The role of station density for predicting daily runoff by top-kriging interpolation in Austria

Juraj Parajka[1*], Ralf Merz[2], Jon Olav Skøien[3], Alberto Viglione[1]

[1] Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Karlsplatz 13/222, Vienna, Austria.
[2] Department for Catchment Hydrology, Helmholtz Centre for Environmental Research – UFZ, Theodor-Lieser-Straße 4, Halle, Germany.
[3] Institute of Environment and Sustainability, Joint Research Centre, European Commission, Via Enrico Fermi 2749, Ispra, Italy.
[*] Corresponding author. E-mail: parajka@hydro.tuwien.ac.at

Abstract: Direct interpolation of daily runoff observations to ungauged sites is an alternative to hydrological model regionalisation. Such estimation is particularly important in small headwater basins characterized by sparse hydrological and climate observations, but often large spatial variability. The main objective of this study is to evaluate predictive accuracy of top-kriging interpolation driven by different number of stations (i.e. station densities) in an input dataset. The idea is to interpolate daily runoff for different station densities in Austria and to evaluate the minimum number of stations needed for accurate runoff predictions. Top-kriging efficiency is tested for ten different random samples in ten different stations densities. The predictive accuracy is evaluated by ordinary cross-validation and full-sample cross-validations. The methodology is tested by using 555 gauges with daily observations in the period 1987–1997. The results of the cross-validation indicate that, in Austria, top-kriging interpolation is superior to hydrological model regionalisation if station density exceeds approximately 2 stations per 1000 km^2 (175 stations in Austria). The average median of Nash-Sutcliffe cross-validation efficiency is larger than 0.7 for densities above 2.4 stations/1000 km^2. For such densities, the variability of runoff efficiency is very small over ten random samples. Lower runoff efficiency is found for low station densities (less than 1 station/1000 km^2) and in some smaller headwater basins.

Keywords: Top-kriging; Interpolation; PUB; Daily runoff; Station density.

INTRODUCTION

Most of the streams worldwide are ungauged and if daily runoff is needed at sites without observations, it has to be estimated from information in similar and/or neighbor basins. Such estimation is termed prediction in ungauged basin (PUB) and it has been centre of interest particularly during the last decade of International Association of Hydrological Sciences (2003–2013) (Sivapalan et al., 2003). The synthesis of results from this initiative (Blöschl et al., 2013) shows that most of the methods used for PUB in the past simulate daily runoff by conceptual hydrologic models. These models need estimates of climate inputs and calibration against observed runoff which is not directly possible in ungauged basins. Hence for PUB, different methods for spatial interpolation of climate characteristics (e.g. Gaál et al., 2008; Szolgay et al., 2009) and transferring model parameters from gauged to ungauged basin has been evaluated (see e.g. Hrachowitz et al., 2013; Merz and Blöschl, 2004; Parajka et al., 2005; Zvolenský et al., 2007). The comparison of results of different transfer methods in Parajka et al. (2013) indicates that all of them show a similar predictive performance with considerable scatter within each method. Evaluation of studies applying large number of basins and dense stream gauge network shows that spatial proximity and geostatistical interpolation of model parameters perform better than other methods (see also Merz and Blöschl, 2004 and Parajka et al., 2005).

An alternative to hydrological modeling is direct interpolation of runoff values. The main advantage of direct interpolation is that it avoids the use of uncertain input variables such as precipitation and potential evaporation. The limitation is that this method is data intensive, i.e., it can only be applied in medium to densely gauged regions, and it is not applicable when one is interested in the causal relationship between precipitation and runoff (Blöschl et al., 2013). There are several studies testing interpolation of runoff characteristics for PUB. Most of them, however, estimate only mean annual and monthly runoff values (see e.g. Bishop and Church, 1995; Gottschalk, 1993; Sauquet et al., 2000, 2008), flow duration curves (e.g., Pugliese et al. 2014) or flood statistics (e.g., Archfield et al., 2013; Salinas et al., 2013). Geostatistical interpolation of daily and hourly runoff values is evaluated in Skøien and Blöschl (2007) and in Skøien et al. (2008) for flood hydrographs. They assess the accuracy of top-kriging interpolation for Innviertel region in Austria and report an improved runoff prediction compared to estimates of regionalized hydrologic model simulations. Such findings are confirmed in a recent study of Viglione et al. (2013). This study shows that, in Austria, a direct interpolation of daily runoff by top-kriging is superior to simulations obtained by hydrologic model driven by transferred model parameters. The question is to what extent this finding is related to the density of gauging stations. The main objective of this study is thus to evaluate predictive accuracy of top-kriging interpolation driven by different number of stations (i.e. station densities) in an input dataset. The idea is to interpolate daily runoff for different station densities and evaluate minimum number of stations needed for accurate runoff prediction in Austria.

DATA

The study region for the analysis is Austria. This region has an area of about 84000 km^2 and is characterized by diverse physiographic and landscape characteristics. Topography varies from 115 m a.s.l. in the eastern lowlands to more than 3000 m a.s.l. in the Alps (Figure 1). Top-kriging interpolation is tested for daily runoff data in the period 1987–1997. The complete

dataset includes measurements of daily runoff at 555 gauging stations. The topography and spatial location of stations with runoff observations are shown in Figure 1. The size of the basins varies from 5 km² to more than 130000 km² and the median basin size is 141 km². The majority of the basins (222 stations) range in area between 5 and 100 km², 167 basins between 100 and 300 km², 93 basins between 300 and 1000 km² and 73 basins have areas of more than 1000 km². The runoff observations and selected meta-data about the basins are accessible through the webportal of the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management (http://ehyd.gv.at/).

METHODS
Top-kriging interpolation method

Top-kriging is a geostatistical interpolation method that allows interpolation of runoff characteristics along the stream network. It was developed by Skøien et al. (2006) and in comparison to traditional deterministic or geostatistical interpolation approaches, it accounts for the river network hierarchy (see also Gottschalk, 1993; Sauquet et al., 2000). Top-kriging combines two processes: local runoff generation, which is continuous in space, and runoff aggregation and routing along the stream network (Viglione et al., 2013). For interpolation, a point variogram, i.e. what can be seen as the expected variance of local runoff generation as a function of separation distance, is needed. The point variogram is integrated over the upstream contributing area for pairs of basin sizes associated with each river cross section. This integrated variogram then depends both on the sizes and the relative positions of basins along the stream river network. Previous studies (Lark, 2000; Skøien et al., 2014) show that the quality of the predictions is relatively insensitive to the choice of the variogram, at least as long as there are several observations within its range. In order to eliminate the uncertainties from variogram fitting in different time periods and from using different number of stations, we apply only one integrated variogram in each computation. The variogram applied here has been fitted, tested and applied for Austrian river network in previous studies of Skøien et al. (2006) and Merz et al. (2008). More details about the theoretical background, mathematical assumptions, description of an implementation of top-kriging in the rtop package (Skøien, 2014) in the statistical environment R (R Development Team, 2014) and an example of daily runoff interpolation in the upper Austria are presented in Skøien et al. (2006 and 2014).

Evaluation of daily runoff accuracy of top-kriging interpolation

Top-kriging interpolation is tested for different number of stations (i.e. station densities) in the input dataset. The input dataset is randomly selected from the complete dataset by using predefined number of stations with runoff observations. We tested ten variants of station densities (Table 1) and in each variant ten different random samples. Ten variants of station density and random samples are selected to cover typical range of station densities in different regions of Europe and to allow an evaluation of uncertainty of the runoff predictions. Figure 2 shows an example of one sample of three different station densities.
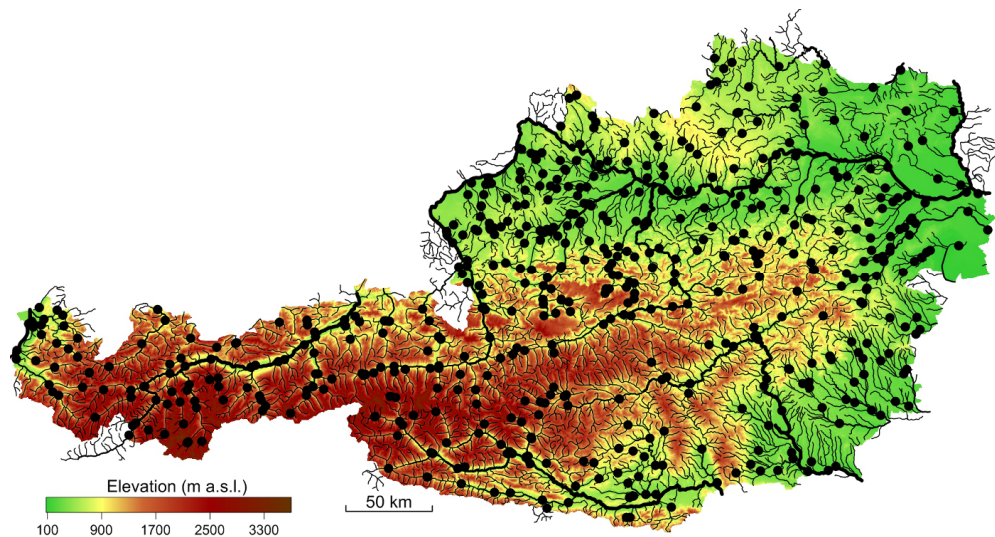
Fig. 1. Topography of Austria and location of 555 stations with daily runoff observations in the period 1987–1997.
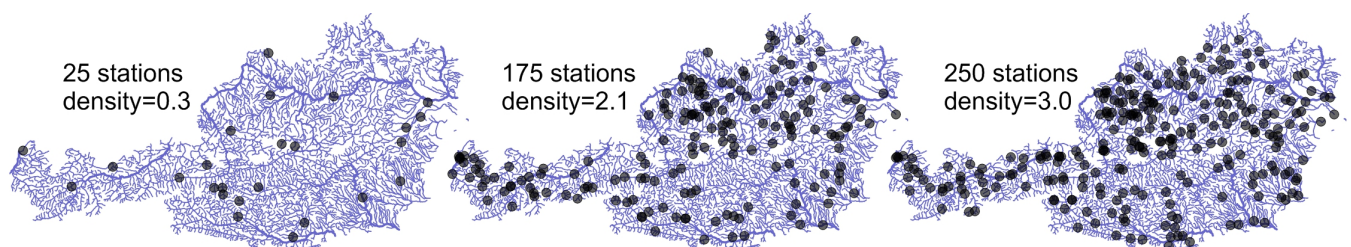
Fig. 2. Example of 3 different station densities (25, 175 and 250 stations).

Table 1. Ten different variants of station densities. Station density is estimated by dividing number of stations by the area of Austria (84000 km^2).

Number of stations	Station density (stations/1000 km^2)
25	0.3
50	0.6
75	0.9
100	1.2
125	1.5
150	1.8
175	2.1
200	2.4
225	2.7
250	3.0

The predictive accuracy of interpolation is then evaluated by two methods. The first is ordinary cross-validation, treating each gauged station in turn as ungauged and interpolating daily runoff from the other gauged stations in the input dataset. The interpolated runoff is then compared with the observed daily runoff for each station of the input dataset. The second method is termed full sample cross-validation here and we use the input data set to interpolate the runoff for all stations of the complete dataset and then compare these results with the observations. For example, if the input dataset consists of 25 randomly select-ed stations from the complete dataset (i.e. 555 stations) then the first method compares the interpolated runoff only at those 25 stations (in turn considered as ungauged), while the second method compares runoff interpolation at each station of the complete dataset, including stations with observations. We decided not to remove stations with observations in the full-sample cross-validation to assure identical datasets in different variants as well as to show the total performance of the interpo-lation (important e.g. in regional water quality modeling).

The predictive accuracy is quantified by the volume error (*VE*) and Nash-Sutcliffe efficiency (*NSE*, Nash and Sutcliffe, 1970) between interpolated and observed daily runoff time series in each basin. These criteria represent typical accuracy measures used in hydrologic modeling (including water quality simulations) and PUB assessment studies (Parajka et al., 2013). The volume error and Nash-Sutcliffe efficiency are defined as:

$$VE = \frac{\sum_{i=1}^{n} Q_{sim,i} - \sum_{i=1}^{n} Q_{obs,i}}{\sum_{i=1}^{n} Q_{obs,i}}, \tag{1}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(Q_{obs,i} - Q_{sim,i} \right)^2}{\sum_{i=1}^{n} \left(Q_{obs,i} - \overline{Q_{obs}} \right)^2} \tag{2}$$

where $Q_{sim,i}$ is the interpolated runoff on day i, $Q_{obs,i}$ is the observed runoff and $\overline{Q_{obs}}$ is the average of the observed runoff over the complete period of 4018 days (1987–1997).

The *NSE* and *VE* efficiencies are calculated separately for both cross-validation strategies.

RESULTS

The results of jack-knife cross-validation are presented in Figure 3. For each random sample a median of *NSE* and *VE* is estimated and its variability over ten random samples is summa-rized in the box-whiskers plots. Left panel (Fig. 3) shows that the variability of *NSE* medians decreases with increasing station density. The average median *NSE* efficiency over ten random samples is smaller than 0.5 and larger than 0.69 for station densities less than 1 and more than 2.1 stations/1000 km^2, re-spectively. The evaluation of *VE* (right panel, Fig.3) indicates that top-kriging tends to overestimate the long-term annual sums of daily runoff (bias). The overestimation and also its variability is larger for lower station densities (less than 0.6 station/ 1000km^2). While the average bias is larger than 7.5% for lower densities, it is less than 2.8% for densities above 1.8 station/1000 km^2.

The full sample cross-validation is presented in Figure 4. Similarly as in Figure 3, the left and right panels show the variability of *NSE* and *VE* medians, respectively. It is clear that evaluation of predictive accuracy over the complete dataset results in a significantly lower variability between medians of 10 samples of *NSE* and *VE*. The average of median *NSE* is below 0.45 for station densities less than 0.6, which is similar to
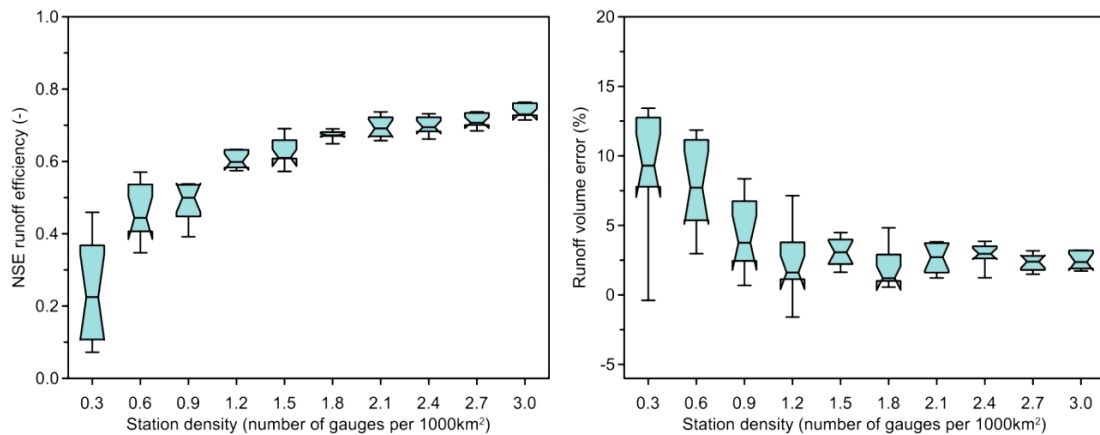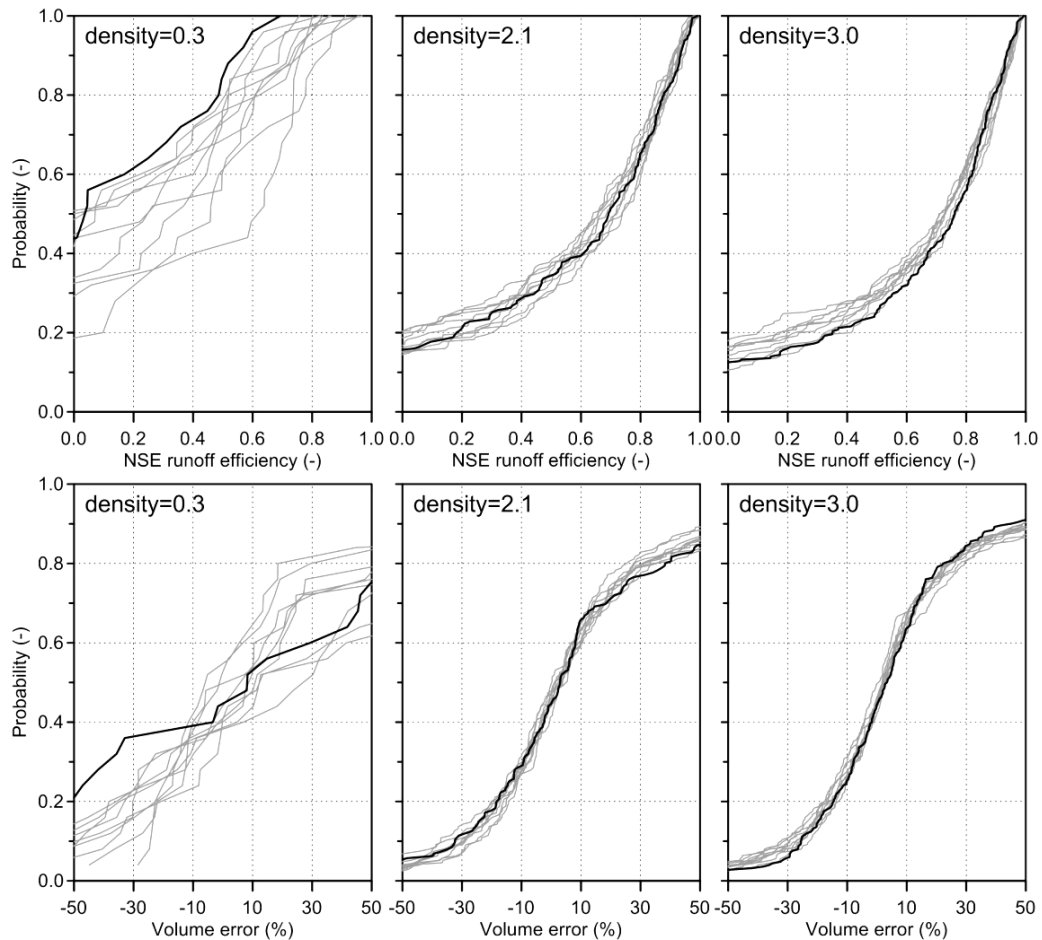

Fig. 3. Medians of Nash Sutcliffe (*NSE*) and runoff volume error (*VE*) efficiencies obtained by ordinary cross-validation for different sta-tion densities. Box (25%, 50% and 75% quantiles) and whiskers (10% and 90% quantiles) show the variability of *NSE* (left panel) and *VE* (right panel) medians over ten random samples.

ordinary cross-validation efficiencies. For larger densities are, however, the average *NSE* significantly larger. The average median of *NSE* for densities above 1.8 is larger than 0.82 and exceeds even 0.9 for densities larger than 2.7 station/1000 km². The assessment of *VE* medians indicates that top-kriging tends to overestimate daily runoff only for lower station densities. For densities above 1.2, top-kriging predictions are essentially unbiased.

While Figure 3 presents only the scatter between medians of random samples, Figure 5 shows the entire distributions of ordinary cross-validation efficiencies (*NSE* and *VE*) for selected station densities (0.3, 2.1 and 3 stations/1000 km²). The bold lines represent random samples presented in Figure 2. Figure 5 indicates that for larger densities (i.e. above 2) the predictive accuracy of top-kriging is indeed very similar not only in terms of medians but also in entire cumulative distributions.
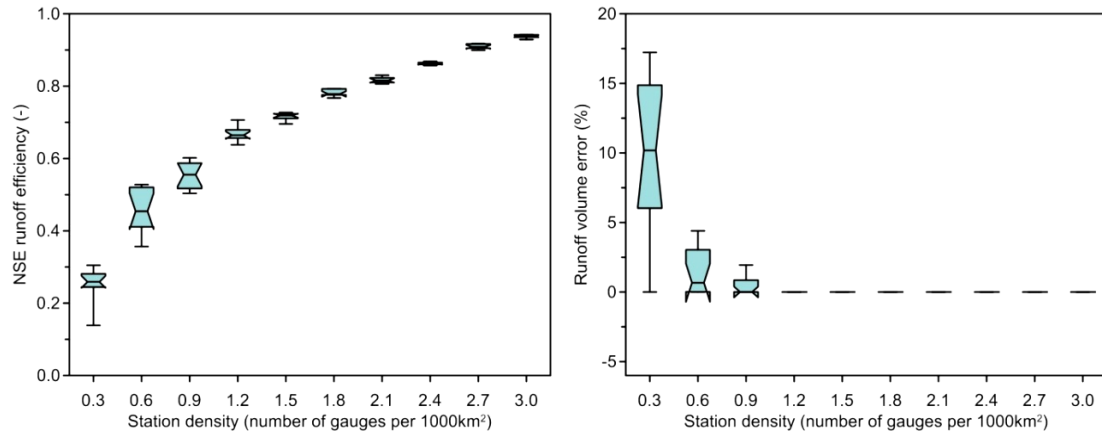
Fig. 4. Medians of Nash Sutcliffe (*NSE*) and runoff volume error (*VE*) efficiencies obtained by full sample cross-validation for different station densities. Box (25%, 50% and 75% quantiles) and whiskers (10% and 90% quantiles) show the variability of *NSE* (left panel) and *VE* (right panel) medians over ten random samples.
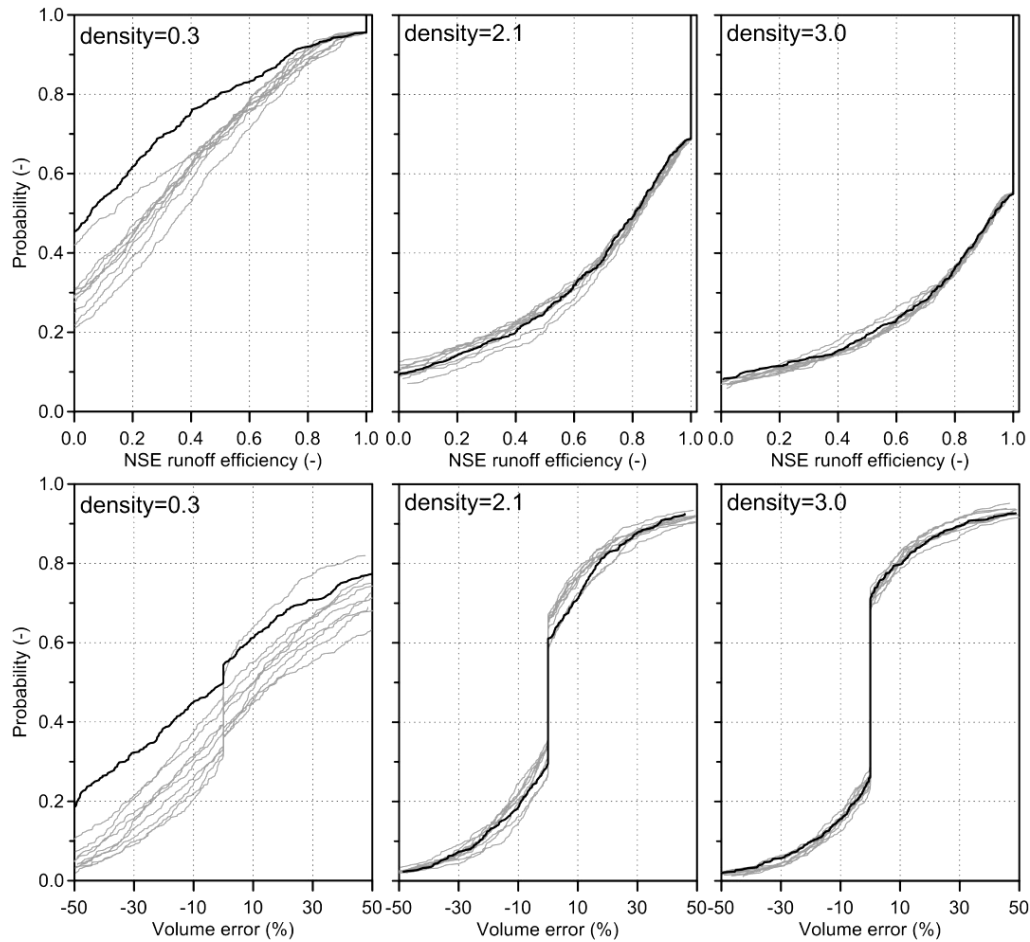
Fig. 5. Distribution of Nash Sutcliffe (*NSE*, top panels) and runoff volume error (*VE*, bottom panels) efficiencies obtained by ordinary cross-validation for ten random samples and selected station densities. Bold lines represent random samples presented in Figure 2.

Noticeable differences between random samples are observed only for densities below 0.6 stations/1000 km².

The comparison of full-sample cross-validation efficiencies (Figure 6) confirms a stable performance of top-kriging for different random samples and larger station densities. Figure 6 evaluates the *NSE* and *VE* efficiencies for the complete dataset, so for different input densities, there is different number of stations with perfect (identical) predictions to observations (straight lines representing $VE = 0$ and $NSE = 1$). Obviously, the best runoff performance is obtained for the largest station density where more than 60% of stations have *NSE* efficiency larger than 0.7 and absolute volume error less than 10%. Interestingly, the loss in top-kriging runoff efficiency is very small for densities larger than 2.1.

Fig. 6. Distribution of Nash Sutcliffe (*NSE*, top panels) and runoff volume error (*VE*, bottom panels) efficiencies obtained by full sample cross-validation for ten random samples and selected station densities. Bold lines represent random samples presented in Figure 2.
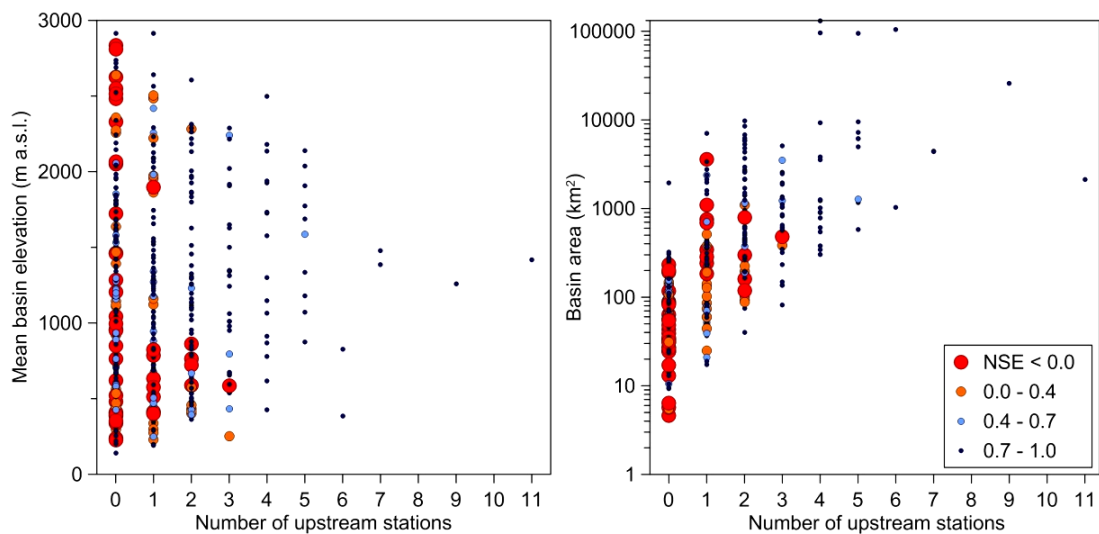
Fig. 7. Relation between the *NSE* full-sample cross-validation efficiency (color and size of symbols) and mean basin elevation, basin area and number of upstream stations for top-kriging interpolation with 250 stations (Figure 2).

The factors that control the accuracy of top-kriging are evaluated in Figure 7. Figure 7 shows an example of the relation between full-sample cross-validation efficiency of *NSE* and basin size, mean basin elevation and number of upstream stations for one random sample size (top-kriging interpolation by 250 stations, Figure 2). Please note that, in order to highlight basins with poor performance, small *NSE* are plotted in large red circles, while larger *NSE* values are plotted in small circles. Our investigation (not shown here) indicates that even if there are some small differences between different random samples, this example shows a typical case for larger station densities, i.e. more than 2 stations/1000 km^2. Left panel of Figure 7 indicates that the lowest top-kriging efficiency is found for some headwater basins. These basins do not have upstream basins with daily runoff observations. They are typically smaller than 200 km^2 (right panel, Figure 7) and are situated at different elevation zones. Mean basin elevation is not found to be a robust factor for discrimination between poor and accurate runoff predictions. On the other hand, the river network hierarchy seems to be important. The largest top-kriging accuracy is observed for basins with at least 4 upstream basins.

DISCUSSION AND CONCLUSIONS

This study evaluates the predictive accuracy of daily runoff interpolated by top-kriging as a function of different station density. Previous studies by Skøien and Blöschl (2007) and Viglione et al. (2013) have already shown that top-kriging can perform better than conceptual hydrologic models with regionalized model parameters. Skøien and Blöschl (2007) even showed that the interpolation method can perform better than a conceptual hydrologic model calibrated at site.

Parajka et al. (2005) analyzed several methods for transferring model parameters, and found that similarity and kriging based approaches, have been the most accurate for PUB in Austria. They found the median *NSE* from cross-validation to be 0.67 for both methods in 330 stations, which corresponds to density almost 4 stations/1000 km^2. Similar top-kriging efficiency is, in this study, found for a density of less than 2.1 stations/1000 km^2. Larger densities thus allow more accurate predictions of daily runoff with top-kriging if the causal relationship between precipitation and runoff is not of interest. A station density of 3 stations/1000 km^2 (250 stations) allows even more accurate runoff predictions than what was obtained by at site model calibration in previous studies. While Viglione et al. (2013) found the median *NSE* from hydrologic model calibration to be 0.72, the average median *NSE* top-kriging exceeds 0.74. Such efficiency is one of the largest found in any of the PUB assessments (Parajka et al., 2013) for larger datasets.

On the other hand, our study also shows that the predictability deteriorates rather quickly when the station density drops below 1 and particularly below 0.5 stations/1000 km^2. It would not be advisable to use top-kriging in Austria or similar regions for such low station densities. The exact limits for low performance of geostatistical interpolation will depend mainly on the hydrological gradient of the study area. For Austria, we assume that this is mainly controlled by the large differences in runoff seasonality (Parajka et al., 2009).

The territory of Austria covers regions with large physiographic and landscape variability and thus represents a wide range of geographical variability in Europe. We believe that our results applies and can be transposed to regions with similar conditions. This study evaluates top-kriging performance in a selected decade and focuses mainly on the runoff bias and *NSE*

model efficiency. In our next study, we plan to extend this assessment to evaluation of top-kriging accuracy for seasonal and extreme runoff predictions, as well as to compare the performance in different hydro-climate regions of Austria. We thus plan to evaluate the effects of seasonally or even daily varying spatial correlations of runoff by testing temporally variable fittings of top-kriging variogram. Our results also indicate that one of the top-kriging challenges is prediction of runoff in headwater basins. We plan to find methods to identify basins with poor predictability in advance, and look into alternative or possible hybrid methods for these. Patil and Stieglitz (2012) used a simple approach with inverse distance weighting for runoff data from the entire USA, and identified some general characteristics for basins with high predictability, such as runoff ratio, baseflow runoff ratio and slope. These might also be helpful indicators on a smaller spatial scale, in addition to the basin area and the number of upstream gauges.

REFERENCES

Archfield, S.A., Pugliese, A., Castellarin, A., Skøien, J.O., Kiang, J.E., 2013. Topological and canonical kriging for design flood prediction in ungauged catchments: an improvement over a traditional regional regression approach? Hydrol. Earth Syst. Sci., 17, 1575–1588.

Bishop, G.D., Church, M.R., 1995. Mapping long-term regional runoff in the eastern United States using automated approaches. Journal of Hydrology, 169, 189–207.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., Savenije, H. (Eds.), 2013. Runoff Prediction in Ungauged Basins Synthesis across Processes, Places and Scales, Cambridge University Press, 500 p.

Gaál, L., Kyselý, J., Szolgay, J., 2008. Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia. Hydrology and Earth System Sciences, 12, 3, 825–839.

Gottschalk, L., 1993. Interpolation of runoff applying objective methods. Stoch. Hydrol. Hydraul., 7, 269–281.

Hrachowitz, M., Savenije, H.H., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB) – a review. Hydrolog. Sci. J., 58, 6, 1198–1255.

Lark, R. M., 2000. A comparison of some robust estimators of the variogram for use in soil survey. European Journal of Soil Science, 51, 1, 137–157.

Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. J. Hydrol., 287, 95–123.

Merz, R., Blöschl, G., Humer, G., 2008. National flood discharge mapping in Austria. Nat. Hazards, 46, 53–72.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Journal of Hydrology, 10, 3, 282–290.

Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. Hydrol. Earth Syst. Sci., 9, 9, 157–171.

Parajka, J., Kohnová, S., Merz, R., Szolgay, J., Hlavcová, K., Blöschl, G., 2009. Comparative analysis of the seasonality of hydrological characteristics in Slovakia and Austria. Hydrological Sciences Journal, 54, 3, 456–473.

Parajka, J., Viglione, A., Rogger, M., Salinas, J.L., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies. Hydrol. Earth Syst. Sci., 17, 1783–1795.

Patil, S., Stieglitz, M., 2012. Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. Hydrol. Earth Syst. Sci., 16, 551–562.

Pugliese, A., Castellarin, A., Brath, A., 2014. Geostatistical prediction of flow–duration curves in an index-flow framework. Hydrol. Earth Syst. Sci., 18, 3801–3816.

R Development Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Salinas, J.L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies. Hydrol. Earth Syst. Sci., 17, 2637–2652.

Sauquet, E., L. Gottschalk, Leblois, E., 2000. Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme. Hydrological Sciences Journal, 45, 6, 799–815.

Sauquet, E., Gottschalk, L., Krasovskaia, I., 2008. Estimating mean monthly runoff at ungauged locations: an application to France. Hydrology Research, 39, 5–6, 403–423.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB) 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal, 48, 6, 857–880.

Skøien, J.O., 2014. Rtop: Interpolation of data with variable spatial support. R package version 0.3-45, http://CRAN.R-project.org/package=rtop.

Skøien, J.O., Blöschl, G., 2007. Spatiotemporal topological kriging of runoff time series. Water Resour. Res., 43, W09419, doi: 10.1029/2006WR005760.

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

Skøien, J.O., Pebesma, E.J., Blöschl, G., 2008. Geostatistics for automatic estimation of environmental variables - some simple solutions. Georisk, 2, 4, 257–270.

Skøien, J.O., Blöschl, G., Laaha, G., Pebesma, E., Parajka, J., Viglione, A., 2014. Rtop: An R package for interpolation of data with a variable spatial support, with an example from river networks. Computers & Geosciences, 67, 180–190.

Szolgay, J., Parajka, J., Kohnová, S., Hlavčová, K., 2009. Comparison of mapping approaches of design annual maximum daily precipitation. Atmospheric Research, 92, 3, 289–307.

Viglione, A., Parajka, J., Rogger, M., Salinas, J.L., Laaha, G., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria. Hydrol. Earth Syst. Sci., 17, 2263–2279.

Zvolenský, M., Hlavčová, K., Kohnová, S., Szolgay, J., 2007. Prediction of rainfall-runoff model parameters in ungauged catchments. IAHS-AISH Publication, 313, 357–364.

Note: Colour version of Figures can be found in the web version of this article.