Hydrol. Earth Syst. Sci. Discuss., 10, 487–517, 2013
www.hydrol-earth-syst-sci-discuss.net/10/487/2013/
doi:10.5194/hessd-10-487-2013

Hydrology and Earth System Sciences Discussions

Disinformative data in large-scale hydrological modelling

A. Kauffeldt[1], S. Halldin[1], A. Rodhe[1], C.-Y. Xu[1,2], and I. K. Westerberg[1,3]

[1]Department of Earth Sciences, Uppsala University, Uppsala, Sweden
[2]Department of Geosciences, University of Oslo, Oslo, Norway
[3]IVL Swedish Environmental Research Institute, Stockholm, Sweden

Received: 14 December 2012 – Accepted: 4 January 2013 – Published: 14 January 2013

Correspondence to: A. Kauffeldt (anna.kauffeldt@geo.uu.se)

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

Abstract

Large-scale hydrological modelling has become an important tool for the study of global and regional water resources, climate impacts, and water-resources management. However, modelling efforts over large spatial domains are fraught with problems of data scarcity, uncertainties and inconsistencies between forcing and evaluation data. Model-independent methods to screen and analyse data for such problems are needed. This study aims at identifying data inconsistencies in global datasets using a pre-modelling analysis, inconsistencies that can be disinformative for subsequent modelling. The consistency between different hydrographic datasets, and between climate data (precipitation and potential evaporation), and discharge data was examined in terms of how well basin areas were represented in the flow networks and the possibility of water-balance closure. It was found that: (i) most basins could be well represented in both gridded basin delineations and polygon-based ones, but some basins exhibited large area discrepancies between flow-network datasets and archived basin areas, (ii) basins exhibiting too-high runoff coefficients were abundant in areas where precipitation data were likely affected by snow undercatch, and (iii) the occurrence of basins exhibiting losses exceeding the energy limit was strongly dependent on the potential-evaporation data, both in terms of numbers and geographical distribution. These results emphasise the need for pre-modelling data analysis to identify dataset inconsistencies as an important first step in any large-scale study. Applying data-screening methods before modelling should also increase our chances to draw robust conclusions from subsequent simulations.

1 Introduction

Large-scale hydrological modelling has become a focal point in hydrological research in recent years and is of fundamental importance for understanding continental and global water balances, impacts of climate and land-use changes, and for water-resources management (Werth and Güntner, 2010; Jung et al., 2012; Lu et al., 2012). However, hydrological modelling and analysis of large spatial domains is severely constrained by data availability and quality (Arnell, 1999; Döll and Siebert, 2002; Güntner, 2008; Hunger and Döll, 2008; Peel et al., 2010; Widén-Nilsson et al., 2009). In addition, the modellers' knowledge of the quality and limitations of large-scale datasets is often inevitably inadequate, which restricts the possibility to distinguish informative from disinformative data.

Several previous studies have emphasised the importance of uncertainties and errors associated with input and evaluation data for robust hydrological inference (Beven and Westerberg, 2011; Beven et al., 2011; McMillan et al., 2011; Montanari and Di Baldassarre, 2012; Thyer et al., 2009). The possibility that data uncertainties may even render combinations of model input and evaluation data disinformative has only recently been discussed (Beven and Westerberg, 2011). Disinformative data in the hydrological context are data that are physically inconsistent and therefore misleading for model inference and hydrological analyses. Beven et al. (2011) use a master recession curve to identify rainfall-runoff events with inconsistent runoff coefficients for a British catchment, i.e. events where the water balance between precipitation input and discharge output is not satisfied, periods that they show are "disinformative" in model evaluation. Westerberg et al. (2011) develop a model evaluation criterion that can be expected to be more robust to some types of moderate disinformation and analyse the effect of some disinformative data periods on model inference in a posterior analysis. Kuczera (1996) shows that rating curve errors can "very substantially, indeed massively" corrupt design-flood estimation. When accounting for precipitation errors in calibration, Vrugt et al. (2008) found that the posterior distribution of watershed-model parameters and model uncertainty were significantly affected. Beven and Westerberg (2011) discuss the difficulties in analysing information/disinformation content in hydrological data given multiple sources of epistemic data errors and their interaction with model-structural errors. They highlight the importance of isolating disinformative data periods independent of a model and then excluding them from model calibration and

HESSD
10, 487–517, 2013

evaluation. Model-independent methods for identifying disinformative data need to be further developed and the effect of different types of disinformation on model inference should also be further investigated. These research questions are particularly relevant for global hydrological models (GHMs) that are severely data-constrained and where model fit is sometimes only anecdotally described. Very substantial correction and tuning factors are reported for GHMs in order to achieve acceptable fit to observed discharge data (e.g. Fekete et al., 1999; Hunger and Döll, 2008; Palmer et al., 2008). At the large scale it is impossible for the modeller to have the same detailed knowledge of the quality and limitations of the modelling datasets as on the local catchment scale. This effectively restricts the possibility to distinguish informative data from disinformative ones, and calls for new types of analysis methods.

GHMs are commonly evaluated against discharge since it represents an aggregated hydrological response of a basin. Selection of basins for calibration/evaluation purposes has previously been done mainly on the grounds of basin size and record-length thresholds (e.g. Fekete et al., 1999; Döll et al., 2003). However, the effect of different thresholds has not been examined.

GHMs typically operate at a spatial resolution of $0.5^\circ \times 0.5^\circ$ longitude and latitude (Arnell, 1999; Döll et al., 2003; Vörösmarty et al., 1989; Widén-Nilsson et al., 2007, 2009) and can therefore not be expected to represent small basins very well. The low resolution of GHMs leads to a trade-off between using discharge stations with small basin areas for spatial coverage and excluding them since their representation in coarse flow networks is restricted. Previous global studies have set minimum area thresholds to 9000 km^2 (Döll et al., 2003; Kaspar, 2004) and 10 000 km^2 (Fekete et al., 1999, 2002), and further reduced the number of basins based on interstation area (i.e. area between river gauges) thresholds of 20 000 and 10 000 km^2, respectively. Hanasaki et al. (2008) use an area threshold of 200 000 km^2, but their model works at a lower spatial resolution ($1^\circ \times 1^\circ$ longitude and latitude). Yet other studies have limited the evaluation to only a few major river basins (e.g. Nijssen et al., 2001).

HESSD
10, 487–517, 2013

Recent development of high-resolution hydrographic datasets such as HydroSHEDS (Lehner et al., 2008) offers the possibility to use high-resolution topographic data in global modelling, e.g. for runoff routing (Gong et al., 2011). This has also led to development of new up-scaled datasets and high-resolution basin delineations. This sparks the question if smaller basins than used in previous studies can be utilised for calibration/evaluation of GHMs and what restrictions to basin size are inflicted by input data, since precipitation for longer periods than the last decades is commonly only available at $0.5^{\circ} \times 0.5^{\circ}$ resolution.

The global-hydrological-modelling community lacks a methodology to evaluate forcing and calibration data independent of a specific model, which hampers comparisons of the results from different models. In order to be right for the right reasons, a global modelling effort should start with an evaluation of data quality and, especially, possible inconsistencies between datasets. This paper presents a basic pre-modelling scrutinisation of large-scale hydrological datasets. The overall goal of the paper was to address the problem of physically inconsistent and therefore disinformative data in large-scale hydrological modelling and to show the importance of a pre-modelling data analysis. The goal was achieved in two steps. The first step was to evaluate how well basin areas were represented in three gridded ($0.5^{\circ} \times 0.5^{\circ}$) hydrography datasets and one high-resolution GIS dataset (derived from 15 arc-second topography) for basins as small as 5000 km^2. The second was to analyse and identify inconsistencies between GHM forcing and evaluation data by comparing four precipitation datasets and three potential-evaporation datasets (all gridded at 0.5° resolution) with observed discharge.

2 Data

Basins were defined using both gridded hydrographies (flow networks) and a GIS-polygon dataset from the Global Runoff Data Centre (GRDC; Lehner, 2011). The gridded hydrographies were DDM30 (Döll and Lehner, 2002), STN-30p (Vörösmarty et al., 2000) and an early version (obtained in May 2011) of the datasets developed by Wu et al. (2012) using the automated dominant-river-tracing algorithm (Wu et al., 2011). This dataset (from here on called DRT) uses a high-resolution baseline, which is a merge between HYDRO1k (USGS EROS, 1996) for high latitudes (above 60° N) and HydroSHEDS (Lehner et al., 2008) for the rest of the land surface. We calculated cell areas for all hydrographies as quadrangles based on the World Geodetic System 1984 ellipsoid.

Discharge data were obtained from GRDC in June 2011, at which time the archive held records for 7763 discharge stations worldwide. Record lengths varied considerably between stations. Only monthly data calculated by GRDC from daily records were used because these data tend to have a higher accuracy than data originally provided (Thomas de Couet, GRDC, personal communication, July 2011). The initial quality control was limited to an elimination of clearly erroneous data (i.e. wrongly set nulls such as 999 instead of the correct missing data value of −999). When these appeared in the daily data, the monthly data were also excluded.

Precipitation datasets included in the study were: the Climate Research Unit's freely available CRU TS 3.10.01 climate data (in preparation at the time of writing; see Mitchell and Jones, 2005, for version 2.1), GPCC Full Data Reanalysis version 6 (Schneider et al., 2011) and both the CRU and the GPCC bias-corrected WATCH forcing data, from here on called $\mathrm{WATCH_{CRU}}$ and $\mathrm{WATCH_{GPCC}}$ (Weedon et al., 2011). Potential evaporation was available both from the CRU and the WATCH datasets. The CRU estimates are based on the FAO (Food and Agricultural Organization) grass reference-evaporation equation (Ekström et al., 2007; Allen et al., 1994), whereas the WATCH dataset provides both Penman–Monteith (Monteith, 1965) and Priestley–Taylor (Priestley and Taylor, 1972) estimates. Cells defined as land in the basin delineations, but not covered by the climate datasets, were set in an iterative manner to an average of the closest surrounding cells covered by the climate datasets until all land areas were covered. For the GIS-polygon dataset, the intersections with the half-degree climate-data grid cells were used to calculate the fraction of precipitation and potential evaporation of each cell contributing to the basin.

HESSD
10, 487–517, 2013

All gridded basin-delineation datasets covered the whole globe, whereas the polygon dataset only covered selected basins (Table 1). Only data within the common period of the climate datasets (1901–2001) were used in the analysis, and periods for individual basins varied depending on the discharge-data records.

3 Method

3.1 Co-registration of discharge stations

The first step was to determine how the different datasets can be connected to one another. The discharge stations had to be connected (co-registered) to the flow networks so that each station was assigned to the cell in the hydrography for which the flow-accumulation area (i.e. the sum of all upstream cell areas as defined by the flow network) best corresponded to the basin area in question. For DDM30, co-registrations of GRDC stations and the flow network were available for 1235 stations (Hunger and Döll, 2008) and for STN-30p for 663 stations (Fekete et al., 2002). GRDC stations were co-registered with the DRT hydrography in three steps. Firstly, each station was assigned to the cell corresponding to the coordinates in the GRDC database. Secondly, an automatic re-assignment was made if the flow-accumulation area of any of the closest eight surrounding cells better corresponded to the basin area reported by GRDC. And thirdly, the symmetric error, ε_{sym} (Fekete et al., 1999), was calculated according to Eq. (1). All stations exhibiting a symmetric error of more than 10% were manually inspected and re-assigned if possible.

$$\varepsilon_{sym} = \frac{A_{Acc} - A_{GRDC}}{\max(A_{Acc}, A_{GRDC})} \cdot 100\,\%, \quad (1)$$

where A_{Acc} is the flow-accumulation area of the assigned cell and A_{GRDC} is the GRDC basin area. Positive and negative symmetric errors mean that calculated basin areas are larger or smaller than the ones reported to GRDC. The inspection was done in

HESSD
10, 487–517, 2013

Google Earth by using online map resources and superimposing the flow network on a 1 : 10 000 000 river network (freely available from www.naturalearthdata.com).

All hydrography datasets were then analysed and compared in terms of how well they represented basin areas close to the ones reported in GRDC based on the symmetric errors.

3.2 Evaluation of consistency between forcing and evaluation data

Similarly to Beven et al. (2011), the basic method suggested in this study was to identify disinformative data as those that violate the conservation equation, i.e. the water balance. In contrast to their event-based approach we analysed the long-term water balance and also analysed data for transgressions of the energy limit, similar to Peel et al. (2010). The change in basin storage can safely be ignored for sufficiently long time periods, except for special cases such as melting of glaciers. The water-balance equation can then be simplified to:

$$P = E_A + R \tag{2}$$

where P is precipitation, E_A actual evaporation and R runoff. For natural basins, runoff should not exceed the precipitation input to the system. Actual evaporation equals the difference between precipitation and runoff and should not exceed the potential evaporation (E_P). These were the fundamental assumptions on which the consistency checks between the forcing data (precipitation and potential evaporation) and the evaluation data (discharge and actual evaporation) were based.

All datasets are affected by different types of uncertainties. Estimating them can be difficult because of lack of knowledge about their nature and magnitude, both temporally and spatially. There is a growing literature on quantification of uncertainties connected to hydrological modelling and McMillan et al. (2012) review the observational uncertainties of some key hydrological variables. In the present study, a relative uncertainty of ±10 % was assumed for the long-term discharge (resulting in a low, a high and

HESSD
10, 487–517, 2013

a best (i.e. the original data) estimate for each time series). Climate data were used as given in their original sources.

The runoff coefficient (RC), i.e. the quotient of runoff to precipitation, is a measure of how precipitation is partitioned into runoff and evaporation in a system. RCs are often calculated on an event basis and for specific surfaces, but can also be determined as a long-term response characteristic for a basin. Long-term RCs were calculated for low, high and best estimate discharge values, resulting in a high, a low, and a best estimate RC value for each basin. Hence, the first test of inconsistency between datasets, that runoff should not exceed precipitation input, stated that RCs should not be higher than one. In reality, a long-term basin RC even close to unity is implausible, but unity was used as a conservative threshold to avoid classifying datasets as inconsistent based on arbitrarily set RC thresholds. When based on low RC values, this threshold could be considered very conservative. In order to investigate when time periods were "sufficiently" long to determine long-term runoff coefficients, an initial analysis was performed of the variation of RCs with regards to record length. A subset (n = 37) was selected of the co-registered GRDC stations with complete data throughout the common period (1901.01–2001.12). For each record length (1 to 15 yr of consecutive data), each discharge record was randomly re-sampled 20 times (overlaps have occurred) and the runoff coefficient for each sample was calculated. For each basin and sample length, the individual RC estimates were divided by the median RC and plotted for all 37 basins (Fig. 1). It was assumed from the spread in the scatter plot that 10 yr of data should suffice to estimate the long-term runoff coefficients.

The discharge datasets analysed in this paper were not screened for anthropogenic influences (e.g. reservoirs and inter-basin transfers), which means that for some basins the water balance according to Eq. (2) could not be expected to be fulfilled.

HESSD
10, 487–517, 2013

4 Results

4.1 Hydrography representation of basin area

Of the 7763 stations available in the GRDC data archive, 245 stations were excluded from the study because of insufficient metadata records, i.e. missing coordinates or basin areas. The remaining 7518 stations with sufficient metadata were first registered in the DRT flow network following only the two automatic steps. Many stations in the database represented basins smaller than a 0.5° cell and clear systematic errors were noticed in this initial stage (Fig. 2). Since manual checking of station locations is very time-consuming, it was decided to limit the study to basins larger than 5000 km^2. This threshold is considerably smaller than those of previous studies but it still meant that most of the large symmetric errors for small areas were excluded. In total, there were 2177 stations in the archive with basins larger than 5000 km^2 for which there were daily data available. The remaining stations with symmetric errors larger than 10 % were subjected to the third, manual co-registration step. Despite this check, many stations could not be relocated to well-fitting cells and the symmetric errors remained large for some stations (Fig. 3b).

Of the stations co-registered in DRT, 558 were commonly co-registered in both DDM30 and STN-30p. The symmetric errors displayed a markedly larger scatter for STN-30p (standard deviation 14.3 %) and DRT (14.6 %) than for DDM30 (8.9 %) (Fig. 3a–c). None of the datasets showed any major tendency to over- or underestimate areas. There was little consistency in the errors between datasets except for a few largely over- and underestimated stations in DDM30 and STN-30p (Fig. 3d–e). Symmetric errors with absolute values over 70 % were observed for all hydrographies.

The GIS-polygon dataset was compared to the stations co-registered in DRT. Of these 2177 stations, 2005 were available in the GIS dataset. The GIS-polygon-based basin areas showed small errors compared to those of the gridded datasets, but some stations exhibited markedly large errors (Fig. 4a). As before, the errors showed little consistency between datasets (only comparison with DRT shown, Fig. 4b). Visual

HESSD
10, 487–517, 2013

inspection of the mapped area discrepancies of the datasets (not shown) revealed no geographical pattern for any of the datasets.

4.2 Consistency between datasets

Long-term runoff coefficients could be calculated for 1561 of the 2005 stations that were available in both the polygon dataset and DRT, given that there should be at least 10 yr of consecutive data. To minimise the effect of area discrepancies, results shown are based on the GIS-polygon basin delineation. The scatter plot of GPCC and CRU precipitation data (Fig. 5a), implies a higher relative difference in precipitation in drier basins. Runoff coefficients for the different precipitation datasets generally show higher relative discrepancies for high runoff coefficients, and implausibly high RCs were mainly found in areas with relatively low precipitation (Fig. 5b and c). The general distributions of RCs did not differ much between the precipitation datasets, and implausibly high runoff coefficients were found for all four even when using the low discharge estimate (Fig. 6). However, RCs larger than one were more common for the CRU and $\mathrm{WATCH}_{\mathrm{CRU}}$ precipitation datasets than for the other two. Basins with high runoff coefficients were almost exclusively located in high-latitude or high-altitude areas (Fig. 7). A particular feature that warrants further study is that the majority of the high-latitude basins with RCs exceeding unity were found on the US side of the border between Alaska and Canada.

The second data-consistency test, that actual evaporation, given as a residual in Eq. (1), should not exceed potential evaporation, was analysed graphically. Calculated actual evaporation was plotted against potential evaporation (a simplified version of the Budyko, 1974, curve) for all combinations of precipitation and potential-evaporation datasets. The geographical patterns were similar for all combinations (Fig. 8 and Table 2 exemplify the results for the CRU precipitation). Uncertainty in the runoff is represented in the colour coding where red represents basins which exhibit evaporation (P-R) higher than the energy limit values (potential evaporation) even for the high discharge estimate (i.e. when the calculated actual evaporation is the lowest, E_{AL}) and

HESSD
10, 487–517, 2013

orange represents basins with losses higher than the energy limit values for discharge estimates between the low (i.e. when the estimated actual evaporation is the highest, E_{AH}) and the high values. One noticeable difference between the different potential-evaporation datasets was the greater frequency of basins exhibiting actual evaporation values higher than the potential evaporation estimates for the two WATCH datasets compared to the CRU dataset. Implausibly high actual evaporation frequently appeared in the Amazon basin for all three datasets, and for the two WATCH datasets also on the east coast of North America, Europe, equatorial Africa and South East Asia. Blue dots in Fig. 8 indicate basins for which the actual evaporation was negative (i.e. RC > 1) for both the high and low discharge estimates and green dots where this occurred only for the low estimates of actual evaporation. The proportion of stations with too-high evaporation and implausibly high RCs were similar for all basin sizes (Fig. 9).

5 Discussion

5.1 Hydrography representation of basin area

A correct basin area is a prerequisite for a correct water balance. The discrepancies between basin-area estimates in the different hydrographies and the area reported in the GRDC database are likely attributable both to deficiencies in the basin representations, and to varying quality of the GRDC metadata. The accuracy of the areas given in the archive is not reported by the different data providers (Ulrich Looser, head of GRDC, personal communication, October 2011). The larger scatter observed for STN-30p and DRT compared to DDM30 can likely be explained by the extensive manual corrections of the flow network (35 % of all cells) performed on the latter (Döll and Lehner, 2002). DDM30 outperforms both STN-30p and DRT in representing basin areas close to the ones reported in GRDC (at least for basins larger than 10 000 km^2). The advantage of using DRT over the other two gridded hydrographies would be the possibility of using the high-resolution baseline for deriving topological basin information.

HESSD
10, 487–517, 2013

Title Page
Abstract Introduction
Conclusions References
Tables Figures
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Some of the basin areas reported in the GRDC archive in June 2011 are likely to be different to the ones reported at the time of the collection of data for co-registration with STN-30p and DDM30 since the GRDC archives have been continuously updated. A comparison of Fig. 5 in Döll and Lehner (2002) and Fig. 3 in this paper showed that at least a few basin areas must have been updated since no absolute symmetric errors above 30 % were reported for the DDM30 stations. Changes in the reported areas were also found between October 2010 (the time of data collection for the GIS polygon dataset) and the time when data were collected for this study. In many cases the changes were small, but increases in basin area even over 100 % were noted for a few stations.

The GIS-polygon delineation of basins matched basin areas in the GRDC archive very well in most cases, but there were some clear discrepancies. Given the extensive manual checks to verify station locations and basin areas during the development of the dataset, it can be argued that the GIS dataset is more likely correct in case of discrepancy. The area discrepancies showed no geographical pattern, even though the GIS dataset is based on a coarser hydrography above 60° N and therefore could be expected to perform worse in those areas.

Among the 2005 stations common between the GIS dataset and the stations co-registered with DRT, 584 stations had a basin area of 10 000 km^2 or less. Of those, 80 % exhibited symmetric errors with absolute values less than 25 % in the gridded hydrography compared to 92 % in the GIS delineation. Corresponding figures for symmetric errors below 10 % were 45 and 84 %, respectively. Hence, many small catchments were well represented even in the 0.5° grid.

5.2 Consistency between datasets

Runoff coefficients greater than unity have been encountered in several global studies (Fekete et al., 2002; Widén-Nilsson et al., 2009; Peel et al., 2010). There could be several reasons for this data mismatch: precipitation underestimation because of poor spatial and temporal resolution and measurement errors, discharge-data uncertainties,

HESSD
10, 487–517, 2013
Disinformative data in large-scale hydrological modelling
A. Kauffeldt et al.

or a failure to identify anthropogenic or subsurface inter-basin transfers (Peel et al., 2010). In addition, poor representation of the basin in the up-scaled hydrography could lead to a mismatch. However, the effect of hydrography errors should be small for most polygon-delineated basins. It was found that the vast majority of the basins with implausible runoff coefficients were located in areas where underestimation of precipitation was likely caused by snowfall undercatch. Wind-induced solid precipitation undercatch can have a substantial effect on precipitation totals in high-latitude areas (Adam and Lettenmaier, 2003). The geographical patterns were similar for all precipitation datasets.

The evaluation of actual and potential evaporation pointed to further inconsistencies between climate and discharge data. Database inconsistencies led to transgression of the energy limit (i.e. $E_A > E_P$) in many basins. Peel et al. (2010) report similar issues when analysing a large set ($n = 861$) of basins worldwide using observed station records rather than gridded data. This could be possible for individual basins by considering e.g. irrigation and inter-basin transfers, not accounted for in this study. However, the clear geographical patterns found in this study indicated that there were whole regions such as the Amazon basin where the inconsistencies were likely a result of systematic problems in the climate data. These problems were more abundant and appeared in more regions when potential evaporation from the WATCH rather than the CRU dataset was used. As many as 8–43 % of the basins exhibited inconsistencies for the best runoff estimate depending on how the datasets were combined. The corresponding figures were 6–35 % when accounting very conservatively for discharge uncertainties and only counting basins falling completely outside the physically reasonable limits.

These violations to fundamental consistency assumptions pose a serious problem for model calibration and evaluation (Beven et al., 2011; Beven and Westerberg, 2011), and could cause severe bias in a subsequent model regionalisation. Depending on the evaluation criteria in calibration, some of these problems could go unnoticed, but model-parameter values would be biased as a result of such long-term inconsistencies.

HESSD
10, 487–517, 2013

It should however be noted that there might be shorter periods of informative data in a dataset even if the long-term averages are disinformative. Similarly, datasets found to be consistent in this analysis might contain data that are disinformative on shorter time scales. Methods for reliable identification of inconsistent events at shorter time scales for these large spatial-scale datasets therefore also need to be developed.

6 Concluding remarks

This study has demonstrated that a pre-modelling data analysis should be an important first step in a large-scale hydrological study. Scrutinising input and evaluation data is vital to reveal inconsistencies between datasets and to highlight basins where one should be cautious when making model inferences based on these data.

– A majority of basin areas larger than 5000 km^2 could be represented in a $0.5^\circ \times 0.5^\circ$ longitude-latitude grid (absolute symmetric error $\leq 25\,\%$). The GIS-polygon delineation derived from a high-resolution hydrographic baseline outperformed all gridded delineations in this study.

– There were inconsistencies between the climate datasets and observed discharge. It was hypothesised that these, because of their clear spatial patterns, were mainly caused by limitations in the forcing/evaluation data. Some inconsistencies could also be caused by anthropogenic influences not considered in this article (e.g. inter-basin transfers, irrigation and reservoirs).

In light of the first point, it could be argued that global hydrological models should use polygon-based basin delineations rather than being limited by input-data resolution, as is common today. This is especially true since large area discrepancies in coarse flow networks can compensate (or aggravate) precipitation-input deficiencies. Even if a model can perform well in such basins, it might be for the wrong reasons. However, this would require development of polygon delineations with global coverage.

HESSD
10, 487–517, 2013

In light of the second point it should be emphasised that inconsistencies between datasets were frequent. Further modelling studies will be required to find out the reasons for these inconsistencies and how they affect model inference. However, a model-independent data analysis, such as the one presented in this study, is a useful tool for identifying and analysing inconsistent datasets – therefore enabling more robust conclusions to be drawn in subsequent hydrological modelling and analyses (Juston et al., 2012).

Acknowledgement. This study has been heavily dependent on the provision of global datasets provided by several groups mentioned in the data section and we would like to thank all the groups that have been involved in the development of those datasets. Special thanks are extended to Huan Wu who provided the early version of the DRT dataset, to Petra Döll who provided the DDM30 dataset, and to the helpful staff at GRDC.

References

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res., 108, 4257, doi:10.1029/2002JD002499, 2003.

Allen, R. G., Smith, M., Pereira, L. S., and Perrier, A.: An update for the calculation of reference evapotranspiration, ICID Bull., 43, 35–92, 1994.

Arnell, N. W.: A simple water balance model for the simulation of streamflow over a large geographic domain, J. Hydrol., 217, 314–335, 1999.

Beven, K. and Westerberg, I.: On red herrings and real herrings: disinformation and information in hydrological inference, Hydrol. Process., 25, 1676–1680, 2011.

Beven, K., Smith, P. J., and Wood, A.: On the colour and spin of epistemic error (and what we might do about it), Hydrol. Earth Syst. Sci., 15, 3123–3133, doi:10.5194/hess-15-3123-2011, 2011.

Budyko, M. I.: Climate and Life, International Geophysics Series No. 18, edited by: Miller, D. H., Academic Press, London, 508 pp., 1974.

Döll, P. and Lehner, B.: Validation of a new global 30-min drainage direction map, J. Hydrol., 258, 214–231, 2002.

Döll, P. and Siebert, S.: Global modeling of irrigation water requirements, Water Resour. Res., 38, 1037–1045, 2002.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, J. Hydrol., 270, 105–134, 2003.

Ekström, M., Jones, P. D., Fowler, H. J., Lenderink, G., Buishand, T. A., and Conway, D.: Regional climate model data used within the SWURVE project – 1: projected changes in seasonal patterns and estimation of PET, Hydrol. Earth Syst. Sci., 11, 1069–1083, doi:10.5194/hess-11-1069-2007, 2007.

Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: Global composite runoff fields of observed river discharge and simulated water balances, Tech. Rep. No. 22, Global Runoff Data Centre, Koblenz, Germany, 39 pp. plus annex, 1999.

Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochem. Cy., 16, 1042, doi:10.1029/1999GB001254, 2002.

Gong, L., Halldin, S., and Xu, C.-Y.: Global-scale river routing – an efficient time-delay algorithm based on HydroSHEDS high-resolution hydrography, Hydrol. Process., 25, 1114–1128, doi:10.1002/HYP.7795, 2011.

Güntner, A.: Improvement of global hydrological models using GRACE data, Surv. Geophys., 29, 375–397, 2008.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K.: An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing, Hydrol. Earth Syst. Sci., 12, 1007–1025, doi:10.5194/hess-12-1007-2008, 2008.

Hunger, M. and Döll, P.: Value of river discharge data for global-scale hydrological modeling, Hydrol. Earth Syst. Sci., 12, 841–861, doi:10.5194/hess-12-841-2008, 2008.

Jung, G., Wagner, S., and Kunstmann, H.: Joint climate-hydrology modeling: an impact study for the data-sparse environment of the Volta Basin in West Africa, Hydrol. Res., 43, 231–248, 2012.

Juston, J. M., Kauffeldt, A., Quesada Montano, B., Seibert, J., Beven, K. and Westerberg, I.: Smiling in the rain: Seven reasons to be positive about uncertainty in hydrological modelling, Hydrol. Process., doi:10.1002/hyp.9625, 2012.

Kaspar, F.: Entwicklung und unsicherheitanalyse eines globalen hydrologischen modells, Ph.D. Thesis, Department of Natural Sciences, Kassel University, Kassel, Germany, 139 pp., 2004.

Kuczera, G.: Correlated rating curve error in flood frequency inference, Water Resour. Res., 32, 2119–2127, 1996.

Lehner, B.: Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network, Tech. Rep. No. 41, Global Runoff Data Centre, Koblenz, Germany, 12 pp., 2011.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, EOS Trans. AGU, 89, 93, 2008.

Lu, L., Ngongondo, C. S., Xu, C.-Y., and Gong, L.: Comparison of the global TRMM and WFD precipitation datasets in driving a large-scale hydrological model in Southern Africa, Hydrol. Res., doi:10.2166/NH.2012.175, in press, 2012.

McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol. Process., 26, 4078–4111, doi:10.1002/HYP.9384, 2012.

Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly climate observations and associated high-resolution grids, Int. J. Climatol., 25, 693–712, 2005.

Montanari, A. and Di Baldassarre, G.: Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty, Adv. Water Resour., doi:10.1016/j.advwatres.2012.09.007, in press, 2012.

Monteith, J. L.: Evaporation and environment, Sym. Soc. Exp. Bio., 19, 205–234, 1965.

Nijssen, B., O'Donnell, G., Hamlet, A., and Lettenmaier, D.: Hydrologic sensitivity of global rivers to climate change, Clim. Change, 50, 143–175, 2001.

Palmer, M. A., Reidy Liermann, C. A., Nilsson, C., Flörke, M., Alcamo, J., Lake, P. S., and Bond, N.: Climate change and the world's river basins: anticipating management options, Front. Ecol. Environ., 6, 81–89, 2008.

Peel, M. C., McMahon, T. A., and Finlayson, B. L.: Vegetation impact on mean annual evapotranspiration at a global catchment scale, Water Resour. Res., 46, W09508, doi:10.1029/2009WR008233, 2010.

Priestley, C. H. B. and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100, 81–92, 1972.

HESSD
10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data, doi:10.5676/DWD_GPCC/FD_M_V6_050, 2011.
Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., and Srikanthan, S.: Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: a case study using Bayesian total error analysis, Water Resour. Res., 45, W00b14, doi:10.1029/2008WR006825, 2009.
USGS EROS: HYDRO1k Elevation Derivative Database: http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro (last access: 30 September 2012), 1996.
Vörösmarty, C. J., Moore III, B., Grace, A. L., Gildea, M. P., Melillo, J. M., Peterson, B. J., Rastetter, E. B., and Steudler, P. A.: Continental scale models of water balance and fluvial transport: an application to South America, Global Biogeochem. Cy., 3, 241–265, 1989.
Vörösmarty, C. J., Fekete, B. M., Meybeck, M., and Lammers, R. B.: Global system of rivers: its role in organizing continental land mass and defining land-to-ocean linkages, Global Biogeochem. Cy., 14, 599–621, 2000.
Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, W00B09, doi:10.1029/2007WR006720, 2008.
Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, J. Hydrometeorol., 12, 823–848, 2011.
Werth, S. and Güntner, A.: Calibration analysis for water storage variability of the global hydrological model WGHM, Hydrol. Earth Syst. Sci., 14, 59–78, doi:10.5194/hess-14-59-2010, 2010.
Widén-Nilsson, E., Halldin, S., and Xu, C.-Y.: Global water-balance modelling with WASMOD-M: parameter estimation and regionalisation, J. Hydrol., 340, 105–118, 2007.
Widén-Nilsson, E., Gong, L., Halldin, S., and Xu, C.-Y.: Model performance and parameter behavior for varying time aggregations and evaluation criteria in the WASMOD-M global water balance model, Water Resour. Res., 45, W05418, doi:10.1029/2007WR006695, 2009.
Wu, H., Kimball, J. S., Mantua, N., and Stanford, J.: Automated upscaling of river networks for macroscale hydrological modeling, Water Resour. Res., 47, W03517, doi:10.1029/2009WR008871, 2011.

HESSD
10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.

Wu, H., Kimball, J. S., Li, H., Huang, M., Leung, L. R., and Adler, R. F.: A new global river network database for macroscale hydrologic modeling, Water Resour. Res., 48, W09701, doi:0.1029/2012WR012313, 2012.

HESSD

10, 487–517, 2013

Table 1. Dataset summary.

Dataset	Temporal resolution	Spatial resolution	Coverage
Basin delineation			
DRT (Wu et al., 2011)	N/A	0.5°	Global
GIS polygons (Lehner, 2011)	N/A	$\sim 15''$	Selected stations
STN-30p (Vörösmarty et al., 2000)	N/A	0.5°	Global
DDM30 (Döll and Lehner, 2002)	N/A	0.5°	Global
Precipitation			
CRU TS 3.10.01 (Mitchell and Jones, 2005)	Monthly	0.5°	Global/1901–2009
$\text{WATCH}_{\text{CRU}}$ (Weedon et al., 2011)	Daily	0.5°	Global/1901–2001
$\text{WATCH}_{\text{GPCC}}$ (Weedon et al., 2011)	Daily	0.5°	Global/1901–2001
GPCC v6 (Schneider et al., 2011)	Monthly	0.5°	Global/1901–2010
Potential evaporation			
CRU TS 3.10.01 (Mitchell and Jones, 2005)	Monthly	0.5°	Global/1901–2009
WATCH Penman–Monteith (Weedon et al., 2011)	Daily	0.5°	Global/1901–2001
WATCH Priestley–Taylor (Weedon et al., 2011)	Daily	0.5°	Global/1901–2001

HESSD
10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.

HESSD
10, 487–517, 2013

Table 2. Percent of basins exhibiting potential data inconsistencies. Each basin is only accounted for in the worst category that applies to it, e.g. a basin for which the lowest of the actual evaporation estimates exceed the potential evaporation is accounted for in column $E_{AL} > E_P$ but not $E_{AH} > E_P$.

Precipitation	Potential evaporation	No remark	$E_{AL} > E_P$	$E_{AH} > E_P$	$P - R_L < 0$	$P - R_H < 0$
CRU	CRU	85.6	4.5	2.9	3.6	3.4
	$WATCH_{PM}$	71.7	12.2	9.1	3.6	3.4
	$WATCH_{PT}$	62.6	19.8	10.6	3.6	3.4

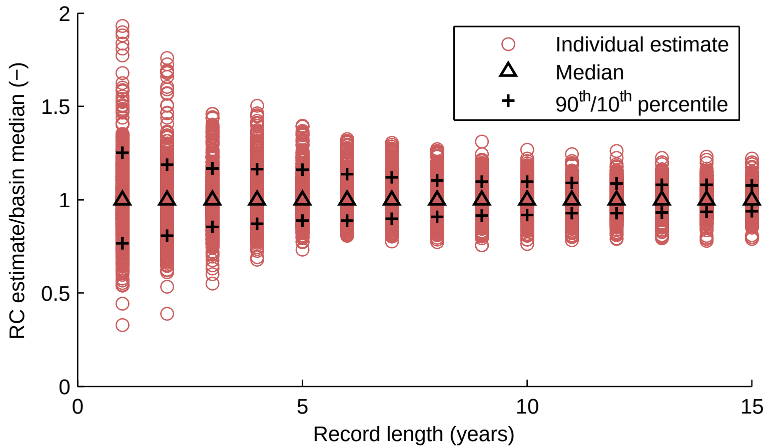

Fig. 1. Variation of runoff coefficient estimates as a function of record length summarised for 37 basins with long data records. Estimates are standardised by division with the basin median for a given record length.

HESSD
10, 487–517, 2013

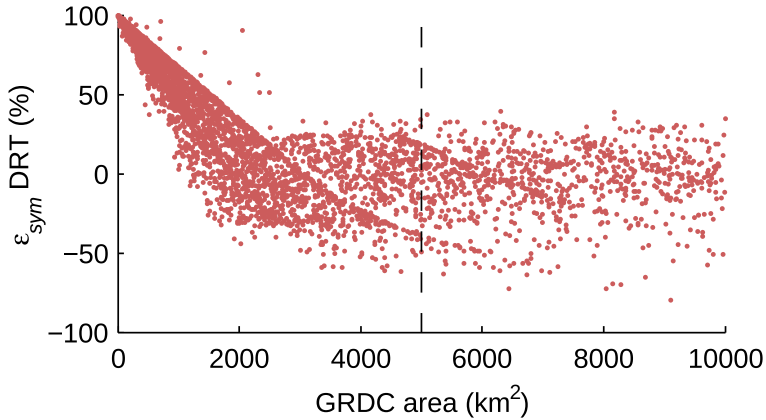

Fig. 2. Symmetric error after automatic relocation for all 7518 GRDC stations with sufficient metadata.

HESSD
10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.

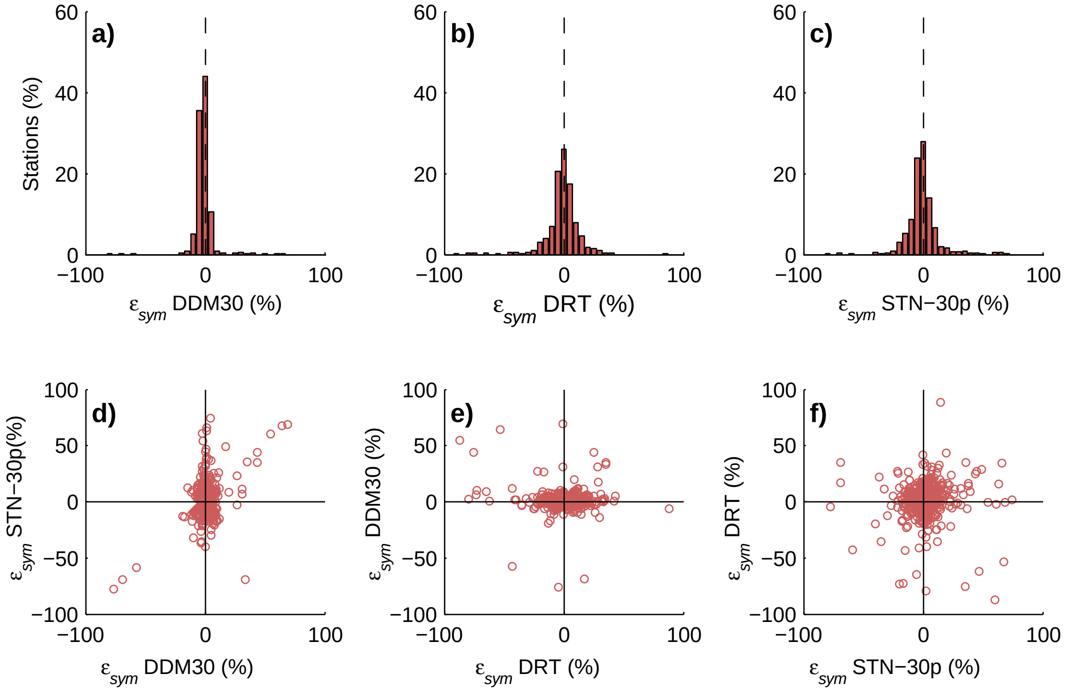

Fig. 3. Histograms of symmetric errors for 558 basins with stations registered in the three gridded flow networks: **(a)** DDM30, **(b)** DRT and **(c)** STN-30p. The lower panel shows comparisons of the symmetric errors for each basin in the different flow networks.

HESSD
10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.

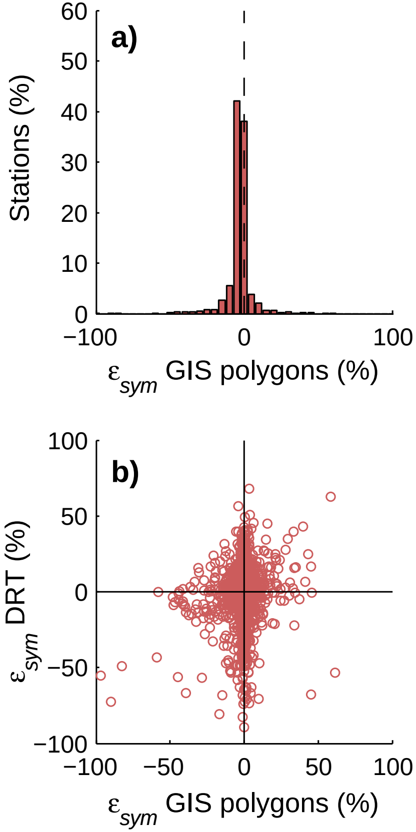

Fig. 4. (a) Symmetric errors for 2005 basins based on GIS-polygon definition and **(b)** comparison to symmetric errors exhibited in DRT.

HESSD
10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.

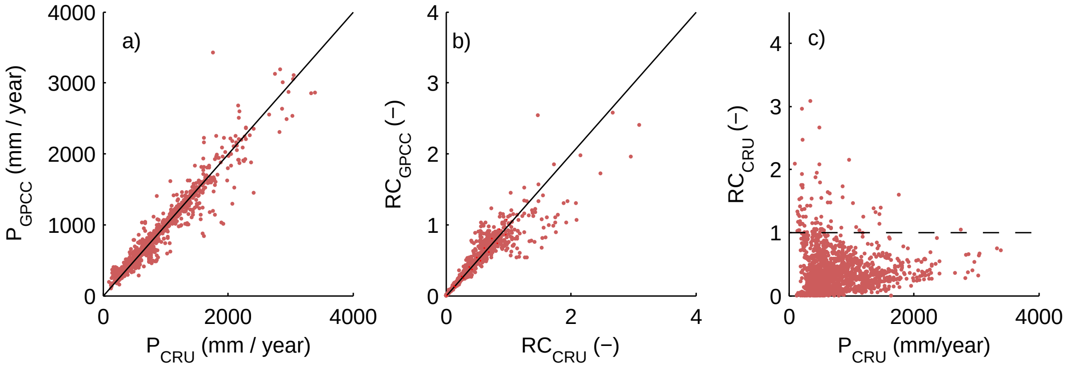

Fig. 5. **(a)** Relation between GPCC and CRU precipitation, **(b)** best RC estimates based on GPCC precipitation versus CRU precipitation (RCs over 1 not shown), and **(c)** best RC estimates based on CRU precipitation versus CRU precipitation. The 45° lines indicate 1 : 1 quotient and the dashed line indicate RC = 1.

HESSD
10, 487–517, 2013

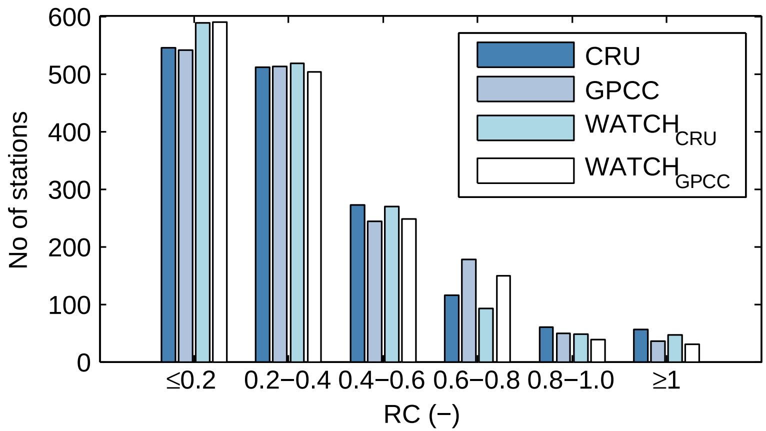

Fig. 6. Distribution of low-estimate RCs for the four precipitation datasets.

HESSD
10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.

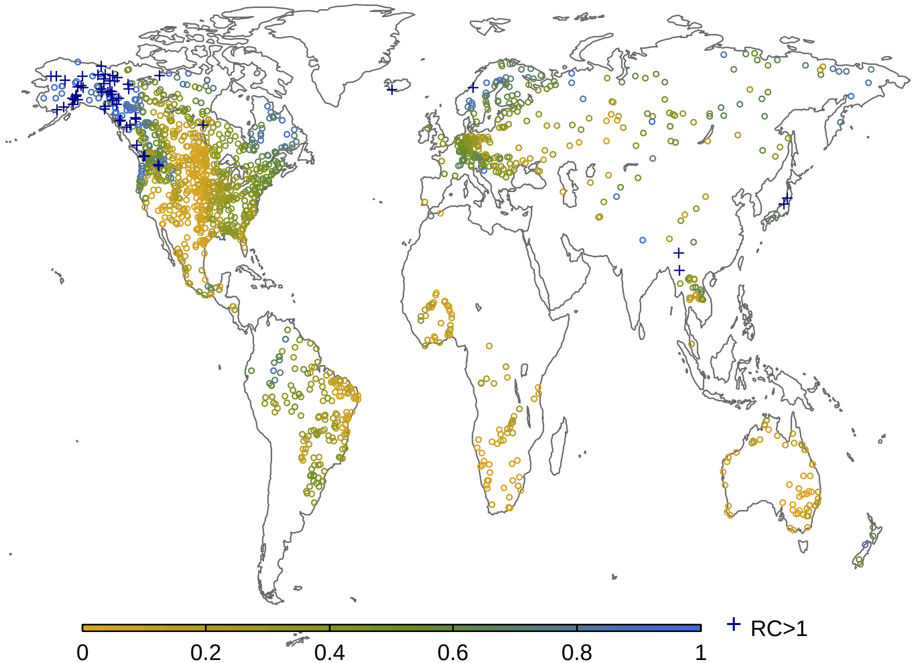

Fig. 7. Spatial pattern of runoff coefficients for CRU precipitation. Circles represent best RC estimates and crosses represent basins with low RC estimates higher than one.

HESSD
10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.

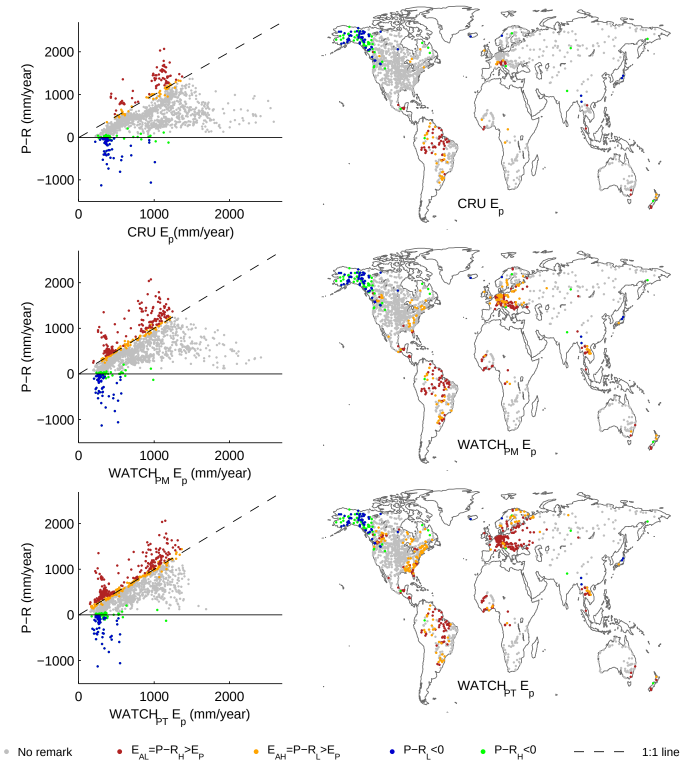

Fig. 8. Mean annual actual evaporation (estimated as P-R using CRU precipitation data) versus potential evaporation from CRU, WATCH Penman–Monteith and WATCH Priestley–Taylor (left panel). Potential evaporation is plotted against actual evaporation estimated using the best runoff estimate, i.e. the y-value of each dot represents the best evaporation estimate. The colour coding is based on the high runoff estimate (R_H, giving low estimate of E_A) and low runoff estimate (R_L, giving high estimate of E_A) as indicated in the legend. The right panel 190 shows geographical distributions.

HESSD
10, 487–517, 2013

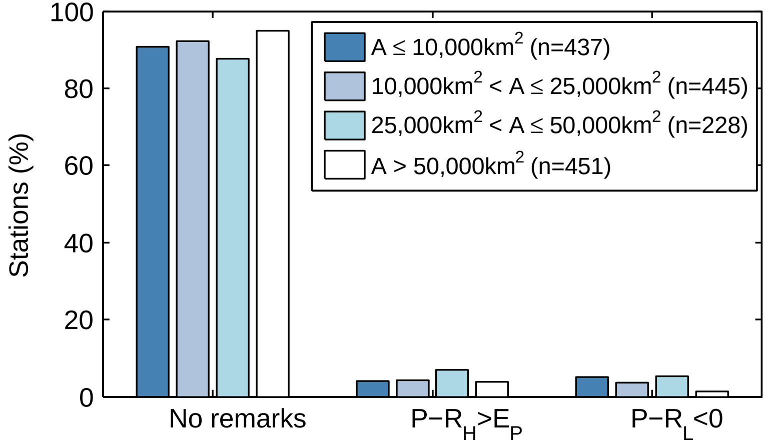

Fig. 9. Basins with and without data inconsistencies, based on CRU precipitation and potential evaporation, in different area categories.

HESSD

10, 487–517, 2013

Disinformative data in large-scale hydrological modelling

A. Kauffeldt et al.