Hydrol. Earth Syst. Sci. Discuss., 12, 2243–2265, 2015
www.hydrol-earth-syst-sci-discuss.net/12/2243/2015/
doi:10.5194/hessd-12-2243-2015
© Author(s) 2015. CC Attribution 3.0 License.

Hydrology and Earth System Sciences
Open Access
Discussions

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Spatially shifting temporal points: estimating pooled within-time series variograms for scarce hydrological data

A. K. Bhowmik[1] **and P. Cabral**[2]

[1]Institute for Environmental Sciences, University of Koblenz-Landau, Germany
[2]NOVA IMS, Universidade Nova de Lisboa, Portugal

Received: 8 January 2015 – Accepted: 9 February 2015 – Published: 20 February 2015

Correspondence to: A. K. Bhowmik (bhowmik@uni-landau.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Estimation of pooled within-time series (PTS) variograms is a frequently used technique for geostatistical interpolation of continuous hydrological variables in spatial data-scarce regions conditional that time series are available. The only available method for estimating PTS variograms averages semivariances, which are computed for individual time steps, over each spatial lag within a pooled time series. However, semivariances computed by a few paired comparisons for individual time steps are erratic and hence they may hamper precision of PTS variogram estimation. Here, we outlined an alternative method for estimating PTS variograms by spatializing temporal data points and shifting them. The data were pooled by ensuring consistency of spatial structure and stationarity within a time series, while pooling sufficient number of data points for reliable variogram estimation. The pooled spatial data point sets from different time steps were assigned to different coordinate sets on the same space. Then a semivariance was computed for each spatial lag within a pooled time series by comparing all point pairs separable by that spatial lag, and a PTS variogram was estimated by controlling the lower and upper boundary of spatial lags. Our method showed higher precision than the available method for PTS variogram estimation and was developed by using the freely available R open source software environment. The method will reduce uncertainty for spatial variability modeling while preserving spatiotemporal properties of data for geostatistical interpolation of hydrological variables in spatial data-scarce developing countries.

1 Introduction

Geostatistical interpolation techniques have been extensively applied to mapping spatially continuous hydrological variables, e.g. precipitation (Carrera-Hernández and Gaskin, 2007; Durão et al., 2009; Haberlandt, 2007), stream flow (Skøien et al., 2006, 2014) and runoff (Skøien et al., 2008). Modeling spatial variability, i.e. spatial variogram

HESSD
12, 2243–2265, 2015

estimation plays a central role in geostatistical interpolation (Webster and Oliver, 2007) while the precision of variogram estimation strongly depends on the number of observations, i.e. spatial data points in a region (Oliver, 2010; Truong et al., 2012). Webster and Oliver (1992, 2007) identified the threshold for satisfactorily precise isotropic and anisotropic variogram estimation as 100 and 250 data points, respectively. Moreover, variograms computed on fewer than 50 data points exhibited little precision, whereas variograms on 400 data points were computed with great precision (reliable) (Webster and Oliver, 1992, 2007).

In developing countries, hydrological data are scarce because of technological and economical constraints (Bhowmik, 2012; Bhowmik and Costa, 2014). Consequently, spatial variograms are often estimated with less than 50 data points and in turn the resulting variograms are mostly imprecise (Bhowmik and Cabral, 2011; Bhowmik and Costa, 2012; Goovaerts, 2000). Moreover, the smallest separation distance between point-pairs for which semivariances are computed, i.e. the smallest-spatial-lag is very high and hence, the uncertainty for short distant spatial variability modeling remains high (Schuurmans et al., 2007).

Estimation of pooled within-time series (PTS) variograms by comparing spatial variability from multiple time steps, e.g. years (similar to pooled within-class (or strata) variograms where spatial variability from multiple attribute classes are compared (Webster and Oliver, 2007)) enables precise variogram estimation in spatial data-scarce regions, conditional that a time series of hydrological data is available (Wagner et al., 2012). PTS variograms have been adapted to cases where the available numbers of data points for individual time steps of a hydrological time series were too few to obtain satisfactory precision (Bhowmik, 2012; Rogelis and Werner, 2012; Wagner et al., 2012). The advantages of PTS variograms over individual variograms are (i) the number of point-pairs could be considerably increased and in turn noise in the semivariances was considerably decreased and hence, the variogram could be estimated with higher precision (Rogelis and Werner, 2012), and (ii) the smallest-spatial-lag was considerably decreased by including spatial variability from the time steps, where point-pairs were

HESSD
12, 2243–2265, 2015

separated by shorter distances, and thus the uncertainty for the short distant spatial variability was substantially reduced (Schuurmans et al., 2007). Moreover, PTS variograms were shown to be more suitable than spatiotemporal variograms (estimated for interpolation in space–time) and mean variograms (averaging estimated non-singular individual variogram parameters, i.e. nuggets, partial sills and ranges within time series) for cases, where the spatial locations and numbers of available data points vary within a time series and do not meet the threshold for precise individual variogram estimation in any time step (Christakos, 2001; Kerry and Oliver, 2004). This is because temporal variability modeling is uncertain for variable spatial locations and numbers of data points within a time series, while estimated spatial variogram parameters for individual time steps are imprecise due to scarce data.

Averaging empirical variograms (semivariances) (AEV), which are computed by paired comparisons in individual time steps, over each spatial lag within a pooled time series represents the only method available for PTS variogram estimation (Gräler et al., 2011). Computation of semivariances for individual time steps, where the numbers of data points do not meet the threshold for precise variogram estimation, is erratic because of a few paired comparisons. Hence, averaging erratic semivariances may lead to an erratic semivariance for a spatial lag within a time series and thus hamper the precision of PTS variogram estimation. Moreover, most studies focused on geostatistical interpolation of hydrological variables in regions with dense spatial data (Haberlandt, 2007; Skøien et al., 2006) whereas there is an increasing need for studies on spatial variability of hydrological variables in spatial data-scarce developing countries (Stocker et al., 2013). Hence, only the AEV method for PTS variogram estimation is insufficient.

We outlined an alternative method for estimating PTS variograms by spatializing temporal data points and shifting them that is called “spatially shifting temporal points (SSTP)”. SSTP was developed using the freely available R (R Core Team, 2014) open source software environment. We apply SSTP to estimate PTS variograms for a hydrological series in a spatial data-scarce developing country and compare it with the AEV method.

HESSD
12, 2243–2265, 2015

2 Materials and methods

2.1 Data and software

SSTP was applied to the PTS variogram estimation for "annual total precipitation in hydrological wet days (PRCPTOT)" in Bangladesh (Peterson et al., 2001, Fig. 1). We used the daily precipitation data from 1948–2007 series (DMICCDMP, 2012). Currently, 32 rain-gauges (data points) report daily precipitation in Bangladesh that classifies the country as data scarce because the number does not meet the threshold for satisfactorily precise variogram estimation (Webster and Oliver, 2007) (Fig. 1). Moreover, the numbers of data points exhibit an increasing coverage from 8 in 1948 to 32 in 2007 and thus indicate imprecise spatial variograms (with < 50 data points) for individual time steps (Fig. S1, Table S1).

The precipitation data were quality controlled and validated using the "RClimdex" routine (Peterson et al., 2001). PRCPTOT was computed for each of the time steps (year) and data points (rain-gauge), where precipitation data were available, following the method described in Bhowmik (2012) and Peterson et al. (2001). In general, high values of PRCPTOT were observed at data points with high longitudes and low latitudes (southeastern part of the country) and vice versa (Fig. S1). The altitudes of all data points were below 50 m and do not significantly ($p = 0.8$) correlate with PRCPTOT in Bangladesh (Fig. 1).

SSTP was developed on R (R Core Team, 2014) using the utilities of the "gstat" package (Pebesma, 2004). The other used packages were "intamap" (Pebesma et al., 2011) and "spacetime" (Pebesma, 2012).

2.2 Pooling hydrological time series

Spatial structure and stationarity indicate the strength and pattern of variability of spatial data, respectively (Kravchenko, 2003). Hence, as a PTS variogram represents a constant variability between data points within a pooled time series, spatial struc-

HESSD
12, 2243–2265, 2015

ture and stationarity require consistency within that time series (Gräler et al., 2011). Moreover, the number of pooled data points should ensure high enough precision for variogram estimation, i.e. the threshold for reliable variogram estimation (400) should be achieved (Webster and Oliver, 2007). Consequently, we first quantified the spatial structure of PRCPTOT in each year by computing its spatial correlation coefficients along the longitudinal and latitudinal gradients as suggested by Kravchenko (2003). Hereafter, the Pettitt–Mann–Whitney test was applied on the correlation coefficients to identify statistically significant change points between 1948 and 2007 and thus to identify changes in the spatial structure (Kiely et al., 1998, Fig. S2). The (sub)time series in between the change points were extracted as time series with consistent spatial structure. Next, we checked for the stationarity of PRCPTOT within the previously extracted time series with consistent spatial structure. For the purpose, we conducted an Augmented Dickey-Fuller test for each series (Said and Dickey, 1984). The null hypothesis of the test was that PRCPTOT has a unit root in each series, where rejecting null hypothesis with statistical significance denotes stationarity. In a final step, the time series with consistent spatial structure and stationarity were checked if the numbers of pooled data points met the threshold for reliable variogram estimation. The data points of the time series that satisfied the above three criteria were pooled and used for the PTS variogram estimation. For comparison, we also pooled the data points from 1948–2007 series, checked for stationarity and number of pooled data points and used for PTS variogram estimation. The Pettitt–Mann–Whitney test and Augmented Dickey-Fuller test were performed using the R packages "cpm" (Ross, 2015) and "tseries" (Trapletti and Hornik, 2012).

2.3 Estimation of pooled within-time series (PTS) variograms

2.3.1 Spatially shifting temporal points (SSTP)

The data point sets from different years (temporal) within a pooled time series were spatialized, i.e. assigned to different sets of coordinates on the same space (Fig. 2).

HESSD
12, 2243–2265, 2015

Given that $\boldsymbol{s}$ is a data point location vector comprised with coordinate vector touples $\boldsymbol{x}, \boldsymbol{y}$, $\boldsymbol{t}$ is a time (year) vector for a pooled time series, $Z(s,t)$ is the vector for computed PRCPTOT value for the data point $\boldsymbol{s}$ in year $\boldsymbol{t}$ and $\boldsymbol{s}_{i,t}, \boldsymbol{s}_{j,t}$ is the separation distance, i.e. spatial lag of the point-pair comprised with points $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ in year $\boldsymbol{t}$, we first assigned the data points from the base year ($\boldsymbol{t}_1$) of a pooled series, e.g. 1948 of the 1948–1975 series, to its original coordinates ($\boldsymbol{x}_{t_1}, \boldsymbol{y}_{t_1}$). Then coordinates for the data points of the latter years were calculated according to Eq. (1), when $(\boldsymbol{t}_1 + 1) + 4n \leq \boldsymbol{t} < (\boldsymbol{t}_1 + 1) + 4(n+1)$; $n \in N$.

$$
\begin{aligned}
\boldsymbol{s}_{(t_1+1)+4n} &= (\boldsymbol{x}_{(t_1+1)+4n} + (n+1)d), \boldsymbol{y}_{(t_1+1)+4n} \\
\boldsymbol{s}_{(t_1+1)+4n+1} &= (\boldsymbol{x}_{(t_1+1)+4n+1} - (n+1)d), \boldsymbol{y}_{(t_1+1)+4n+1} \\
\boldsymbol{s}_{(t_1+1)+4n+2} &= \boldsymbol{x}_{(t_1+1)+4n+2}, (\boldsymbol{y}_{(t_1+1)+4n+2} + (n+1)d) \\
\boldsymbol{s}_{(t_1+1)+4n+3} &= \boldsymbol{x}_{(t_1+1)+4n+3}, (\boldsymbol{y}_{(t_1+1)+4n+3} - (n+1)d).
\end{aligned} \tag{1}
$$

For example, for the years $\boldsymbol{t} = \{1949, 1950, 1951, 1952\}$ within the pooled series of 1948–1975, $n = 0$ because $(1948 + 1) + 4 \cdot 0 \leq \boldsymbol{t} < (1948 + 1) + 4(0 + 1)$ and hence,

$$
\begin{aligned}
\boldsymbol{s}_{1949} &= (\boldsymbol{x}_{1949} + d), \boldsymbol{y}_{1949} \\
\boldsymbol{s}_{1950} &= (\boldsymbol{x}_{1950} - d), \boldsymbol{y}_{1950} \\
\boldsymbol{s}_{1951} &= \boldsymbol{x}_{1951}, (\boldsymbol{y}_{1951} + d) \\
\boldsymbol{s}_{1952} &= \boldsymbol{x}_{1952}, (\boldsymbol{y}_{1952} - d).
\end{aligned} \tag{2}
$$

d in Eqs. (1) and (2) is a shift distance that is bigger than two-fold the largest-spatial-lag available within the pooled time series, i.e. $d > 2 \cdot \max(\boldsymbol{s}_{i,t}, \boldsymbol{s}_{j,t})$ and shifts the data point sets of different years from each other. This shift distance was chosen because it prevents the influence of data point sets from different years on each other while estimating PTS variograms, i.e. the peripheral data points of the sets from neighboring years are separated by a distance outside of the range of the largest-spatial-lag available within the pooled time series (Fig. 2). Thus the shift distance is a spatially rescaled

HESSD
12, 2243–2265, 2015

temporal distance (1 year) between data point sets from two consecutive years that preserves the spatiotemporal properties of PRCPTOT. We selected the shift distance as in Eq. (3), but the users can choose any distance that is $> 2 \cdot \max(\boldsymbol{s}_{i,t}, \boldsymbol{s}_{j,t})$.

$$d = 2 \cdot \max(\boldsymbol{s}_{i,t}, \boldsymbol{s}_{j,t}) + \max(\boldsymbol{s}_{i,t}, \boldsymbol{s}_{j,t})/100. \quad (3)$$

Spatial shifting of the temporal data points was performed using the R package "spacetime" (Pebesma, 2012).

2.3.2 Computation of empirical variograms

The empirical variograms (semivariances) were computed by comparing all point-pairs of the spatially shifted points within a pooled time series using the commonly applied Methods of Moments (MoM) (Webster and Oliver, 2007). For the point-pair $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, the semivariance $\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is a function of the spatial lag $\boldsymbol{s}_i, \boldsymbol{s}_j$ (that is not affected by actual location of data points) and was computed by Eq. (4).

$$\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \frac{1}{2M(\boldsymbol{s}_i, \boldsymbol{s}_j)} \sum_{i,j} (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2. \quad (4)$$

$M(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is the number of point pairs that can be separated by the spatial lag $\boldsymbol{s}_i, \boldsymbol{s}_j$. The upper and lower boundary boundaries of $\boldsymbol{s}_i, \boldsymbol{s}_j$ were set to the smallest- and largest-spatial-lags available within the pooled time series, respectively.

$$\begin{aligned} \boldsymbol{s}_1, \boldsymbol{s}_2 &= \min(\boldsymbol{s}_{i,t}, \boldsymbol{s}_{j,t}) \\ \boldsymbol{s}_{n-1}, \boldsymbol{s}_n &= \max(\boldsymbol{s}_{i,t}, \boldsymbol{s}_{j,t}). \end{aligned} \quad (5)$$

These (Eq. 5) were done to reduce the uncertainty of modeling short distant spatial variability, i.e. variability was modeled for the smallest-spatial-lag within the time series and to avoid inclusion of a spatially shifted point-pair in semivariance computation that contains points from two different years.

HESSD
12, 2243–2265, 2015

In the next step, we checked for anisotropy in the spatial variability of PRCPTOT within the pooled time series. In case that anisotropy was detected, we computed the ratio between the major (A) and minor (B) axes of the anisotropy ellipse and the angle of the anisotropy (φ). Computation of semivariances and anisotropy parameters were performed using "gstat" (Pebesma, 2004) and "intamap" (Pebesma et al., 2011) packages of R.

2.3.3 Estimation of pooled within-time series (PTS) variograms

We estimated PTS variograms, i.e. fitted variogram models to the PTS empirical variograms (semivariances) for each pooled time series. The available variogram models were fitted to the computed semivariances by a weighted least square approach providing $M(\boldsymbol{s}_i, \boldsymbol{s}_j)/(\boldsymbol{s}_i, \boldsymbol{s}_j)^2$ as weights (see Pebesma, 2004, for details). However, variogram models can also be fitted by the maximum likelihood approach as described in Marchant and Lark (2007) or by providing different weights than ours if using weighted least square approach (Pebesma, 2004). The parameters of the fitted models, i.e. nugget and sill variances, and range (a) were extracted. In case that anisotropy was detected, a was replaced by the anisotropy parameter where geometric anisotropy was made isotropic according to Eq. (5) through a linear transformation of coordinates with reference to the anisotropy ellipse described above (Oliver, 2010).

$$a = \sqrt{\left(A^2\cos^2\varphi + B^2\sin^2\varphi\right)}. \quad (6)$$

2.4 Precision of variograms

Precision of the estimated PTS variograms was evaluated by (i) variogram model-fit to the empirical variograms and (ii) cross-validation of an ordinary kriging (OK) interpolation of PRCPTOT using the best-fit models (Webster and Oliver, 1992, 2007). We computed the "weighted sum of squared error (SSE)" as a model-fit statistic (Pebesma, 2004). The previously fitted variogram models were compared by the corresponding

HESSD
12, 2243–2265, 2015

SSEs and the best-fit model with the lowest SSE was identified for each pooled series. Then the best-fit model was used in a leave-one-out cross-validation of the OK interpolation of PRCPTOT in each year of the each pooled series. The OK interpolation method was chosen because it gives unbiased evaluation of how well the variogram model fits the data (Oliver, 2010). Finally, the root means squared error (RMSE) was computed for each model by comparing the observed and OK interpolated PRCPTOT values through the cross-validation (Pebesma, 2004).

For comparison, we also estimated PTS variograms for the above pooled series by applying the averaging empirical variogram (AEV) method following the steps described in Gräler et al. (2011) and Pebesma and Gräler (2014). The SSEs and RMSEs were also computed for the AEV variograms following the method described above and compared with the SSEs and RMSEs of the SSTP variograms.

We provide a commented R-script as a Supplement (SM) detailing the SSTP method for PTS variogram estimation (SM2). The sample data for reproducibility is also provided as a Supplement (SM3) that will be permanently archived in PANAGEA.

3 Results

Statistically significant change points were detected in 1976 and 1993, and in 1976 for the spatial correlation coefficients of PRCOTOT along the longitudinal and latitudinal gradients, respectively, within the 1948–2007 series (Fig. S2). These change points indicated changes in spatial structure from 1976 and 1993. Consequently, spatial structure within the entire 1948–2007 series was inconsistent whereas the (sub)time series 1948–1975, 1976–1992 and 1993–2007 showed consistent spatial structure. The Dickey-Fuller statistics obtained for the 1948–1972, 1976–1992, 1993–2007 and 1948–2007 series were −4.5, −3.4, −5.0 and −4.0 respectively, and they were statistically significant at $p < 0.01$. Therefore, for each of these series null hypothesis was rejected and thus PRCPTOT showed stationarity. Moreover, the number of total data points within the 1948–1975, 1976–1992, 1993–2007 and 1948–2007 series met the

HESSD
12, 2243–2265, 2015

threshold for reliable variogram estimation (Table 1). PRCPTOT values did not vary much between the pooled series though the spatial variation of PRCPTOT within the pooled time series were high (CV ≥ 41 %) (Tables 1 and S1).

The distance d used for spatial shifting in each of the pooled series was 1111 km (~ 10°) because the largest-spatial-lag available within these series was approximately 550 km (~ 5°) (Fig. 2, Tables 1 and S1). Thus the shifted peripheral data points of sets from neighboring years showed a distance > 550 km, i.e. ≥ (1111–550) km (Fig. 2), and thus the spatiotemporal properties of PRCPTOT were preserved. The smallest-spatial-lags available within the three pooled series were similar and allowed for modeling spatial variability of PRCPTOT at ≤ 29 km (Tables 1 and S1).

Anisotropy was detected in the spatial variability of PRCPTOT for all pooled series in the northwest-southeast direction ($90° > \varphi > 0°$ from normal north to anticlockwise) indicating a strong variability of PRCPTOT in that direction (Figs. 3 and S1). Moreover, 1948–1975 series depicted weak anisotropy ($A : B = 0.8$), i.e. relatively weak variability whereas 1976–1992 and 1993–2007 series depicted strong anisotropy ($A : B = 0.4$), i.e. relatively strong variability (Fig. 3).

The SSTP computed semivariances were much less noisy than the semivariances computed by AEV, especially for large spatial lags (Fig. 3). Consequently, the estimated PTS variograms by SSTP showed better model-fit (lower SSE) and in turn entailed better performance of OK interpolation in cross-validation, and thus showed higher precision than the PTS variograms estimated by AEV (Table 2). The "Power" (Pow) model showed the best fit for both methods in all pooled series except for the SSTP in 1948–2007 series, where the "Hole" (Hol) model showed the best fit (Fig. 3). The PTS variograms estimated for the time series with inconsistent spatial structure, i.e. 1948–2007 by both methods showed lower precision than the variograms estimated for the time series with consistent spatial structure (Table 2). For the time series with consistent spatial structure, precision of PTS variogram estimation increased with the increasing number of pooled data points (Table 2).

HESSD
12, 2243–2265, 2015

4 Discussion

In this paper, we developed and implemented spatially shifting temporal points (SSTP), an alternative method for estimating pooled within-time series (PTS) variograms in spatial data-scare regions. Contrasting with the available method of averaging empirical variograms (AEV) computed for individual time steps, SSTP computed empirical variograms (semivariances) by comparing all point pairs separable by a spatial lag within a time series. Consequently, when compared to the PTS variograms estimated by AEV, SSTP variograms showed higher precision (Table 2). The numbers of available data points did not meet the threshold for satisfactorily precise variogram estimation in any of the individual time steps (year) within 1948–2007 series and hence the available numbers of point-pairs for comparisons were not sufficient for reliable semivariance computation (Table S1). As a result, computed semivariances for those years were likely erratic that induced noisy and erratic semivariances when averaged by AEV method (Fig. 3). Thus model fitting to AEV semivariances showed a lower goodness-of-fit and ordinary kriging (OK) interpolation of PRCPTOT using the AEV variograms showed worse performance than the SSTP variograms (Fig. 3, Table 2). By contrast, SSTP computed semivariances were reliable because of much higher number (that also met the threshold for reliable variogram estimation) of comparisons than by AEV and thus entailed higher precision in PTS variogram estimation. These results are in line with Webster and Oliver (1992, 2007).

Semivariances computed for small spatial lags by SSTP and AEV methods were similar whereas semivariances for large spatial lags were largely different (Fig. 3). Moreover, semivariances computed by AEV showed much more noise at large spatial lags than small spatial lags. The number of erratic semivariances averaged by AEV for large spatial lags were higher than for small spatial lags because point-pairs from more years were separable by large spatial lags than by small spatial lags due to data availability (Table S1). For example, point-pairs from only one year (1973) were separable by the smallest-spatial-lag for 1948–1975 series whereas point-pairs from 20 years

HESSD
12, 2243–2265, 2015

were separable by the largest-spatial-lag (Table S1). In addition, the numbers and spatial locations of available data points are highly variable within the pooled series and spatial variability of PRCPTOT was high (Tables 1 and S1, Fig. S1). Hence, we argue that the average semivariances computed by AEV was representative of the small number of semivariances at small spatial lags but unrepresentative of the large number of semivariances at large spatial lags because of the variable number and spatial location of data points and high spatial variability of PRCPTOT. As a result, semivariances for large spatial lags computed by SSTP and AEV could be similar if the numbers and spatial locations of data points were the same for all time steps and spatial variability of PRCPTOT was low (Gräler et al., 2011). Moreover, for variable number and spatial locations of data points, the noise in the semivariances computed by AEV can be partly reduced if the average of the semivariances per spatial lag is weighted by the corresponding number of data points available per time step.

The PTS variograms estimated for the 1948–2007 series (inconsistent spatial structure) showed lower precision than the variograms estimated for the series with consistent spatial structure, although PRCPCTOT was stationary within 1948–2007 series and the number of data points (higher than for the series with consistent spatial structure) met the threshold for reliable variogram estimation (Webster and Oliver, 1992, 2007) (Tables 1 and 2). Moreover, higher precision was obtained for PTS variogram estimation with higher number of pooled data points among the series with consistent spatial structure (Tables 1 and 2). Furthermore, the Hole model showed the best fit for the series with inconsistent spatial structure that did not represent the variability for individual time steps (Power variability was representative as depicted by the models for consistent spatial structure). These results suggest that the consistency of spatial structure, i.e. the strength of spatial variabilty within pooled time series is crucial for PTS variogram estimation (Kravchenko, 2003) and increasing the number of pooled data points can increase the precision of PTS variogarm estimation if the spatial structure is persistent. Many studies pooled data points only by assuming the consistency of spatial structure within time series (Bhowmik, 2012; Gräler et al., 2011; Rogelis

HESSD
12, 2243–2265, 2015

and Werner, 2012; Wagner et al., 2012). We recommend that time series should be checked for consistency of spatial structure before pooling. Notwithstanding, if the required number of data points for reliable variogram estimation is unavailable users should comply with the threshold for precise isotropic (100) and anisotropic (250) variogram estimation (Webster and Oliver, 2007).

A weaker anisotropy, i.e. variability was detected in the northwest-southeast direction for the 1948–1975 series than for 1976–1992 and 1993–2007 series (Fig. 3). This is presumably because of the lower number of spatial points per year in the 1948–1975 series than in 1976–1992 and 1993–2007 series, and thus a loss of anisotropy information (Table S1). However, higher PRCPTOT values were observed in the southeast than the northeast of Bangladesh and a high spatial variation (average CV = 42 %) was observed for 1948–1975 series (Fig. S1, Table S1). Hence, it can be claimed that the anisotropy, i.e. variability of PRCPTOT was equally strong for 1948–1975 series although not captured due to lower number of spatial points per year.

The PTS variograms allowed for modeling spatial variability at ≤ 29 km distance for all time steps (constant) within the pooled series although the smallest-spatial-lags available for many years, e.g. 1948–1950 were much higher (Tables 1 and S1). However, the smallest-spatial-lag for which spatial variability can be modeled for a pooled series depends on the availability of spatial lags in individual time steps, i.e. at least one point-pair should be separated by the smallest-spatial-lag in a time step. Thus, modeling short distant spatial variability by PTS variograms can be further improved if smaller spatial lags are available or more point pairs are available for comparison, i.e. more point pairs in individual time steps are separable by the smallest-spatial-lags (Rogelis and Werner, 2012; Schuurmans et al., 2007).

SSTP was developed on the freely available open source R software environment (R Core Team, 2014), and thus ensures reproducibility and wide spread application to geostatistical interpolation for resource constraint developing countries (Pebesma et al., 2012). The method is also applicable to PTS variograms estimation for geostatistical interpolation of non-hydrological spatially continuous variables in data-scarce

HESSD
12, 2243–2265, 2015

regions. It reduces uncertainty for spatial variability modeling at both short and long distances by including variability of smallest-spatial-lag within a time series and comparing many point-pairs for large distances. Inclusion of external variables that correlate with the variable for interpolation, e.g. altitude with precipitation (although did not correlate in our case) will increase the precision of PTS variogram estimation by SSTP (Diodato, 2005; Pebesma, 2006). To conclude, SSTP method can be further improved by integrating with the expert elicitation technique (Truong et al., 2013).

The Supplement related to this article is available online at doi:10.5194/hessd-12-2243-2015-supplement.

Author contributions. A. K. Bhowmik conceived the study. A. K. Bhowmik developed the method under supervision of P. Cabral. A. K. Bhowmik drafted the manuscript. A. K. Bhowmik and P. Cabral revised the manuscript.

Acknowledgements. The study was carried out within the framework of the European Commission, Erasmus Mundus Programme, project no. 2007-0064. Edzer Pebesma and Benedikt Gräler partly supervised the method development. Ralf B. Schäfer gave valuable comments that helped to improve the manuscript.

References

Bhowmik, A.: A comparison of Bangladesh climate surfaces from the geostatistical point of view, ISRN Met., 2012, 353408, doi:10.5402/2012/353408, 2012.

Bhowmik, A. and Cabral, P.: Statistical evaluation of spatial interpolation methods for small-sampled region: a case study of temperature change phenomenon in Bangladesh, in: Computational Science and its Applications – ICCSA 2011: Lecture Notes in Computer Science, Springer, Heidelberg, Dordrecht, London, New York, 44–59, doi:10.1007/978-3-642-21928-3_4, 2011.

HESSD
12, 2243–2265, 2015

Bhowmik, A. and Costa, A.: A geostatistical approach to the seasonal precipitation effect on Boro rice production in Bangladesh, Int. J. Geosci., 3, 443–462, doi:10.4236/ijg.2012.33048, 2012.
Bhowmik, A. and Costa, A.: Representativeness impacts on accuracy and precision of climate spatial interpolation in data-scarce regions, Meteorol. Appl., doi:10.1002/met.1463, online first, 2014.
Carrera-Hernández, J. and Gaskin, S.: Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico, J. Hydrol., 336, 231–249, doi:10.1016/j.jhydrol.2006.12.021, 2007.
Christakos, G.: Modern Spatiotemporal Geostatistics, Oxford University Press, New York, 2001.
Diodato, N.: The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain, Int. J. Climatol., 25, 351–363, doi:10.1002/joc.1131, 2005.
Disaster Management Information Center of Comprehensive Disaster Management Program (DMICCDMP): Bangladesh Meteorological Department, available at: http://www.bmd.gov.bd/index.php (last access: 25 July 2014), 2014.
Durão, R., Pereira, M. J., Costa, A. C., Côrte-Real, J. M., and Soares, A.: Indices of precipitation extremes in Southern Portugal – a geostatistical approach, Nat. Hazards Earth Syst. Sci., 9, 241–250, doi:10.5194/nhess-9-241-2009, 2009.
Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, J. Hydrol., 228, 113–129, doi:10.1016/S0022-1694(00)00144-X, 2000.
Gräler, B., Gerharz, L., and Pebesma, E.: Spatio-temporal analysis and interpolation of PM_{10} measurements in Europe, Technical paper 2011/10, European Topic Center on Air Pollution and Climate Change Mitigation, Bilthoven, the Netherlands, 2011.
Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, J. Hydrol., 332, 144–157, doi:10.1016/j.jhydrol.2006.06.028, 2007.
Kerry, R. and Oliver, M.: Average variograms to guide soil sampling, Int. J. Appl. Earth Obs., 5, 307–325, doi:10.1016/j.jag.2004.07.005, 2004.
Kiely, G., Albertson, J., and Parlange, M.: Recent trends in diurnal variation of precipitation at valentina on the West Coast of Ireland, J. Hydrol., 207, 270–279, 1998.

HESSD
12, 2243–2265, 2015

Kravchenko, A.: Influence of spatial structure on accuracy of interpolation methods, Soil Sci. Soc. Am. J., 67, 1564–1571, doi:10.2136/sssaj2003.1564, 2003.
Marchant, B. and Lark, R.: Robust estimation of the variogram by residual maximum likelihood, Geoderma, 140, 62–72, doi:10.1016/j.geoderma.2007.03.005, 2007.
Oliver, M.: The variogram and kriging, in: Handbook of Applied Spatial Analysis, Springer-Verlag, Berlin, Heidelberg, 319–352, doi:10.1007/978-3-642-03647-7_17, 2010.
Pebesma, E.: Multivariable geostatistics in S: the gstat package, Comput. Geosci., 30, 683–691, doi:10.1016/j.cageo.2004.03.012, 2004.
Pebesma, E.: The role of external variables and GIS databases in geostatistical analysis, Trans. GIS, 10, 615–632, doi:10.1111/j.1467-9671.2006.01015.x, 2006.
Pebesma, E.: spacetime: spatio-temporal data in R, J. Stat. Softw., 51, 1–30, 2012.
Pebesma, E. and Gräler, B.: Spatio-temporal geostatistics using gstat, available at: http://cran.r-project.org/web/packages/gstat/index.html, last access: 10. April, 2014.
Pebesma, E., Cornford, D., Dubois, G., Heuvelink, G., Hristopulos, D., Pilz, J., Stöhlkerg, U., Morin, G., and Skøien, J.: INTAMAP: the design and implementation of an interoperable automated interpolation web service, Comput. Geosci., 37, 343–352, doi:10.1016/j.cageo.2010.03.019, 2011.
Pebesma, E., Nüst, D., and Bivand, R.: The R software environment in reproducible geoscientific research, EOS T. Am. Geophys. Un., 93, 163–163, doi:10.1029/2012EO160003, 2012.
Peterson, T., Folland, C., Gruza, G., Hogg, W., Mokssit, A., and Plummer, N.: Report on the activities of the Working Group on Climate Change Detection and Related Rapporteurs 1998–2001, Report WCDMP-47, WMO-TD 1071, World Meteorological Organization, Geneva, 2001.
R Core Team: R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, available at: http://www.R-project.org, last access: 30 December, 2014.
Rogelis, M. and Werner, M.: Spatial interpolation for real-time rainfall field estimation in areas with complex topography, J. Hydrometeorol., 14, 85–104, doi:10.1175/JHM-D-11-0150.1, 2012.
Ross, G.: Parametric and nonparametric sequential change detection in R: the cpm package, J. Stat. Softw., in press, 2015.
Said, S. and Dickey, D.: Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika, 71, 599–607, doi:10.1093/biomet/71.3.599, 1984.

HESSD
12, 2243–2265, 2015

Schuurmans, J., Bierkens, M., and Pebesma, E.: Automatic prediction of high-resolution daily rainfall fields for multiple extents: the potential of operational radar, J. Hydrometeorol., 8, 1204–1224, doi:10.1175/2007JHM792.1, 2007.

Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, Hydrol. Earth Syst. Sci., 10, 277–287, doi:10.5194/hess-10-277-2006, 2006.

Skøien, J. O., Pebesma, E. J., and Blöschl, G.: Geostatistics for automatic estimation of environmental variables – some simple solutions, Georisk, 2, 259–272, doi:10.1080/17499510802086769, 2008.

Skøien, J. O., Blöschl, G., Laaha, G., Pebesma, E., Parajka, J., Viglione, A.: rtop: an R package for interpolation of data with a variable spatial support, with an example from river networks, Comput. Geosci., 67, 180–190, doi:10.1016/j.cageo.2014.02.009, 2014.

Stocker, T., Dahe, Q., and Plattner, G.: Climate Change 2013: The Physical Science Basis, Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Summary for Policymakers, Intergovernmental Panel on Climate Change (IPCC), 2013.

Truong, P., Heuvelink, G., and Gosling, J.: Web-based tool for expert elicitation of the variogram, Comput. Geosci., 51, 390–399, doi:10.1016/j.cageo.2012.08.010, 2013.

Wagner, P., Fiener, P., Wilken, F., Kumar, S., and Schneider, K.: Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions, J. Hydrol., 464–465, 388–400, doi:10.1016/j.jhydrol.2012.07.026, 2012.

Webster, R. and Oliver, M.: Sample adequately to estimate variograms of soil properties, J. Soil Sci., 43, 177–192, doi:10.1111/j.1365-2389.1992.tb00128.x, 1992.

Webster, R. and Oliver, M.: Geostatistics for Environmental Scientists, John Wiley and Sons Ltd., Chichester, 2007.

HESSD

12, 2243–2265, 2015

Table 1. Number of data points, smallest- and largest-spatial-lags, and summary statistics, i.e. minimum (Min.), mean, maximum (Max.) and coefficient of variation (CV) of annual total precipitation in hydrological wet days (PRCPTOT) within the pooled time series.

Pooled time series	Number of pooled data points	Spatial lag		PRCPTOT			
		Smallest (km)	Largest (km)	Min. (mm)	Mean (mm)	Max. (mm)	CV (%)
1948–1975	441	29.16	550	17	1659	4036	42
1776–1992	465	26.61	550	84	1759	4499	42
1993–2007	475	27.51	550	29	1789	4516	41
1948–2007*	1381	26.61	550	17	1738	4516	41

* Pooled time series with inconsistent spatial structure.

HESSD
12, 2243–2265, 2015

Table 2. Precision statistics of the pooled within-time series (PTS) variograms estimated by spatially shifting temporal points (SSTP) and averaging empirical variograms (AEV) methods. The weighted sum of squared errors (SSE) as the variogram model-fit statistics and root means squared error (RMSE) as the ordinary kriging interpolation performance statistics are presented.

	SSE		RMSE	
Pooled time series	SSTP	AEV	SSTP	AEV
1948–1975	4.66×10^8	1.19×10^9	524.82	634.15
1776–1992	4.10×10^8	9.44×10^8	511.29	624.40
1993–2007	3.86×10^8	8.68×10^8	501.17	612.97
1948–2007*	2.27×10^9	3.29×10^9	572.06	683.32

* Pooled time series with inconsistent spatial structure.

HESSD
12, 2243–2265, 2015

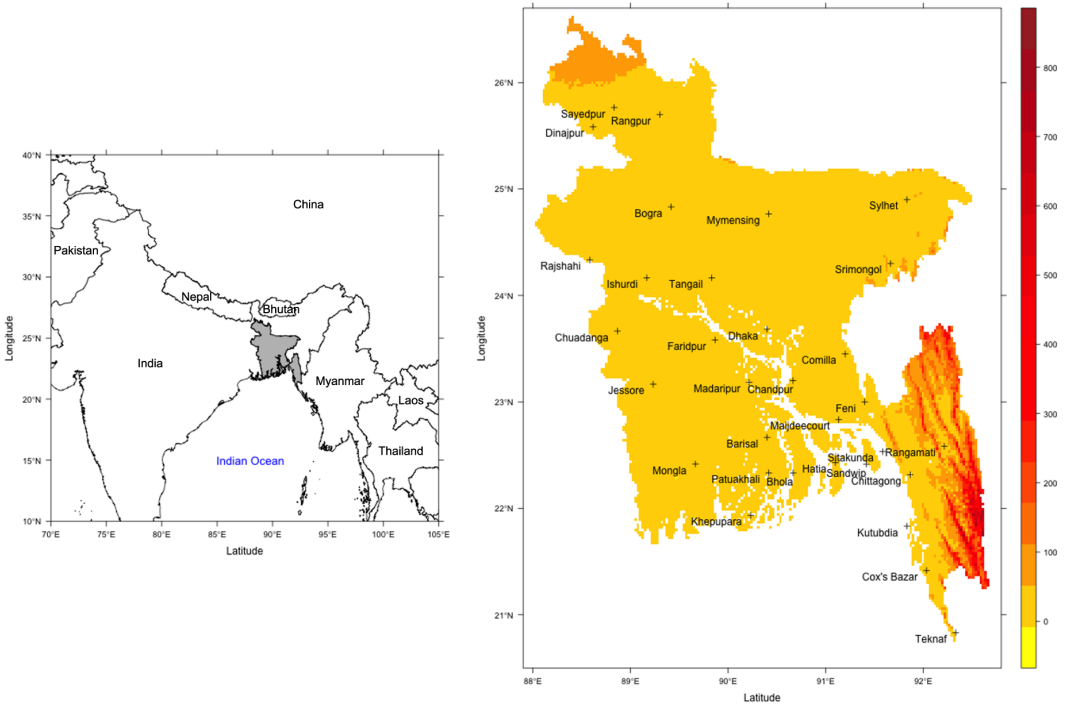


Figure 1. Geographic location of Bangladesh (left) in Southeast Asia within the coastal belt of Indian Ocean and the spatial distribution of currently active 32 rain-gauges (right) with altitudes (m above mean sea level) in the background. The coordinate reference system is WGS 1984.

HESSD
12, 2243–2265, 2015

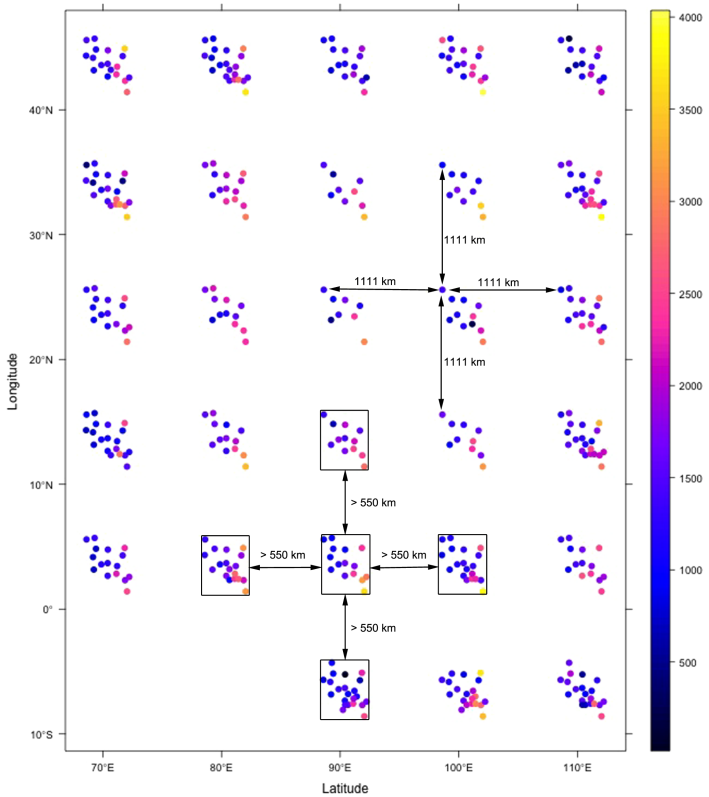
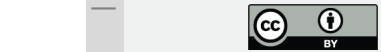


Figure 2. Spatially shifted (according to Eq. 1) temporal data points for the pooled 1948–1975 series. Shift distance (d = 1111 km) is calculated based on the largest-spatial-lag (550 km) available within the series (Eq. 3). The data point sets from neighboring years are shifted by 1111 km (~ 10°), which ensures that the peripheral points of the sets are shifted by > 550 km (~ 5°). The rectangles and legend indicate peripheries (convex hull) of data points in a year and PRCPTOT in mm, respectively.

HESSD
12, 2243–2265, 2015



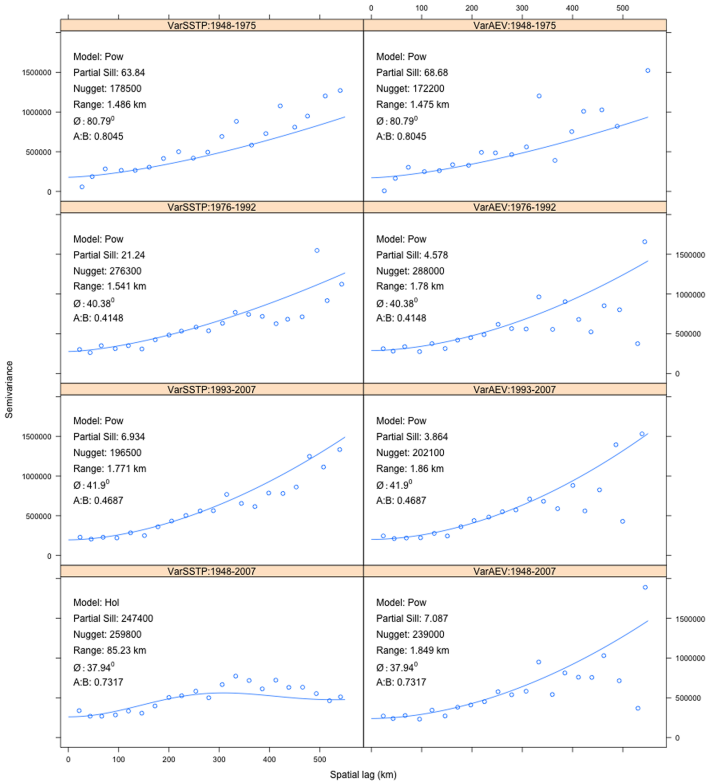


Figure 3. Estimated pooled within-time series (PTS) variograms (fitted best models to empirical variograms) estimated by spatially shifting temporal points (SSTP) and averaging empirical variograms (AEV) methods. Fitted variogram models ("Power" (Pow) and "Hole" (Hol)), partial sill and nugget variance, range, anisotropy angle (φ) and the ratio between major and minor axes of the anisotropy ellipse ($A : B$) are presented. Figure captions depict variogram(Var) estimation method: pooled series.

HESSD
12, 2243–2265, 2015