Hydrol. Earth Syst. Sci. Discuss., 9, 6829–6856, 2012
www.hydrol-earth-syst-sci-discuss.net/9/6829/2012/
doi:10.5194/hessd-9-6829-2012

Hydrology and Earth System Sciences Discussions

Data-based discharge extrapolation: estimating annual discharge for a partially gauged large river basin from its small sub-basins

L. Gong

Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, Sweden

Bert Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

Department of Geosciences, University of Oslo, Oslo, Norway

Received: 29 March 2012 – Accepted: 17 May 2012 – Published: 1 June 2012

Correspondence to: L. Gong (lebing@gmail.com)

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

Abstract

Large-scale hydrological models and land surface models are by far the only tools for accessing future water resources in climate change impact studies. Those models estimate discharge with large uncertainties, due to the complex interaction between climate and hydrology, the limited quality and availability of data, as well as model uncertainties. A new purely data-based scale-extrapolation method is proposed, to estimate water resources for a large basin solely from selected small sub-basins, which are typically two-orders-of-magnitude smaller than the large basin. Those small sub-basins contain sufficient information, not only on climate and land surface, but also on hydrological characteristics for the large basin In the Baltic Sea drainage basin, best discharge estimation for the gauged area was achieved with sub-basins that cover 2–4 % of the gauged area. There exist multiple sets of sub-basins that resemble the climate and hydrology of the basin equally well. Those multiple sets estimate annual discharge for gauged area consistently well with 5 % average error. The scale-extrapolation method is completely data-based; therefore it does not force any modelling error into the prediction. The multiple predictions are expected to bracket the inherent variations and uncertainties of the climate and hydrology of the basin. The method can be applied in both un-gauged basins and un-gauged periods with uncertainty estimation.

1 Introduction

The interests in understanding the impact of climate change to future water recourses have driven the rapid development of large-scale hydrological models (e.g. Arnell, 1999, 2003, 2004; Vörösmarty et al., 1989, 2000a, 2004). Those models are heavily data-dependent; however, input and calibration data for large river basins are usually associated with large uncertainties, which gives rise to uncertainties in model estimations. Several data-related issues include: (1) large percentage of the land surface are

HESSD
9, 6829–6856, 2012

un-gauged or poorly gauged; (2) the number of meteorological observations and discharge gauges are decreasing around the world; (3) available climate and hydrological data mismatch in time; (4) river regulations change the seasonality of observed discharge and make it difficult to calibrate hydrological models. Water resource projections made by large-scale models are an important basis for socio-economical analyses and decision-making processes (e.g. Vörösmarty et al., 2000a). Projections of water resources are believed to be associated with large uncertainty, especially in un-gauged basins that cover around 50 % of the global land area. For instance, Global runoff estimates from various models differ between 29 000 $km^3 yr^{-1}$ and 43 000 $km^3 yr^{-1}$, i.e. around 30 %, and continental estimates differ up to 70 % (Widén-Nilsson et al., 2007). Besides uncertainty in the forcing (climate) and the validation (discharge) data, model uncertainties also contribute significantly to the uncertainties of simulated discharge (Widén-Nilsson et al., 2009). For instance, large-scale hydrological models formulate essential processes with effective, i.e. non-physically-based parameter values to bypass the difficulty of describing the large-scale dynamic. A number of regionalisation methods were developed to extend the prediction capability of hydrological models into the un-gauged area. Commonly used regionalisation methods utilise spatial proximity and catchment similarity to transfer model parameters from gauged to un-gauged basins (e.g. Kokkonen et al., 2003; Huang et al., 2003; Xu, 1999, 2003; Kim and Kaluarachchi, 2008; McIntyre et al., 2005). Model averaging, i.e. average of model outputs from different proximity or similarity approaches was found to provide more robust results in regionalisation (e.g. McIntyre et al., 2005). Hydrological models inherently have limited parameter transferability over different spatial scales; therefore regionalisation methods use gauged large river basins as potential donors so that the parameter sets are transferable. Transferring of parameter sets is based on similarity analysis of various averaged basin characteristics. However, averaged basin characteristics often cannot sufficiently summarise the spatial variability and nonlinearity of the large basin, which in turn, limits the prediction accuracy of the regionalisation methods. In recent years, small river basins are shown to be useful in large-scale hydrological applications.

For instance, the European Water Archive dataset, which consists of streamflow data from hundreds of small basins, was used extensively in mapping the hydrology of Europe, especially as a mean to validate large-scale models (e.g. Stahl et al., 2011). This paper utilises a few small sub-basins to mimic the average water and energy balance of the whole basin; therefore average discharge of those sub-basins can be extrapolated to estimate the average basin discharge. In the following sections, the study area and datasets were presented, followed by the description of the scale-extrapolation method; results of both theoretical tests and the case study in the Baltic Sea drainage basin were then presented, and finally potential future applications of the method was discussed.

2 Study area and data

2.1 The Baltic Sea drainage basin

The scale-extrapolation was tested in the Baltic Sea drainage basin (Fig. 1a). The Baltic Sea is one of the largest brackish seas in the world; the Baltic Sea drainage basin lies between maritime temperate and continental subarctic climate zones. With a surface area of 415 000 km^2, The drainage basin spans 14 countries and 85 million people, a majority of them living in big cities like St. Petersburg, Copenhagen, Helsinki, Tallinn, Riga, Vilnius, Warsaw, and Stockholm. The Baltic Sea is semi-enclosed; therefore, it is vulnerable to pollutions; its environmental status is one of the major concerns for the Northern European countries. The Baltic Sea is affected by various sources including fresh water and nutrients input from rivers, pollutions from industries, and direct atmospheric depositions (Wulff et al., 2001). Many of these factors are dependent on the climate and hydrology in the basin.

HESSD
9, 6829–6856, 2012

2.2 Datasets

Monthly precipitation for the period of 1975–2001 was taken from the 30-min monthly CRU TS 2.1 database (Mitchell and Jones, 2005). The number of stations used by the CRU TS 2.1 dataset has significant temporal variations (Mitchell and Jones, 2005). Spatial density of CRU Precipitation stations decreased in the Baltic Sea drainage basin after 1990. Monthly precipitation data from 1984 SHMI (Sweden's Meteorological and Hydrological Institute) precipitation stations for the period of 1961–2002 were interpolated to a regular 30-min grid, and the quality of the CRU precipitation data was validated against the SHMI data prior to the analysis. Results (figure now shown) showed that the spatial differences between CRU and SHMI annual average precipitation were similar for the period of 1961–1990 and 1991–2002. In another word, the reduction of precipitation stations after 1990 has not significantly changed the long-term mean precipitation and its spatial distribution. Differences between 1991–2002 and 1961–1990 mean annual precipitation as calculated by CRU data and SMHI data also agreed well in their general spatial pattern, although those calculated with SMHI data showed much higher spatial variability at smaller scales.

WATCH forcing data (WFD; Weedon et al., 2010) for the period between 1975 and 2001 at 30-min spatial resolution were used to derive potential evaporation. The WFD provides bias-corrected variables based on the ERA-40 reanalysis product of the European Centre for Medium Range Weather Forecasting (ECMWF) as described by Uppala et al. (2005). Specific humidity, atmospheric pressure, 2-m air temperature, 10-m wind speed, downward shortwave radiation and net long-wave radiation were used to calculate reference evaporation using the Penman-Monteith FAO-56 equation (Allen et al., 1998). Specific humidity was first converted to relative humidity using a mixing-ratio method, and 10-m wind speed was converted to 2-m wind speed using a logarithmic relationship (Allen et al., 1998). Prior to the calculation of reference evaporation, the quality of the WFD air temperature, wind speed, and WFD derived relative humidity was tested in a comparison with daily weather data (Global Surface Summary

HESSD
9, 6829–6856, 2012

of the Day, or GSOD) from the National Climatic Data Center (NCDC, 2011). In the Penman-Monteith FAO-56 equation, surface albedo is fixed at 0.23; however we found this value is too high for the Baltic Sea basin. Therefore, the albedo values were taken directly from the ERA-Interim dataset (Simmons et al., 2007). The daily WATCH forcing data were aggregated to obtain annual values for each 30-min grid cell. The monthly CRU precipitation data were also aggregated to annual.

STN-30P dataset (Vörösmarty et al., 2000b) was used to identify 1386 cells on a regular 30-min global grid that belong to the Baltic Sea drainage basin. HYDRO1k (USGS, 1996) was used to delineate the upstream area of the discharge stations. The discharge data were taken from the Global Runoff Data Centre database (GRDC, 2012), and the SHMI Vatten Web (http://vattenweb.smhi.se/). Among 425 available sub-basins, 100 sub-basins were selected under the following criteria: (1) do not contain nested sub-basins; (2) when registered in the Hydro1k river network, the register area does not differ more than 20 % with the reported area from GRDC or SMHI; (3) have complete daily data coverage from 1975 to 2001. Figure 1a shows the location of the 100 sub-basins. The sizes of the sub-basins vary between 5 to 109 564 km^2.

3 Theories

This paper proposes a novel way of hydrological extrapolation, without the involvement of modelling, but purely with available data. The extrapolation is based on the hydrological similarity across spatial scales. Hydrological characteristics are dictated by a number of controlling factors. If regions A and B share the same controlling factors, they shall exhibit same hydrological characteristics. In another word, the hydrological characteristics of Region B, if unknown, can be estimated by Region A.

The controlling factors change with the time scale on which the hydrological characteristics are measured. On annual or seasonal time scale, it has long been recognised that the interaction between climate and hydrology controls the nonlinear partitioning of precipitation (e.g. L'vovich, 1979; Budyko, 1974; Wagener et al., 2007). L'vovich (1979)

HESSD
9, 6829–6856, 2012

and Budyko (1974) are among the first to characterized climate and hydrology using long-term average water and energy balance variables. The aridity index, as expressed by the ratio of long term average potential evapotranspiration over that of precipitation, has long been used as a useful index describing the interaction between climate and hydrology of a region (e.g. Wagener et al., 2007). When time scale is shorter than a month, or when statistics and extremes of discharge are of interests, land surface characteristics, for instance, river transport delays (e.g. Gong et al., 2009), soil storage capacity (e.g. Gong et al., 2011), etc., shall also be considered.

It is hard to find two large river basins that share the same climate and land surface characteristics. Therefore, the scale-extrapolation method allows Region A, to be spatially discrete and two-orders-of-magnitude smaller than Region B. The scale-extrapolation method uses a number of small regions, each of which overlaps with a gauged small river basin, to summaries the average dynamic for the large region B. Those small regions, selected by Monte-Carlo simulations, have mutually independent climate patterns. The scale-extrapolation method builds upon a multitude of data-ready small river basins that cover a large range of climate patterns and land surface types. Using small river basins has the following advantages:

1. Maximize the predictability

   Both climate and hydrology exhibit larger spatial variability at smaller scales. A large spectrum of climate patterns can be obtained by combining several small basins. The abundance and diversity of small river basins ensure good matching of climate variations in any un-gauged large river basin of concern; therefore increases the chance of reliable discharge extrapolation.

2. Robust ensemble estimation

   Except for data-sparse regions, there are usually abundant small-gauged river basins that can be used as candidates for extrapolation. A large number of equally good extrapolations from different sets of small basins offer statistically robust uncertainty estimation.

HESSD
9, 6829–6856, 2012

3. Preserve the nonlinearity

   The hydrological system is nonlinear. One prominent feature of a spatially distributed nonlinear system is that the average of input does not lead to the average of output. Similarly, if a single small basin shares similar climate pattern with a large basin, it usually does not well resemble the hydrology of the large basin. Theoretical tests and experiment with observational data in the following sections would show that a number of small sub-basins, although only count for less than 5 % areal percentage of a large basin, are able to account for the nonlinearity of the system, and therefore can be used to estimate the hydrology of the large basin accurately.

The concept of the scale-extrapolation shall be distinguished with several similar but different concepts:

1. Spatial interpolation and extrapolation, as in geostatistics

   Spatial interpolation and extrapolation commonly refer to methods estimate a variable at unobserved point or area based on the values at observed locations. The scale-extrapolation method refers to the use of river discharge from a number of small river basins to represent the discharge of a large basin or region, only if the small river basins, when weighted combined, can well resemble both climate and land surface characteristics of the large region.

2. Self-similarity

   River system shows strong self-similarity. The scale-extrapolation method uses instead of one, but several small river basins to provide an estimation for a large region.

3. Up-scaling

   "Up-scaling" is defined as transforming information from small scale to large scale. Up-scaling is usually achieved by aggregating information at small scale either numerically or analytical, in order to summaries the heterogeneity and to maintain

HESSD
9, 6829–6856, 2012

Data-based discharge extrapolation

L. Gong

the nonlinearity of the system at larger scales. Similarly, the scale-extrapolation method also aims at preserving the nonlinearity of the system, by only using a small part of the spatial dataset.

4 Method

4.1 Definitions of gauged basin area and active sub-basins

The area covered with the 100 gauged sub-basins was defined as "gauged basin area". The gauged basin area was used to validate the scale-extrapolation method. Precipitation of the gauged area closely resembles the whole basin area, whereas potential evaporation in the gauged basin area is about 30 mm yr^{-1} lower than the whole basin (Fig. 2).

The successfulness of the scale-extrapolation depends on the abundance of discharge data from small river basins. It is critical to select river basins within a suitable size range for the extrapolation to perform well. The resolution of available global or regional climate dataset defines the lower limit for the size of the small river basins that can be used for extrapolation, i.e. the size of a river basin should be at least comparable with the climate grid, so reliable climate data can be obtained for the basin. Preliminary results showed that river basins between 500 and 5000 km^2 are most useful for discharge extrapolation at the global scale, considering that the resolution of most global climate dataset is 0.5 degree. Therefore, 51 sub-basins between 500 and 5000 km^2 are selected as "active sub-basins". Only those 51 sub-basins can be selected for discharge extrapolation.

4.2 Represent basin average with subsets of spatial data

Climate of a large region usually exhibits a wide range of variation. Annual precipitation, for instance, ranges from 242 mm to 2420 mm in the Baltic Sea drainage basin, and average annual potential evaporation ranges from 120 mm to 675 mm. The feasibility

HESSD
9, 6829–6856, 2012

of using a subset of cells to resemble the annual precipitation and potential evaporation for the whole basin was tested by the following method: the number of cells was allowed to change from 1 % to 15 % of the total with a step of 1 %. For each percentage, cells were selected randomly within the basin in 100 tries. In each try, spatial average of annual precipitation and potential evaporation from selected cells were calculated, and compared to the spatial averages of the whole basin by calculating the standardised RMSE error:

$$\mathrm{SRMSE} = \frac{1}{\overline{x}} \cdot \sqrt{\frac{\sum_{i=1}^{n} (x_i' - x_i)^2}{n}}. \tag{1}$$

Three scenarios were considered when choosing the final set of cells from the 100 tries for each percentage tested: (1) the selected cells should produce the smallest SRMSE for precipitation, i.e. the best possible resemble of average basin precipitation that can be achieved with a certain percentage of cells; (2) the selected cells should produce the smallest SRMSE for potential evaporation, and (3) the selected cells should produce smallest SRMSE for both precipitation and potential evaporation. The test was performed twice. In the first time, cells were selected within the entire Baltic Sea drainage basin; in the second time, cells were selected only from the active sub-basin.

4.3 Represent nonlinearity with subsets of spatial data

The Budyko's equation describes the nonlinear partitioning of precipitation at annual scale:

$$\frac{e}{p} = \left\{ \frac{\mathrm{pet}}{p} \cdot \tanh\left(\frac{p}{\mathrm{pet}}\right) \cdot \left[1 - \exp\left(-\frac{\mathrm{pet}}{p}\right)\right] \right\}^{0.5} \tag{2}$$

HESSD
9, 6829–6856, 2012

where e, p and pet are long term averaged evaporation, precipitation and potential evaporation. If the variation of storage is neglectable from year to year, subtracting evaporation from precipitation gives the annual discharge:

$$d = p - p \cdot \left\{ \frac{\text{pet}}{p} \cdot \tanh\left(\frac{p}{\text{pet}}\right) \cdot \left[1 - \exp\left(-\frac{\text{pet}}{p}\right)\right] \right\}^{0.5}. \quad (3)$$

Equation (3) describes the annual discharge as a two-dimensional non-linear function of precipitation and potential evaporation (Fig. 1b). Figure 1b shows the range of precipitation and potential evaporation covered by the gauged basin area. The dotted plot shows annual discharge against precipitation and potential evaporation from the 100 gauged sub-basins. It can be seen that the dotted plot agreed well with the contour lines derived from the Budyko's curve. Therefore, it is reasonable to assume that (1) precipitation and potential evaporation are two of the main controlling factor for the annual variation of discharge, and (2) the Budyko's equation resembles well the interaction between climate and hydrology for sub-basins of various sizes. In this paper, Budyko's equation is only used to illustrate the nonlinear relationship, but is not be used to make discharge extrapolation. The actual functional relationship between annual climate and hydrology is much more complex, and, as a matter of fact, unknown. However, this nonlinear functional relationship is embedded in the small-scale climatic and hydrological data. The scale-extrapolation method utilises the embedded nonlinear functional relationship to predict for the un-gauged basin area as well as the entire basin area. Therefore, it is only assumed that each cell of the Baltic Sea drainage basin is a nonlinear responding unit, for which annual discharge is an unknown nonlinear function of precipitation and potential evaporation. It would produce significant bias if only the mean of precipitation or potential evaporation is used to estimate the average dynamic of discharge, due to the nonlinear nature of the function. On the other hand, it is argued that it is not necessary to use the whole set of spatial data. A theoretical test was constructed to represent the mean of Eq. (3) with a number of points. Precipitation was ranged from 300–1000 mm yr^{-1} and potential evaporation 150–600 mm yr^{-1}. The

HESSD
9, 6829–6856, 2012

average discharge was obtained numerically prior to the test. The number of points varied from 1 to 100 with a step of 1. Because Eq. (3) is a two-dimensional function, each point is a unique combination of annual precipitation and potential evaporation. Four different scenarios were considered when selecting the points: (1) selected points should best resemble the average precipitation; (2) selected points should best resemble the average potential evaporation; (3) selected points should best resemble both precipitation and potential evaporation; (4) points are selected randomly.

4.4 Scale-extrapolation procedure

The scale-extrapolation was carried out in the following steps:

1. Identify active sub-basins according to the resolution of the climate dataset; in this study 51 small sub-basins between 500 and 5000 km^2 were selected (Sect. 4.1).
2. Collect grid cells that overlap with the selected active sub-basins; those cells form the "active basin area".
3. Randomly select cells from the active basin area in a Monte-Carlo simulation. In each Monte-Carlo run, a small basin is flagged if it overlaps with the randomly selected cells. The Monte-Carlo simulation continues until one particular combination of cells meets the following criteria:
   a. For each flagged small sub-basin, the overlapping cells give good resemblance of the climate variation (and land surface characteristics if extrapolate on shorter time scale) of it, according to a predefined criterion. (So that the part of the small sub-basin, which overlap with selected cells, has the same hydrological characteristic as the small sub-basin itself).
   b. The combination of cells gives good resemblance of the climate variation (and land surface characteristics if extrapolate on shorter time scale) of the entire large basin, according to a predefined criterion. (So that the combination of cells has the same hydrological characteristic as the large region).

HESSD
9, 6829–6856, 2012

4. From the small river basins, find those that overlap with the selected cells. Calculate the weighted average of discharge according to the overlapping area. This discharge will be used to estimate the discharge for the large region.

4.5 Ensemble prediction for the whole basin

There exist multiple sets of cells that could equally well resemble the climate of the basin; therefore they are expected to resemble basin hydrology equally well. An ensemble of discharge estimations for the basin can be derived, by repeating step 3 and 4 in Sect. 4.4. Such ensemble estimation brackets the uncertainties of the estimated discharge with inherent variability and uncertainty of the data within the basin.

5 Result

5.1 Represent average basin climate with subsets of spatial data

A good resemblance of average basin climate can be achieved with as few as 1 % of the cells, as shown by the small SRMSE (3 % for precipitation and 1.5 % for potential evaporation) values in Fig. 3. The SRMSE value dropped below 1 % when more than 5 % of cells were selected. It was also shown that the SRMSE value for precipitation is slightly higher than that of potential evaporation. Potential evaporation is better spatially correlated, even at annual scale, compared with precipitation, it is reasonable that it is easier to mimic its variation with less cells. It appeared to be slightly more difficult for a subset of cells to mimic both annual precipitation and potential evaporation of the basin; nevertheless the SRMSE still remains as low as 1 %. In general, the error reduced when more cells were included, when cells were selected randomly within the whole basins. When the selection of cells was limited to the active basin area, it was harder to mimic the average of the whole basin if a larger portion of the area was selected (Fig. 3b). Annual precipitation time series from selected cells were more mutually independent (Fig. 4) if fewer cells were selected to represent the basin average.

HESSD
9, 6829–6856, 2012

5.2 Represent nonlinearity with subsets of spatial data

The predicted discharge showed the largest bias when no control was applied on the selection of points (Fig. 5d). The bias was slightly reduced if selected points were forced to mimic the average of potential evaporation (Fig. 5b). A significant reduction of bias was seen when points were also forced to mimic precipitation (Fig. 5a). The bias was the smallest if points were forced to mimic both precipitation and potential evaporation (Fig. 5c). In the last case, when the number of points exceeded 10 the bias quickly converge to $\pm 10\,\mathrm{mm\,yr^{-1}}$. This result offered theoretical support for using a few small sub-basins to mimic the nonlinear interaction between climate and hydrology over the whole basin. It also indicated that ensemble estimation using multiple sets of cells might better bracket the uncertainty.

5.3 Discharge extrapolation

A baseline test was performed prior to the discharge extrapolation. Gauged sub-basins sized between $500\,\mathrm{km^2}$ and $15\,000\,\mathrm{km^2}$ were allowed to be randomly combined and annual discharge for those combined area were calculated and compared to the discharge of the entire gauged basins area by calculating SRMSE values (Fig. 6). Although certain combinations can give small error values, significant errors were generally expected by randomly combining sub-basins. The SRMSE values were consistently low when 77 multiple sets of sub-basins were selected using the scale-extrapolation method (Fig. 6). Slightly better overall performance can be seen when the upper area limit for the selection of active sub-basins was set to $10\,000\,\mathrm{km^2}$ instead of $5000\,\mathrm{km^2}$.

Figure 7 illustrated the map of the Baltic Sea drainage basin with boundaries of 100 gauged-sub-basins. The active sub-basins were marked with red, and the rest of the sub-basins were marked with blue. Figure 8a, b showed two examples of totally different sets of selected sub-basins, together with selected cells within each sub-basin. In the first example, the selected sub-basins represented precipitation and potential evaporation of the whole basin area with SRMSE values of 3.5 % and 2 %, respectively (Fig. 8c, d); and the extrapolated discharge well resembled discharge of the whole basins with SRMSE of 6 % (Fig. 8g). In the second example, precipitation and potential evaporation of the whole basin area were represented with SRMSE values of 3.3 % and 2.6 %, respectively (Fig. 8e, f), and the extrapolated discharge resembled discharge for the whole basins also with SRMSE of 6 % (Fig. 8h).

Figure 9 showed the comparison of spatially averaged annual discharge from the gauged area of Baltic Sea drainage basin and predicted discharge for the entire basin area with data-extrapolation method. All of the multiple extrapolations predicted less specific discharge for the whole basin compared to the gauged basin area, which is expected because the potential evaporation for the whole basin is higher compared to the gauged area.

6 Discussions and conclusion

A new data-based scale-extrapolation method is proposed to estimate annual water resources for large river basins. The new method builds upon the fact that the dynamic interaction between climate and hydrology of a large river basin can be equally well resembled by multiple small regions, each characterized by a number of small river basins that typically count for less than 5 % areal percentage of the large basin. Therefore, those multiple small regions can provide an ensemble of water recourse estimations for the large basin. The new method, being purely data-based, makes it possible for regional water resource estimations to benefit from a multitude of readily available measurements from small river basins.

The scale-extrapolation method provides both new methodology and new data into the field of large-scale hydrology. It allows regional water resources to be estimated directly from small river basins that are typically two-orders-of-magnitude smaller; therefore better preserves the small-scale dynamics and nonlinearity, which are vital for the credible predictions. The extrapolation is modelling-free; therefore the estimation is free

HESSD
9, 6829–6856, 2012

of modelling uncertainties that usually contribute significantly to large-scale estimation uncertainties; but the method is sensitive to error of discharge data. The method is not sensitive to the bias of climate dataset because climate dataset are only used for sub-basin selection but not directly used for extrapolation.

The scale-extrapolation methods made it possible to study the interaction between climate and hydrology, and the climate change impact in un-gauged or partially gauged large river basins from data alone. In the same time, the method offers ensemble predictions that bracket the estimation uncertainty. Because the scale-extrapolation uses completely different data and method compared to the modelling approach, it provides a unique opportunity to be compared with modelling results.

The scale-extrapolation method is simpler, and easier to operate compared with modelling approaches. In the same time, the method is easy to validate. Currently it only uses publically available dataset, so it is easy for many other researchers to apply the method and compare the results.

Acknowledgements. Discharge data were provided by The Global Runoff Data Centre, 56068 Koblenz, Germany and the Sweden's Meteorological and Hydrological Institute. The author had discussions with Fritjof Fagerlund and Anna Kauffeldt and received valuable insights.

References

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration – guidelines for computing crop water requirements, FAO Irrigation and Drainge Paper, UN-FAO, Rome, 1998.

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

Arnell, N. W.: Effects of IPCC SRES* emissions scenarios on river runoff: a global perspective, Hydrol. Earth Syst. Sci., 7, 619–641, doi:10.5194/hess-7-619-2003, 2003.

Arnell, N. W.: Climate change and global water resources: SRES emissions and socio-economic scenarios, Global Environ. Change, 14, 31–52, 2004.

Budyko, M. I.: Climate and Life, Academic Press, New York, 18 pp., 1974.

Gong, L., Widén-Nilsson, E., Halldin, S., and Xu, C.-Y.: Large-scale runoff routing with an aggregated network-response function, J. Hydrol., 368, 237–250, doi:10.1016/j.jhydrol.2009.02.007, 2009.

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.

GRDC: Global Runoff Data Centre, available online: http://grdc.bafg.de, last access: 1 April 2012.

Huang, M., Liang, X., and Liang, Y.: A transferability study of model parameters for the variable infiltration capacity land surface scheme, J. Geophys. Res., 108, 8864, doi:10.1029/2003JD003676, 2003.

Kokkonen, T. S., Jakeman, A. J., Young, P. C., and Koivusalo, H. J.: Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina, Hydrol. Process., 17, 2219–2238, 2003.

Kim, U. and Kaluarachchi, J. J.: Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia, J. Hydrol., 362, 39–56, 2008.

L'vovich, M. I.: World Water Resouces and Their Future, Litho Crafters Inc., Chelsea, UK, 1979.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, Water Resour. Res., 41, W12434, doi:10.1029/2005WR004289, 2005.

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.

NCDC: Global Surface Summary of the Day, National Climatic Data Center (NCDC), Asheville, NC, available online: http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=climvisgsod.html, last access: 1 February 2011.

Simmons, A., Uppala, S., Dee, D., and Kobayashi, S.: ERA-Interim: new ECMWF reanalysis products from 1989 onwards, ECMWF Newsletter, Reading, UK, 110, 25–35, 2007.

Stahl, K., Tallaksen, L. M., Gudmundsson, G., and Christensen, J. H.: Streamflow data from small basins: a challenging test to high resolution regional climate modeling, J. Hydrometeorol., 12, 900–912, 2011.

Uppala, S. M., Kaållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteorol. Soc., 131, 2961–3012, 2005.

USGS – US Geological Survey: HYDRO 1K Elevation Derivative Database, available online: http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/HYDRO1K, last access: February 2009, from the Earth Resources Observation and Science (EROS) Data Center (EDC), Sioux Falls, South Dakota, USA, 1996.

Vörösmarty, C. J., Moore, 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., Green, P., Salisbury, J., and Lammers, R. B.: Global water resources: vulnerability from climate change acid population growth, Science, 289, 284–288, 2000a.

Vörösmarty, C. J., Fekete, B. M., Meybeck, M., and Lammers, R. B.: Globalsystem of rivers: its role in organizing continental land mass and defining landto-ocean linkages, Global Biogeochem. Cy., 14, 599–621, 2000b.

Vörösmarty C., Lettenmaier D., Levqeue C., Meybeck M., and Pahl-Wostl C.: Humans transforming the global water system, Eos Trans. AGU, 85, 509–514, 2004.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment classification and hydrologic similarity, Geogr. Compass, 1, 901–931, 2007.

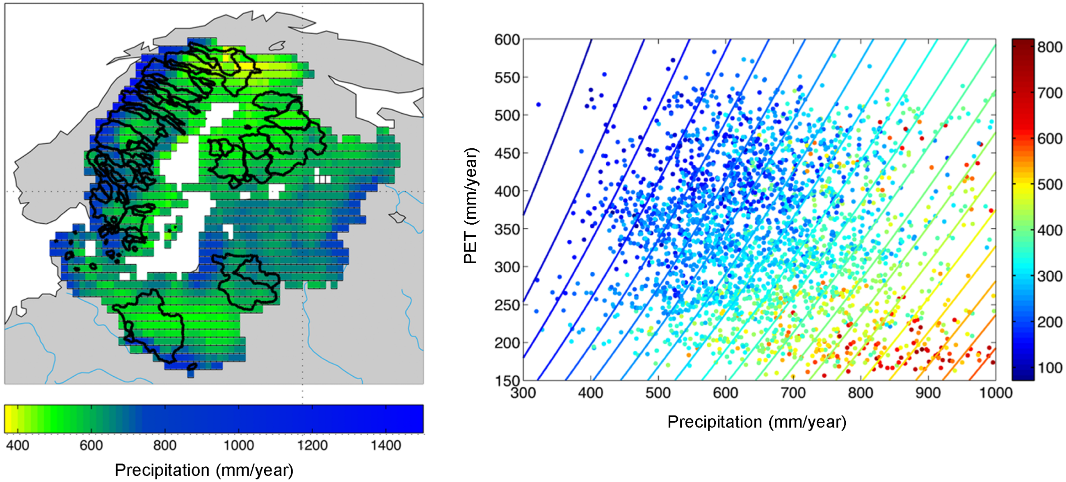
Weedon, G. P., Gomes, S.,Viterbo, P., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: The WATCH forcing data 1958–2001: a meteorological forcing dataset for land surface- and hydrological-models, WATCH Technical Report, 2010.

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.

Wulff, F., Rahm, L., Hallin, A.-K., and Sanberg, S.: A SystemsAnalysis of the Baltic Sea, edited by: Wulff, F., Rahm, L., and Larsson, P., Springer-Verlag, Berlin, 353–372, 2001.
Xu, C.: Testing the transferability of regression equations derived from small sub-catchments to a large area in central Sweden, Hydrol. Earth Syst. Sci., 7, 317–324, doi:10.5194/hess-7-317-2003, 2003.
Xu, C.-Y.: Estimation of parameters of a conceptual water balance model for ungauged catchments, Water Resour. Manage., 13, 353–368, 1999.

HESSD
9, 6829–6856, 2012



(a)

(b)

Fig. 1. (a) Map of the Baltic Sea drainage basin as shown by 0.5 degree STN-30p global grid cells, with boundaries of 100 gauged-sub-basins shown by back lines. Background color shows annual average precipitation during 1975–2001. **(b)** Annual discharge as a function of precipitation and potential evaporation, according to Budyko's curve (contour lines); and dotted plot of annual discharge against precipitation and potential evaporation from 100 gauged sub-basins of the Baltic Sea drainage basin during 1975–2001 (dots).

HESSD

9, 6829–6856, 2012

Data-based discharge extrapolation

L. Gong

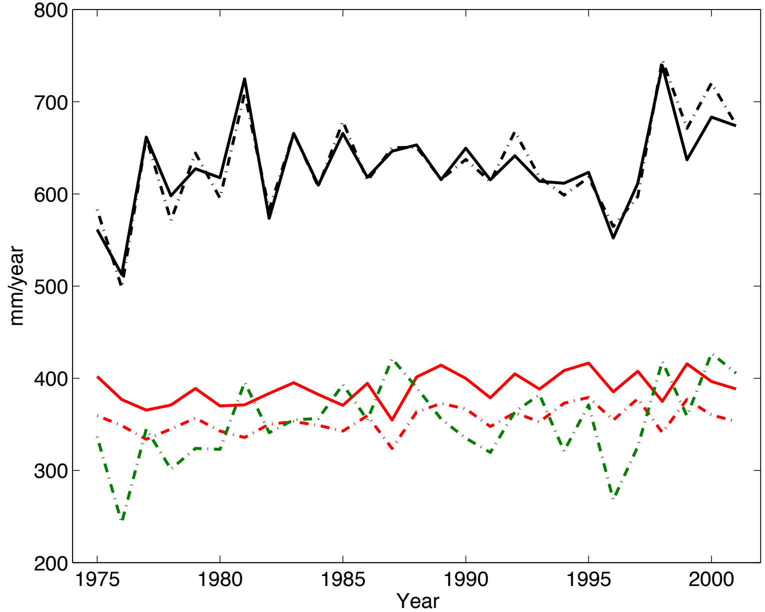


Fig. 2. Annual time series precipitation (black), potential evaporation (red) and discharge (green) for the entire Baltic Sea drainage basin (solid lines) and for the partial basin area (dashed lines).

HESSD
9, 6829–6856, 2012

Data-based discharge extrapolation

L. Gong

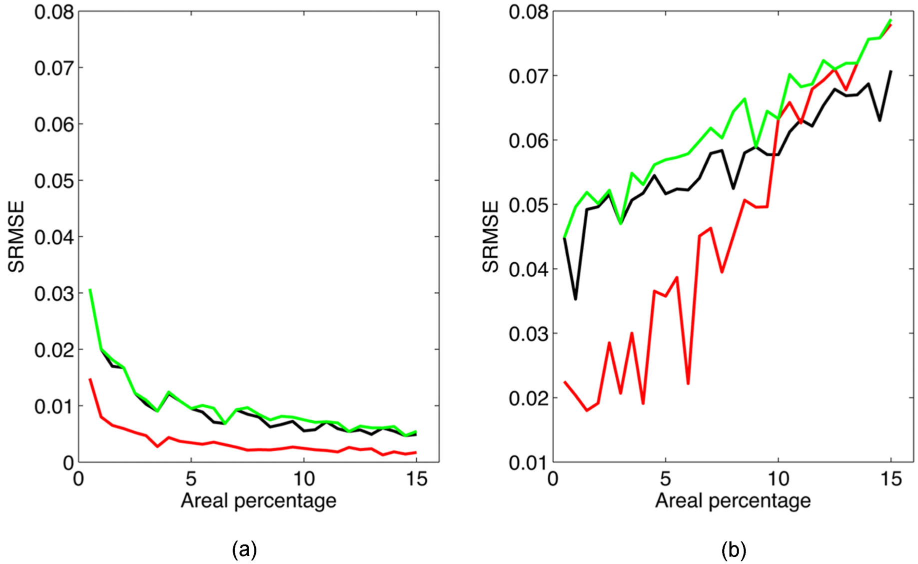


Fig. 3. Standardized RMSE errors (y-axis) between annual climate variables (precipitation as black and potential evaporation as red) calculated from selected cells and from the entire Baltic Sea drainage basin. x-axis shows areal ratio between selected cells and the total area of the basin. The green lines show the standardized RMSE when the set of cells are forced to mimic both precipitation and potential evapotranspiration. Cells are selected within the whole basin area **(a)**, or only within active basin area **(b)**.

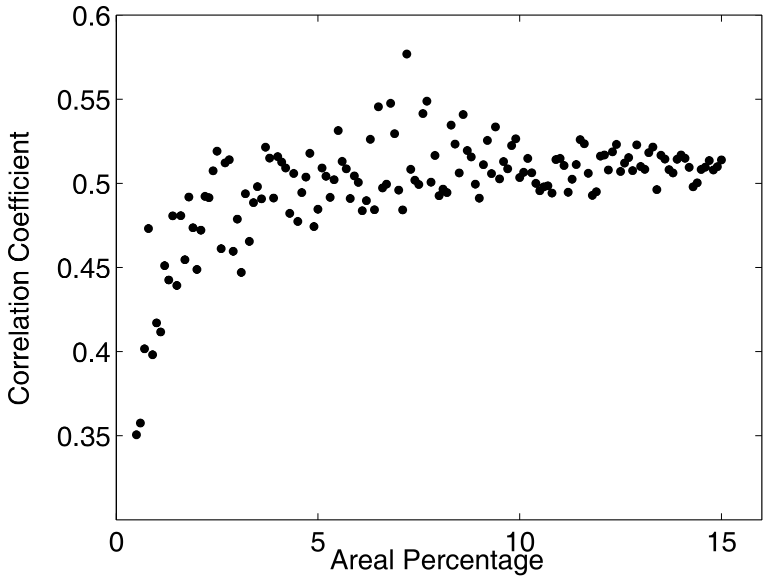


Fig. 4. Correlation coefficients (black dots) between annual precipitation time series of selected cells within the active basin area (y-axis), as a function of the areal ratio between selected cells and the area of Baltic Sea drainage basin (x-axis).

HESSD
9, 6829–6856, 2012

Data-based discharge extrapolation

L. Gong

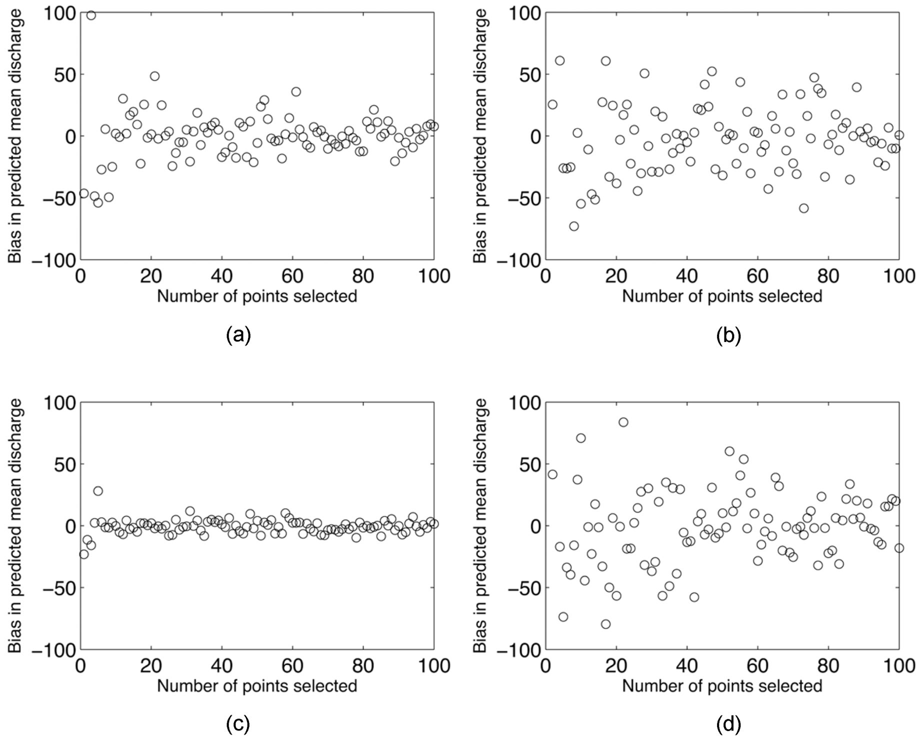


Fig. 5. Representation of the mean of the 2-dimentional nonlinear function Eq. (3), with 1–100 points. Precipitation ranges from 300–1000 $\mathrm{mm\,yr^{-1}}$ and potential evaporation 150–600 $\mathrm{mm\,yr^{-1}}$. Points are randomly sampled, but in the mean time to fulfil the following criteria: mimic the mean of precipitation **(a)**; mimic the mean of potential evaporation **(b)**; mimic mean of both precipitation and potential evaporation **(c)**; and no control **(d)**. x-axis shows the number of points used, and y-axis shows the bias of predicted mean discharge value.

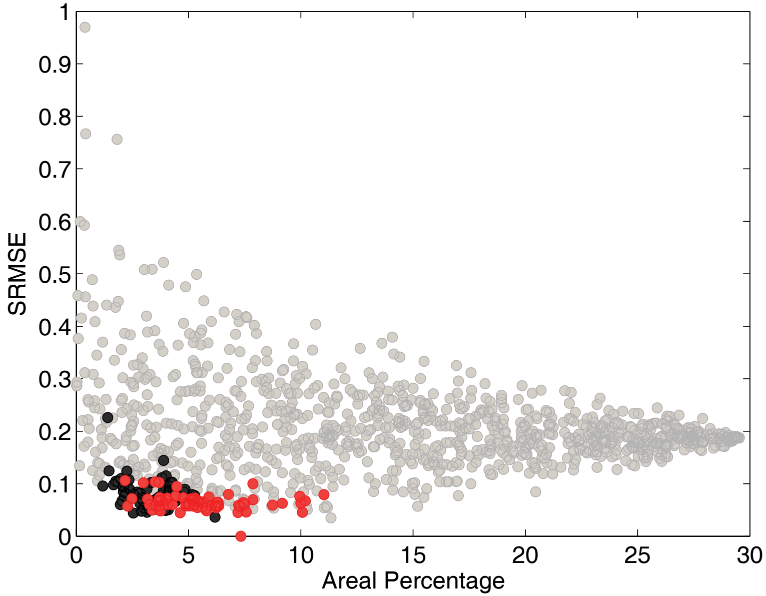


Fig. 6. Represent the spatially averaged annual discharge of the gauged basin area withselected sub-basins. Three different selection methods are plotted: (1) randomly selectsub-basins that are between 500 km^2 and 15 000 km^2 (grey); (2) select sub-basins sized between 500 km^2 and 10 000 km^2 with the scale-extrapolation method (red); (3) select sub-basins sized between 500 km^2 and 5000 km^2 with the scale-extrapolation method (black); x-axis shows arealratio of selected sub-basins to the area of the gauged basin area, y-axis shows standardized RMSE of predicted discharge compared with observed discharge in the gauged basin area.

HESSD
9, 6829–6856, 2012

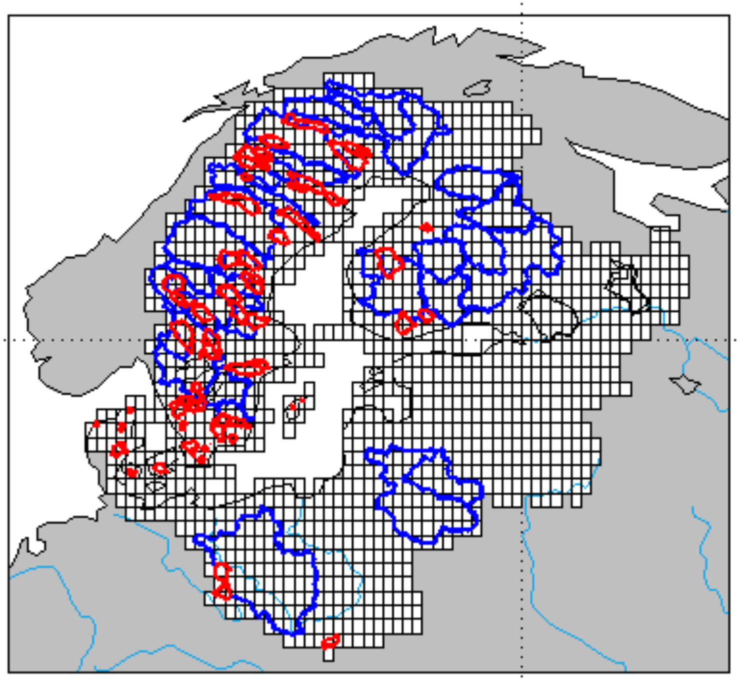


Fig. 7. Map of the Baltic Sea drainage basin as shown by 0.5 degree STN-30p global grid cells, with boundaries of 100 gauged-sub-basins shown by lines. Active basin area marked with red and the rest marked with blue.

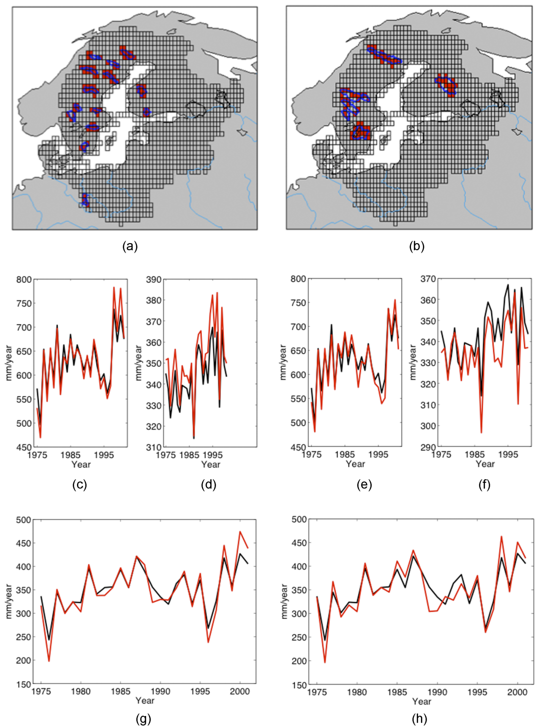


Fig. 8. Two examples of the validation of the discharge extrapolation in the Baltic Sea drainage basin. **(a)**, **(b)**: Map of the Baltic Sea drainage basin as shown by 0.5 degree STN-30p global grid cells, with two different examples of selected sub-basins shown by blue lines and selected cells marked with red. **(c)**, **(d)**: Spatially averaged annual precipitation **(c)** and potential evaporation from **(d)** from the selected sub-basins (red) and basin average (black) for example 1. **(e)**, **(f)**: Spatially averaged annual precipitation **(e)** and potential evaporation from **(f)** from the selected sub-basins (red) and basin average (black) for example 2. **(g)**, **(h)**: Spatially averaged annual discharge from the gauged basin area (black) and predicted discharge using data-extrapolation method (red) using selected sub-basins from example 1 **(g)** and example **(h)**.

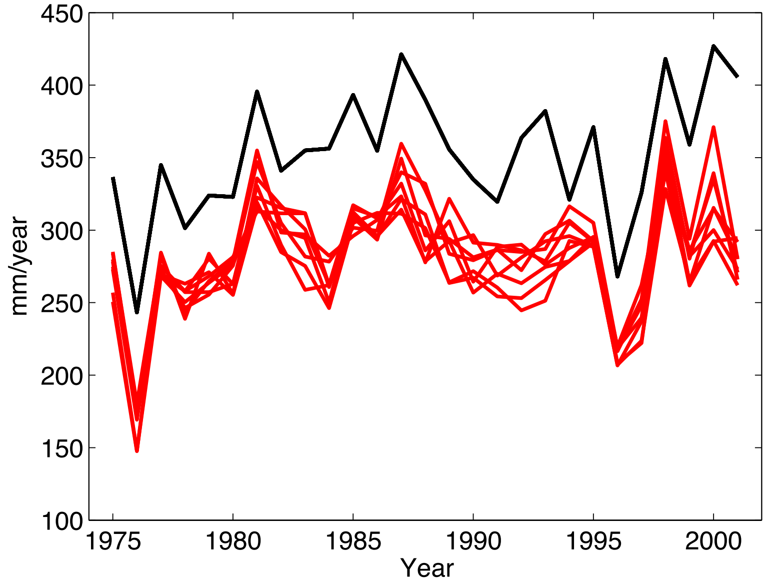


Fig. 9. Comparison of average annual discharge from the gauged basin area (black), and multiple predictions of spatially averaged annual discharge for the entire Baltic Sea drainage basin using data-extrapolation method (red).

HESSD

9, 6829–6856, 2012