Hydrological Sciences Journal – Journal des Sciences Hydrologiques, 2015
http://dx.doi.org/10.1080/02626667.2014.959956
Special issue: *Modelling Temporally-variable Catchments*

Performance of the COSERO precipitation–runoff model under non-stationary conditions in basins with different climates

Harald Kling[1], Philipp Stanzel[1], Martin Fuchs[1] and Hans-Peter Nachtnebel[2]

[1]*Water Resources Division, Pöyry Energy GmbH, Vienna, Austria*
harald.kling@poyry.com

[2]*Department of Water-Atmosphere-Environment, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria*

Received 5 December 2013; accepted 16 May 2014

Editor Z.W. Kundzewicz; **Guest editor** G. Thirel

Abstract This study is a contribution to a model intercomparison experiment initiated during a workshop at the 2013 IAHS conference in Göteborg, Sweden. We present discharge simulations with the conceptual precipitation–runoff model COSERO in 11 basins located under different climates in Europe, Africa and Australia. All of the basins exhibit some form of non-stationary conditions, due, for example, to warming, droughts or land-cover change. The evaluation of the daily discharge simulations focuses on the overall model performance and its decomposition into three components measuring temporal dynamics, mean flow volume and distribution of flows. Calibration performance is similarly high as in previous COSERO applications. However, when looking at evaluation periods independent of the calibration, the model performance drops considerably, mainly due to severely biased discharge simulations in semi-arid basins with strong non-stationarity in rainfall. Simulations are more robust in European basins with humid climates. This highlights the fact that hydrological models frequently fail when simulations are required outside of calibration conditions in basins with non-stationary conditions. As a consequence, calibration periods should be sufficiently long to include both wet and dry periods, which should yield more robust predictions.

Key words precipitation–runoff modelling; COSERO model; model performance; non-stationarity

Performance du modèle précipitations–ruissellement COSERO dans des conditions non-stationnaires dans des bassins de différents climats

Résumé Cette étude est une contribution à une expérience de comparaison de modèles entreprise durant un atelier lors de la conférence 2013 de l'AISH à Göteborg, en Suède. Nous présentons des simulations de débits réalisées avec le modèle pluie-débit conceptuel COSERO sur 11 bassins subissant soumis à différents climats en Europe, Afrique et Australie. Tous les bassins montrent une forme de conditions non-stationnaires, dues par exemple à une augmentation de la température, à des sécheresses ou à des changements d'occupation du sol. L'évaluation des simulations de débits journaliers a été focalisée sur la performance globale du modèle et sur sa décomposition en trois éléments mesurant la dynamique temporelle, le volume moyen, et la distribution des débits. La performance du calage est bonne comme dans les applications antérieures de COSERO. Toutefois, lorsque l'on quitte le calage pour des périodes d'évaluation indépendantes, la performance du modèle diminue considérablement, ce qui est principalement dû à des simulations de débits sévèrement biaisées dans des bassins semi-arides avec des fortes non-stationnarités des précipitations. Les simulations les plus robustes ont été obtenues pour les bassins européens de climats humides. Cela souligne le fait que les modèles hydrologiques échouent fréquemment lorsque les simulations sont réalisées dans des domaines en dehors de leurs conditions de calage dans des bassins avec des conditions non-stationnaires. Les périodes de calage devraient par conséquent être suffisamment longues pour inclure à la fois des conditions humides et sèches, ce qui devrait permettre d'obtenir des prévisions plus robustes.

Mots clefs modélisation pluie–débit ; modèle COSERO ; performance du modèle ; non-stationnarité

INTRODUCTION

Precipitation–runoff models—also known as hydrological models—are widely used in hydrological science and water resources engineering. Early applications in the 1940s and 1950s included water balance modelling (e.g. Thornthwaite and Mather 1955), and since then there have been many examples where hydrological models provided accurate

estimates of runoff, relative changes in soil moisture, evapotranspiration, etc. (Gleick 1987). However, hydrological models may fail to provide robust simulations if they are applied outside of calibration conditions (Seibert 2003, Andréassian *et al*. 2009), or if no observation data are available for calibration, as is the case for prediction in ungauged basins (Hrachowitz *et al*. 2013). In recent years there is also increased awareness about the challenge of hydrological prediction in basins experiencing non-stationarity due to climate change and/or land-use change (Milly *et al*. 2008).

Even though the same models are used in science and engineering, the focus of the application is quite different. For the scientific community hydrological models typically are an avenue for hypothesis testing (e.g. to learn more about the hydrological processes in a basin), whereas for the engineering community hydrological models are mere tools that are used within the context of water resources projects (e.g. to provide best estimates of water resources).

There has been lively debate in the scientific community about which model concepts and calibration approaches to use. The discussion included (but was not limited to):

- Physically based *vs* conceptual process representation (e.g. Butts *et al*. 2004, Georgakakos *et al*. 2004).
- Detail of spatial discretization, i.e. distributed *vs* lumped catchment representation (e.g. Beven 2001, Boyle *et al*. 2001, Ajami *et al*. 2004, Andréassian *et al*. 2004, Smith *et al*. 2004).
- Impact of spatial and temporal discretization on the calibration process (e.g. Kavetski *et al*. 2006, Gallagher and Doherty 2007, Littlewood and Croke 2008, Kling and Gupta 2009).
- Manual *vs* automatic parameter calibration (e.g. Bergström *et al*. 2002, Seibert and McDonnell 2002).
- Choice of objective function to measure model performance, i.e. agreement between simulations and available observations (e.g. Yapo *et al*. 1998, Gupta *et al*. 2008, 2009).

While this discussion has deepened the understanding of the intricacies of precipitation–runoff modelling, it contributed little to engineering applications, where the modelling is often limited by constraints regarding availability of input data, budget and time. Furthermore, our personal experience is that the specific demands for each engineering project require the adaptation of available models to the peculiarities of the catchment, the data constraints, and the simulation objective. For example, if the objective is to use discharge simulations for subsequent long-term hydropower assessments, then unbiased simulations are required for long-term mean discharge, whereas floods are of less importance. On the contrary, in operational flood forecasting systems the objective is the accurate prediction of flood peaks and timing. This affects the choice of model concept (e.g. routing method) and the choice of calibration objective.

One precipitation–runoff model that has been applied frequently in research studies as well as engineering projects is the COSERO model (see references in Table 1). The main reason for choosing COSERO was that the modellers had considerable previous experience with it, thereby facilitating (a) adaptation of the model to different data constraints and (b) linking of COSERO to other models (e.g. reservoir simulation models). As the model concept of COSERO is quite similar to other well-known precipitation–runoff models, we do not expect COSERO to perform significantly better or worse than other models. However, a model intercomparison with other widely-used and tested models using the same basins and datasets has not until now been sufficiently analysed. Therefore, the objective of this paper is to present the results with COSERO in the model-intercomparison study *"Testing simulation and forecasting models in non-stationary conditions"* initiated by IAHS (Thirel *et al*. 2015).

The paper is structured as follows: The Methods section gives a description of the COSERO model, the study basins and the experiment set-up. Results are presented as performance statistics for two experiment levels, as well as an intercomparison with one other precipitation–runoff model. After a discussion of the results the paper closes with a summary and conclusions.

Because of the design of the model-intercomparison study, there are some limitations to overcome in real applications:

- the basins are not well known to the modeller (apart from one or two sentences of basin description in the metadata set of the experiment);
- the focus on multiple basins (on different continents) reduces the available time for detailed analysis in each basin, as compared to studies focusing solely on one basin; and

Table 1 Examples of previous COSERO applications. Only those studies are listed where the first author of this paper had direct access to the executable models (to compute the data for this table). P: mean annual precipitation in calibration period. PET: mean annual potential evapotranspiration in calibration period. Q: mean annual runoff-depth in calibration period. All performance statistics (KGE′, r, β, γ, NSE) are dimensionless.

Country[a]	River	Area (km^2)	Snow	P (mm)	PET/P (-)	Q/P (-)	Calibration		Calibration performance (-)					Application	Reference
							period	Objective	KGE′	r	β	γ	NSE		
Monthly time step															
Austria	Ybbs[b]	507	Yes	1776	0.34	0.68	1961-1990	Manual[e]	0.84	0.87	0.98	0.90	0.76	Hydrological atlas	Kling and Nachtnebel (2009a)
Austria	Gail	1305	Yes	1588	0.33	0.69	1961-1990	Manual[e]	0.87	0.93	1.00	0.89	0.87	Research	Kling and Nachtnebel (2009b)
Germany	Danube	101 810	Yes	1130	0.52	0.53	1961-1990	KGE′	0.94	0.94	1.00	1.00	0.87	Research	Kling *et al.* (2012)
Zambia	Zambezi	519 399	No	942	1.71	0.08	1961-1990	KGE′	0.92	0.94	1.00	0.96	0.88	Decision support	Kling *et al.* (2014)
Daily time step															
Austria	Enns	2116	Yes	1215	0.42	0.74	1971-1983	Manual	0.94	0.95	1.03	0.99	0.89	Research	Nachtnebel and Fuchs (2004)
Austria	Gail	1305	Yes	1577	0.36	0.66	1971-1995	Manual	0.86	0.92	0.98	0.89	0.84	Research	Eder *et al.* (2005)
Austria	Glan[c]	432	Yes	902	0.67	0.35	1973-1988	NSE	0.84	0.89	1.03	0.89	0.79	Research	Gupta *et al.* (2009)
Portugal	Paiva	647	(Yes)	1676	0.58	0.64	1960-1990	KGE	0.95	0.96	0.98	0.99	0.91	Hydro power	Unpublished[f] (2011)
Turkey	Seyhan	13 264	Yes	514	1.81	0.45	2001-2006	KGE′	0.84	0.84	0.99	1.00	0.69	Hydro power	Unpublished[f] (2013)
Zambia	Zambezi	519 399	No	1012	1.62	0.07	1998-2009	KGE′	0.91	0.91	1.00	1.00	0.83	Flood forecasting	Unpublished[f] (2013)
Mozambique	Revubue	16 263	No	967	1.51	0.12	1998-2012	KGE′	0.83	0.83	1.01	1.00	0.65	Flood forecasting	Unpublished[f] (2013)
Laos	Kading	8668	No	2428	0.42	0.62	2000-2007	Manual	0.72	0.82	1.04	0.79	0.67	Hydro power	Unpublished[f] (2014)
Mali	Niger	1.6×10^6	No	488	4.63	0.05	1998-2005	KGE′	0.96	0.96	1.00	1.00	0.92	Flood forecasting	Unpublished[f] (2014)
Hourly time step															
California, US	Carson	922	Yes	1150	0.87	0.44	1990-1997	NSE	0.88	0.96	1.06	0.91	0.92	Research	Kling *et al.* (2008)
Arizona, US	Marshall	1.5	(Yes)	696	0.72	0.43	2007-2010[d]	KGE	0.86	0.87	1.04	0.97	0.74	Research	Unpublished[g] (2011)
Papua Guinea	Frieda	1036	No	7678	0.20	0.81	1995-2009	KGE′	0.87	0.87	1.00	0.99	0.73	Flood assessment	Unpublished[f] (2011)
15-minutes time step															
Austria	Traisen	729	Yes	1243	0.50	0.52	2003-2011	Manual	0.90	0.94	1.07	1.03	0.84	Flood forecasting	Stanzel *et al.* (2008)

[a]Only the country where the majority of the basin is located is given.
[b]Data for Ybbs River basin, which has median NSE performance of 140 Austrian basins studied.
[c]Data for Glan River basin, which is a selected representative example of 49 Austrian basins studied.
[d]No reliable input data in winter, evaluation only from May to October.
[e]Regional calibration, no local adjustment of parameters in individual basin.
[f]Technical report for client, not available for general public.
[g]Unpublished research study, see Heidbüchel et al. (2012) for basin description.

- the model is applied without testing alternative model structures and calibration parameters.

The last point above, in particular, may considerably limit the achievable model performance. If the peculiarities of a basin are not well-known to the modeller, then there is the risk that sensitive parameters are left out of the calibration process.

METHODS

This section describes the hydrological model, the study basins, the applied model structure for the different basins, and the experiment set-up (calibration and evaluation procedure).

The COSERO model

This study uses the COntinous SEmi-distributed RunOff model (COSERO), which was developed at the University of Natural Resources and Life Sciences, Vienna (Austria) in the early 1990s, originally for the simulation of discharge of the alpine Enns River in Austria (Nachtnebel *et al*. 1993). Several Austrian climate change impact studies made use of the COSERO model (e.g. Nachtnebel and Fuchs 2004, Stanzel and Nachtnebel 2010, Kling *et al*. 2012). In the last decade the COSERO model was also implemented as the hydrological core in operational discharge forecasting systems for the Austrian rivers Traisen, Salzach, Ybbs, Enns and Mur (see e.g. Stanzel *et al*. 2008).

In addition to Austrian basins, previous applications of the COSERO model also included research and engineering studies in other European countries, the USA, Africa and South-East Asia (see examples in Table 1). The climate varied from humid (e.g. annual rainfall above 7000 mm/year in Papua New Guinea) to semi-arid (e.g. 500 mm/year in Turkey). As a consequence, the runoff ratio—i.e. the ratio between long-term mean annual runoff and precipitation—varied from 81% in the wet basins to below 10% in the dry basins. As the examples in Table 1 show, the model performance for discharge simulation in the calibration period usually yielded sufficient results, in both wet and dry basins.

The model structure of COSERO is similar to the well-known HBV model (Bergström 1995), which belongs to the group of conceptual (as opposed to physically-based) precipitation–runoff models. Inputs to the COSERO model are precipitation, temperature and potential evapotranspiration. The model determines rainfall and snowfall from precipitation and temperature data. Snowmelt is simulated with a temperature-index approach. Actual evapotranspiration is a function of potential evapotranspiration and soil moisture. Further components of actual evapotranspiration are interception losses and sublimation of snow. Runoff generation due to rainfall and meltwater is a nonlinear function of soil moisture, using the identical model equation as in the HBV model. The generated runoff is split into three components representing surface flow, interflow and baseflow by means of linear reservoirs. Alternatively, instead of three components only two components may be used, representing a fast component and baseflow. Three runoff components are usually used for short (e.g. daily) time steps in alpine basins, whereas two components are usually used for monthly time steps and/or in lowland basins.

The COSERO model can use variable lengths of time steps, ranging from monthly to hourly or shorter (see application examples in Table 1). In the case of monthly time steps, intra-monthly variability is considered for the simulation of interception, snow processes and soil moisture accounting (see Kling 2006 for more information).

Figure 1 shows the model structure of COSERO; the basic model equations are given in the Appendix. A detailed snow module is available in COSERO, including e.g. log-normal distribution of snow-depth, cold-content of snow-pack, water holding capacity of snow-pack, refreezing of retained melt-water, settlement of snow-pack, etc. However, the detailed snow modelling is of course only important in high alpine, snow dominated basins. Therefore, for the sake of brevity the lengthy model equations of the snow module are excluded from the Appendix.

In most previous applications of COSERO a semi-distributed spatial discretization with hydrological response units (HRUs) was used. The HRUs may be delineated according to elevation bands, vegetation maps, soil maps, regular raster, etc. If only one HRU is used then this corresponds to a lumped modelling of the catchment. Several upstream and downstream sub-catchments (each consisting of its own set of HRUs) can be linked in COSERO to simulate discharge in larger river basins with multiple tributaries.

For the current study the COSERO model is applied with a lumped spatial discretization in most basins, i.e. these basins are simulated with solely one HRU and one single sub-basin. As temporal discretization daily time-steps are used. Since there is no

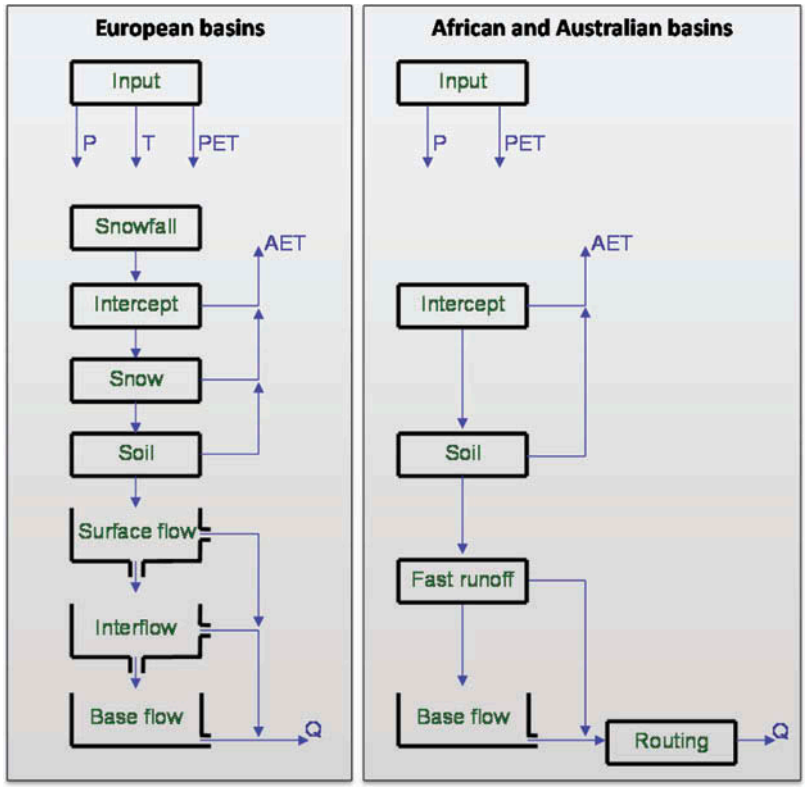

Fig. 1 Effective model structure of COSERO for European (left) and African and Australian (right) basins. P: precipitation. T: air temperature. PET: potential evapotranspiration. AET: actual evapotranspiration. Q: Runoff.

snow in several of the catchments, the model structure differs between catchments (see below). Furthermore, discharge routing with a simple lag-method for temporal shifting of the hydrograph is considered in only some of the catchments.

Study basins and applied model structure

The study basins are described in detail by Thirel *et al*. (2015). Our simulations focused on 11 basins, which were grouped into two categories:

- European basins in humid climates: Allier (France), Durance (France), Garonne (France), Rimbaud (France), Kamp (Austria), Obyan (Sweden).
- African and Australian basins in semi-arid climates: Bani (West Africa), Axe (Australia), Flinders (Australia), Gilbert (Australia), Wimmera (Australia)

Table 2 lists the main characteristics of the basins. The size varies between 1.4 km^2 (Rimbaud River) and 103 391 km^2 (Bani River). The sources for the daily input data time-series of precipitation, air temperature and potential evapotranspiration as well as the discharge measurements are given in Thirel *et al*. (2015). The climate ranges from humid to semi-arid, with runoff-ratios of about 50% in the European basins but only about 10% in the African and Australian basins. This means that the runoff-ratios cover a similar range as in previous COSERO applications (compare to Table 1).

All of the 11 study basins were specifically selected for the experiment because they exhibit some form of non-stationarity, including warming, changes in precipitation patterns, and land-cover change (Table 2, for a full description see Thirel *et al*. 2015). Therefore, it can be expected that the basins are difficult to model.

Even though the same model code was applied in all 11 basins, the parameterization of the model emphasized or shut-off different model components.

Snow processes are only simulated in the European basins. For the Allier, Durance and Garonne basins – which are located in hilly or alpine regions – five elevation bands (with equal area share) are considered to account for temperature variability within the basin, which is important for snow modelling. The temperature in each elevation band is

Table 2 Study basins and calibration performance with COSERO in complete period (Level 1 experiment). P: mean annual precipitation in calibration period; PET: mean annual potential evapotranspiration in calibration period; Q: mean annual runoff depth in calibration period. All performance statistics (KGE′, r, β, γ, NSE) are dimensionless. Basins are sorted from wet to dry (according to PET/P data).

Country[(a)]	River	Area (km^2)	Type of change	Snow	Complete period	P (mm)	PET/P (-)	Q/P (-)	Period	Calibration performance				
										KGE′	r	β	γ	NSE
European basins														
France	Durance	2170	warming	yes	1904–2010	1145	0.36	0.68	1904–2010	0.88	0.88	1.02	1.00	0.76
France	Garonne	9980	warming	yes	1961–2008	1134	0.63	0.50	1961–2008	0.90	0.90	1.00	1.00	0.80
Sweden	Obyan	97	forest loss	yes	1984–2010	794	0.64	0.45	1984–2010	0.92	0.93	1.00	1.00	0.85
France	Allier	2267	warming	yes	1961–2008	895	0.71	0.45	1961–2008	0.87	0.87	0.98	0.98	0.75
Austria	Kamp	622	warming	yes	1978–2008	798	0.76	0.35	1978–2008	0.88	0.88	1.01	0.99	0.76
France	Rimbaud	1.4	forest fire	yes	1968–2006	1061	0.94	0.63	1968–2006	0.92	0.93	0.99	0.99	0.85
African and Australian basins														
Mali	Bani	103 391	drought	no	1961–1990	1085	1.59	0.10	1961–1990	0.96	0.96	1.00	1.00	0.92
Australia	Axe	237	drought	no	1973–2011	624	1.92	0.09	1973–2011	0.85	0.85	1.00	1.00	0.70
Australia	Wimmera	2000	drought	no	1965–2009	557	2.07	0.08	1965–2009	0.89	0.89	1.00	1.00	0.78
Australia	Gilbert	1907	major flood	no	1969–1988	762	2.27	0.13	1969–1987	0.84	0.84	1.00	1.00	0.67
Australia	Flinders	1912	major floods	no	1973–2011	624	2.76	0.10	1973–2010	0.77	0.77	1.01	0.99	0.53

[(a)]Only country where the majority of the basin is located is given.

computed from the basin average temperature by assuming a vertical temperature gradient of –.0065°C/m. All other basins are simulated lumped. For the sake of simplicity, snow parameters are not calibrated, even though slightly higher performance would probably be achieved with calibrated snow parameters.

In the Allier basin the Naussac dam was constructed in 1982, which is also apparent in the observed hydrograph. Therefore, a reservoir component is inserted in the model to consider the reservoir operation since 1982, with a minimum release of 6 m^3/s during low-flow periods and refilling of the reservoir in high-flow periods.

The European basins use the standard model structure of COSERO consisting of surface flow, interflow and baseflow components. In the African and Australian basins the alternative model structure is used, only consisting of a fast runoff component and baseflow (i.e. no separate distinction between surface flow and interflow). Routing is only considered in the African and Australian basins.

Experiment set-up

The experiment set-up is described in detail by Thirel *et al*. 2015). We focused on the Level 1 and Level 2 experiments:

- Level 1: Calibration on the complete period, evaluation in five sub-periods.
- Level 2: Separate calibration in each of the five sub-periods, evaluation in the remaining four sub-periods (not used for calibration).

Level 1 yields one parameter-set for each basin, whereas Level 2 yields five parameter-sets for each basin.

The complete period and sub-periods (P1 to P5) were pre-defined in the experiment set-up (Thirel *et al*. 2015). The longest complete period was defined for the Durance River (1904–2010), whereas the shortest sub-periods were defined for the Gilbert River (3 years each sub-period). A typical length of the sub-periods was about 7 years in most basins.

As the applied model structure of COSERO differs between the European basins on the one hand and the African and Australian basins on the other, different sets of calibration parameters had to be defined (see parameter description in Table A1 in the Appendix):

(a) calibrated parameters in European basins: $S0_{max}$, Beta, $H1$, $K1$, KV1, $K2$, KV2, $K3$; and
(b) calibrated parameters in African and Australian basins: SI_{max}, $S0_{max}$, $S0_{crit}$, Beta, KF, $K3$, KL.

Thus, eight parameters were selected for calibration in the European basins and seven for calibration in the African and Australian basins.

Parameters were calibrated with the shuffled complex evolution algorithm (Duan *et al*. 1992), using a slightly modified version of the KGE statistic (Gupta *et al*. 2009) as objective function. Equation (1) gives the modified version of KGE (Kling *et al*. 2012).

$$\mathrm{KGE}' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{1}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{2}$$

$$\gamma = \frac{\mathrm{CV}_s}{\mathrm{CV}_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{3}$$

where KGE′ is the modified KGE statistic (-), r is the correlation coefficient between simulated and observed discharge (-), β is the bias ratio (-), γ is the variability ratio (-), μ is the mean discharge (m^3/s), CV is the coefficient of variation of discharge (-), σ is the standard deviation of discharge (m^3/s), and the subscripts s and o represent simulated and observed discharge values, respectively.

The KGE′ statistic and its three sub-components (r, β and γ) were also used to analyse the results of the experiment. In addition, the well-known Nash-Sutcliffe efficiency criterion (NSE, Nash and Sutcliffe 1970) was computed. For a critical discussion of these measures of model performance see Gupta *et al*. (2009).

RESULTS

Level 1 experiment

Table 2 lists the performance statistics for the calibration period of the Level 1 experiment (i.e. calibration on complete period). For the European basins as well as the African basin (Bani River) the overall model performance—as measured by KGE′ and NSE—is similar as in previous COSERO applications (compare to Table 1). However, in general, the model

performance in the Australian basins is lower. The sub-components of the model performance show that this is due to lower correlation (r), whereas the bias ratio (β) and variability ratio (γ) are close to the optimum. A lower correlation means that the model has problems in accurately simulating the temporal dynamics of discharge.

The main focus of the experiments is on the evaluation in the five sub-periods. The performance statistics were computed for each of the five periods separately. Figure 2 summarizes these results with boxplots for each basin, arranged from wet (left) to dry (right) according to the PET/P ratio (see Table 2). The boxplots show the minimum, maximum, lower quartile, upper quartile and median performance.

The performance in the evaluation sub-periods generally decreases from wet to dry basins. In all of the European basins the performance does not vary significantly between sub-periods (which is indicated by small boxplots in Fig. 2) and, in general, is high. This relates to the overall performance (KGE′) and its three components. This means that temporal dynamics are simulated well (high values of r), the simulations are unbiased (β close to 1.0) and variability is simulated reasonably well (γ close to 1.0) in all sub-periods.

More ambiguous results are obtained in the dry basins. In the Bani River basin in Africa the correlation is above 0.95 in all sub-periods and the variability ratio is close to the optimum. However, the bias becomes significant in some of the sub-periods, with values of ±20%.

Worse results are obtained in the Australian basins, especially for the Gilbert River. Here, the overall performance (KGE′) shows large variations between sub-periods, which is caused by variations in all three sub-components (i.e. r, β, γ).

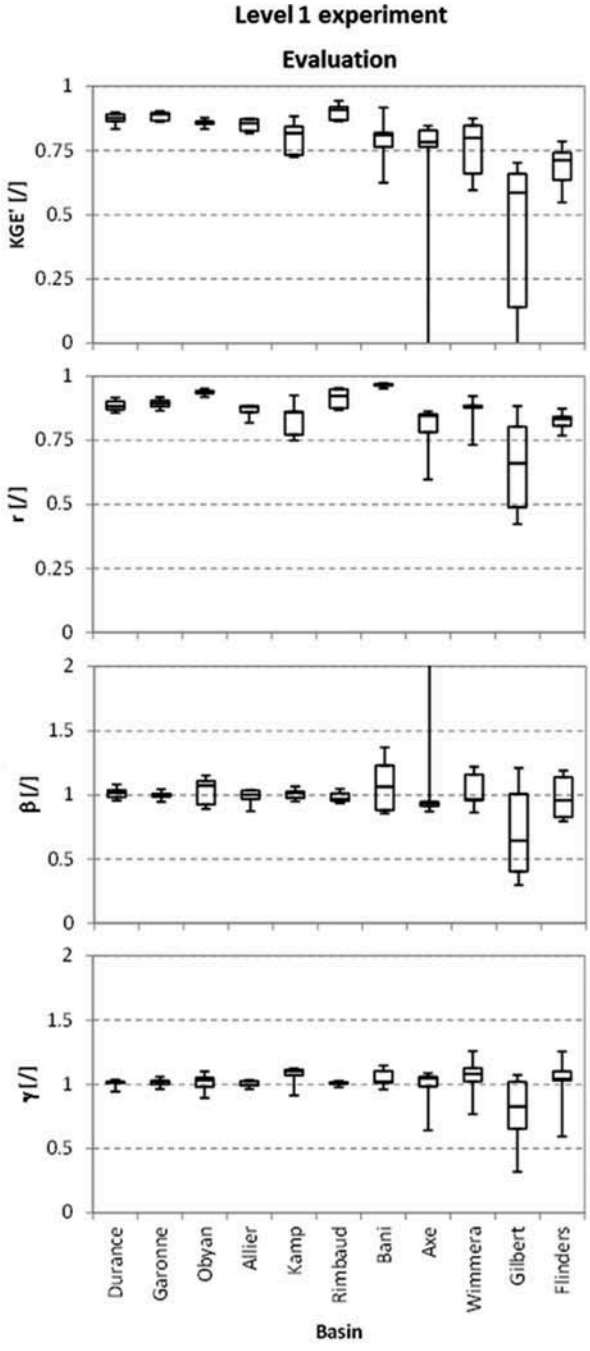

Fig. 2 Evaluation results in Level 1 experiment (see text for description). Basins are sorted from wet (left) to dry (right) according to PET/P data (Table 2). Top: overall model performance (KGE′). Second row: correlation (r). Third row: bias ratio (β). Bottom: variability ratio (γ). Note that for Level 1 experiment each box plot is based on only five values.

Level 2 experiment

Figure 3 presents the results for the Level 2 experiment, with the calibration results (performance in the sub-period used for calibration) on the left and the evaluation results (performance in the remaining four sub-periods) on the right. More specifically, each boxplot for calibration is based on five values (i.e. five sub-periods) and each boxplot for evaluation is based on 20 values (i.e. 5 parameter sets × 4 independent sub-periods). As in Fig. 2, the basins are arranged from wet (left) to dry (right).

In general, the model performance in the calibration period is high (Fig. 3, left), albeit with a tendency of higher performance in the European basins (wet) and lower performance in the Australian basins (dry). The best calibration performance is obtained in the Bani River basin. In all basins the overall performance (KGE′) is determined

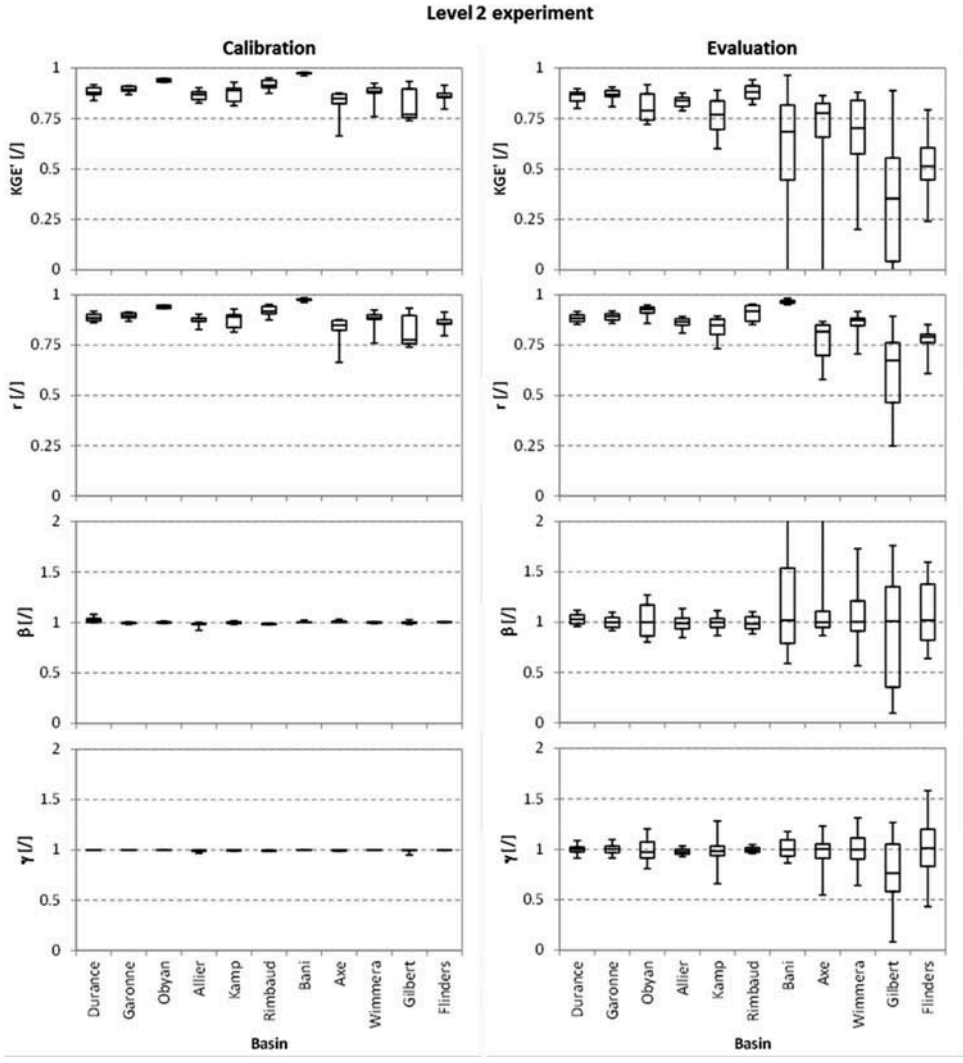

Fig. 3 Calibration and evaluation results in Level 2 experiment (see text for description). Basins are sorted from wet (left) to dry (right) according to PET/P data (Table 2). Top: overall model performance (KGE′). Second row: correlation (r). Third row: bias ratio (β). Bottom: variability ratio (γ). Boxplots for calibration results are based on five values, whereas boxplots for evaluation results are based on 20 values.

solely by the correlation (r), as the bias ratio (β) and variability ratio (γ) are close to the optimum at unity after calibration.

In the independent evaluation periods, the model performance drops slightly in the European basins (Fig. 3, right). This drop is markedly more significant in the dry African and Australian basins. The lower performance is mainly caused by poor bias ratios (β), but also the variability ratio (γ) and—apart from the Bani River—poor correlation (r) become important for the overall model performance (KGE′). Frequently, simulations are obtained with a bias of about ±50%. These are simulations where the model fails to yield robust results.

Model intercomparison

This section provides a brief comparison of the COSERO simulation results with another precipitation–runoff model, the conceptual MORDOR6 model, which is a simplified version of the original MORDOR model (Mathevet 2005). Results of the performance evaluation of MORDOR6 for the same 11 basins with the same calibration protocol were provided by the model-intercomparison project (Thirel *et al.* 2015). A difference is that MORDOR6 was calibrated on the NSE statistic (square root transformation was used for observed and simulated discharge values before computing NSE), whereas COSERO was calibrated on the

KGE′ statistic. Furthermore, for MORDOR6 only six parameters were calibrated, whereas for COSERO seven and eight parameters were calibrated in the African/Australian and European basins, respectively.

Figure 4 plots the difference between the KGE′ performance achieved by COSERO and MORDOR6 for the 11 basins and each possible combination of calibration and evaluation periods (complete period, as well as five sub-periods, yielding a 6 × 6 comparison matrix). The diagonal line from the lower left box to the upper right box represents the comparison in the calibration period, whereas all other boxes represent independent evaluation periods.

The difference in performance between COSERO and MORDOR6 is quite small for the European basins (top row in Fig. 4), with a tendency of slightly higher performance by COSERO. For the African and Australian basins (bottom row in Fig. 4) ambiguous results are obtained. Performance of COSERO is higher for the Bani, Axe and Wimmera rivers, but lower for the Flinders River and especially the Gilbert River. Interestingly, COSERO yields poorer performance in most evaluation periods for the Gilbert River even though the performance in the calibration period (diagonal line) is always slightly higher than with the MORDOR6 model. Close examination of all 11 basins depicted in Fig. 4 shows that in the calibration period (diagonal line) the performance of COSERO is always higher than or equal to MORDOR6, but never lower. Here it has to be considered that COSERO was calibrated on KGE′ (i.e. the same performance statistic that the comparison in Fig. 4 is based on), whereas MORDOR6 was calibrated on NSE (with transformed discharge data).

Figure 5 shows the same type of comparison as in Fig. 4, but this time using the NSE statistic (computed with un-transformed discharge data) as measure of model performance. In the European basins the difference in model performance between COSERO and MORDOR6 is again quite small (top row in Fig. 5), but this time MORDOR6 shows slightly higher performance. For the Bani River in Africa the performance of COSERO is in general higher than that of MORDOR6. For the Axe and Wimmera rivers in Australia the performance of COSERO is higher than that of MORDOR6 in most evaluation periods. However, for the Gilbert and Flinders rivers, MORDOR6 shows higher performance than COSERO. When comparing solely the performance in calibration periods (diagonal line), COSERO and MORDOR6 are quite similar in all 11 basins.

DISCUSSION

As found in previous studies with other models (Coron *et al*. 2012, Alfieri *et al*. 2013) the reliability of the COSERO simulations decreases from the humid European basins, to the semi-arid African and Australian basins. Similar conclusions apply to the reliability of the MORDOR6 model, which was used here for a model intercomparison.

The obtained performance of the two models is quite similar in the European basins. A ranking of the two models would depend on choice of performance measure, where a ranking based on the KGE′ statistic would favour the COSERO model and a ranking based on the NSE statistic would favour the MORDOR6 model. For the African basin (Bani River) and two Australian basins (Axe and Wimmera), COSERO simulations generally yield higher performance (regardless of whether using KGE′ or NSE), whereas, in the remaining two Australian basins (Gilbert and Flinders), MORDOR6 simulations yield higher performance. For the latter two basins, COSERO achieves similar performance to MORDOR6 only in the calibration period. This means that the drop in model performance when considering independent evaluation periods rather than calibration periods is more pronounced for the COSERO simulations than for the MORDOR6 simulations. This points at the possible problem of over-parameterization of COSERO, meaning that the information contained in the calibration data of sub-periods is not sufficient for robust parameter estimation.

The main purpose of the comparison between the COSERO and MORDOR6 models was not to determine which model is superior, but rather to analyse under which circumstances the model performance drops significantly. The results show that both models show very similar behaviour, with robust simulations in European basins. In contrast, neither model gives robust simulations in semi-arid Australian basins and it will be interesting to see the results of other models of the IAHS intercomparison project.

To better understand the differences in performance results between European, African and Australian basins, the remainder of this discussion focuses on the Durance River (France), the Bani River (Mali) and the Gilbert River (Australia). Each basin shows some peculiarities of non-stationarity (Fig. 6).

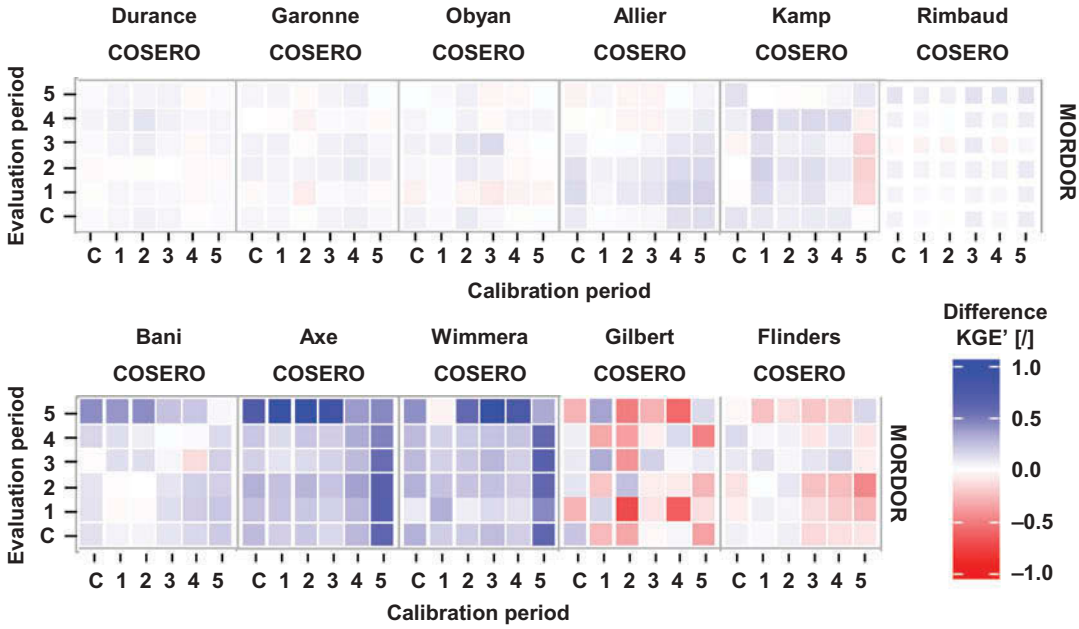

Fig. 4 Comparison of model performance for streamflow simulations with COSERO *vs* MORDOR6. Evaluation of performance based on KGE′ statistic in 11 basins and six different periods (C: complete period, 1–5: five sub-periods). Horizontal axis gives calibration period. Vertical axis gives evaluation period. Blue indicates higher performance of COSERO, whereas red indicates higher performance of MORDOR6. A detailed description of this plot type is given by Thirel *et al*. 2015).

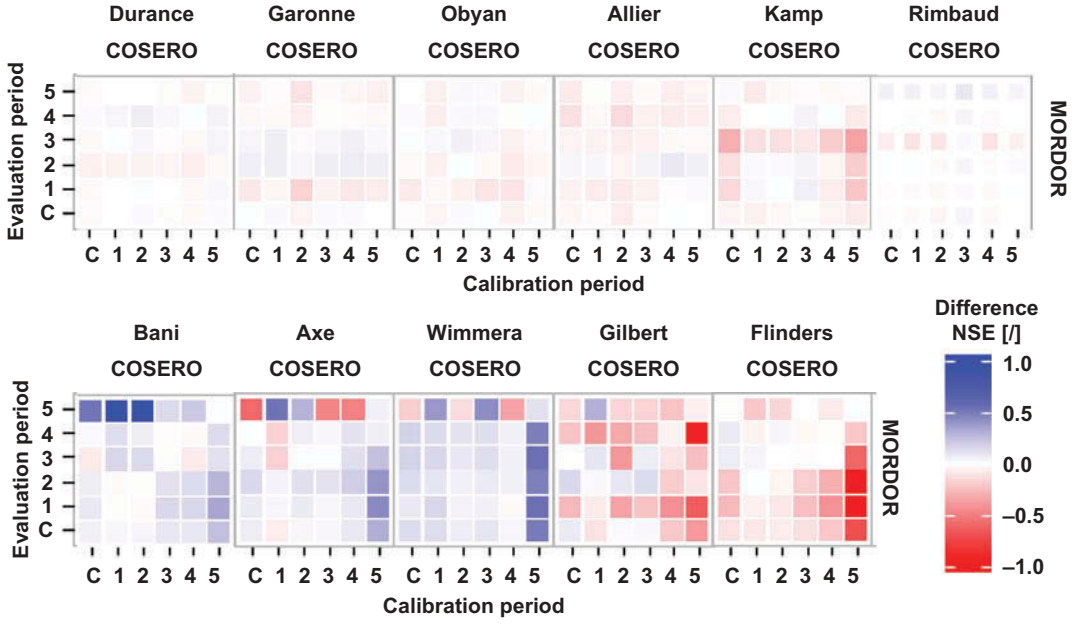

Fig. 5 Same as Fig. 4, but using the NSE statistic as performance measure.

For the Durance River, the simulation period covers more than 100 years, with a significant warming trend (Fig. 6, top). This warming trend mainly affects snow processes and thereby the seasonality in discharge. During winter, precipitation is stored in the snow layer and discharge only consists of baseflow (Fig. 7, top). A warming climate causes shifts in snow processes, which results in increase in winter discharge and decrease during the summer period (see also Kling *et al*. 2012). A second, but less significant impact is on increase in actual evapotranspiration (due to higher potential

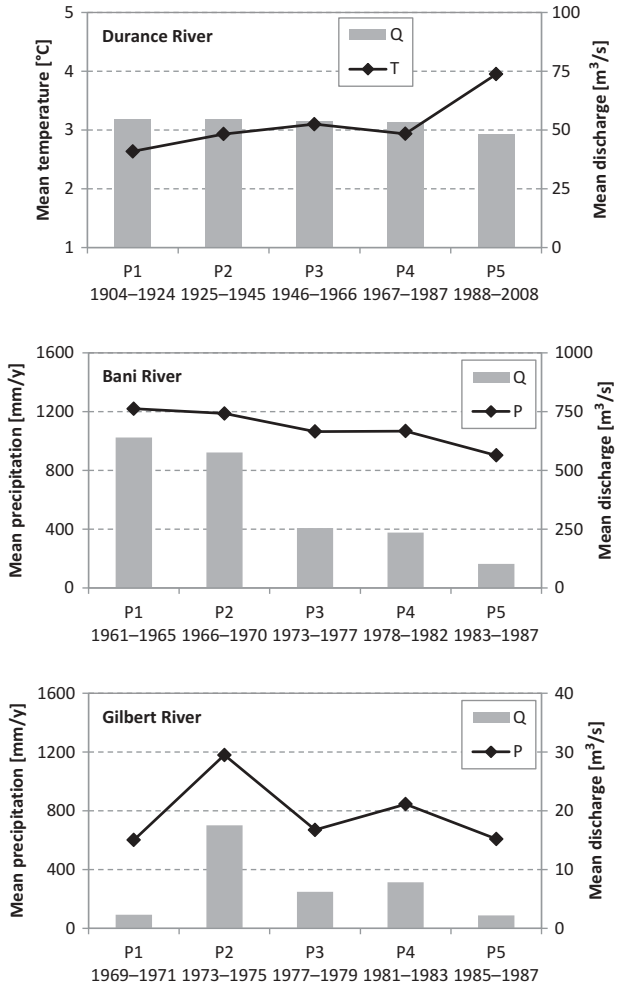

Fig. 6 Three examples for basins with non-stationarity in five different sub-periods. Q: mean observed discharge. T: mean annual air temperature. P: mean annual precipitation. Top: Durance River (France, warming trend in temperature). Middle: Bani River (Mali, decrease in precipitation). Bottom: Gilbert River (Australia, variations in precipitation).

evapotranspiration and longer growing season after earlier snowmelt) and therefore the mean annual discharge decreases. However, the increase in temperature due to climate change (about +1.5°C in the last century in the Durance basin) is rather small compared to the seasonal variations of daily temperature in each year (hot summer days are about 30°C warmer than cold winter days). In the experiment all five sub-periods have similar variations in daily temperatures and therefore represent similar hydrological conditions. This ensures that the simulations are also robust in independent evaluation periods. Hence, long-term warming trends are deemed to pose no serious challenges for hydrological modelling in this basin.

Simulations in semi-arid regions with large variations in rainfall are more problematic. In the Bani River basin, mean annual rainfall decreased from 1200 to 900 mm/year between the 1960s and the 1980s (Fig. 6, middle). This caused a sharp decline in discharge, with the result that mean observed discharge was six times larger (i.e. +500%) in the 1960s than in the 1980s. The runoff ratio was about 15% in the first two sub-periods (1960s), decreased to 7% in the third and fourth sub-periods (1973–1982), and plummeted to an extremely low value of 3% in the fifth sub-period (1983–1987). This is a challenging situation for hydrological modelling, where the experiment set-up includes calibration on extremely dry periods and evaluation in wet periods and *vice versa*. Here, the calibration and evaluation periods are not at all hydrologically similar.

Probably also of importance is the fact that in the large Bani River basin (103 941 km^2) there is only one single flood peak during each rainy season (see Fig. 7, middle). This effectively means that the model is calibrated only on five independent events during the 5-year sub-periods. This is quite a small number for calibration, especially when all five years were unusually dry or unusually wet. Due to the high sensitivity of discharge to small variations in annual precipitation, a proper calibration of the model should ensure that both wet and dry years are included in the calibration period, for example as shown by Kling *et al*. (2014) for the Zambezi River basin in southern Africa. Therefore, the performance results for the Bani River are considerably worse in the Level 2 experiment than in the Level 1 experiment. For the Level 1 experiment the bias was about ±20% in the evaluation sub-periods, whereas the bias was about ±50% in the Level 2 experiment (compare Fig. 3 to Fig. 2).

Similar considerations apply for the interpretation of the results in the semi-arid Gilbert River basin in Australia, with high inter-annual variability and intermittency of the flow regime. A difference to the Bani River is that flood events are rare in the Gilbert River, with a short and flashy response (Fig. 7, bottom). Here, rainfall showed large variations between the five sub-periods (Fig. 6, bottom), with each sub-period covering only three years. The mean observed discharge in the second sub-period was eight times larger (i.e. +700%) than in the first and fifth sub-periods. Actually, discharge was zero most of the time in the first and fifth sub-periods, whereas the second sub-period includes a flood peak

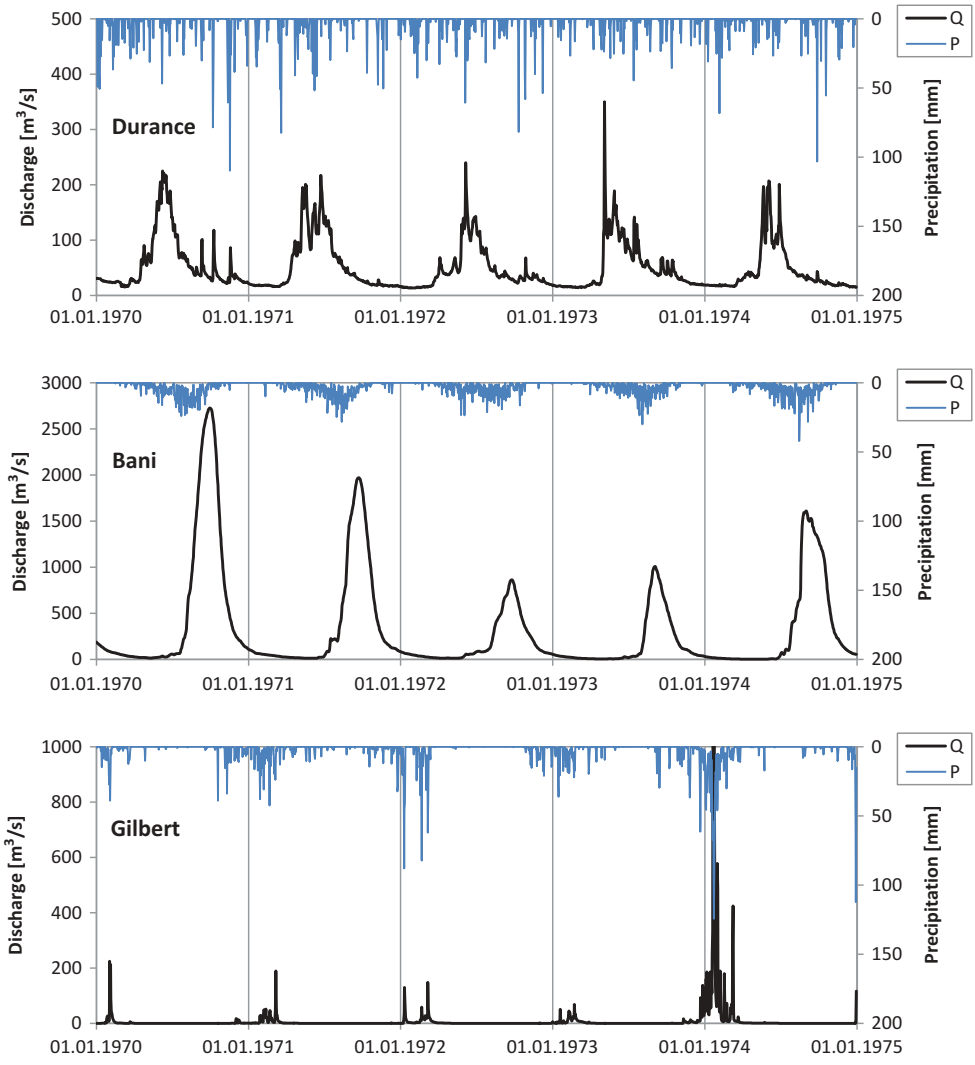

Fig. 7 Three examples for observed hydrographs under different climates. Q: observed daily discharge. P: daily precipitation. Top: Durance River (France, humid climate, snow processes). Middle: Bani River (Mali, semi-arid climate, seasonal flood peaks). Bottom: Gilbert River (Australia, semi-arid climate, intermittent hydrograph with flashy flood peaks, maximum discharge in 1974 at 3000 m^3/s).

of about 3000 m^3/s. It is clear that any hydrological simulation will fail if the model is calibrated on extremely dry conditions and then applied on periods with extreme floods.

Thus, the three examples discussed here for the Durance, Bani and Gilbert rivers show that it can be quite problematic to apply a hydrological model outside of calibration conditions. This is actually well-known to hydrologists (see e.g. Seibert 2003) and, therefore, calibration periods are usually chosen such that both dry and wet periods are included. However, in engineering applications in data-sparse, semi-arid regions the observation records may be quite short, which then poses a highly problematic situation for the application of hydrological models.

Apart from the length of observation record available for calibration, a critical step for reliability of simulations is the selection of parameters to include in the calibration. As the COSERO model includes a long list of parameters (19 in total, see Table A1), it is up to the modeller to select calibration parameters and to define parameter bounds. Typically, these decisions are based on previous experience in similar basins. As the COSERO model was applied here in Australian basins for the first time, previous experience was missing. Therefore, some parameter sensitivity tests were performed, but ideally an in-depth, global parameter sensitivity analysis would be required to learn more about the model behaviour in Australian basins.

However, this would have been outside the scope of the current study, where a straightforward approach was required for automatic parameter calibration in multiple catchments. We believe that a modeller's thorough knowledge of the basin's peculiarities can significantly improve the reliability of the simulation results (see also the "Chicken Creek" experiment, Holländer *et al*. 2009). Therefore, it may be that the lower performance in Australian basins is also affected by the selection of calibration parameters and their boundaries during optimization.

In the current study the COSERO model was applied without any modifications. In contrast, in previous COSERO applications, the modellers typically spent a considerable amount of time to learn about the peculiarities of the basin and—if required—adapted the model to these conditions. For example, for the semi-arid Australian basins it was observed that the model has problems in simulating flood peaks after prolonged dry periods. This may be a consequence of the daily computational time steps for soil moisture accounting and runoff generation. In COSERO, the runoff generation from daily rainfall is a function of soil moisture on the previous day (see model equations in the Appendix). If the soil moisture was dry on the previous day then no runoff is generated, even for high daily rainfall. The model code could be refined by applying a similar internal temporal discretization as is already used for monthly time step modelling in COSERO.

The lower performance in semi-arid basins may also be related to deficiencies in the model structure (Stanzel and Nachtnebel 2012). This includes inaccurate representation of threshold processes that affect overland flow. Here, soil hydrophobicity also affects the complex rainfall–runoff processes in semi-arid regions (see e.g. Burch *et al*. 1989). In general, daily time-step modelling is problematic if the underlying rainfall–runoff processes occur on shorter time scales (e.g. short, intense rainfall events).

In the interpretation of results it has to be considered that only non-stationarities in the climate data (precipitation, air temperature, potential evapotranspiration) are considered in the modelling exercise. Other sources of non-stationarity—such as land-cover change—are ignored. However, land-cover change actually occurred in several of the 11 basins studied (e.g. Bani River, see complete basin description in Thirel *et al*. 2015), which had an impact on runoff that is not modelled.

As a closing remark, one should be aware that every model will be falsified if it is investigated in sufficient detail (Beven 2002, Refsgaard and Henriksen 2004). Therefore, the pragmatic statement of Dooge (1972)—"*Models are to be used, not to be believed in!*"—is often a guiding principle in engineering applications. Hopefully, this experiment can highlight that hydrological simulations frequently fail if applied under extreme situations, which are encountered if the calibration period is short and the model is applied outside of calibration conditions due to strong non-stationarity in the basin.

SUMMARY AND CONCLUSIONS

This study is a contribution to a model-intercomparison study initiated by the IAHS (Thirel *et al*. 2015). The experiment consists of applying different hydrological models under non-stationary conditions in basins with different climates. The calibration and evaluation of the discharge simulations are performed in several sub-periods.

In this contribution the results with the COSERO model are presented for 11 basins. The performance in the calibration period is high in most basins; these include humid basins in Europe (with important snow processes in some of the basins), as well as a semi-arid basin in Africa. The calibration performance obtained in the current study is similar to previous applications of COSERO, even though in the current study a straightforward automatic calibration protocol was followed. Such an approach has the drawback that the modeller does not learn about and adapt the model to the peculiarities of the individual basins.

The calibration performance in Australian basins was in general lower than in the other basins, which may be related to:

- lower predictability of the semi-arid Australian basins, with high intermittency in streamflow;
- insufficient model structure of COSERO, especially for threshold processes during rainfall events; and
- modeller's selection of calibration parameters and definition of parameter bounds. As the COSERO model was applied for the first time in Australian basins, previous experience was missing to guide the model calibration.

The main focus of the analysis was on evaluation in five independent sub-periods, with the following main conclusions:

- Some of the basins—especially the semi-arid African and Australian basins—show extreme

changes (e.g. +500%) in discharge between the sub-periods, which is related to variations in precipitation. Furthermore, some sub-periods are quite short (e.g. 3 years), which poses a serious challenge for any hydrological modelling.

- Hydrological simulations frequently fail to yield robust results if the calibration period differs starkly from the evaluation period. For example, calibration on an extremely dry period will yield highly biased simulations in extremely wet periods, or *vice versa*. Therefore, a proper model calibration should always include both wet and dry periods, especially for model applications in semi-arid regions.
- The drop in performance in evaluation periods independent of the calibration is mainly related to large biases in mean discharge (bias ratio), but poor simulation of the distribution of flow (variability ratio) can also become important. The simulated temporal dynamics (correlation) are less problematic. As the bias ratio and variability ratio represent the first two moments of the flow–duration curve, this has important implications for the achievable accuracy of water resources predictions in basins with non-stationary conditions.
- In the European basins with humid climates, the different sub-periods exhibit rather similar hydrological conditions. Therefore, simulations are robust also in basins with a warming trend in temperature, because the intra-annual variability of temperature is much higher than the long-term warming trend. As a consequence, the model is not applied outside calibration conditions in independent sub-periods.
- The comparison with the MORDOR6 model showed that, in general, both models yield rather similar results, especially in the humid European basins. The COSERO simulations are more robust in the semi-arid African basin (Bani River), whereas in two of the four Australian basins the MORDOR6 model yields a higher performance.

Overall, this study showed that in basins with non-stationary conditions a high calibration performance does not guarantee sufficient performance in separate, independent evaluation periods. This is especially problematic in semi-arid basins, whereas simulations were more robust in the humid basins. As intended in the experiment design, the frequent failures of the hydrological model represented a true "crash test" under extreme conditions.

Acknowledgements This study is a contribution to the model-intercomparison project initiated by the IAHS for the IAHS Gothenburg conference in July 2013. All data were provided via Guillaume Thirel, who also computed the performance statistics from the simulation results and provided the plot-data for the comparison between the COSERO and MORDOR6 models. Many thanks go to Vazken Andréassian for the interesting discussions. The comments of Guillaume Thirel and two anonymous reviewers helped to improve the manuscript. We are grateful to Steve McClain who assisted in English proof-reading of the manuscript.

Disclosure statement No potential conflict of interest was reported by the author(s).

REFERENCES

Ajami, N.K., *et al*., 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology*, 298, 112–135. doi:10.1016/j.jhydrol.2004.03.033

Alfieri, L., *et al*., 2013. GloFAS – global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, 17 (3), 1161–1175. doi:10.5194/hess-17-1161-2013

Andréassian, V., *et al*., 2004. Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall–runoff models: a theoretical study using chimera watersheds. *Water Resources Research*, 40, W05209. doi:10.1029/2003WR002854

Andréassian, V., *et al*., 2009. *HESS Opinions* "Crash tests for a standardized evaluation of hydrological models". *Hydrology Earth System Sciences*, 13, 1757–1764. doi:10.5194/hess-13-1757-2009

Bergström, S., 1995. The HBV model. *In*: V.P. Singh, ed. *Computer models of watershed hydrology.* Highlands Ranch, CO: Water Resources Publications, ISBN No: 0–918334–91–8

Bergström, S., Lindström, G., and Pettersson, A., 2002. Multi-variable parameter estimation to increase confidence in hydrological modelling. *Hydrological Processes*, 16, 413–421. doi:10.1002/hyp.332

Beven, K., 2001. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, 5 (1), 1–12. doi:10.5194/hess-5-1-2001

Beven, K., 2002. Towards a coherent philosophy for modelling the environment. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 458 (2026), 2465–2484. doi:10.1098/rspa.2002.0986

Boyle, D.P., *et al*., 2001. Toward improved streamflow forecasts: value of semidistributed modeling. *Water Resources Research*, 37 (11), 2749–2759. doi:10.1029/2000WR000207

Burch, G.J., Moore, I.D., and Burns, J., 1989. Soil hydrophobic effects on infiltration and catchment runoff. *Hydrological Processes*, 3, 211–222. doi:10.1002/hyp.3360030302

Butts, M.B., *et al*., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, 298 (1–4), 242–266. doi:10.1016/j.jhydrol.2004.03.042

Coron, L., *et al*., 2012. Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. *Water Resources Research*, 48, 5552. doi:10.1029/2011WR011721

Dooge, J.C., 1972. Mathematical models of hydrological systems. *In*: *Proceedings of the symposium on modelling techniques in water resources systems*, Vol. 1, Ottawa: Environment Canada, 171–189.

Duan, Q., Sorooshian, S., and Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resources Research*, 28 (4), 1015–1031. doi:10.1029/91WR02985

Eder, G., *et al*., 2005. Semi-distributed modelling of the monthly water balance in an alpine catchment. *Hydrological Processes*, 19, 2339–2360. doi:10.1002/hyp.5888

Gallagher, M.R. and Doherty, J., 2007. Parameter interdependence and uncertainty induced by lumping in a hydrologic model. *Water Resources Research*, 43, W05421. doi:10.1029/2006WR005347

Georgakakos, K.P., *et al*., 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, 298, 222–241. doi:10.1016/j.jhydrol.2004.03.037

Gleick, P.H., 1987. The development and testing of a water balance model for climate impact assessment: modeling the Sacramento Basin. *Water Resources Research*, 23 (6), 1049–1061. doi:10.1029/WR023i006p01049

Gupta, H.V., *et al*., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377, 80–91. doi:10.1016/j.jhydrol.2009.08.003

Gupta, H.V., Wagener, T., and Liu, Y., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes*, 22, 3802–3813. doi:10.1002/hyp.6989

Heidbüchel, I., *et al*., 2012. The master transit time distribution of variable flow systems. *Water Resources Research*, 48 (6), 1–19. doi:10.1029/2011WR011293

Holländer, H.M., *et al*., 2009. Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. *Hydrology and Earth System Sciences*, 13, 2069–2094. doi:10.5194/hess-13-2069-2009

Hrachowitz, M., *et al*., 2013. A decade of predictions in ungauged basins (PUB) – a review. *Hydrological Sciences Journal*, 58 (6), 1198–1255. doi:10.1080/02626667.2013.803183

Kavetski, D., Kuczera, G., and Franks, S.W., 2006. Calibration of conceptual hydrological models revisited: 1. *Overcoming numerical artefacts. Journal of Hydrology*, 320, 173–186.

Kling, H., 2006. *Spatio-temporal modelling of the water balance of Austria*. Dissertation, University of Natural Resources and Applied Life Sciences, 234 pp. Available from: http://iwhw.boku.ac.at/dissertationen/kling.pdf [Accessed 1 December 2013].

Kling, H. and Gupta, H., 2009. On the development of regionalization relationships for lumped watershed models: the impact of ignoring sub-basin scale variability. *Journal of Hydrology*, 373, 337–351. doi:10.1016/j.jhydrol.2009.04.031

Kling, H. and Nachtnebel, H.P., 2009a. A method for the regional estimation of runoff separation parameters for hydrological modelling. *Journal of Hydrology*, 364, 163–174. doi:10.1016/j.jhydrol.2008.10.015

Kling, H. and Nachtnebel, H.P., 2009b. A spatio-temporal comparison of water balance modelling in an Alpine catchment. *Hydrological Processes*, 23, 997–1009. doi:10.1002/hyp.7207

Kling, H., Fuchs, M., and Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424-425, 264–277. doi:10.1016/j.jhydrol.2012.01.011

Kling, H., Stanzel, P., and Preishuber, M., 2014. Impact modelling of water resources development and climate scenarios on Zambezi River discharge. *Journal of Hydrology: Regional Studies*, 1, 17–43.

Kling, H., Yilmaz, K., and Gupta, H., 2008. Diagnostic evaluation of a distributed precipitation–runoff model for snow dominated basins. *In*: *American geophysical union: AGU joint assembly*, 27–30 May, Ft. Lauderdale, FL.

Littlewood, I. and Croke, B., 2008. Data time-step dependency of conceptual rainfall-streamflow model parameters: an empirical study with implications for regionalisation. *Hydrological Sciences Journal*, 53 (4), 685–695. doi:10.1623/hysj.53.4.685

Mathevet, T., 2005. *Quels modèles pluie-débit globaux au pas de temps horaire? Développements empiriques et comparaison de modèles sur un large échantillon de bassins versants*. Thesis (PhD). Ecole Nationale du Génie Rural, des Eaux et Forêts, Paris, 463 pp.

Milly, P.C.D., *et al*., 2008. Stationarity is dead: whither water management? *Science*, 319, 573–574. doi:10.1126/science.1151915

Nachtnebel, H.P. and Fuchs, M., 2004. Assessment of hydrological changes in Austria due to possible climate change (in German). *Österreichische Wasser- und Abfallwirtschaft*, 56 (7/8), 79–92.

Nachtnebel, H.P., Baumung, S., and Lettl, W., 1993. Abflussprognosemodell für das Einzugsgebiet der Enns und Steyr (in German). Report, Institute of Water Management, Hydology and Hydraulic Engineering, University of Natural Resources and Applied Life Sciences Vienna.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I. A discussion of principles. *Journal of Hydrology*, 10, 282–290. doi:10.1016/0022-1694(70)90255-6

Refsgaard, J.C. and Henriksen, H.J., 2004. Modelling guidelines - terminology and guiding principles. *Advances in Water Resources*, 27, 71–82. doi:10.1016/j.advwatres.2003.08.006

Seiber, J. and McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resources Research*, 38 (11), 1241.

Seibert, J., 2003. Reliability of model predictions outside calibration conditions. *Nordic Hydrology*, 34, 477–492.

Smith, M.B., *et al*., 2004. Runoff response to spatial variability in precipitation: an analysis of observed data. *Journal of Hydrology*, 298, 267–286. doi:10.1016/j.jhydrol.2004.03.039

Stanzel, P., *et al*., 2008. Continuous hydrological modelling in the context of real time flood forecasting in alpine Danube tributary catchments. IOP Conference Series. *Earth and Environmental Science*, 4, 012005.

Stanzel, P. and Nachtnebel, H.P., 2010. Mögliche Auswirkungen des Klimawandels auf den Wasserhaushalt und die Wasserkraftnutzung in Österreich. *Österreichische Wasser- und Abfallwirtschaft*, 62 (9–10), 180–187. doi:10.1007/s00506-010-0234-x

Stanzel, P. and Nachtnebel, H.P., 2012. Comparative evaluation of different rainfall runoff models. European Geosciences Union: EGU General Assembly, Geophysical Research Abstracts 14: EGU2012–8704

Thirel, G., *et al*., 2015. Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. *Hydrological Sciences Journal*, 60 (7–8), doi:10.1080/02626667.2014.967248

Thornthwaite, C.W. and Mather, J.R., 1955. The water balance. *In*: *Publications in Climatology*, 8 (1), 104 pp., Centerton, NJ: Laboratory of Climatology, Drexel Institute of Technology.

Yapo, P.O., Gupta, H.V., and Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, 204, 83–97. doi:10.1016/S0022-1694(97)00107-8

APPENDIX

The equations, parameters and variables of the COSERO model are given for application at daily or shorter time steps. The code of the COSERO model is also applicable for monthly time steps, but here internal disaggregation and discretization schemes come into effect (see Kling 2006).

The parameters of the model are listed in Table A1 and the variables in Table A2, with Δt in hours. Variables use the index t to identify the time step. The model uses hydrological response units (HRUs) to spatially discretize a basin into homogenous, computational units. To improve readability, the index for identifying HRUs is omitted for all parameters and variables.

Snowfall module

if $T_t \geq \mathrm{RAINTRT}$:

$$\mathrm{PRAIN}_t = P_t \tag{A1}$$

$$\mathrm{PSNOW}_t = 0$$

if $T_t \leq \mathrm{SNOWTRT}$:

$$\mathrm{PRAIN}_t = 0 \tag{A2}$$

$$\mathrm{PSNOW}_t = P_t$$

if $\mathrm{SNOWTRT} < T_t < \mathrm{RAINTRT}$:

$$\mathrm{PRAIN}_t = P_t \frac{T_t - \mathrm{SNOWTRT}}{\mathrm{RAINTRT} - \mathrm{SNOWTRT}} \tag{A3}$$

$$\mathrm{PSNOW}_t = P_t - \mathrm{PRAIN}_t$$

Table A1 Parameters of the COSERO model.

Parameter	Units	Description
RAINTRT	°C	Transition temperature above which precipitation is pure rainfall
SNOWTRT	°C	Transition temperature below which precipitation is pure snowfall
TF	-	Through-fall fraction for interception of rainfall
SI_{max}	mm	Interception storage capacity
CTMAX	mm/(°C.d)	Maximum melt-factor on 21 June
CTMIN	mm/(°C.d)	Minimum melt-factor on 21 December
EVPSNO	-	Correction factor for snow sublimation
$\mathrm{S0}_{max}$	mm	Soil storage capacity
SM_{crit}	-	Critical soil moisture below which evapotranspiration is reduced
Beta	-	Soil exponent for runoff generation
H1	mm	Outlet height for surface flow
K1	h	Recession coefficient for surface flow
KV1	h	Recession coefficient for vertical percolation to interflow reservoir
H2	mm	Outlet height for interflow
K2	h	Recession coefficient for interflow
KV2	h	Recession coefficient for vertical percolation to baseflow reservoir
KF	-	Fraction for separation of fast runoff component
K3	h	Recession coefficient for baseflow
KL	h	Time-shift for lag-routing

Interception module

$$X1 = \mathrm{SI}_{t-1} + (1 - \mathrm{TF})\,\mathrm{PRAIN}_t \tag{A4}$$

$$X2 = \max\{0, X1 - \mathrm{SI}_{max}\} \tag{A5}$$

$$X3 = X1 - X2 \tag{A6}$$

$$\mathrm{ETPI}_t = \mathrm{ETP}_t \frac{1.5\ \mathrm{SI}_{max}}{1.5\ \mathrm{SI}_{max} + 1} \tag{A7}$$

$$\mathrm{ETAI}_t = \min\{\mathrm{ETPI}_t, \mathrm{X3}\} \tag{A8}$$

$$\mathrm{ETPRI}_t = \mathrm{ETP}_t - \mathrm{ETAI}_t \tag{A9}$$

$$SI_t = X3 - \mathrm{ETAI}_t \tag{A10}$$

$$\mathrm{PNETRAIN}_t = \mathrm{X2} + \mathrm{TF}\ \mathrm{PRAIN}_t \tag{A11}$$

Snow module

For snow modelling a basic version and a detailed version are available. The detailed version of the snow module includes the following concepts:

- log-normal distribution of snowfall into several classes to account for variability of snow-depth;
- melt of snow by rain;
- water holding capacity of snowpack;
- cold content of snowpack;
- refreezing of retained melt water;
- simulation of snow depth and settlement of the snow pack (snow density);
- reduction of the melt factor after snowfall;
- immediate melt of snowfall on bare, warm soil;
- fraction of area that is always free of snow due to steep slopes; and
- consideration of vertical transport of snow due to avalanches and wind.

The equations for the basic version of the snow module are given below (the equation for CTA is valid for the Northern Hemisphere):

Table A2 Variables of the COSERO model. Flux variables represent sums over the time step. State variables give the water storage at the end of the time step.

Variable	Units	Type	Description
P	mm	Input	Precipitation (sum over time step)
T	°C	Input	Air temperature (average over time step)
ETP	mm	Input	Potential evapotranspiration (sum over time step)
PRAIN	mm	Flux	Rainfall
PSNOW	mm	Flux	Snowfall
SI	mm	State	Water stored in interception reservoir
X1 to X3	mm	-	Temporary variables for interception module
ETPI	mm	-	Available potential evapotranspiration for interception
ETAI	mm	Flux	Actual evaporation from interception storage
ETPRI	mm	-	Remaining potential evapotranspiration after interception
PNETRAIN	mm	Flux	Net-rainfall after interception
CTA	mm/(C.d)	-	Actual melt factor
JDAY	-	-	Julian day of the year starting on 22 December
SMTPOT	mm	-	Potential snowmelt
SMT	mm	Flux	Actual snowmelt
SWE	mm	State	Snow water equivalent
PSOIL	mm	Flux	Rainfall plus snowmelt input to soil
ETAS	mm	Flux	Snow sublimation
ETPRS	mm	-	Remaining potential evapotranspiration after sublimation
SM	-	State	Soil moisture
S0	mm	State	Water stored in soil reservoir
ETAG	mm	Flux	Actual evapotranspiration from soil reservoir
ETA	mm	Flux	Total actual evapotranspiration
Q0	mm	Flux	Runoff generation from soil reservoir
S1	mm	State	Water stored in surface flow reservoir
Q1	mm	Flux	Surface flow
QV1	mm	Flux	Percolation to interflow reservoir
S2	mm	State	Water stored in interflow reservoir
Q2	mm	Flux	Interflow
QV2	mm	Flux	Percolation to baseflow reservoir
S3	mm	State	Water stored in baseflow reservoir
Q3	mm	Flux	Baseflow
QSIM	mm	Flux	Runoff

$$\mathrm{CTA}_t = -\cos\left(\mathrm{JDAY}_t \frac{2\pi}{365.25} \frac{\mathrm{CTMAX}-\mathrm{CTMIN}}{2}\right) + \frac{\mathrm{CTMAX}+\mathrm{CTMIN}}{2} \tag{A12}$$

$$\mathrm{SMTPOT}_t = \max\{0, T_t\}\ \mathrm{CTA}_t \frac{\Delta t}{24} \tag{A13}$$

$$\mathrm{SMT}_t = \min\{\mathrm{SMTPOT}_t, \mathrm{SWE}_{t-1} + \mathrm{PSNOW}_t\} \tag{A14}$$

$$\mathrm{PSOIL}_t = \mathrm{SMT}_t + \mathrm{PNETRAIN}_t \tag{A15}$$

$$\mathrm{ETAS}_t = \min\{\mathrm{ETPRI}_t\,\mathrm{EVPSNO}, \mathrm{SWE}_{t-1} + \mathrm{PSNOW}_t - \mathrm{SMT}_t\} \tag{A16}$$

$$\mathrm{ETPRS}_t = \mathrm{ETPRI}_t - \mathrm{ETAS}_t \tag{A17}$$

$$\mathrm{SWE}_t = \mathrm{SWE}_{t-1} + \mathrm{PSNOW}_t - \mathrm{SMT}_t - \mathrm{ETAS}_t \tag{A18}$$

Soil module

if $\mathrm{SM}_{t-1} < \mathrm{SM}_{crit}$:

$$\mathrm{ETAG}_t = \mathrm{ETPRS}_t \cdot \mathrm{SM}_{t-1}/\mathrm{SM}_{crit} \tag{A19}$$

if $SM_{t-1} \geq SM_{crit}$:

$$ETAG_t = ETPRS_t \tag{A20}$$

$$ETA_t = ETAI_t + ETAS_t + ETAG_t \tag{A21}$$

$$Q0_t = PSOIL_t \cdot (SM_{t-1})^{Beta} \tag{A22}$$

$$S0_t = S0_{t-1} + PSOIL_t - ETAG_t - Q0_t \tag{A23}$$

$$SM_t = S0_t / S0_{max} \tag{A24}$$

In addition, there is an option to consider slow outflow from the soil reservoir with a linear reservoir equation.

Surface flow module

The following differential equation describes the storage S1 in the surface flow module (x is the time within the computational time step, whereas t is the index for the modelling time step). The actual model code uses the analytical solution of the differential equation (and accounts for change in system properties if the outlet at H1 falls dry).

$$\frac{dS1}{dx} = \frac{Q0_t}{\Delta t} - \frac{S1 - H1}{K1} - \frac{S1}{KV1} \tag{A25}$$

$$Q1_t = \int_0^x \frac{S1 - H1}{K1} dx \tag{A26}$$

$$QV1_t = \int_0^x \frac{S1}{KV1} dx \tag{A27}$$

$$S1_t = S1_{t-1} + Q0_t - Q1_t - QV1_t \tag{A28}$$

As an option, instead of the linear reservoir equation, a nonlinear reservoir equation can be used.

Interflow module

The following differential equation describes the storage S2 in the interflow module (x is the time within the computational time step, whereas t is the index for the modelling time step). The actual model code uses the analytical solution of the differential equation (and accounts for change in system properties if the outlet at H2 falls dry).

$$\frac{dS2}{dx} = \frac{QV1_t}{\Delta t} - \frac{S2 - H2}{K2} - \frac{S2}{KV2} \tag{A29}$$

$$Q2_t = \int_0^x \frac{S2 - H2}{K2} dx \tag{A30}$$

$$QV2_t = \int_0^x \frac{S2}{KV2} dx \tag{A31}$$

$$S2_t = S2_{t-1} + QV1_t - Q2_t - QV2_t \tag{A32}$$

There is an option to simulate the surface flow and interflow reservoirs as a coupled system with limited storage capacities, i.e. if the interflow reservoir is full then there is no percolation from the surface flow reservoir. Similarly, if the surface flow reservoir is full, then all additional inputs immediately generate surface flow. A further option is to reduce percolation to the baseflow reservoir if the soil is frozen.

Fast runoff component module

Instead of simulation of surface flow and interflow with the linear reservoir equations given above, the model alternatively can use a constant fraction to separate a fast runoff component.

$$Q1_t = KF\, Q0_t \tag{A33}$$

$$QV1_t = (1 - KF)\, Q0_t \tag{A34}$$

$$Q2_t = 0 \tag{A35}$$

$$QV2_t = QV1_t \tag{A36}$$

Baseflow module

The following differential equation describes the storage S3 in the baseflow module (x is the time within the computational time step, whereas t is the index for the modelling time step). The actual model code uses the analytical solution of the differential equation.

$$\frac{dS3}{dx} = \frac{QV2_t}{\Delta t} - \frac{S3}{K3} \tag{A37}$$

$$Q3_t = \int_0^x \frac{S3}{K3} \mathrm{d}x \quad \text{(A38)}$$

$$S3_t = S3_{t-1} + \mathrm{QV2}_t - Q3_t \quad \text{(A39)}$$

Runoff

$$\mathrm{QSIM}_t = \mathrm{Q1}_t + \mathrm{Q2}_t + \mathrm{Q3}_t \quad \text{(A40)}$$

Routing

Different options are available for downstream routing of QSIM, including:

- linear reservoir;
- Kalinin-Miljukov; and
- lag-routing (constant time shift with parameter KL).