

12th International Conference on Hydroinformatics, HIC 2016

Assessing the performance of several rainfall interpolation methods as evaluated by a conceptual hydrological model

M. Mendez^a, L. Calvo-Valverde^{b,*}

^a*Escuela de Ingenieria en Construccion, Instituto Tecnologico de Costa Rica, 159-7050, Cartago, Costa Rica*

^b*Escuela de Ingenieria en Computacion, Instituto Tecnologico de Costa Rica, 159-7050, Cartago, Costa Rica*

Abstract

The objective of this study was to assess the performance of several rainfall interpolation methods as evaluated by a conceptual hydrological model. To this purpose, the upper Toro River catchment (43.15 km²) located in Costa Rica was selected as case study. Deterministic and geostatistical interpolation methods were selected to generate time-series of daily and hourly average rainfall over a period of 10 years (2001-2010). These time-series were used as inputs for the HBV-TEC hydrological model and were individually calibrated against observed streamflow data. Based on model results, the performance of the deterministic methods can be said to be comparable to that of the geostatistical methods at daily time-steps. However, at hourly time-steps, deterministic methods considerably outperformed geostatistical methods.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the organizing committee of HIC 2016

Keywords: HBV-TEC; hydrological model; Kriging, PEST; R; rainfall interpolation

1. Introduction

Hydrological models are important tools in operational hydrology, water resources management and planning. For such purposes, conceptual hydrological models are frequently used. These models approximate the general physical mechanisms governing the hydrological processes through simplified equations. For instance, they are typically less

* Corresponding author. Tel.: +506-2550-2425; fax: +506-2552-5333.

E-mail address: mamendez@itcr.ac.cr

demanding in terms of model input data [1]. On the other hand, spatial interpolation of rainfall data is of great relevance for modelling purposes, as it has a direct impact on runoff generation and catchment response. This is particularly true in mountainous areas, where patterns of rainfall spatial distribution are influenced by factors such as topography and orographic characteristics of the catchment [2]. As rainfall data are commonly collected by a network of point rain-gauges, rainfall-input should be prepared as spatially distributed data before being forced into the hydrological model [3]. Traditionally, spatial interpolation methods have been classified in two main groups; deterministic and geostatistical [4]. The fundamental principle behind deterministic methods is that the relative weight of an observed value decreases as the distance from the prediction location increases. Geostatistical however, are based on the theory of regionalized variables, and provide a set of statistical tools for incorporating the spatial correlation of observations in data processing [5]. In this context, the objective of this study is to assess the performance of various deterministic and geostatistical rainfall interpolation methods as evaluated by a conceptual hydrological model in a mountainous tropical catchment in Costa Rica.

2. Methodology

2.1. Study area

The upper Toro River catchment (43.15 km²) is located in the province of Alajuela in north-western Costa Rica (Fig.1). The topography is mountainous with elevations ranging from 2593 to 1334 m. The slope is steep with a mean value of 23%. The mean annual rainfall of the area is 4200 mm and the mean annual temperature range is between 17.2 and 32.8 °C. The land use in the catchment is dominated by forest (62%) and grassland (35%) with minor contributions from other uses; mainly water and urban. The catchment has a highly complex precipitation pattern and its temporal and spatial distribution is influenced by factors such as El Niño southern oscillation (ENOS), geomorphology, rugged terrain and microclimates.

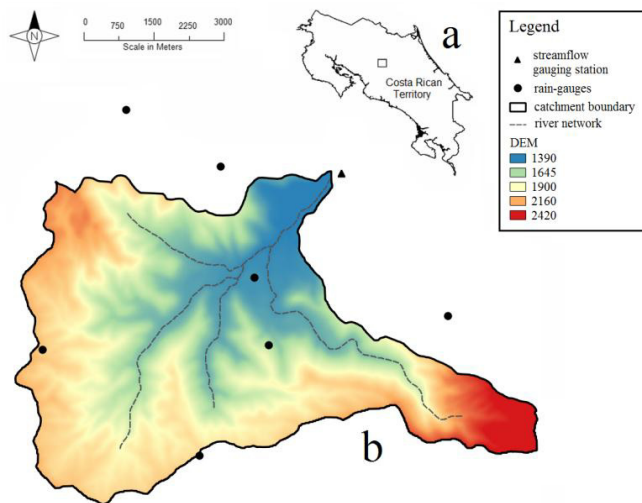


Fig. 1. (a) Position of the upper Toro River catchment in Costa Rica; (b) Upper Toro River catchment boundary, river network, RapidEye digital elevation model (DEM), rain-gauges and streamflow gauging station.

2.2. Data sources

Rainfall and temperature were calculated from daily and hourly measurements at 8 rain-gauges (Fig.1). Daily and hourly observed streamflow data for the same period were obtained from ICE-12-6 streamflow gauging station. The catchment boundary was delineated using a RapidEye 10 m digital elevation model (DEM). The monthly long-term mean potential evapotranspiration records were calculated using the Penman-Monteith method.

2.3. Interpolation methods

Deterministic and geostatistical interpolation methods (Table 1) were selected to produce spatially continuous rainfall time-series based on the 8 available rain-gauges (Fig.1) over a period of 10 years (2001-2010). These time-series were subsequently averaged over the entire catchment domain due to the next reasons: (a) the lumped/semi-distributed nature of the HBV-TEC hydrological model; and (b) the relatively small area of the catchment. Average rainfall time-series were generated for both daily and hourly temporal resolutions. Regardless of the interpolation method, spatial resolution was kept constant at 100 m, mostly on the grounds of computational costs. Selected interpolation methods were chosen on the basis of: (a) previous use in literature, (b) continuity of recorded data, (c) topography and geomorphology, (d) location and distribution of available rain-gauges and (e) computational cost for each method. Further description of these interpolation methods can be found in appropriate references [5]-[8]. All spatial data processing was executed using the R programming language [9], along with specialized R packages (Table 1). All spatial products followed the official Transverse Mercator projection system (CRTM05). In the case of ORK and KED, automatic variogram fitting was performed by R package *automap* at each time-step (daily or hourly) by minimizing the deviations between experimental data and mechanical models. Mechanical models included Spherical (Sph), Gaussian (Gau), Exponential (Exp) and M. Stein's parameterization (Ste). Spatially distributed slope, aspect and elevation raster maps were used as secondary variables (covariates) for the KED interpolation method. An Intel® Core™ i7-930, 2.80 GHz multi-core multi-threaded processor with 24 GB RAM memory was used to run all interpolation methods. ORK and KED were run in parallel by means of the *doParallel* and *foreach* R packages due to highly demanding computational requirements.

Table 1. Selected interpolation methods and relevant R packages.

Abbr.	Method	Method type	R packages
NN	Nearest neighbour	Deterministic	base, gstat, sp, raster
IDW(2-5)	Inverse distance weighting (IDW) powers 2 to 5	Deterministic	base, gstat, sp, raster
TS2	Trend surface, second order polynomial surface	Deterministic	base, gstat, sp, raster
TS2PARA	Trend surface, second order parabolic polynomial surface	Deterministic	base, gstat, sp, raster
TS2LINEAR	Trend surface with second order planar surface	Deterministic	base, gstat, sp, raster
ORK	Ordinary Kriging	Geostatistical	automap, doParallel, foreach, base, gstat, sp, raster
KED	Kriging with External Drift	Geostatistical	automap, doParallel, foreach, base, gstat, sp, raster

2.4. The HBV-TEC model

In this study, the HBV-TEC hydrological model [10] was selected for its simplicity, local development, parsimony, robustness and ease of use. The HBV-TEC is a redesign of the HBV (*Hydrologiska Byråns Vattenbalansavdelning*) model [11]-[12] developed using the R programming language [9]. Similar to the original HBV version, the HBV-TEC is a semi-distributed conceptual rainfall-runoff model for continuous calculation of runoff. The basic concept is that discharge is related to storage through a conservation of mass equation and a transformation routine. The hydrologic response is easily modelled due to the use of lumped/semi-distributed data and a simplified conceptual representation of flow processes. The structure of the HBV-TEC model consists of routines for precipitation, soil moisture, response function and transformation. The model can be run using daily or hourly time-steps; input data are precipitation, air temperature and long-term estimates of monthly potential evapotranspiration.

2.5. Optimization process and model performance

The non-linear parameter estimation and optimization package PEST [13] was used to calibrate the HBV-TEC model. Daily and hourly time-series produced by each interpolation method were individually calibrated against observed streamflow data. PEST uses the gradient-based Gauss-Marquardt Levenberg (GML) algorithm which searches for optimum values of model parameters by minimizing the deviations between field observations and modelled values. A total of 9 HBV-TEC model parameters were included in the optimization process (Table 2). The

selection of these parameters followed the next reasons: (a) the equifinality problem needed to be reduced by constraining the parameter space; and (b) only parameters having a direct influence on runoff generation were to be considered. The chosen parameters control the total volume and shape of the hydrographs and are associated with the response, transformation and soil moisture routines of the model. As individual calibration facilitates compensating errors [14], the HBV-TEC model was later rerun using average optimal parameters from all interpolation methods at daily and hourly time-steps. Parameter optimization ranges were selected based on recommended literature values. [11]-[12]. Table 3 shows the various objective functions used to evaluate the performance of the HBV-TEC model. The Correlation coefficient (R^2) was also included in the analysis.

Table 2. HBV-TEC model parameters used in PEST optimization.

Parameter	Description	Units	Function	PEST Parameter Ranges	
				Min.	Max.
perc	Percolation - upper to lower zone	mm/ Δt	Response	0.1	10
uzl	Threshold for quick flow	mm	Response	10	100
k0	Recession coefficient - upper zone	1/ Δt	Response	1E-05	1
k1	Recession coefficient - upper zone	1/ Δt	Response	1E-05	1
k2	Recession coefficient - lower zone	1/ Δt	Response	1E-05	1
maxbas	Length of weighting function	Δt	Transformation	1	100
fc	Maximum soil moisture storage	mm	Soil moisture	100	800
lp	Soil moisture threshold	-	Soil moisture	0.1	1
beta	Contribution to runoff from rain	-	Soil moisture	1	4

Table 3. Objective functions used to evaluate the performance of the HBV-TEC model.

Abbr.	Objective function	Equation
NS _{eff}	The Nash and Sutcliffe efficiency criterion (expressed as fraction)	$NS_{eff} = 1 - \frac{\sum_{i=1}^n (Q_i^{obs} - Q_i^{mod})^2}{\sum_{i=1}^n (Q_i^{obs} - \bar{Q}^{obs})^2}$
LNNS _{eff}	The Nash and Sutcliffe efficiency with logarithmic values (expressed as fraction)	$LNNS_{eff} = 1 - \frac{\sum_{i=1}^n (LNQ_i^{obs} - LNQ_i^{mod})^2}{\sum_{i=1}^n (LNQ_i^{obs} - LN\bar{Q}^{obs})^2}$
PBIAS	The Percent Bias (expressed as percentage)	$PBIAS = \frac{\sum_{i=1}^n (Q_i^{obs} - Q_i^{mod}) \cdot 100}{\sum_{i=1}^n (Q_i^{obs})}$
APB	The Absolute Percent Bias (APB) (expressed as percentage)	$APB = \frac{\sum_{i=1}^n Q_i^{obs} - Q_i^{mod} }{\sum_{i=1}^n (Q_i^{obs})}$

where i is the timestep, n is the total number of time-steps, Q is the discharge and subscripts *obs* and *mod* refer to observed and modeled correspondingly.

3. Results and discussion

3.1. Interpolation results

Daily time-series analysis shows similar descriptive statistics for all interpolation methods except for ORK (Table 4). This method produced noticeably higher maximum, median and mean values as compared with the remaining methods (Fig. 2(a)). This tendency is also reflected in the sum of the values (48490.85 mm), which represents the rainfall volume produced during the period 2001-2010. On the other hand, all three TS methods behave in a related way, with slightly lower rainfall volumes as compared with IDW methods and KED.

Table 4. Descriptive statistics for each interpolation method at daily time-steps for the period 2001-2010.

Parameter	NN	IDW2	IDW3	IDW4	IDW5	TS2	TS2PARA	TS2LINEAR	ORK	KED
max	207.86	214.07	209.74	207.98	207.25	196.23	196.30	193.49	235.50	201.30
median	4.33	4.50	4.38	4.30	4.30	4.25	4.18	4.38	5.15	4.61
mean	11.86	12.17	11.95	11.87	11.84	11.59	11.48	11.63	13.29	12.03
sum	43269.22	44388.91	43605.11	43309.04	43198.75	42262.50	41860.41	42441.88	48490.85	43877.21

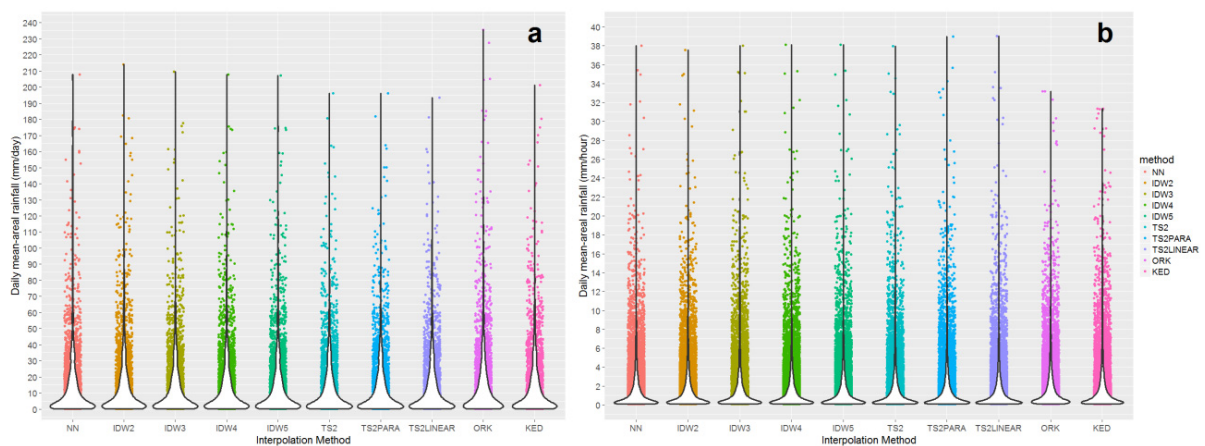


Fig. 2. Violin plots of rainfall interpolation methods results for the upper Toro River catchment at daily (a) and hourly (b) time-steps.

At hourly time-steps though, ORK and KED yielded the highest rainfall volumes values (sum) but the lowest maximum values; which can graphically be appreciated by the shallower base and the shorter extension of the violin-plot for these two methods (Fig. 2(b)). Nonetheless, median and mean values for ORK and KED are not significantly different from the remaining methods (Table 5). Regardless of the temporal resolution, no improvement was achieved by adding secondary spatial information (slope, aspect and elevation) to KED.

Table 5. Descriptive statistics for each interpolation method at hourly time-steps for the period 2001-2010.

Parameter	NN	IDW2	IDW3	IDW4	IDW5	TS2	TS2PARA	TS2LINEAR	ORK	KED
max	37.99	37.54	38.02	38.11	38.09	37.95	38.96	39.02	33.20	31.34
median	0.47	0.50	0.50	0.51	0.51	0.52	0.51	0.50	0.53	0.48
mean	1.36	1.40	1.40	1.40	1.41	1.42	1.40	1.40	1.47	1.38
sum	39708.48	40198.26	39586.05	39384.97	39327.91	39045.07	38410.60	38824.69	45032.16	42576.83

3.2. Model performance

HBV-TEC model performance at daily time-steps does not considerably differ among all interpolation methods, as stated by the outcome stability of all objective functions (Table 6). The NS_{eff} efficiency, a normalized statistic that determines the relative magnitude of the residual variance compared to the observed data variance, reaches a mean value of 0.866. The $LNNS_{eff}$ efficiency, aimed to reduce NS_{eff} sensitivity to extreme values (mainly peak flows), also returns a high mean value of 0.841 (Fig. 3(a)). The PBIAS, commonly used to quantify water balance errors, represents the objective function with the highest variance, reaching a mean value of -5.897% (Fig. 3(e)). Nonetheless, even when all interpolation methods return a negative PBIAS value (overestimation of the water balance); the TS methods yield the lowest values. This trend can somehow be graphically seen from the corresponding violin-plot (Fig. 2. (a)). The APB, a measure of the timing-difference between observations and modelled values, remains insensitive among all interpolation methods (Fig. 3(b)). Correlation coefficient (R^2) follows a similar pattern (Fig. 3(a)). To minimize compensating errors, the HBV-TEC model was later rerun using average optimal parameters from all interpolation methods; which ultimately returned similar results at daily time-steps (Fig. 3(b) and 3(f)). Based on these results, the performance of the deterministic methods can be said to be comparable to that of the geostatistical interpolation methods at daily time-steps, with a slightly better performance from the TS methods.

Table 6. Performance of various objective functions for each interpolation method at daily time-steps for the period 2001-2010.

Obj.Fuc	NN	IDW2	IDW3	IDW4	IDW5	TS2	TS2PARA	TS2LINEAR	ORK	KED	mean
NS_{eff}	0.867	0.848	0.862	0.866	0.868	0.876	0.878	0.874	0.859	0.861	0.866
$LNNS_{eff}$	0.840	0.822	0.837	0.843	0.841	0.854	0.856	0.854	0.835	0.829	0.841
PBIAS	-6.112	-9.662	-7.143	-6.146	-5.841	-2.977	-1.319	-3.486	-7.732	-8.557	-5.897
APB	18.822	20.302	19.163	18.810	18.739	17.884	17.679	18.119	19.380	19.654	18.855
R^2	0.876	0.870	0.876	0.876	0.877	0.879	0.879	0.878	0.873	0.876	0.876

At hourly time-steps nevertheless, all objective functions show a lower model performance as compared to daily time-steps, particularly $LNNS_{eff}$ (Table 7). This suggests that (a) either the HBV-TEC structure is unable to properly describe important hydrologic dynamics at higher temporal resolution (daily vs. hourly) or (b) the adopted spatially-lumped configuration of the model (which includes all interpolation methods) is incapable to properly describe the spatial pattern of rainfall distribution itself. In spite of that, ORK and KED exhibit the highest $LNNS_{eff}$ as compared to the remaining interpolation methods. At the same time, ORK and KED have the lowest NS_{eff} and the highest APB (Fig. 3(c) and 3(g)). In general, PBIAS is the most variable objective function, with ORK and KED marked as outliers in the corresponding boxplot (Fig. 3(g)).

Table 7. Performance of various objective functions for each interpolation method at hourly time-steps for the period 2001-2010.

Obj.Fuc	NN	IDW2	IDW3	IDW4	IDW5	TS2	TS2PARA	TS2LINEAR	ORK	KED	mean
NS_{eff}	0.750	0.750	0.748	0.747	0.747	0.740	0.732	0.731	0.712	0.683	0.734
$LNNS_{eff}$	0.313	0.315	0.305	0.297	0.300	0.318	0.291	0.387	0.387	0.423	0.334
PBIAS	0.296	-1.015	0.700	1.204	1.404	2.027	3.896	2.811	-13.850	-7.670	-1.020
APB	28.526	28.510	28.622	28.704	28.632	29.055	29.873	28.733	31.613	29.816	29.208
R^2	0.756	0.757	0.754	0.753	0.753	0.747	0.740	0.739	0.744	0.757	0.750

Once more, the addition of secondary spatial information did not improve KED performance. A reason for this might reside in the complex relationship between rainfall spatial distribution and the effect of steep terrain topography. The quantity and spatial distribution of the available rain-gauges (Fig. 1) might not have fulfilled all the typical assumptions of a Kriging-based geostatistical method. Furthermore, it must be remarked that automatic variogram fitting was performed by R package **automap** at each time-step, regardless of effective rainfall measurement in all or just one of the available rain-gauges. In consequence, if only 2 or 3 of the available 8 rain-

gauges recorded rainfall at one time-step, this could have created an unbalance variogram and therefore produced an unrealistic average rainfall over the entire catchment domain. Similar to daily time-steps, the HBV-TEC model performance did not vary substantially when it was rerun using average optimal parameters from all interpolation methods (Fig. 3(d) and 3(h)). Consequently, for the upper Toro River catchment, deterministic methods generally outperformed geostatistical methods at hourly time-steps.

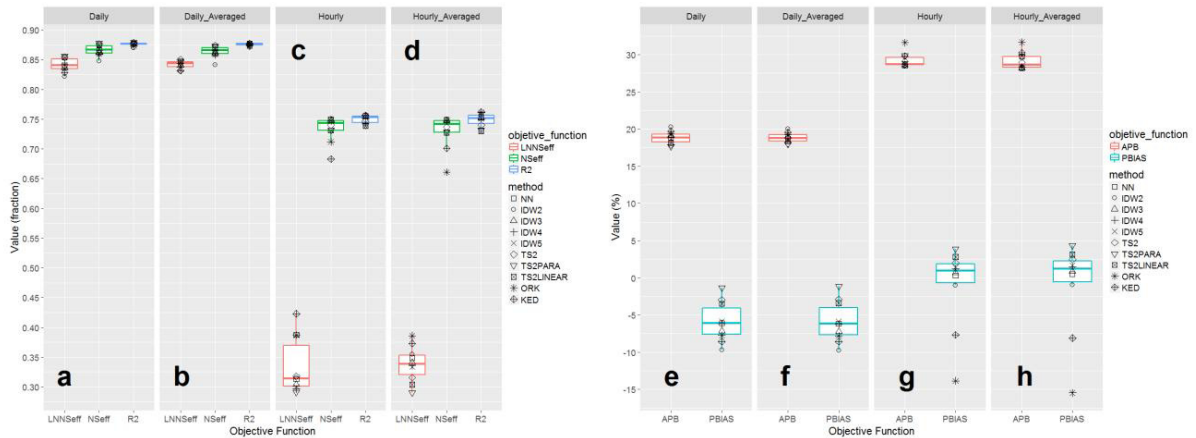


Fig. 3. HBV-TEC model performance boxplots of selected rainfall interpolation methods for the upper Toro River catchment at daily (a and e), daily-averaged (b and f), hourly (c and g) and hourly-averaged (d and h) time-steps.

3.3. Parameter optimization

PEST optimization of the HBV-TEC model parameters seems to confirm the findings made through the objective-functions analysis. As model parameters are intrinsically dependent on time resolution (Table 2), a substantial difference in the proportion of optimized parameters is to be expected between daily and hourly time-steps (Fig. 4(a), 4(b), 4(c) and 4(d)). However, parameter variation among interpolation methods (except ORK and KED) is relatively low, irrespectively of the time-step. Additionally, ORK and KED appear as outliers in the hourly time-step boxplots, mainly for those parameters controlling HBV-TEC response function; $k0$, $k1$ and $k2$ (Fig. 4(d)).

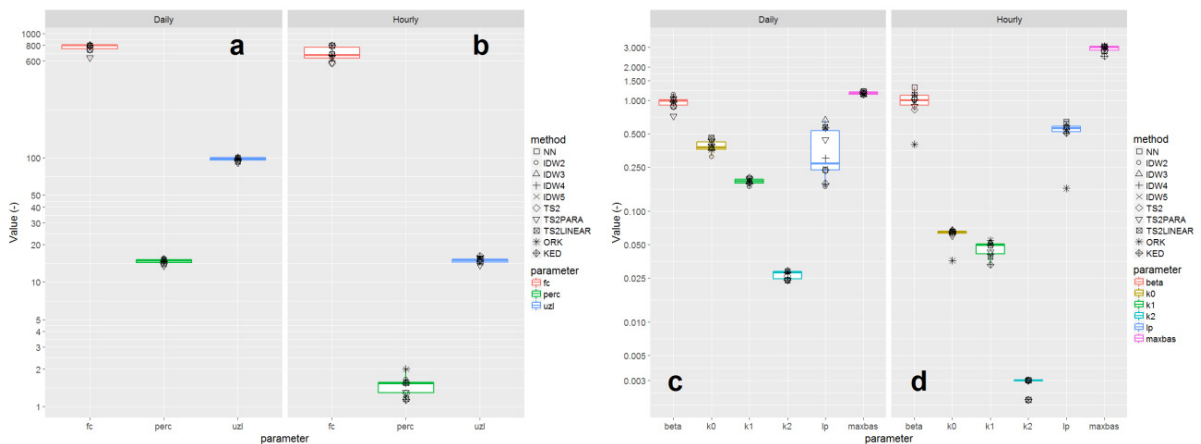


Fig. 4. HBV-TEC parameters-variation boxplots of selected rainfall interpolation methods for the upper Toro River catchment at daily (a and c), and hourly (b and d) time-steps.

At the same time, these three parameters, along with *beta* and *lp* from the HBV-TEC soil moisture function, appear to be the most sensitive parameters to changes in rainfall input volumes (Table 5), which could also help to explain the poor performance of ORK and KED at hourly time-steps. Further investigation is needed to confirm this assumption.

4. Conclusions and recommendations

The performance of several rainfall interpolation methods as evaluated by the HBV-TEC model was assessed for the upper Toro River catchment, Costa Rica. The following conclusions can be drawn:

1. The deterministic methods can be said to be comparable to that of the geostatistical interpolation methods at daily time-steps, with a slightly better performance from the TS methods.
2. Deterministic methods generally outperformed geostatistical methods at hourly time-steps.
3. Parameters controlling HBV-TEC response function seem to be the most sensitive parameters to changes in rainfall input volumes.
4. HBV-TEC model parameter optimization analysis represents a complementary indicator of the quality of the interpolated rainfall.

Future investigation is needed to fully understand the effect of interpolated rainfall over the response of the HBV-TEC model. This includes: (a) the addition of more geostatistical and machine learning methods, (b) further spatial discretization of the model domain (lumped to semi-distributed), (c) evaluation of various spatial resolutions, (d) deeper sensitivity analysis of model-parameters and (e) selection of additional secondary spatial variables (covariates).

Acknowledgements

This research was supported by Vicerrectoría de Investigación & Extensión, Instituto Tecnológico de Costa Rica (TEC). The authors are grateful to Centro de Servicio de Estudios Básicos de Ingeniería & Construcción of Instituto Costarricense de Electricidad (ICE) for providing observed data for the Upper Toro River catchment.

References

- [1] Q. Duan, S. Sorooshian, V.K. Gupta, Effective and efficient global optimization for conceptual rainfall runoff models, *Water Resour. Res.* 24 (1992) 1163–1173.
- [2] A. Mair, A. Fares, Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island, *J Hydrol Eng* 16(2011) 371–383.
- [3] S. Ly, C. Charles, A. Degré, Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review, *Biotechnol. Agron. Soc. Environ.* 17 (2013) 392–406.
- [4] A. Di Piazza, F. Lo Conti, L.V. Noto, F. Viola, G. La Loggia, Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy, *Int J Appl Earth Obs Geoinf.* 13(2011) 396–408.
- [5] P. Goovaerts, Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *J. Hydrol.* 228(2000), 113–129.
- [6] P. Goovaerts, *Geostatistics for natural resources evaluation*, New York, USA: Oxford University Press, 1997.
- [7] S. Ly, C. Charles, A. Degré, Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium, *Hydrol. Earth Syst. Sci.* 15 (2001) 2259–2274.
- [8] ITC, ILWIS 3.0, Academic User's Guide. International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede, The Netherlands, 2001.
- [9] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2015. URL <http://www.R-project.org/>
- [10] M. Mendez, L.A. Calvo, Development of the HBV-TEC hydrological model, 12th International Conference on Hydroinformatics, HIC 2016, August 2016, Unpublished conference paper, *Procedia Eng.* (2016).
- [11] S. Bergström, The HBV model. In: Singh VP (ed) *Computer Models of Watershed Hydrology*, Water Resources Publications, Highlands Ranch, Colorado, USA, 1995, pp. 443–476.
- [12] G. Lindström, B. Johansson, M. Persson, M. Gardelin, S. Bergström, Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.* 201(1997), 272–288.
- [13] J. Doherty, PEST: model independent parameter estimation: User's Manual, Watermark Numerical Computing, Brisbane, Australia, 2005.
- [14] M. Mendez, J.A. Araya, L.D. Sanchez, Automated parameter optimization of a water distribution system, *J HYDROINFORM.* 15 (2013) 71–85.