahl, K., Moore, R., Floyer, J., Asplin, M., and McKendry, I.: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density, Agricultural and Forest Meteorology, 139, 224–236, 2006.

Tabios, G. Q. and Salas, J. D.: A comparative analysis of techniques for spatial interpolation of precipitation1, 1985.

Thiessen, A. H.: Precipitation averages for large areas, Monthly weather review, 39, 1082–1089, 1911.

Todini, E.: Influence of parameter estimation uncertainty in Kriging: Part 1-Theoretical Development, Hydrology and Earth System Sciences Discussions, 5, 215–223, 2001.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-310
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 25 January 2018

Van Ittersum, M. K., Ewert, F., Heckelei, T., Wery, J., Olsson, J. A., Andersen, E., Bezlepkina, I., Brouwer, F., Donatelli, M., Flichman, G., et al.: Integrated assessment of agricultural systems–A component-based framework for the European Union (SEAMLESS), Agricultural systems, 96, 150–165, 2008.

Verfaillie, E., Van Lancker, V., and Van Meirvenne, M.: Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas, Continental Shelf Research, 26, 2454–2468, 2006.

WMO: Guide to hydrological practices, 1994.