

* References to lines and pages, coloured in blue in the following, refer to the revised version of the manuscript.

** Tracked changes version of the manuscript is attached at the end of this document.

❖ Reply to referee W.H. Farmer

Referee comment:

The authors have presented a very well-prepared manuscript exploring the value of post-processing or “bias-correcting” daily streamflow simulations with independently derived regional information. By using an independently developed flow duration curve, the authors present an approach to customize macro-scale models to local conditions. The approach is very valuable. The manuscript is well-written and thorough. Below, I will provide some minor comments, but I see no impediment to swift publication.

Authors reply:

We thank Reviewer #1 W. H. Farmer for his valuable and precise comments. We are pleased of the appreciation for the overall study, and we hope that the replies to his comments presented below might resolve the issues arisen.

Referee comment:

Was the performance of GAE-HYPE always better than the performance of E-HYPE? At the top of page 10, the authors cite the best and worst improvements. Were there any sites that showed decreases in performance? If not, would you expect universal improvement in other regions?

Authors reply:

Yes, we obtained improvements in each one of the 11 pairs, however it is worth recalling that EHYPE was not calibrated in those sites, which results in locally poor performance of EHYPE. Overall, improvements are always to be expected when FDCs are well predicted by the geostatistical interpolator.

ACTION: In the revised version of the manuscript we will underline that all selected pairs show improvements in the performances.

Done. See P12 L8.

Referee comment:

Was the degree of change in LNSE between GAE-HYPE and E-HYPE a function of the accuracy with which the TNDTK FDC was produced? That is, if the TNDTK FDC were poorly produced, it seems like the improvement in LNSE might be reduced or reversed. Some exploration of this might be useful.

Authors reply:

The reviewer is right, the whole procedure relies on good prediction of regional FDC in the study area. When FDCs are poorly predicted across the area, detriments of the geostatistical assimilation procedure are to be expected. Focusing on the 11 EHYPE prediction nodes, FDCs are actually very well predicted, with an average LNSE of 0.988 in cross-validation (with minimum and maximum values about 0.967 and 0.998, respectively) compared to the average performance in the whole study area, which is 0.898. Therefore, we cannot show here the effects of poor geostatistical interpolation of FDCs on the assimilation procedure.

ACTION: we will specifically highlight in the revised manuscript that biased regional model for predicting FDCs are expected to lead to poor performances of the proposed geostatistical assimilation procedure.

Done. See P6 L14-17.

Referee comment:

What method of interpolation was used to map the residuals to simulated streamflows? On line 13 of page 9, it is stated that the TNDTK FDC is resampled to 20 points. So, if a simulated streamflow (E-HYPE) produced a duration that did not fall at one of the 20 resampled points, how was the residual estimated? That is, how was the eDC resolved between these 20 points?

Authors reply:

We thank the reviewer for letting us explain further this technical aspect. Observed FDCs are resampled to 20 equally spaced points, in the normal space, i.e. $d \sim \{0.0004; 0.0013; 0.0040; 0.0108; 0.0259; 0.0558; 0.1080; 0.1884; 0.2979; 0.4298; 0.5702; 0.7021; 0.8116; 0.8920; 0.9442; 0.9741; 0.9892; 0.9960; 0.9987; 0.9996\}$. The extreme values of the resampling scheme depends on the length of the shorter observed streamflow series in the dataset (i.e. 7 yrs), so that, those values extruding from such extremes are excluded in the resampled curves. Each EHYPE simulation covers a period of 30 yrs. with no gap, therefore we adopted the same resampling interval, with 20 points, used for the observed series. As a result, the residual duration curves are defined in the same 20 points of the resampling scheme and a linear interpolation is employed in between each data point. Finally, according to Pugliese et al. (2014), regional interpolation of TNDs, which is the weighting scheme for the curves as well, does not require any resampling scheme beforehand.

ACTION: in the revised version, we will add more information on the resampling strategy we used and we will clearly underline that a finer resolution of the curves could be necessary in specific applications, in which very high (low) duration are of particular interest.

Done. See P8 L21-25, and P13 L14-17.

Referee comment:

Does the simulated FDC resulting from the GAE-HYPE series match the TNDTK FDC? It seems as if it could (should?), by the nature of the method. This suggests that this method could also be considered as a re-scaling of the simulated streamflow distribution. Essentially, this means that the volumes from E-HYPE are discarded while the sequencing of E-HYPE (durations, relative values) is retained. I do not see anything wrong with this, but wonder if it is another way to think about the procedure. If accurate, does this way of thinking provide any further insight?

Authors reply:

The reviewer rises a good point here.

ACTION: In the revised version of the manuscript we will add two panels in fig. 9 where we show FDCs of either empirical, E-HYPE, or GAE-HYPE next to their streamflow series counterparts, since we believe that the result is worth showing. We will also include in the discussion an analysis resulting from this comparison and possible insights on the aspect highlighted by the reviewer (the proposed geostatistical assimilation procedure discards simulated volumes, while retaining simulated sequencing).

Done. See the revised fig. 9, see also P13 L27-30.

Referee comment:

What was the significance of the changes in LNSE? Firstly, equation 10 can be simplified as the fractional change in root-mean-squared error of logarithms. This, can, of course, be interpreted as a percent. (Line 1 of page 11 uses percent, but the figure does not.) More importantly, the LNSE values could be compared in a pairwise test to determine if the improvement in LNSE is statistically significant (Wilcoxon). I imagine it is, but demonstrating this would provide stronger evidence.

Authors reply:

We thank the reviewer for these recommendations, which we will add in the revised version of the manuscript. We are aware that the proposed $LNSE_{ratio}$ can be simplified as the fractional change in root-mean-squared error of logarithms, therefore, we cannot see much difference in using one or the other, however, for the sake of consistency, we would prefer to keep this assessment metric in terms of LNSE.

ACTIONS:

- 1) We will clearly highlight that the proposed index derives from the fractional change in root-mean-squared error of logarithms.
- 2) We will improve the consistency between notation and units in figures and text.
- 3) In addition, for each gauge density scenario, we will perform a pairwise test (Wilcoxon test) between LNSE values in order to determine whether improvements are statistically significant in terms of LNSE.

Done:

- 1) See P9 L26.
- 2) See e.g. equations subscripts switched from i to j whenever they refer to streamgauge spatial index.
- 3) See P9 L27-29 and P12 L25-27.

Referee comment:

How were streamflow values of zero handled? The authors measure performance in terms of the LNSE. What was the frequency of zeros? How were there logarithms taken?

Authors reply:

Although the alpine climate might be at risk for the presence of zero flows within the series, especially during the winter seasons, we observed no zero flows in the selected 11 sites with either empirical, EHYPE or GAE-HYPE.

ACTION: We will report such information in section 4.

Done. See P10 L22-23.

Referee comment:

How many bins were used to discretize the variograms? Binning is described on line 27 of page 7, but it might be worthwhile to be explicit.

Authors reply:

We set the number of bins as the default setting of the “rtop” package, which is 1 bin per order of magnitude of drainage area. In our case, areas span over 4 orders of magnitude, therefore we used 3 bins for variogram pairs.

ACTION: We will add this information too.

We double-checked the results and the number of bins used, relative to three orders of magnitude in the drainage areas range, are actually 2. We reported this information at P7 L13-15.

Referee comment:

Editorial: On line 16 of page 1 the authors use “macro-scale” while line 17 uses “macroscale”. Both are used throughout the manuscript; select one.

Authors reply:

Thanks.

ACTION: We will use “macro-scale” and change accordingly.

Done.

Referee comment:

Editorial: The figures seem to be out of order. Fig. 9 is mentioned after Fig. 7 and before Fig. 8.

Authors reply:

Thanks.

ACTION: We will change accordingly.

Done.

❖ Reply to anonymous referee#2

Referee comment:

General Comments: This paper “A geostatistical data-assimilation technique for enhancing macro-scale rainfall-runoff simulations” develops a geostatistical method for enhancing streamflow simulation performance of large-scale rainfall-runoff models. The proposed method has proved to be effective for Tyrol area and shows great potential for basins with few gauges or even ungauged basins.

Referee comment:

Some major comments: (1) Organizations of Section 5 seem to be not logical. Following “Section 4: study area”, some new method (e.g. LOOCV) and new metric (e.g. LNSE) are introduced in Section 5, and then followed by the findings and discussions. This may confuse readers as they may fail to grasp the intention and the core of this paper. It had better reorganize the manuscript in ‘Method-Results’ order.

Author reply:

Although we see reviewer’s comment, we did not introduce LOOCV and LNSE earlier in the text as these are a common cross-validation procedure (see e.g. Kroll and Song, 2013; Salinas et al., 2013; Wan Jaafar et al., 2011; Srinivas et al., 2008) and a widely used metric (see e.g. Farmer, 2016; Castellarin, 2014). Yet, we do agree that this may generate some confusion.

ACTION: We decided to introduce a new section in the manuscript, in between “Study area” and “Results” sections, titled “4 Assessment of the geostatistical assimilation algorithm” in which we present the structure of the analysis, the cross-validation strategies and the performance index. This section will report the following subsections:

4.1 Structure of the analysis

4.2 Cross-validation strategy

4.3 Performance indices

Nevertheless, before changing the sectioning, and thus the whole structure of the paper, we would like wait for editor decision.

The structure has been fully revised. We decided to modify the sectioning by moving the structure proposed above before the description of study area. Please, pay attention to changes applied at sections 2, 3, and 4 whereas the results have been blended to a single section.

Referee comment:

(2) The concept of total negative deviation (TND) is very essential to this paper’s research. So, it is recommended that TND be explained more clearly with both words and figure. Particularly, Figure 1 should be interpreted in detail, for instance, what does the symbol ‘1’ exactly stand for?

Authors reply:

(2) We are grateful to the reviewer for highlighting this lack of clarity behind the idea of TND. We will add more information about this novel metric in the revised version of the paper. TND is conceived to mimic the slope of a standardised FDC, where the standardisation is an arbitrary reference value (e.g. mean annual flow; see also Pugliese et al., 2014, for other standardisation methods). Therefore, “1” on the y-axis in Fig. 1 represents the equality between a given streamflow record and the reference value.

ACTION: We will add this sentence in P4 L25 : “[...] The equality between a given streamflow value and the reference value Q^* is represented by an horizontal dashed line in Fig. 1, i.e. the threshold given by the equation $Q/Q^*=1$. [...]”

Done. See P4 L25-26.

Referee comment:

(3) Why choose the mean annual flow (MAF) as the reference value Q^* in applying equation (3)? Does this mean that you didn’t consider the flow above MAF when applying “TNDTK”, according to the definition of TND by the shaded area in Figure 1? If this is the case, more should be elaborated on this.

Authors reply:

(3) We understand that this paper neither explains carefully the idea behind the TND, nor it presents a thorough assessment of its reliability as regional metric summarising FDC’s, however we think that this is out of the scope of this paper. Indeed, such assessments have been already carried out in other two independent research studies, where we proposed the TNDTK method and we contrasted its performances against other regional models, such as regional regression and statistical models, which are known to be the state-of-the-art for regional FDC predictions in ungauged basins (see Pugliese et al., 2016, 2014).

ACTION: We will better clarify this in the revised manuscript, providing an interested reader with indications on original literature sources.

Done. See P5 L12-15.

Referee comment:

(4) Page 8 line 28. It stated that “each pair includes one of the E-HYPE catchments depicted in Fig. 4 and its corresponding gauged catchment.” How to determine the corresponding gauged catchment for a certain E-HYPE catchment? By Equation 4 or other method? Please explain in detail.

Authors reply:

(4) The selection criteria of the 11 sites used for comparing the proposed technique with EHYPE simulations have been explained in section 4 Study area (which will be Section 5 in the revised manuscript).

ACTION: We will change the sentences on P7 L10 “[...] We selected only those E-HYPE prediction nodes located within Tyrol, which were the closest to one of the available streamgauges. In terms of selection criteria, we selected the E-HYPE catchments that showed limited differences in terms of (1) drainage areas (<14%) and (2) distance between catchment centroids (<15km). These criteria resulted in 11 E-HYPE prediction nodes that are evenly distributed in the study region (see red lines in Fig. 4). [...]”

with

“[...] Among all E-HYPE prediction nodes available in Tyrol we selected only those whose catchments were the closest to gauged ones, i.e. differences in terms of drainage areas <14% and distance between catchment centroids <15km. These criteria resulted in the selection of 11 E-HYPE prediction nodes that are evenly distributed in the study region (see red lines in Fig. 4). [...]”

Finally, it is worth noting that Eq. (4) does not deal with catchment selection, but with empirical TND computation only.

Done. See P10 L12-15.

Referee comment:

Minor Comments: (1) I notice that Equation 1 takes the symbol ‘ i ’ as indicator of catchments, while in Equation 4 it represents the quantiles of q_i . This should be avoided. Please use a different symbol. (2) In Figure 9, the black dashed line indicates the observed streamflow series, which is not in line with its legend, where the black solid line is plotted. This is a minor mistake that could have been avoided. (3) In Figure 10, the meaning of the black triangles is interpreted in the title. This is not a good idea. Please use legends instead. (4) What does ‘TNDTK’ mean in the title of Section 2.1? DO NOT use abbreviation before it is defined.

Authors reply:

(1) The reviewer is right. We will substitute i with j in Eq. (1).

(2) Thanks. We will change accordingly.

(3) Thanks. We will add a legend for both the black triangles and catchment boundaries.

(4) Thanks. We will drop both TNDTK and GAE-HYPE from the title of the sections.

1) Done. See eq. 1.

2) Done. See fig. 9.

3) Done. See fig. 10.

4) Done.

❖ Reply to anonymous referee#3

Referee comment:

(1) The symbol should confirm, such as equation (6).

Authors reply:

Thanks.

ACTION: the missing cap over $\hat{Q}^*(x_0)$ will be added in the revised version.

Done. See eq. 6.

Referee comment:

(2) This author exploring a technique for the daily streamflow simulation post processing, whether the name “dataassimilation technique” is ok or another name is better

Authors reply:

There are some similarities between our technique and applications in climate modelling (see e.g. Komma et al., 2007). We would like to keep the title as it is now.

Referee comment:

This method used the information from local information (such as rainfall, landuse and topography) and neighbour watershed data (observation runoff data for FDC), Can we get the effect of local/ neighbour information on different part of runoff simulation (such as peak flow and baseflow) combined this technique.

Authors reply:

The power of the proposed technique relies on the fact that no further observations than mere discharges are needed for enhancing streamflow simulations. Surely, it would be interesting to investigate how other hydrological features, such as baseflow index or peakflow data, might be assimilated in the method.

ACTION: We will underline in the discussion section that future research studies will deal this problem.

Done. See P14 L24-26.

Referee comment:

Whether the method can be applied in real time forecasting, and hope the author give us some perspective.

Author comment:

In principle it could be used by blending this assimilation technique to e.g. long term forecast, even though the proposed method cannot be applied without any locally observed streamflow series.

ACTION: We will add in the discussion section that future analyses will assess the reliability of the method with the final aim to provide better simulations for practitioners at operational level, e.g. applications in civil protection management strategies, climate change trends, safety of river structures, etc.

Done. Here we revised the conclusions rather than the discussion section. See P14 L21-24.

Referee comment:

I tried to do the example, but I cannot download the Observed daily streamflow series. <http://www.water-switchon.eu/sip-webclient/sip-beta/#/resource/12072>.

Authors reply:

(5) Thanks. We will fix the broken link.

Links fixed. Please, refer to the one in the revised version of the paper.

References

- Castellarin, A., 2014. Regional prediction of flow-duration curves using a three-dimensional kriging. *J. Hydrol.* 513, 179–191. <https://doi.org/10.1016/j.jhydrol.2014.03.050>
- Komma, J., Reszler, C., Blöschl, G., Haiden, T., 2007. Ensemble prediction of floods – catchment non-linearity and forecast probabilities. *Nat Hazards Earth Syst Sci* 7, 431–444. <https://doi.org/10.5194/nhess-7-431-2007>
- Kroll, C.N., Song, P., 2013. Impact of multicollinearity on small sample hydrologic regression models. *Water Resour. Res.* 49, 3756–3769. <https://doi.org/10.1002/wrcr.20315>
- 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. <https://doi.org/10.5194/hess-18-3801-2014>
- Pugliese, A., Farmer, W.H., Castellarin, A., Archfield, S.A., Vogel, R.M., 2016. Regional flow duration curves: Geostatistical techniques versus multivariate regression. *Adv. Water Resour.* 96, 11–22. <https://doi.org/10.1016/j.advwatres.2016.06.008>
- 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. <https://doi.org/10.5194/hess-17-2637-2013>
- Srinivas, V.V., Tripathi, S., Rao, A.R., Govindaraju, R.S., 2008. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. *J. Hydrol.* 348, 148–166. <https://doi.org/10.1016/j.jhydrol.2007.09.046>
- Wan Jaafar, W.Z., Liu, J., Han, D., 2011. Input variable selection for median flood regionalization. *Water Resour. Res.* 47, W07503. <https://doi.org/10.1029/2011WR010436>

A geostatistical data-assimilation technique for enhancing macro-scale rainfall-runoff simulations

Alessio Pugliese¹, Simone Persiano¹, Stefano Bagli², Paolo Mazzoli², Juraj Parajka³, Berit Arheimer⁴, René Capell⁴, Alberto Montanari¹, Günter Blöschl³, and Attilio Castellarin¹

¹Department DICAM, Univeristy of Bologna, Bologna, Italy

²GECOSistema srl, Cesena, Italy

³Institute for Hydraulic and Water Resources Engineering, TU Wien, Wien, Austria

⁴Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

Correspondence to: Attilio Castellarin (attilio.castellarin@unibo.it)

Abstract. Our study develops and tests a geostatistical technique for locally enhancing macro-scale rainfall-runoff simulations on the basis of observed streamflow data that were not used in calibration. We consider Tyrol (Austria and Italy) and two different types of daily streamflow data: macro-scale rainfall-runoff simulations at 11 prediction nodes and observations at 46 gauged catchments. The technique consists of three main steps: (1) period-of-record flow-duration curves (FDCs) are geostatistically predicted at target ungauged basins, for which macro-scale model runs are available; (2) residuals between geostatistically predicted FDCs and FDCs constructed from simulated streamflow series are computed; (3) the relationship between duration and residuals is used for enhancing simulated time series at target basins. We apply the technique in cross-validation to 11 gauged catchments, for which simulated and observed streamflow series are available over the period 1980-2010. Our results show that (1) the procedure can significantly enhance macro-scale simulations (regional $NSE-LNSE$ increases from nearly zero to ≈ 0.7 in our study area ≈ 0.7) and (2) improvements are significant for low gauging network densities (i.e. 1 gauge per ≈ 2000 km²).

1 Introduction

The steady increase in computational capabilities together with the expanding accessibility of regional and global datasets (e.g. soil properties, land-cover, morphology, climate characteristics, satellite-based gridded precipitation, etc.) trigger the development of regional/continental- and global-scale hydrological models (Archfield et al., 2015), hereafter referred to as macro-scale models.

During the last decade, several of these $macro\ scale$ $macro\ scale$ models have become operational and thus continuously provide data automatically for decision-making. For instance, the distributed rainfall-runoff-routing model LISFLOOD (De Roo et al., 2000) provides daily forecast for operational warning services through the systems of EFAS (Pappenberger et al., 2013) and GLOFAS (Alfieri et al., 2013); the LARSIM models (Haag and Luce, 2008) are used operationally for simulating streamflow at large areas in southern Germany, Luxembourg, Austria, Switzerland, and the eastern part of France; the WATFLOOD, developed at the University of Waterloo is used operationally in Canada (Kouwen et al., 1993); and the S-HYPE

model (Strömqvist et al., 2012) is running operationally for flood or drought forecasting and water quality assessments for the Swedish landmass, providing high resolution information to authorities and citizens (Hjerdt et al., 2011).

Other macro-scale models are used for off-line water assessments and research purposes. For instance, the global WaterGAP Global Hydrological model (Alcamo et al., 2003) assists in water accounting; the SAFRAN-ISBA-MODCOU model (Habets et al., 2008) has been applied over the entire French territory to combine a meteorological analysis system, a land surface model, and a hydrogeological model; the PGB-IPH model (Pontes et al., 2017) has been applied to many South American basins; and the SWIM model (Krysanova et al., 1998) couples water balance simulations with water quality for small to mid-size watersheds, i.e. regional meso-scale.

The macro-scale hydrological models are getting more and more popular due to three main reasons: (1) they can provide users with a large-scale representation of hydrological behavior, a fundamental information for effectively addressing several water resources planning and management problems (e.g. surface water availability assessment, instream water quality studies, ecohydrological studies, etc.); (2) they can be used to compute a variety of hydrological signatures everywhere along the stream network at the resolution of the model; (3) model outputs in some cases are open-access and freely distributed, so that their regional runs represent a wealth of information for addressing the problem of hydrological predictions in data scarce regions of the world (Pechlivanidis and Arheimer, 2015; Donnelly et al., 2016; Beck et al., 2016). Accurate regional hydrological simulations undoubtedly foster and support the implementation of improved large-scale and trans-boundary policies for water resources system management and flood-risk mitigation or climate change adaptation (de Paiva et al., 2013; Sampson et al., 2015; Falter et al., 2016; Arheimer et al., 2017).

However, improved accuracy in terms of average regional performance does not necessarily imply homogeneous improvements in local performance. In fact, due to the difficulties to perform local calibrations and validations of macro-scale models over the entire modelled regions, local performance can be rather diverse (see e.g. de Paiva et al., 2013; Donnelly et al., 2016). Factors controlling the heterogeneity of local performance may be various, for instance: the quality of macro-scale input data, local water management, representativeness of model structure chosen, the influence of local geophysical and micro-climatic factors, etc.

There is a recognized and noteworthy value of readily available and easy accessible simulated daily streamflow series for scarcely gauged, or ungauged, areas of the world to enhance awareness and decision-making (e.g. Arheimer et al., 2011; Hjerdt et al., 2011). Nevertheless, the harmonization and enhancement of local performances of macro-scale models is still a scientific challenge that is worth addressing in operational hydrology, and which raises different research questions, such as:

- How to deal with locally biased simulations?
- Can we assimilate additional data to improve model performance without re-calibration?
- Is there a minimum gauging network density that makes the post-modelling data-assimilation viable and effective?

Recent literature shows the significant potential of kriging-based techniques for performing regional prediction of streamflow indices in ungauged locations (Skøien et al., 2006; Castiglioni et al., 2011; Pugliese et al., 2014). Among such techniques,

Topological kriging, or Top-kriging (see Skøien et al., 2006) has shown high prediction accuracy and excellent adaptability to a variety of water-related applications, such as prediction of low-flow indices (Castiglioni et al., 2009), interpolation of river temperatures (Laaha et al., 2013), estimation of flood quantiles (Alferi et al., 2013), regionalization of flow-duration curves (Castellarin, 2014; Pugliese et al., 2014, 2016), estimation of daily runoff in ungauged basins (Parajka et al., 2015) and reconstruction of historical daily streamflow series (Farmer, 2016).

Our study aims at developing and testing a geostatistical data-assimilation procedure for better agreement between locally observed streamflow and model results from macro-scale rainfall-runoff models. The procedure employs Top-kriging for geostatistically interpolating empirical period-of-record flow-duration curves (FDCs) along the stream network available at gauged basins. Interpolated FDCs are assimilated into simulated daily streamflow series at ungauged stream-network nodes, enhancing the local accuracy of simulated daily streamflow series. We test our method by improving European HYPE-model simulations (E-HYPE; Donnelly et al., 2016; Hundecha et al., 2016), which provides approximately 30-year of simulated daily streamflows freely and openly accessible for [35408 prediction nodes in Europe](http://hypeweb.smhi.se/europehype/time-series/) (mean catchment size of 215 km²) ~~prediction nodes in Europe (see, see also~~ <http://hypeweb.smhi.se/europehype/time-series/>). We address the Tyrolean region, as this area gives particularly poor simulations in the HYPE version 2.1 and thus would benefit from statistical enhancement of results. For the geostatistical interpolation, we use a group of 46 gauged catchments obtained from Austrian and Italian water services, and not used when setting up E-HYPE. With the observed streamflows we construct and interpolate empirical FDCs. Then, we assess the value and potential of assimilating this streamflow information into E-HYPE simulated series. In particular, (1) we cross-validate the proposed data-assimilation procedure for 11 E-HYPE prediction nodes located nearby an existing stream-gauge, and (2) we assess the enhancement of simulated series resulting from the geostatistical data-assimilation under different hypotheses on the spatial density of the streamgauging network. Section ~~??~~ [2](#) presents the proposed procedure ~~in detail, while Sections ?? and ?? illustrate E-HYPE modelling and the study area, respectively. Section ??, while Section 3~~ details the system of ~~cross-validation cross-validations~~ and sensitivity analyses we adopted for assessing the procedure ~~and present the results. Section 4 illustrates the study area and E-HYPE simulation data.~~ The last ~~two sections report three sections report~~ [results](#), discussion and conclusions, [respectively](#).

25 **2 ~~Geostatistical~~ [A new geostatistical streamflow-data assimilation method](#)**

2.1 Geostatistical interpolation of empirical flow-duration curves (TNDTK)

Top-kriging (or topological kriging) is a powerful geostatistical procedure developed by Skøien et al. (2006) for the prediction of hydrological variables. Like all kriging approaches, Top-kriging produces predictions of hydrological variables at ungauged sites with a linear combination of the empirical information collected at neighbouring gauging stations. Through this method, the unknown value of the streamflow index of interest at prediction location x_0 , $Z(x_0)$, can be estimated as a weighted average of the ~~variable~~ [regionalised variable](#), measured within the neighbourhood:

$$\underline{Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)}$$

where λ_i

$$\underline{Z(x_0) = \sum_{j=1}^n \lambda_j Z(x_j)}, \quad (1)$$

where λ_j is the kriging weight for the empirical value $Z(x_i)$ at location x_i $Z(x_j)$ at location x_j , and n is the number of neighbouring stations used for interpolation. Kriging weights λ_i λ_j can be found by solving the typical ordinary kriging linear system (2a2a) with the constraint of unbiased estimation (2b2b):

$$\sum_{j=1}^n \gamma_{i,j} \lambda_j + \theta = \gamma_{0,i} \quad i = 1, \dots, n \quad (2a)$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (2b)$$

where θ is the Lagrange parameter and $\gamma_{i,j}$ is the semi-variance between catchment i and j (Isaaks and Srivastava, 1990). The semi-variance, or variogram, represents the spatial variability of the regionalised variable Z . Unique from any other method of kriging, Top-kriging considers the variable defined over a non-zero support S , the catchment drainage area (Cressie, 1993; Skjøien et al., 2006). The kriging system of equations (2) remains the same, but the semi-variances between the measurements need to be obtained by regularization, i.e. smoothing the point variogram over the support area. The point variogram can then be back-calculated by fitting aggregated variogram values to the sample variogram (Skjøien et al., 2006). Pugliese et al. (2014) proposed a method for using Top-kriging to predict FDCs at ungauged locations that they termed Total Negative Deviation Top-kriging (TNDTK). The authors reduce the dimensionality of the problem by seeking a unique index of site-specific FDCs. Unlike other regional approaches (e.g. regional regression of streamflow quantiles, see e.g. Castellarin et al., 2013), the Top-kriging-based interpolates the entire curve, therefore ensuring its monotonicity (see e.g. Pugliese et al., 2014; Castellarin, 2014). This is accomplished by first standardising the empirical FDCs at site x , $\Psi(x, d)$, for some reference value, $Q^*(x)$, to yield a dimensionless FDC:

$$\underline{\psi(x, d) = \frac{\Psi(x, d)}{Q^*(x)},}$$

$$\underline{\psi(x, d) = \frac{\Psi(x, d)}{Q^*(x)} \quad (3)}$$

where d denotes a specific duration. Pugliese et al. (2014) identified an overall point index that effectively summarizes the entire curve. This index, which the authors termed total negative deviation (TND), is derived by integrating the area between

the lower limb of the FDC and the reference streamflow value Q^* (see Fig. 1). Empirical TND values are computed as:

$$TND(x) = \sum_{i=1}^m |q_i(x) - 1| \delta_i,$$

where $q_i = Q_i/Q^*$

$$TND(x) = \sum_{i=1}^m |q_i(x) - 1| \delta_i \quad (4)$$

- 5 where $q_i = \frac{Q_i}{Q^*}$ represents the i -th empirical dimensionless quantile standardised for the selected reference value Q^* , δ_i is half of the frequency interval between the $(i+1)$ -th and $(i-1)$ -th $(i+1)$ th and $(i-1)$ th quantile and the summation involves only the m standardised quantiles lower than 1. The equality between a given streamflow value and the reference value Q^* is represented by an horizontal dashed line in Fig. 1, i.e. the threshold given by the equation $\frac{Q}{Q^*} = 1$. The range of the summation, m , in Eq. (4) is a function of the maximum duration d_{max}^* , which is itself a function of that sample with minimum length
- 10 across gauged sites in the study region. Having calculated empirical TNDs, Pugliese et al. (2014) propose using the TNDs as a regionalised variable to develop site-specific weighting schemes. The same weights, derived through the solution of the linear kriging system (2), are used for a batch prediction of the continuous, dimensionless FDC for the ungauged site x_0 :

$$\hat{\psi}(x_0, d) = \sum_{i=1}^n \lambda_i \psi(x_i, d) \quad \forall d \in (0, 1),$$

where λ_i , with $i = 1, \dots, n$

$$15 \quad \hat{\psi}(x_0, d) = \sum_{j=1}^n \lambda_j \psi(x_j, d) \quad \forall d \in (0, 1) \quad (5)$$

where λ_j , with $j = 1, \dots, n$, are the weights resulting from the kriging interpolation of TNDs; $\psi(x_i, d)$ is the dimensionless empirical FDC at the donor site x_i and $\hat{\psi}(x_0, d)$ is the predicted dimensionless FDC. It is worth highlighting that the computation of the linear kriging system (2) depends on n , the number of neighbouring sites on which to base the spatial interpolation, a fact that will be explored below.

- 20 If a reliable model for predicting Q^* at the ungauged site x_0 can be developed, the prediction of the dimensional FDC, $\hat{\Psi}(x_0, d)$, is obtained as:

$$\hat{\Psi}(x_0, d) = \hat{Q}^*(x_0) \hat{\psi}(x_0, d) \quad \forall d \in (0, 1),$$

where $Q^*(x_0)$

$$\hat{\Psi}(x_0, d) = \hat{Q}^*(x_0) \hat{\psi}(x_0, d) \quad \forall d \in (0, 1) \quad (6)$$

- 25 where $\hat{Q}^*(x_0)$ is the prediction of Q^* at the ungauged site x_0 and $\hat{\psi}(x_0, d)$ has the same meaning as in Eq. (5).

2.2 New algorithm for geostatistical assimilation of local streamflow data

The main objective of our study is to improve rainfall-runoff model simulations, whereas an assessment of the reliability of the regional metric TND is out of the scope of this paper. Nevertheless, further details on TNDTK, i.e. cross-validations in different geomorphological and climatic regions, sensitivity analyses, comprehensive assessments and comparisons to state of the art models for predicting FDCs in ungauged basins, are reported in recent studies to which an interested reader is referred to (see e.g. Pugliese et al., 2014, 2016; Kim et al., 2017).

2.2 Algorithm for assimilation of local streamflow data

Following the approach proposed by Smakhtin and Masse (2000), we presents-present a novel procedure for predicting the model residuals that may be associated with macro-scale rainfall-runoff-rainfall runoff model simulations (e.g. from LIS-FLOOD, HYPE, PGB-IPH, etc., see Introduction). This method relies on a regional prediction of the long-term flow-duration curve (FDC) in the same site where these simulations are available.

For instance, let $\Psi(x_0, d)$ be the “true” unknown FDC for a given catchment x_0 and $\hat{\Psi}_{SIM}(x_0, d)$ be its prediction constructed on the basis of the daily streamflows simulated through the macro-scale model. We can assume that a general relationship between the two curves exists and reads:

$$\Psi(x_0, d) = \hat{\Psi}_{SIM}(x_0, d) + \varepsilon(x_0, d) \quad d \in (0, 1),$$

$$\Psi(x_0, d) = \hat{\Psi}_{SIM}(x_0, d) + \varepsilon(x_0, d) \quad \forall d \in (0, 1) \quad (7)$$

where $\varepsilon(x_0, d)$ are the model residuals defined over the duration domain d , which we may term residual-duration curve (ε DC). Evidently, the “true” residual-duration curve is unknown at ungauged basins, nevertheless one can estimate such a curve on the basis of geostatistically interpolated flow-duration curves $\hat{\Psi}_{TNDTK}(x_0, d)$ introduced in SectSec. 2.1,

$$\hat{\varepsilon}(x_0, d) = \hat{\Psi}_{TNDTK}(x_0, d) - \hat{\Psi}_{SIM}(x_0, d) \quad d \in (0, 1).$$

$$\hat{\varepsilon}(x_0, d) = \hat{\Psi}_{TNDTK}(x_0, d) - \hat{\Psi}_{SIM}(x_0, d) \quad \forall d \in (0, 1). \quad (8)$$

The estimated residual-duration curve obtained from the regional prediction of the long-term flow duration curve can then be used for assimilating local streamflow information into the simulated daily streamflow series. The procedure is sketched in Fig. 2: (1) given a simulated streamflow series (red line in the top-right), select a specific day t , and the corresponding discharge $Q(t)$; (2) retrieve the duration d associated with $Q(t)$ from the flow-duration curve constructed from simulated data (red line in the top-left quadrant); (3) read the estimated residual $\hat{\varepsilon}(t)$ off of the predicted residual-duration curve (blue line in the bottom-left quadrant); and (4) assimilate the residual into the simulated series as $Q(t) + \hat{\varepsilon}(t)$. The iteration of the algorithm through all time steps leads to an enhanced simulated series (blue line in the top-right quadrant).

This new assimilation procedure ~~share shares~~ some analogies with a technique called “quantile mapping”, used in the context of bias corrections for Global Climate Model predictions (see e.g. Komma et al., 2007). The procedure we propose in this context, though, is a rather general tool that can be applied to e.g. any macro-scale rainfall-runoff model for locally enhancing long simulated streamflow series without the need to re-run computationally intensive simulations, provided that the model itself is behavioural and validated on the basis of streamflow data that were not available for model calibration. The performance of the assimilation procedure depends on a variety of drivers, e.g. the quality of the simulated streamflows, which can be severely impacted by the local quality of input data even for a behavioural and well calibrated model, the ~~accuracy of the chosen regional model for predicting FDCs, the quality of locally available quality of~~ streamflow data and the density of the streamgauging network. ~~This latter element is specifically investigated in Sect. ?? (see Sec. 3.1.2), the accuracy of the chosen regional model for predicting FDCs~~ (we refer to TNDTK herein, but there are other viable options, see e.g. Castellarin et al., 2013; Castellarin, 2014). Regarding the latter element, indeed, the proposed procedure deeply relies onto accurate regional FDC predictions by means of an unbiased regional model, which necessarily must be validated beforehand for the area of interest. Otherwise, detriments are likely to be expected.

3 Pan-European rainfall-runoff simulation: E-HYPE

~~The HYdrological Prediction for the Environment (HYPE) model is a hydrological model for small-scale and large-scale assessments of water resources and water quality, developed at the Swedish Meteorological and Hydrological Institute (SMHI) during 2005–2007. The European application, E-HYPE, has been proved to be a powerful tool for water resources managers and practitioners, addressing nutrient concentration in river flow as well as water forecasts on short or seasonal time scale. It is also widely used to estimate snow storage and accumulated TWh of water inflow to hydropower dams and in climate change impact analysis. The website Hypeweb () provides visualisation and free downloading of 30 years of continuous and consistent daily streamflow simulations across the European river network at rather fine scale (i. e. the average size of elementary catchments is equal to 215) as well as forecasts and climate change impact analysis.~~

3 Testing the proposed algorithm: cross-validation procedures and sensitivity analysis

3.1 Structure of the analysis

~~The HYPE model is open access and can be downloaded with documentation and model set-up guidelines from the model website (). It simulates water flow and substances on their way from precipitation through soil, river and lakes to the river outlet. River basins are divided into subbasins, which in turn are divided into classes (the finest calculation units) depending on land use, soil type and elevation (Fig. 4). The soil is modelled as several layers, which may have different thickness for each class. In E-HYPE, each subbasin can have up to some 40 soil and land-use classes, which are lumped within the subbasins, while the watercourses are routed through the river network. The model parameters can be associated with land use (e.g.~~

evapotranspiration), soil type (The rationale of the analyses implemented in this study drives a sequence of operations, which can be summarised as follows:

1. streamflow simulations from a given model, e.g. water content in soil), or be common for the whole catchment or a region with geophysical similarities. This way of coupling the parameters with geographic information makes the model better suited for simulations in ungauged catchments. a rainfall-runoff macro-scale model, are available in a well-defined study area;
2. suppose that recorded streamflow data are made available for a reasonable number of streamgauge stations within the study area, apply a suitable model for the regionalisation of FDCs;
3. validate the regional model with respect to available streamflow observations. In this case we adopted a leave-one-out cross-validation (see e.g. Kroll and Song, 2013; Salinas et al., 2013; Wan Jaafar et al., 2011; Srinivas et al., 2008), even though different validation schemes might be preferred in other regions (see e.g. Pugliese et al., 2016);
4. validate the assimilation procedure by sequentially (a) neglecting all the streamflow information at a given nodes of the river network, (b) predicting FDCs, and (c) applying the assimilation method illustrated in Sec. 2.2;
5. evaluate the sensitivity of the assimilation procedure to the streamgauge density of the study area;

15 4 Study area

We focus on a large alpine region located in Tyrol (Italy, Austria and, for small portion only, Switzerland), for which the E-HYPE model show particularly poor results. Our analyses consider two types of data, observed daily streamflows and E-HYPE simulated daily streamflows, representing different catchments (Fig. 3). This alpine area is particularly suitable for hydro-power generation, indeed the presence of dams along the stream network might significantly alter the streamflow regime downstream producing a significant alteration of the natural flow conditions. E-HYPE only simulates the dams present in the global database of GrandD, which might not be representative for hydropower production at the local scale. Therefore, we removed from the initial group of gauged catchments all basins for which the streamflow regime is highly or significantly altered by upstream dams. Table 1 reports the main characteristics of streamflow regimes for the set of 46 gauged basins and the 11 selected E-HYPE prediction nodes. We selected only those E-HYPE prediction nodes located within Tyrol, which were the closest to one of the available streamgauges. In terms of selection criteria, we selected the E-HYPE catchments that showed limited differences in terms of (1) drainage areas (<14%) and (2) distance between catchment centroids (<15). These criteria resulted in 11 E-HYPE prediction nodes that are evenly distributed in the study region (see red lines in Fig. 3). We addressed the limited differences existing between drainage areas of E-HYPE and gauged basins by adopting the drainage-area ratio technique, that is by rescaling daily streamflows according to drainage areas of the corresponding catchment. Such method

assumes the same unit daily streamflow for any pair of hydrologically similar catchments i and j , which reads,

$$\frac{Q_i(t)}{A_i} = \frac{Q_j(t)}{A_j},$$

where $Q_i(t)$ represents the daily streamflow at day t for catchment i and $Q_j(t)$ is the daily streamflow at day t for catchment j . In our application, i and j could correspond to any given pair (streamgauge, E-HYPE prediction node), and A_i and A_j the corresponding drainage areas. The methodologies adopted for addressing points 3., 4. and 5., are illustrated in Sec. 3.1.1, 3.1.2, and 3.1.2, in this order; the accuracy of predictions (e.g. regional FDCs, assimilated and simulated streamflow series, etc.) relative to their empirical or observed counterparts is quantified through performance indices described in Sec. 3.2.

4 Cross-validated predictions: description of resampling procedures and results of the analyses

3.1 Cross-validation of the FDC geostatistical interpolator (TNDTK)

3.1.1 Cross-validation of the FDC geostatistical interpolator (point 3 above)

We assessed proposed to assess the accuracy of the geostatistical predictor of FDCs (i.e. TNDTK, see [Sec. 2.1](#)) in cross-validation over the entire study area by focusing on the 46 with respect to available streamflow observations from a sufficient number of gauged catchments. We chose the mean annual flow (MAF), computed as the average flow of recorded historical streamflow series, as the reference value Q^* (see details in [Sec. 2.1](#)).

TNDTK operates by first applying Top-kriging to empirical TND values (see [Sec. 2.1](#)), which we performed by calculating binned sample variogram first, and then by modelling binned empirical data with a 5-parameter “modified” exponential theoretical variogram (a combination of exponential and a fractal model, see details in Skøien et al., 2006). The fitted theoretical point variogram, and its five parameters were obtained through the weighted least squares (WLS) regression method from Cressie (1993) by fitting simultaneously all regularised binned variograms that were computed for various area classes (in this case study we employed 2 variogram bins as a result of the range of drainage areas, which spans over 3 orders of magnitude, see details on binning methods in Skøien, 2014). Recent applications of TNDTK indicate $n = 6$ as an optimal number of neighbouring donor stations, thus we chose the same value for this case study as well (see details in Pugliese et al., 2014, 2016). Then, TNDTK uses the kriging weights obtained for predicting TND values for interpolating the dimensionless FDCs at the location of interest as the weighted average of dimensionless empirical FDCs constructed from the $n = 6$ neighbouring gauged sites (see Eq. (5), in which λ_i , with $i = 1, \dots, n$ λ_j , with $j = 1, \dots, 6$ and $n = 6$, are the kriging weights).

While the computation of TNDs does not require any specific resampling scheme of the FDCs, the prediction of the curves in ungauged locations is carried out using a fixed number of quantiles which should be selected to thoroughly represent the variability from high to low flows. Thus, observed FDCs are resampled to 20 equally spaced points, in the normal space, leading to the widest range of durations compatible with the shorter observed streamflow series in the dataset. We adopted a

leave-one-out cross-validation (LOOCV) procedure (see e.g. Pugliese et al., 2014, 2016) to test the accuracy and uncertainty associated with FDCs predictions in the study area. This simulates the ungauged conditions at each and every gauged site in

the study area by (1) removing it in turn from the dataset and (2) referring to the $n = 6$ neighbouring gauges for predicting its dimensionless FDC. Given that the geostatistical assimilation procedure uses dimensional FDCs, we also tested the suitability of standard ~~Topkriging~~ Top-kriging for predicting MAF at ungauged locations in the study area, still through a LOOCV procedure (general validity of Top-kriging for predicting mean annual flows is also described in Blöschl et al., 2013). For

5 MAF interpolation, we adopted the same settings used for predicting TND values (i.e. a 5-parameter “modified” exponential theoretical variogram and $n = 6$ neighbouring sites).

We then used cross-validated dimensionless FDCs and MAF predictions at each and every gauging station in the study area to obtain cross-validated predictions of dimensional FDCs for each measuring node through Eq. ~~(6)~~.

~~We assessed TNDTK performance by means of Nash-Sutcliffe Efficiency computed for log-transformed streamflows (LNSE). The application of the geostatistical method TNDTK through an LOOCV procedure reveals a good agreement between empirical values and predictions as shown in Fig. 5 and 6. Specifically, Fig. 5 reports empirical (x -axis) against geostatistically predicted (y -axis) MAF values as well as LNSE obtained in cross-validation (i.e. 0.96). The left panel of Fig. 6 shows a scatter diagram between observed (x -axis) and predicted streamflows (y -axis) from FDCs. This good agreement is confirmed also by the distributions of at-site LNSE values (see box-plots in the right panel of Fig. 6); median LNSEs is equal to 0.97, while mean LNSE is e.a. 0.90. The good performance obtained in cross-validation legitimate the use of TNDTK for predicting FDCs in the study area at the 11 E-HYPE prediction nodes of interest, for which TNDTK delivers very high performances, with LNSE values above 0.97 (see the spatial distribution of efficiency values in the left panel of Fig. 10).~~

10

15

3.1.2 Cross-validation of the geostatistical assimilation procedure (point 4 above)

3.2 ~~Cross-validation of the geostatistical assimilation procedure (GAE-HYPE)~~

20 We applied the proposed assimilation procedure as outlined in ~~Sect. ??, focusing on 11 pairs of catchments, i.e. each pair includes one of the E-HYPE catchments depicted in Fig. 3 and its corresponding gauged catchment. We first assessed the~~ Sec. 2.2, by firstly, assessing the efficiency of the procedure ~~in through a~~ leave-one-out cross-validation. We predicted the FDC associated with each ~~E-HYPE catchment simulation node~~ by using TNDTK and by, ~~also,~~ neglecting the hydrological information coming from the closest (i.e. immediately upstream or downstream) gauged catchment, therefore assuming that no

25 streamflow information is available near the ~~E-HYPE prediction simulation~~ node. The ~~work-flow workflow~~ of the validation algorithm is as follows:

-
- a) we select one ~~(E-HYPE prediction node, streamgauge) pair, among $n_{pair} = 11$ pair among n_p possible pairs, let us term it pair i, j~~ i, j , where i stands for ~~E-HYPE prediction simulation~~ node and j stands for the corresponding streamgauge;
- 30 b) we drop the daily streamflow series observed at streamgauge j from the set of observed series;
-

- c) we interpolate FDC at ~~E-HYPE prediction simulation~~ node i by using ~~TNDTK through TNDTK~~ (as illustrated in ~~Sec. 2.1 and focusing on the remaining $n_{gauges} - 1$~~) using the remaining $n_g - 1$ gauged sites, where ~~$n_{gauges} = 46 - 1$~~ n_g is total number of streamgauges in the study region ~~minus streamgauge j~~ ;
- i
- 5 d) we apply the assimilation procedure outlined in ~~Sec. ??~~ Sec. 2.2 and depicted in Fig. 2 to the streamflow series simulated for the ~~E-HYPE prediction simulation~~ node i ;
-
- e) we compare the original ~~E-HYPE simulated~~ daily streamflow series and the geostatistically enhanced one at prediction node i with the daily streamflow series observed at streamgauge j times the corresponding area ratio A_i/A_j (i.e. A_i is the drainage area of ~~E-HYPE prediction simulation~~ node i , A_j is the drainage area streamgauge j);
- 10 – ; see also Sec. 4.1);
- f) we repeat all previous steps for each one of the remaining ~~$n_{pairs} - 1 = 10 - n_p$~~ pairs.

~~It~~ For the sake of consistency, we anticipate here that we will refer to the procedure presented above as GAE-HYPE (i.e. Geostatistically Assimilated E-HYPE) in the remainder of the manuscript. The acronym clearly recalls the rainfall-runoff model used in this study (see sec. 4.2), however the procedure disregards from the use of a specific rainfall-runoff model.

15

Finally, it is worth highlighting here that either ~~TNDTK FDCs or E-HYPE FDCs have been~~ FDCs obtained from either the geostatistical model or the rainfall-runoff simulation model are resampled to 20 ~~non-equally-spaced points across duration~~ equally-spaced points across the normally-transformed duration intervals (see details in Pugliese et al., 2014). Thus, as a result, the produced ~~$\hat{\epsilon}DC$ - ϵDCs~~ reflect the same sampling scheme of the curves. Nevertheless, the procedure does not foresee any

20 restriction to the resolution of the resampled curve, allowing for a finer resampling scheme in other analyses.

3.1.3 Streamgauging network density and effectiveness of geostatistical data-assimilation (point 5 above)

Since the proposed geostatistical data-assimilation procedure (GAE-HYPE) relies upon the local availability of streamgauges records, understanding to what extent the performances of the assimilation method are driven by gauging network density is a fundamental of paramount importance. Therefore, we performed a sensitivity analysis and assessed the degree of enhancement

25 of simulated daily streamflow sequences associated with different scenarios of streamflow data availability, repeating for each scenario the procedure described in Sec. 3.1.2. Thus, we randomly discarded some of the gauges available over the study area and varied the total number of available stations continuously from the lowest to the highest gauge density. At each density scenario, we performed exactly the same kriging settings and same LOOCV illustrated in detail in Sec. 3.1.1 and 3.1.2, respectively.

3.2 Performance indices

We assessed the performances for predicting regional FDCs by means of Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) computed for log-transformed streamflows (LNSE). These indices are computed as follows:

$$LNSE_{FDC,j} = 1 - \frac{\sum_{k=1}^{n_d} (\ln \Psi(x_j, d_k) - \ln \hat{\Psi}(x_j, d_k))^2}{\sum_{k=1}^{n_d} (\ln \Psi(x_j, d_k) - \mu_j)^2} \quad j = 1, \dots, n_g, \quad (9)$$

- 5 where $\Psi(x_j, d_k)$ and $\hat{\Psi}(x_j, d_k)$ are the empirical and the predicted k th streamflow quantiles at site x_j , respectively, μ_j is the mean of empirical log-transformed streamflow quantiles at site x_j , n_d is the number of discretisation points throughout duration range and n_g is the number of streamgauges.

Another useful metric of performance for the assessment of FDC predictions is the overall absolute curve error (see Ganora et al., 2009), which reads as follows:

$$10 \quad \delta_{FDC,j} = \sum_{k=1}^{n_d} |\Psi(x_j, d_k) - \hat{\Psi}(x_j, d_k)| \quad j = 1, \dots, n_g, \quad (10)$$

where $\Psi(x_j, d_k)$ and $\hat{\Psi}(x_j, d_k)$ have the same meaning as in Eq. (9).

Similarly, concerning streamflow time series, the assessment of modelled streamflows is carried out with LNSE, but in this case, it reads as follows:

$$LNSE_{mod,j} = 1 - \frac{\sum_{t=1}^{t_s} (\ln Q_{emp,j}(t) - \ln Q_{mod,j}(t))}{\sum_{t=1}^{t_s} (\ln Q_{emp,j}(t) - \omega_j)^2} \quad j = 1, \dots, n_g, \quad (11)$$

- 15 where, $Q_{emp,j}(t)$ and $Q_{mod,j}(t)$ are the empirical and predicted streamflow at site x_j and time t , respectively, ω_j is the mean of empirical log-transformed streamflow at site x_j , and n_s is the number of selected simulation nodes.

Furthermore, we assessed the efficiency of the data-assimilation procedure through the following metric,

$$LNSE_{ratio,j} = \frac{LNSE_{mod2,j} - LNSE_{mod1,j}}{1 - LNSE_{mod1,j}} \quad j = 1, \dots, n_p. \quad (12)$$

- 20 $LNSE_{ratio}$ quantifies the degree of enhancement of “model 2” relative to “model 1” standardized by the maximum possible improvement (i.e. $1 - LNSE_{mod1}$). An $LNSE_{ratio}$ close to zero means no significant enhancement (detriment of original sequences in case of negative values), whereas an $LNSE_{ratio}$ close to 1 indicates that no further enhancement is possible. Such index derives from the reciprocal root-mean-squared error ratio between the two models.

- Finally, in order to verify whether or not the proposed assimilation procedure GAE-HYPE outperforms rainfall-runoff simulations, for different gauge density scenario (see Sec. 3.1.2), we used the Wilcoxon signed-rank test with the null hypothesis that simulation model LNSEs are greater than GAE-HYPE ones at 5% significance level (Hollander and Wolfe, 1999).
- 25

4 Data

4.1 Study region

We focus on a large alpine region located in Tyrol (Italy, Austria and, for small portion only, Switzerland). Our analyses consider two types of data, observed daily streamflows and E-HYPE simulated daily streamflows (see Sec. 4.2), representing different sets of catchments (Fig. 3). In this study, E-HYPE represents the rainfall-runoff model selected to evaluate the procedure presented in Sec. 2.2, which has shown significantly poor results in this region. Indeed, this alpine area is particularly suitable for hydro-power generation, therefore the presence of dams along the stream network might likely alter the streamflow regime downstream producing a significant alteration of the natural flow conditions. E-HYPE only simulates the dams present in the global database of Grand (Lehner et al., 2011), which might not be representative for hydropower production at the local scale (Arheimer et al., 2017). Thus, we removed from the initial group of gauged catchments all basins for which the streamflow regime is highly or significantly altered by upstream dams. Table 1 reports the main characteristics of streamflow regimes for 46 gauged basins and 11 selected E-HYPE prediction nodes. Among all E-HYPE prediction nodes available in Tyrol we selected only those whose catchments were the closest to gauged ones, i.e. difference in terms of drainage areas $< 14\%$ and distance between catchment centroids $< 15\text{km}$. These criteria resulted in the selection of 11 E-HYPE prediction nodes that are evenly distributed in the study region (see red lines in Fig. 3). We addressed the limited differences existing between drainage areas of E-HYPE and gauged basins by adopting the drainage-area ratio technique (DAR, see e.g. Farmer and Vogel, 2013), that is by rescaling daily streamflows according to drainage areas of the corresponding catchment. Such method assumes the same unit daily streamflow for any pair of hydrologically similar catchments i and j , which reads,

$$\frac{Q_i(t)}{A_i} = \frac{Q_j(t)}{A_j}, \quad (13)$$

where $Q_i(t)$ represents the daily streamflow at day t for catchment i , and $Q_j(t)$ is the daily streamflow at day t for catchment j . In our application, i and j could correspond to any given pair (streamgauge, E-HYPE prediction node), and A_i and A_j the corresponding drainage areas. Finally, it is worth pointing out that neither the observed nor the E-HYPE series present zero values in their recording (or simulation) periods for each of the 11 selected nodes.

4.2 Pan-European rainfall-runoff simulation: E-HYPE

The HYdrological Prediction for the Environment (HYPE) model is a hydrological model for small-scale and large-scale assessments of water resources and water quality, developed at the Swedish Meteorological and Hydrological Institute (SMHI) during 2005–2007 (Lindström et al., 2010). The European application, E-HYPE, has been proved to be a powerful tool for water resources managers and practitioners, addressing nutrient concentration in river flow as well as water forecasts on short or seasonal time scale. It is also widely used to estimate snow storage and accumulated TWh (terawatt hour) of water inflow to hydropower dams and in climate change impact analysis (Donnelly et al., 2017). The website Hypeweb (<http://hypeweb.smhi>).

se) provides visualisation and free downloading of 30 years of continuous and consistent daily streamflow simulations across the European river-network at rather fine scale (i.e. the average size of elementary catchments is equal to 215km²) as well as forecasts and climate change impact analysis.

5 The HYPE model is open access and can be downloaded with documentation and model set-up guidelines from the model website (<http://hypecode.smhi.se/>). It simulates water flow and substances on their way from precipitation through soil, river and lakes to the river outlet (Lindström et al., 2010). River-basins are divided into sub-basins, which in turn are divided into classes (the finest calculation units) depending on land use, soil type and elevation (Fig. 4).

10 The soil is modelled as several layers, which may have different thickness for each class. In E-HYPE, each sub-basin can have up to some 40 soil and land-use classes, which are lumped within the sub-basins, while the watercourses are routed through the river network. The model parameters can be associated with land use (e.g. evapotranspiration), soil type (e.g. water content in soil), or be common for the whole catchment or a region with geophysical similarities (Hundecha et al., 2016). This way of coupling the parameters with geographic information makes the model better suited for simulations in ungauged catchments.

5 Results

15 The application of the geostatistical method TNDTK through an LOOCV procedure reveals a good agreement between empirical values and predictions as shown in Figs. 5 and 6. Specifically, Fig. 5 reports empirical (x -axis) against geostatistically predicted (y -axis) MAF values as well as LNSE obtained in cross-validation (i.e. 0.96). The left panel of Fig. 6 shows a scatter diagram between observed (x -axis) and predicted streamflows (y -axis) from FDCs. This good agreement is confirmed also by the distributions of at-site LNSE values (see box-plots in the right panel of Fig. 6); median LNSEs is equal to 0.97, while mean LNSE is c.a. 0.90. The good performance obtained in cross-validation legitimate the use of TNDTK for predicting FDCs in the study area at the 11 E-HYPE prediction nodes of interest, for which TNDTK delivers very high performances, with LNSE values above 0.97 (see the spatial distribution of efficiency values in the left panel of Fig. 10).

Figure 7 reports the results obtained by applying the aforementioned cross-validation algorithm. Circles in the left panel represent the cumulative absolute error δ (see Eq. (10)) computed for each catchment pair i - j , between empirical FDCs vs. predicted FDCs for either E-HYPE (δ_{E-HYPE} on the y -axis) or TNDTK (δ_{TNDTK} on the x -axis) predictions. This figure clearly shows that for 9 out of 11 target sites the geostatistical method TNDTK outperforms E-HYPE in predicting FDCs. Moreover, one of the two sites for which E-HYPE outperforms TNDTK shows nearly the same performance as TNDTK (i.e. the circle is very close to the 1:1 line), while the other one, i.e. site 3675-9001070 highlighted with a black dot in the figure, is associated with the worst performance of TNDTK relative to E-HYPE (see also [SeetSec. 6](#))-

30 [on this](#)). Right panel of Fig. 7 reports estimated residual duration curves ($\hat{\epsilon}$ DCs) for the selected sites. For the sake of representation, we report standardised residuals in the y -axis, i.e. residuals divided by the corresponding streamflow quantiles predicted via TNDTK; we referred to TNDTK quantiles for standardization since the real empirical FDC is supposed to be unknown (see cross-validation algorithm illustrated [above in Sec. 3.1.2](#)). Overall, $\hat{\epsilon}$ DCs show negative values for lower du-

rations and positive values for higher durations (see also Fig.9 9). This means that, in Tyrol, E-HYPE tends to overestimate streamflow in wet periods as well as to underestimate streamflows in drier ones relative to the geostatistically predicted FDCs (i.e. TNDTK, see the left panels in Fig. 9). We eventually used the $\hat{\epsilon}$ DC curves, which are estimates of E-HYPE residuals, to assimilate locally available streamflow data into E-HYPE simulated series as illustrated in Fig. 2, obtaining what we **term**
5 termed GAE-HYPE simulations ~~-(see right panels of Fig. 9).~~

Representativeness of simulations (i.e. E-HYPE and GAE-HYPE simulated daily streamflows) is assessed through Nash-Sutcliffe Efficiency of log-flows (LNSE) computed by referring to ~~the whole E-HYPE simulation period - recorded streamflow time period of the paired streamgauge (see pairing method adopted in Sec. 4.1)~~. We found remarkable improvements obtained with the proposed data-assimilation procedure (Fig. 8). Indeed, one can notice we obtained a substantial enhancement of LNSE
10 values of GAE-HYPE simulations relative to the original E-HYPE ones ~~;-which in the best case increases for all the 11 selected sites, which show LNSE increments from -0.462 to 0.594 in the best case (catchment pair IDs: 201236 - 9608296) and in the worst case from 0.527 to 0.690 in the worst case (catchment pair IDs: 3675 - 9001070, see also Tab. 2).~~ The median at-site LNSE value increases from 0.045 to 0.685, which ultimately underline the benefits introduced with the proposed method. Figure 8 8, also, illustrates the impact of geostatistical data-assimilation for the two E-HYPE prediction nodes mentioned
15 above (i.e. the one characterized by the best improvements in terms of overall LNSE value, and the one associated with the most limited improvement). In both cases the advantages associated with the data-assimilation procedure are strikingly evident.

Finally Moreover, looking at the spatial distribution of LNSE values across the 11 selected prediction nodes within the study area depicted in Fig. 10, it is clear how the proposed enhancement strategy benefits from an unbiased estimations of FDCs. In fact, TNDTK shows homogeneous and rather good performance for predicting FDCs (left panel of Fig. 10); also, middle and
20 right panels of Fig. 10 reveal that the enhancement capabilities of the assimilation procedure are lower for those catchments where E-HYPE performs better (see elementary catchments filled in yellow to green in Fig. 10), whereas the assimilation procedure proves to be very powerful when E-HYPE performs worse (see elementary catchment catchments coloured in orange to red), bringing efficiencies from negative to positive values in all cases (from green to blue).

5.1 ~~Streamgauging network density and effectiveness of geostatistical data-assimilation~~

25 ~~Since the proposed geostatistical data-assimilation procedure relies upon the local availability of streamgauges records, understanding to what extent the performance of the assimilation method are driven by gauging network density is of paramount importance. Therefore, we performed a sensitivity analysis and assessed the degree of enhancement of simulated daily streamflow sequences associated with different scenarios of streamflow data availability, repeating for each scenario the procedure described in Sect. ??.~~

30 ~~We randomly discarded some of the gauges available over the study area (i.e. 46 in total) and varied the total number of available stations continuously from 7 (lowest density) to 46 (highest density) gauges (The assessment of gauge density impacts on the proposed procedure reveals remarkable enhancement even with the lowest gauge density scenario (i.e. 40 iterations in total), that means approximately from 4 to 29 gauges per - . We repeated these 40 iterations for each and every E-HYPE prediction node, 11 in total, by using exactly the same settings illustrated in detail in Sect. ?? and 2.1 for implementing~~

the data-assimilation procedure (e.g. 6 donor sites, exponential-fractal theoretical variogram, MAF is the index-flow discharge, etc.). We assessed the efficiency of the data-assimilation procedure through the following metric,

$$LNSE_{ratio} = \frac{LNSE_{GAE-EHYPE} - LNSE_{EHYPE}}{1 - LNSE_{EHYPE}},$$

where $LNSE_x$ is the log-flows Nash-Sutcliffe efficiency for simulated daily streamflows, and x could either refer to E-HYPE (original) or GAE-HYPE (after data-assimilation).

$LNSE_{ratio}$ quantifies the degree of enhancement of GAE-HYPE relative to E-HYPE standardized by the maximum possible improvement (i.e. $1 - LNSE_{GAE-EHYPE}$). An $LNSE_{ratio}$ close to zero means no significant enhancement (detriment of original sequences in case of negative values), whereas an $LNSE_{ratio}$ close to 100% indicates that no further enhancement is possible.

Figure 11 shows $LNSE_{ratio}$ values we obtained under different scenarios of streamflow data availability and displays a clear pattern, which confirms the showing an improvement of the degree of enhancement associated with an increasing gauge density. Moreover, Fig. 11 shows how the degree of enhancement flattens out in cases in which there are more than 20 approximately 25 gauges available per 10000 km². Another noteworthy feature of Fig. 11 is the remarkable enhancement that can be obtained for the scenario with the lowest gauge density (i.e. 7 gauging stations, minus 1 because the sensitivity analysis is performed in leave-one-out cross validation, is equal to 6, which is exactly the number of donor sites we chose for the implementation of Top-kriging). Finally, the p -values resulting from the Wilcoxon signed-rank tests (see Sec. 3.2) highlight that the assimilation procedure outperforms E-HYPE: the null hypothesis is rejected, with p -values always lower than 0.04%, regardless a particular streamgauge density scenario.

6 Discussion

This new geostatistical procedure enables practitioners and water resources managers and planners to profit from the wealth of hydrological information, by adjusting open data products with local observations. We enhanced the streamflow series simulated by macro- and continental scale rainfall-runoff models at ungauged prediction nodes by assimilating streamflow observations, which are locally available in the region of interest, without having to redo the original hydrological model calculations. This is a recurrent condition since local streamflow data are released under different license terms and policies: some of them could be public and open-access, while some other might not be openly and freely accessible by the broad public. The E-HYPE model obviously lack storage capacity in the Tyrol region and the proposed approach to enhance the results should be seen as temporary until a new model version accounting for this is available. We do not propose the procedure as a general fix for structurally unsuitable (or non-behavioural, see Beven and Binley, 1992) models, which have been proved to be unfit for either the area of interest or the water-problem at hand. For a more sustainable solution, we suggest to use another model structure or re-calibrate the model, instead of post-processing the output. However, this procedure makes sense for making a first assessment of water issues in regions where information is otherwise missing, but only macro-scale models are readily available.

Our study shows for Tyrol that ~~indeed~~ it is possible to significantly enhance rainfall-runoff simulations resulting from macro-scale, regional or continental hydrological models by geostatistically assimilating (geographically sparse) streamflow observations (see e.g. ~~Fig~~Figs. 8, 9 and 10); provided that available streamflow series are long enough to obtain a good empirical approximation of the long-term flow-duration curve (FDC) for the site of interest (i.e. 5-10 years, see Castellarin et al., 2013). Indeed, series length of the observed streamflow dataset controls the magnitude of duration extremes (i.e. duration boundary interval), which, in turn, might affect the adopted resampling scheme needed for predicting FDCs at simulation nodes. Nonetheless, in some specific application, very high or very low durations might be of particular interest (e.g. studies focused on flood or drought only), therefore a preliminary investigation on the resampling scheme (e.g. duration extremes, duration intervals, resolution of the points to represent correctly the whole curve, etc.) should be always taken into account.

One of the main advantages of the proposed method is that the end-user can get enhanced streamflow simulations without any further model calibration or refinement. Even though one could argue that when additional streamflow data become available at neighbouring gauges, it should be used for improving the performance of the model at the site of interest, calibrating and validating macro-scale and regional model could be a time-consuming and computationally demanding task. The proposed procedure, instead, is neither computational nor data intensive, and is implemented only using observed streamflow data and a GIS vector layer with catchment boundaries (see e.g. Fig. 3). The application requires the identification of a suitable regional model for predicting FDC in ungauged basis (see e.g. Fig. 6). However, it has advantages, such as: (a) a regional model can be a very informative and useful tool for water resources managers and planners; and (b) the subsequent advantages obtained from the data-assimilation procedure is transferred downstream in the entire regional river-network (see Fig. 11).

One important limitation of the proposed method is that, once a target prediction node is considered, any given simulated streamflow value is associated with a single duration, which corresponds to a particular estimated residual, that will be used in turn for correcting the streamflow value itself (see ~~Fig-s~~Figs. 2 and 9). ~~This~~ Essentially, this means that the volumes from E-HYPE are discarded while the sequencing of E-HYPE simulations is retained. Moreover, this algorithm cannot possibly account for seasonal (or interannual) modifications in the hydrological behaviour of the catchment. Indeed, as shown in time series comparison of Fig. 9, when the geostatistical prediction of FDCs is unreliable, ~~then neither improvements nor detriments are delivered by the proposed procedure~~ the assimilation procedure reflects such low accuracy (i.e. the procedure fails to correctly capture high-flow and low-flow regimes, see e.g. the resulting FDC from GAE-HYPE simulations at the catchment pair 3675-9001070 in Fig. ~~7 and~~ 9), propagating this bias throughout the whole simulated series.

Finally, designing a theoretical framework that combines statistical data-driven approaches with deterministic process-driven ones is seen by many as the correct way for tackling the ‘Prediction in Ungauged Basins (PUB)’ problem and further advancing the scientific research in this area (see e.g. Di Prinzio et al., 2011). We believe that our geostatistical data-assimilation procedure for macro-scale hydrological models is one example in this direction. Future analyses will focus on the relaxation of the main limitation of the approach (i.e. the incorporation of seasonal patterns in data-assimilation procedure) and on the extension of its applicability to anthropogenically anthropologically altered streamflow regimes.

7 Conclusions

This research work focuses on the development of an innovative method for enhancing streamflow series simulated by macro-, continental-, and global-scale rainfall-runoff models by means of a geostatistical prediction of model residuals. We focus on Tyrol as study region and E-HYPE (European - HYdrological Predictions for the Environment, from the Swedish Meteorological and Hydrological Institute, SMHI) as a macro-scale hydrological model, respectively; nevertheless, the geostatistical data-assimilation procedure is general and can be applied to simulated ~~streamflow-series~~ streamflow series coming from other macro-scale rainfall-runoff models. The proposed data-assimilation procedure utilises ~~streamflow-data~~ streamflow data that are locally available for the area of interest, which were not considered in the implementation of the macro-scale hydrological model; it (1) adopts Top-kriging for regionally interpolating empirical period-of-record flow-duration curves (FDCs) that can be constructed from locally available streamflow data; (2) constructs residual-duration relationships at any prediction node in the study region where simulated streamflow series are available, by comparing FDCs resulting from geostatistical interpolation (~~Topkriging~~ Top-kriging) and rainfall-runoff simulation (E-HYPE); (3) uses the error-duration curve to enhance macro-scale simulated streamflows.

The ~~eross-validation~~ cross-validation tests of the proposed approach with different scenarios of streamflow data availability shows the significant advantages of geostatistical data-assimilation even for very low stream-gauging network densities (i.e. c.a. 1 gauge per ~~2000~~ km²). It can become a standalone numerical tool to be used for enhancing results from macro-scale models anywhere along the stream network of a given region. Potential applications are envisaged for a variety of water resources management and planning problems that require accurate streamflow series (e.g. regional assessment of hydropower potential, habitat suitability studies, surface water allocation, civil protection management strategies, climate change trends, safety of river structures, etc.). Future analyses will address the main limitation of the proposed geostatistical data-assimilation procedure, aiming at incorporating observed seasonal and inter-annual variations of the hydrological behaviour of the study region as well as other hydrological features (e.g. baseflow index or peakflow data) into the geostatistical regionalization of model residuals.

Code and data availability. The analysis was carried out in the Virtual Water Science Lab developed within the FP7 funded research project SWITCH-ON (grant agreement no. 603587). We invite the interested reader to explore the experiment protocol here: <http://dl-ng005.xtr.deltares.nl/view/462/>

Acknowledgements. The analyses presented in the study are the main outcomes of the international experiment Geostatistical ~~Enhancement of European Hydrological Predictions~~ enhancement of European hydrological predictions (GEEHP), which was performed in the Virtual Water-Science Lab developed within the European Commission FP7 funded research project SWITCH-ON (Sharing Water-related Information to Tackle Changes in the Hydrosphere –for Operational Needs, grant agreement no. 603587). The overall aim of the project is to promote data sharing to exploit open data sources. The study also contributes to developing the framework of the ~~“Panta Rhei”~~ “Panta Rhei” Research

Initiative of the International Association of Hydrological Sciences (IAHS). ~~The majority of figures included in the study were produced by the use of free and open source softwares (i.e. Quantum GIS Geographic Information System—Open Source Geospatial Foundation Project, <http://qgis.osgeo.org>, and the R Project for Statistical Computing, <https://www.R-project.org/>)~~

References

- Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, *Hydrological Sciences Journal*, 48, 317–337, <https://doi.org/10.1623/hysj.48.3.317.45290>, 2003.
- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrol. Earth Syst. Sci.*, 17, 1161–1175, <https://doi.org/10.5194/hess-17-1161-2013>, 2013.
- Archfield, S. A., Pugliese, A., Castellarin, A., Skøien, J. O., and Kiang, J. E.: 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, <https://doi.org/10.5194/hess-17-1575-2013>, 2013.
- Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., and Over, T.: Accelerating advances in continental domain hydrologic modeling, *Water Resour. Res.*, 51, 10078–10091, <https://doi.org/10.1002/2015WR017498>, 2015.
- Arheimer, B., Wallman, P., Donnelly, C., Nyström, K., and Pers, C.: E-HypeWeb: Service for Water and Climate Information - and Future Hydrological Collaboration across Europe?, in: *Environmental Software Systems. Frameworks of eEnvironment, IFIP Advances in Information and Communication Technology*, pp. 657–666, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-22285-6_71, 2011.
- Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., and Strömqvist, J.: Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea basin – influence of input-data quality and scale, *Hydrology Research*, 43, 315–329, <https://doi.org/10.2166/nh.2012.010>, 2012.
- Arheimer, B., Donnelly, C., and Lindström, G.: Regulation of snow-fed rivers affects flow regimes more than climate change, *Nature Communications*, 8, <https://doi.org/10.1038/s41467-017-00092-8>, 2017.
- Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, *Water Resour. Res.*, 52, 3599–3622, <https://doi.org/10.1002/2015WR018247>, 2016.
- Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrological Processes*, 6, 279–298, <https://doi.org/10.1002/hyp.3360060305>, 1992.
- Blöschl, G., Sivapalan, M., Thorsten, W., Viglione, A., and Savenije, H.: *Runoff prediction in ungauged basins: synthesis across processes, places and scales*, Cambridge University Press, 2013.
- Castellarin, A.: Regional prediction of flow-duration curves using a three-dimensional kriging, *Journal of Hydrology*, 513, 179–191, <https://doi.org/10.1016/j.jhydrol.2014.03.050>, <http://www.sciencedirect.com/science/article/pii/S0022169414002352>, 2014.
- Castellarin, A., Botter, G., Hughes, D. A., Liu, S., Ouarda, T. B. M. J., Parajka, J., Post, M., Sivapalan, M., Spence, C., Viglione, A., and Vogel, R.: Prediction of flow duration curves in ungauged basins, chap. 7, pp. 135–162, Cambridge University Press, 2013.
- Castiglioni, S., Castellarin, A., and Montanari, A.: Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation, *J. Hydrol.*, 378, 272–280, <https://doi.org/10.1016/j.jhydrol.2009.09.032>, 2009.
- Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J. O., Laaha, G., and Blöschl, G.: Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging, *Hydrol. Earth Syst. Sc.*, 15, 715–727, <https://doi.org/10.5194/hess-15-715-2011>, 2011.
- Cressie, N. A. C.: *Statistics for spatial data*, Wiley series in probability and mathematical statistics: Applied probability and statistics, J. Wiley, 1993.

- de Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., and Bulhões Mendes, C. A.: Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin, *Water Resour. Res.*, 49, 1226–1243, <https://doi.org/10.1002/wrcr.20067>, 2013.
- De Roo, A. P. J., Wesseling, C. G., and Van Deursen, W. P. A.: Physically based river basin modelling within a GIS: the LISFLOOD model, *Hydrological Processes*, 14, 1981–1992, [https://doi.org/10.1002/1099-1085\(20000815/30\)14:11/12<1981::AID-HYP49>3.0.CO;2-F](https://doi.org/10.1002/1099-1085(20000815/30)14:11/12<1981::AID-HYP49>3.0.CO;2-F), 2000.
- Di Prinzio, M., Castellarin, A., and Toth, E.: Data-driven catchment classification: application to the pub problem, *Hydrol. Earth Syst. Sc.*, 15, 1921–1935, <https://doi.org/10.5194/hess-15-1921-2011>, 2011.
- Donnelly, C., Andersson, J. C. M., and Arheimer, B.: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, *Hydrological Sciences Journal*, 61, 255–273, <https://doi.org/10.1080/02626667.2015.1027710>, 2016.
- Donnelly, C., Greuell, W., Andersson, J., Gerten, D., Pisacane, G., Roudier, P., and Ludwig, F.: Impacts of climate change on European hydrology at 1.5, 2 and 3 degrees mean global warming above preindustrial level, *Climatic Change*, 143, 13–26, <https://doi.org/10.1007/s10584-017-1971-7>, 2017.
- Falter, D., Dung, N., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., Apel, H., Theisselmann, F., and Merz, B.: Continuous, large-scale simulation model for flood risk assessments: proof-of-concept, *J. Flood Risk Manage*, 9, 3–21, <https://doi.org/10.1111/jfr3.12105>, 2016.
- Farmer, W. H.: Ordinary kriging as a tool to estimate historical daily streamflow records, *Hydrol. Earth Syst. Sci.*, 20, 2721–2735, <https://doi.org/10.5194/hess-20-2721-2016>, 2016.
- Farmer, W. H. and Vogel, R. M.: Performance-weighted methods for estimating monthly streamflow at ungauged sites, *Journal of Hydrology*, 477, 240–250, <https://doi.org/10.1016/j.jhydrol.2012.11.032>, <http://www.sciencedirect.com/science/article/pii/S0022169412010116>, 2013.
- Ganora, D., Claps, P., Laio, F., and Viglione, A.: An approach to estimate nonparametric flow duration curves in ungauged basins, *Water Resour. Res.*, 45, <https://doi.org/10.1029/2008WR007472>, 2009.
- Haag, I. and Luce, A.: The integrated water balance and water temperature model LARSIM-WT, *Hydrol. Process.*, 22, 1046–1056, <https://doi.org/10.1002/hyp.6983>, 2008.
- Habets, F., Boone, A., Champeaux, J. L., Etchevers, P., Franchistéguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., Noilhan, J., Quintana Seguí, P., Rousset-Regimbeau, F., and Viennot, P.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, *J. Geophys. Res.*, 113, D06 113, <https://doi.org/10.1029/2007JD008548>, 2008.
- Hjerdt, N., Arheimer, B., Lindström, G., Westman, Y., Falkenroth, E., and Hultman, M.: Going Public with Advanced Simulations, in: Environmental Software Systems. Frameworks of eEnvironment: 9th IFIP WG 5.11 International Symposium, ISESS 2011, Brno, Czech Republic, June 27-29, 2011. Proceedings, edited by Hřebíček, J., Schimak, G., and Denzer, R., pp. 574–580, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-22285-6_62, 2011.
- Hollander, M. and Wolfe, D. A.: *Nonparametric Statistical Methods*, Wiley, 1999.
- Hundecha, Y., Arheimer, B., Donnelly, C., and Pechlivanidis, I.: A regional parameter estimation scheme for a pan-European multi-basin model, *Journal of Hydrology: Regional Studies*, 6, 90–111, <https://doi.org/10.1016/j.ejrh.2016.04.002>, 2016.
- Isaaks, E. H. and Srivastava, R. M.: *Applied Geostatistics*, OUP USA, 1990.

- Kim, D., Jung, I. W., and Chun, J. A.: A comparative assessment of rainfall–runoff modelling against regional flow duration curves for ungauged catchments, *Hydrol. Earth Syst. Sci.*, 21, 5647–5661, <https://doi.org/10.5194/hess-21-5647-2017>, <https://www.hydrol-earth-syst-sci.net/21/5647/2017/>, 2017.
- Komma, J., Reszler, C., Blöschl, G., and Haiden, T.: Ensemble prediction of floods – catchment non-linearity and forecast probabilities, *Nat. Hazards Earth Syst. Sci.*, 7, 431–444, <https://doi.org/10.5194/nhess-7-431-2007>, 2007.
- 5 Kouwen, N., Soulis, E., Pietroniro, A., Donald, J., and Harrington, R.: Grouped Response Units for Distributed Hydrologic Modeling, *J. Water Resour. Plan. Manage.-ASCE*, 119, 289–305, [https://doi.org/10.1061/\(ASCE\)0733-9496\(1993\)119:3\(289\)](https://doi.org/10.1061/(ASCE)0733-9496(1993)119:3(289)), wOS:A1993KY26700001, 1993.
- Kroll, C. N. and Song, P.: Impact of multicollinearity on small sample hydrologic regression models, *Water Resour. Res.*, 49, 3756–3769, <https://doi.org/10.1002/wrcr.20315>, <http://dx.doi.org/10.1002/wrcr.20315>, 2013.
- 10 Kryanova, V., Müller-Wohlfeil, D.-I., and Becker, A.: Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds, *Ecological Modelling*, 106, 261–289, [https://doi.org/10.1016/S0304-3800\(97\)00204-4](https://doi.org/10.1016/S0304-3800(97)00204-4), 1998.
- Laaha, G., Skøien, J., Nobilis, F., and Blöschl, G.: Spatial Prediction of Stream Temperatures Using Top-Kriging with an External Drift, *Environ. Model Assess.*, pp. 1–13, <https://doi.org/10.1007/s10666-013-9373-3>, 2013.
- 15 Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management, *Frontiers in Ecology and the Environment*, 9, 494–502, <https://doi.org/10.1890/100125>, 2011.
- Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, *Hydrology Research*, 41, 295–319, <https://doi.org/10.2166/nh.2010.007>, <http://hr.iwaponline.com/content/41/3-4/295>, 2010.
- 20 Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I - A discussion of principles, *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Pappenberger, F., Stephens, E., Thielen, J., Salamon, P., Demeritt, D., van Andel, S. J., Wetterhall, F., and Alfieri, L.: Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication, *Hydrol. Process.*, 27, 132–146, <https://doi.org/10.1002/hyp.9253>, 2013.
- 25 Parajka, J., Merz, R., Skøien, J. O., and Viglione, A.: The role of station density for predicting daily runoff by top-kriging interpolation in Austria, *Journal of Hydrology and Hydromechanics*, 63, <https://doi.org/10.1515/johh-2015-0024>, 2015.
- Pechlivanidis, I. G. and Arheimer, B.: Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case, *Hydrol. Earth Syst. Sci.*, 19, 4559–4579, <https://doi.org/10.5194/hess-19-4559-2015>, 2015.
- 30 Pontes, P. R. M., Fan, F. M., Fleischmann, A. S., de Paiva, R. C. D., Buarque, D. C., Siqueira, V. A., Jardim, P. F., Sorribas, M. V., and Collischonn, W.: MGB-IPH model for hydrological and hydraulic simulation of large floodplain river systems coupled with open source GIS, *Environmental Modelling & Software*, 94, 1–20, <https://doi.org/10.1016/j.envsoft.2017.03.029>, 2017.
- Pugliese, A., Castellarin, A., and Brath, A.: Geostatistical prediction of flow–duration curves in an index-flow framework, *Hydrol. Earth Syst. Sci.*, 18, 3801–3816, <https://doi.org/10.5194/hess-18-3801-2014>, <http://www.hydrol-earth-syst-sci.net/18/3801/2014/>, 2014.
- 35 Pugliese, A., Farmer, W. H., Castellarin, A., Archfield, S. A., and Vogel, R. M.: Regional flow duration curves: Geostatistical techniques versus multivariate regression, *Advances in Water Resources*, 96, 11–22, <https://doi.org/10.1016/j.advwatres.2016.06.008>, 2016.

- Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, *Hydrology and Earth System Sciences*, 17, 2637–2652, <https://doi.org/10.5194/hess-17-2637-2013>, <http://www.hydrol-earth-syst-sci.net/17/2637/2013/>, 2013.
- Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., and Freer, J. E.: A high-resolution global flood hazard model, *Water Resour. Res.*, 51, 7358–7381, <https://doi.org/10.1002/2015WR016954>, 2015.
- Skøien, J. O.: rtop: Interpolation of data with variable spatial support, <http://CRAN.R-project.org/package=rtop>, r package version 0.3-45, 2014.
- Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging - geostatistics on stream networks, *Hydrol. Earth Syst. Sci.*, 10, 277–287, <https://doi.org/10.5194/hess-10-277-2006>, 2006.
- 10 Smakhtin, V. Y. and Masse, B.: Continuous daily hydrograph simulation using duration curves of a precipitation index, *Hydrol. Process.*, 14, 1083–1100, [https://doi.org/10.1002/\(SICI\)1099-1085\(20000430\)14:6<1083::AID-HYP998>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1099-1085(20000430)14:6<1083::AID-HYP998>3.0.CO;2-2), [http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1099-1085\(20000430\)14:6<1083::AID-HYP998>3.0.CO;2-2/abstract](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-1085(20000430)14:6<1083::AID-HYP998>3.0.CO;2-2/abstract), 2000.
- Srinivas, V., Tripathi, S., Rao, A. R., and Govindaraju, R. S.: Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering, *Journal of Hydrology*, 348, 148–166, <https://doi.org/10.1016/j.jhydrol.2007.09.046>, <http://www.sciencedirect.com/science/article/pii/S0022169407005537>, 2008.
- 15 Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C., and Lindström, G.: Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale, *Hydrological Sciences Journal*, 57, 229–247, <https://doi.org/10.1080/02626667.2011.637497>, 2012.
- Wan Jaafar, W. Z., Liu, J., and Han, D.: Input variable selection for median flood regionalization, *Water Resour. Res.*, 47, W07503, <https://doi.org/10.1029/2011WR010436>, <http://dx.doi.org/10.1029/2011WR010436>, 2011.
- 20

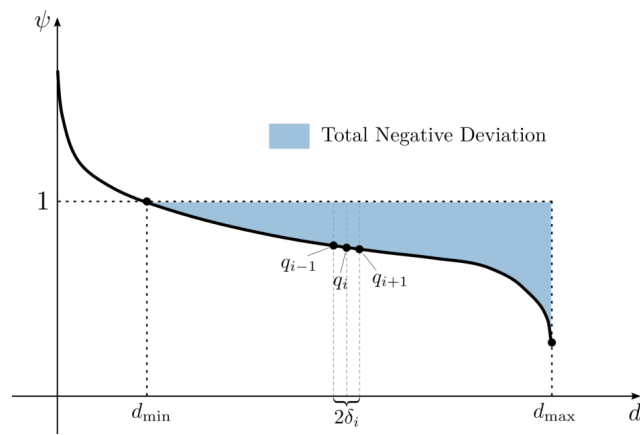


Figure 1. A sketch of the Total Negative Deviation (TND).

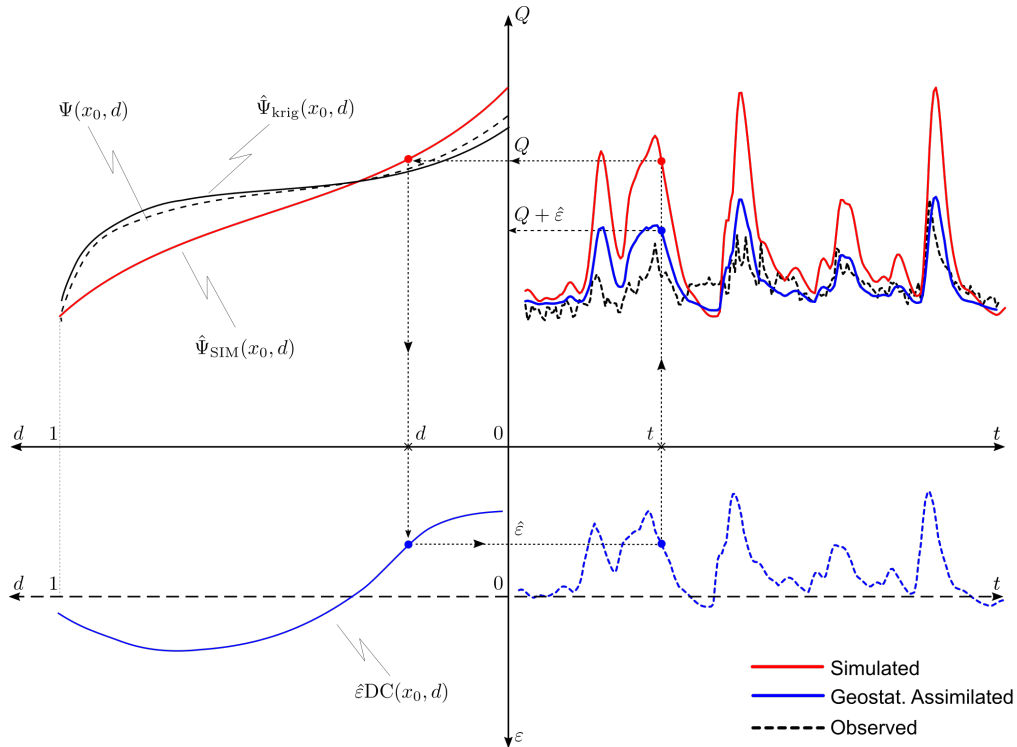


Figure 2. Illustration of the proposed data-assimilation procedure for a given simulated time series. Top-right panel: real streamflow series (unknown, since the basin is ungauged, black dashed line); macro-scale model simulation (red solid line); geostatistically-enhanced streamflow series (blue solid line). Top-left panel: FDCs predictions obtained from simulated streamflows (red solid line) and via geostatistical interpolation (black solid line); real (unknown) FDC (black dashed line). Bottom-left panel: estimated residual-duration curve (blue solid line) computed as the difference between the two predicted FDCs in the top-left panel. Bottom-right panel: time series of residuals (blue dashed line).

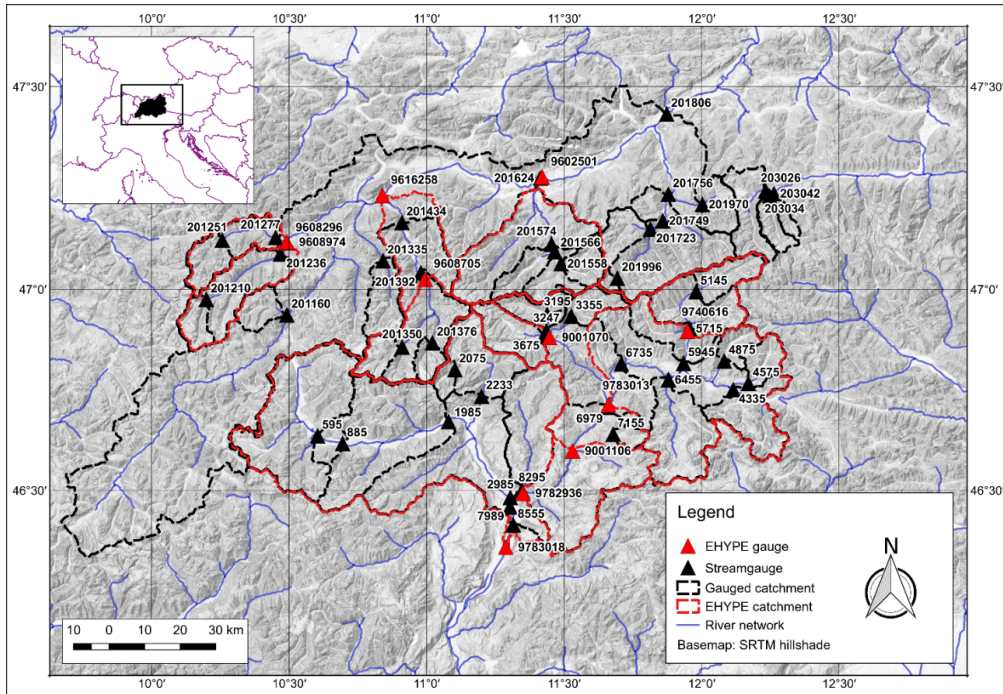


Figure 3. Study area: Tyrol. Catchment boundaries for 11 E-HYPE prediction nodes (red) and 46 stream gauges (black).

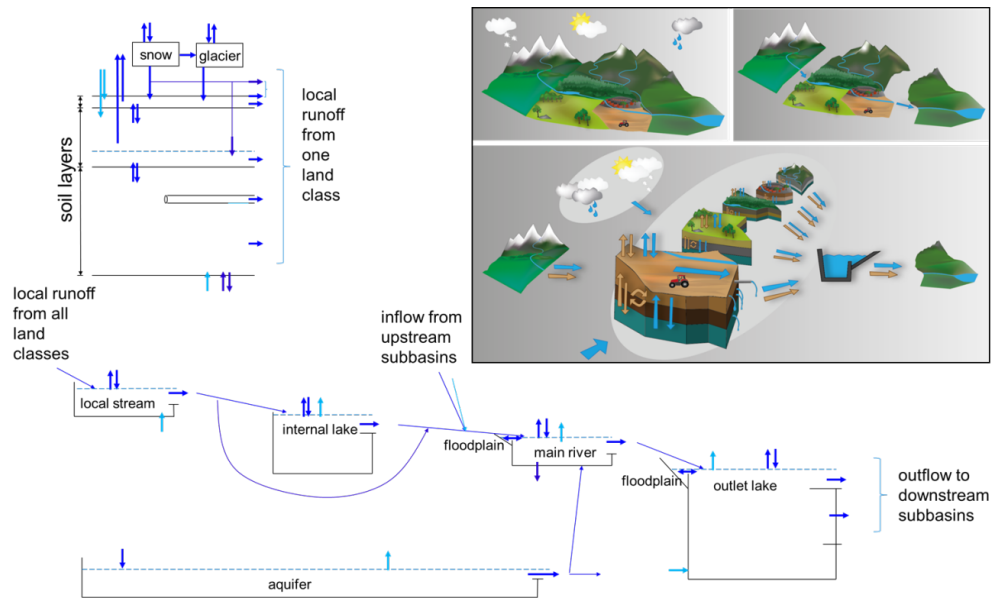


Figure 4. Schematic concept of the HYPE model (all equations are available at <http://hypecode.smhi.se/>).

Study area: Tyrol. Catchment boundaries for 11 E-HYPE prediction nodes (red) and 46 stream gauges (black):
LOOCV MAFTK

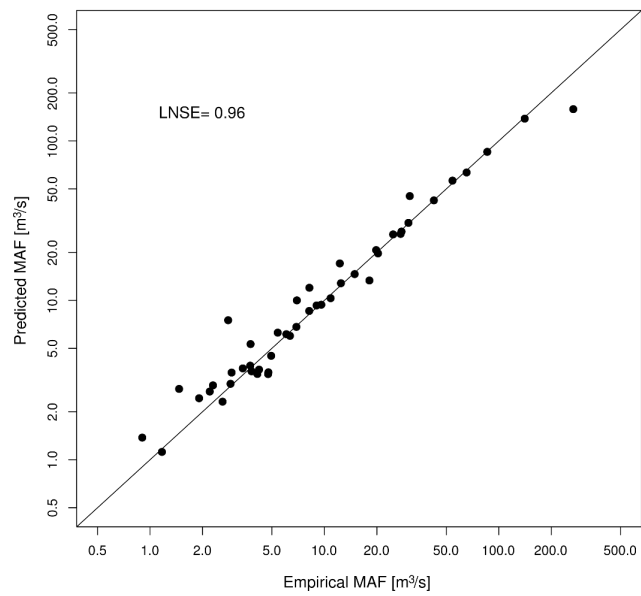


Figure 5. Top-kriging predictions of mean annual flow (MAF) in cross-validation mode.

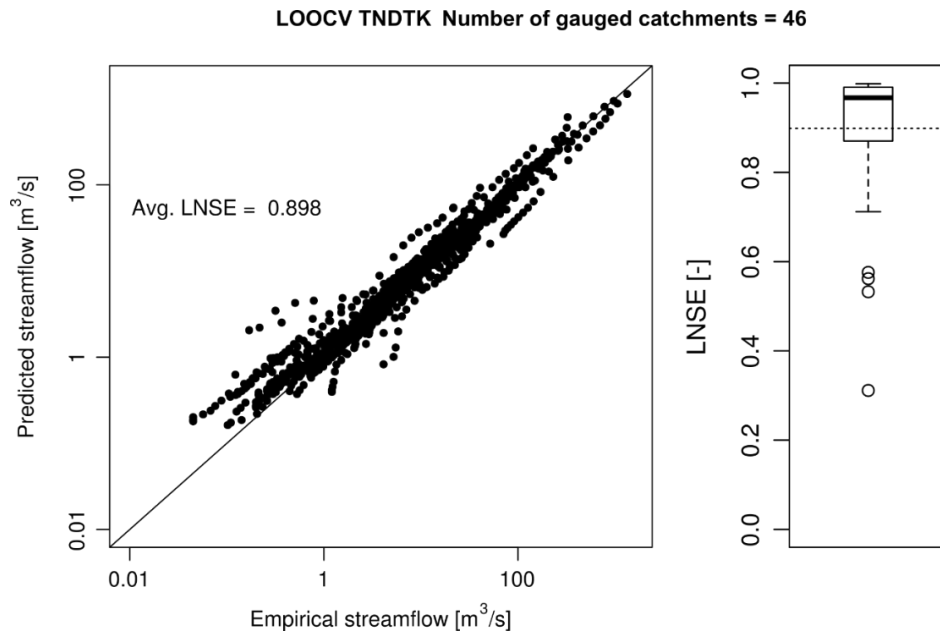


Figure 6. Left panel: scatter diagrams of empirical (*x*-axis) vs. predicted (*y*-axis) streamflows. Right panel: box-plot representation of at-site LNSE values, summarizing the 1st, 2nd (median) and 3rd quartiles along with whiskers extending to the most extreme non-outlying data point (outliers are highlighted as circles and lay at more than 1.5 times the interquartile-range from the nearest quartile); average at-site LNSE value is reported in the left panel and illustrated as a dashed line in the right one.

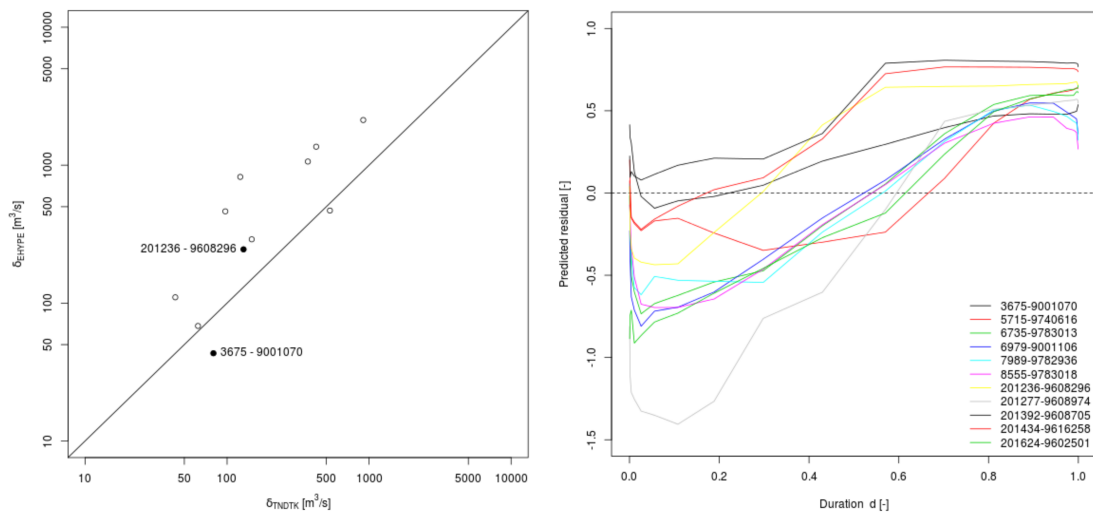


Figure 7.11 E-HYPE prediction nodes: left panel, comparison between TNDTK (x -axis) and E-HYPE (y -axis) in terms of distances δ s between empirical and predicted FDCs; the 1:1 line represents equivalent performance for TNDTK and E-HYPE; right panel, standardized standardised residual-duration curves computed as illustrated in [eqEq. \(88\)](#) (TNDTK streamflow quantiles are used for standardization).

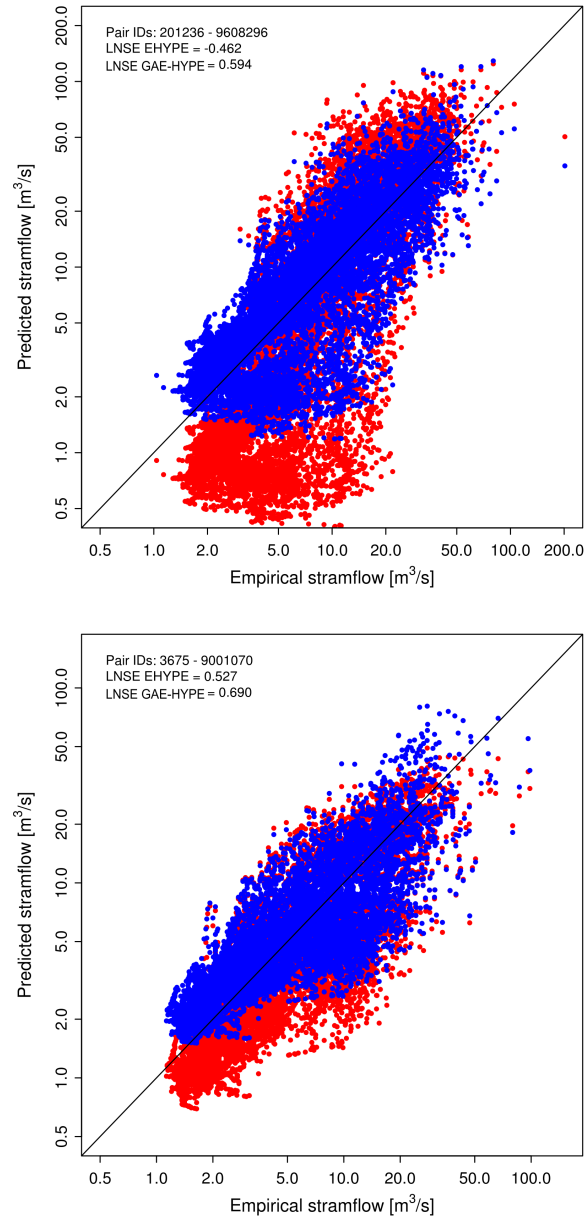


Figure 8. Scatter diagrams of empirical vs. simulated daily streamflows for either E-HYPE (red dots) and GAE-HYPE (blue dots) for two representative sites, showing the cases in which the data-assimilation procedure respectively produced the largest (upper panel) and smallest (lower panel) degree of enhancement for the study area, respectively.

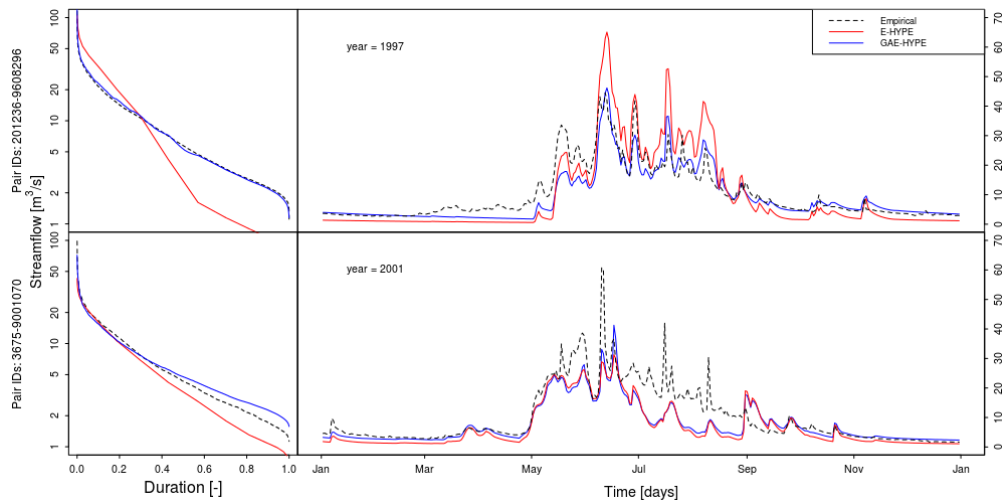


Figure 9. Examples of comparison between observed streamflow series (back dashed lines) and simulated daily streamflows via E-HYPE (red solid lines) and GAE-HYPE (blue solid lines) for two representative sites and a given year, showing two cases for which the geostatistical assimilation procedure resulted in sizeable (top panel) and limited (bottom panel) improvements, respectively.

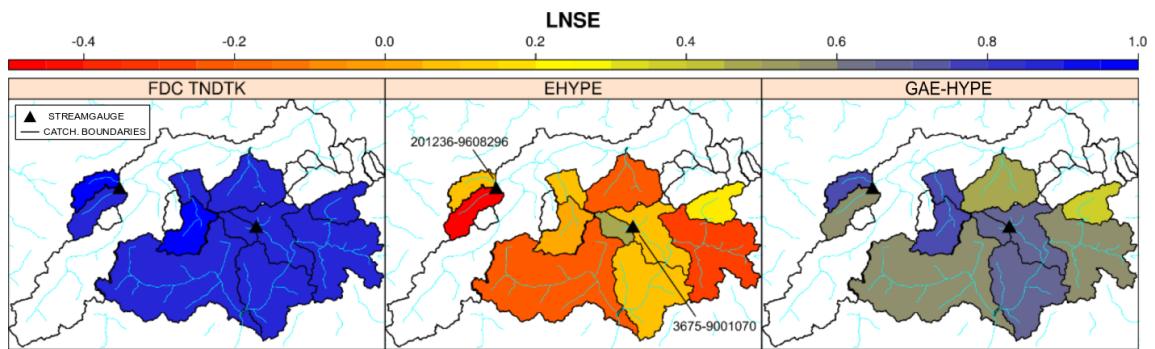


Figure 10. Spatial distribution of Nash-Sutcliffe Efficiency computed for log-transformed streamflows (LNSE) at the 11 E-HYPE prediction nodes considered in the study; geostatistically predicted flow-duration curves (FDC TNDTK, left); predicted daily streamflow time series (E-HYPE, centre, and GAE-HYPE, right, respectively); the locations of the two sites considered in [Figure 9 Figs. 8 and 9](#) are highlighted with black triangles.

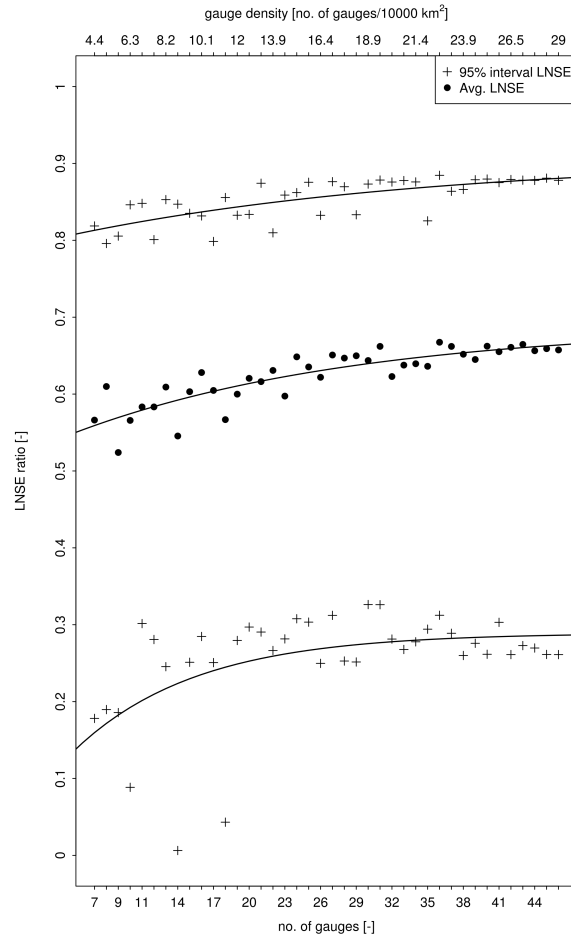


Figure 11. LNSE ratio (see [eqEq. \(1912\)](#)) as a function of streamgauge availability: black dots represent the average of 11 LNSE ratio values, while crosses indicate their 95% confidence interval.

Table 1. Study catchments: streamflow properties standardised by drainage area [$\text{m}^3\text{s}^{-1}\text{km}^{-2}$] for either gauged catchments or E-HYPE catchments, mean annual flow (qMAF), 50% and 95% streamflow quantiles (q50 and q95, respectively).

	Gauged catchments (46)			E-HYPE catchments (11)		
	qMAF	q50	q95	qMAF	q50	q95
Min.	0.0147	0.0078	0.0023	0.0261	0.0057	0.0009
75th percentile	0.0205	0.0158	0.0046	0.0276	0.0117	0.0022
Median	0.0309	0.0188	0.007	0.0294	0.0169	0.0032
Mean	0.0315	0.0195	0.0066	0.034	0.0168	0.0028
25th percentile	0.0369	0.0221	0.008	0.0351	0.0223	0.0035
Max.	0.0588	0.043	0.0116	0.0622	0.0275	0.0047

Table 2. Nash-Sutcliffe Efficiencies computed on log-transformed daily streamflows for E-HYPE and GAE-HYPE: median values for the 11 prediction nodes considered in the study; smallest enhancement (IDs 3675 - 9001070), largest enhancement (IDs 201236 - 9608296).

LNSE			E-HYPE	GAE-HYPE
Median			0.045	0.685
Pair	GAUGE ID	E-HYPE ID		
Smallest enhancement	3675	9001070	0.527	0.69
Largest enhancement	201236	9608296	-0.462	0.594