Journal of Hydrology 544 (2017) 479–487

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Developing and testing a global-scale regression model to quantify mean annual streamflow

Valerio Barbarossa [a,*], Mark A.J. Huijbregts [a,b], A. Jan Hendriks [a], Arthur H.W. Beusen [c], Julie Clavreul [d], Henry King [d], Aafke M. Schipper [b]

[a] Radboud University, Institute for Water and Wetland Research, Department of Environmental Science, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[b] PBL Netherlands Environmental Assessment Agency, Department of Nature and Rural Areas, P.O. Box 303, 3720 AH Bilthoven, The Netherlands
[c] PBL Netherlands Environmental Assessment Agency, Department of Information, Data and Methodology, P.O. Box 303, 3720 AH Bilthoven, The Netherlands
[d] Unilever R&D, Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedfordshire MK44 1LQ, UK

ARTICLE INFO

Article history:
Received 20 July 2016
Received in revised form 20 September 2016
Accepted 24 November 2016
Available online 25 November 2016
This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Subashisa Dutta, Associate Editor

Keywords:
Mean annual discharge
River discharge
Global hydrology
Empirical modelling
Predictions in ungauged basins
Scaling relationships
Model comparison
PCR-GLOBWB
Spatial error model

ABSTRACT

Quantifying mean annual flow of rivers (MAF) at ungauged sites is essential for assessments of global water supply, ecosystem integrity and water footprints. MAF can be quantified with spatially explicit process-based models, which might be overly time-consuming and data-intensive for this purpose, or with empirical regression models that predict MAF based on climate and catchment characteristics. Yet, regression models have mostly been developed at a regional scale and the extent to which they can be extrapolated to other regions is not known. In this study, we developed a global-scale regression model for MAF based on a dataset unprecedented in size, using observations of discharge and catchment characteristics from 1885 catchments worldwide, measuring between 2 and 10^6 km^2. In addition, we compared the performance of the regression model with the predictive ability of the spatially explicit global hydrological model PCR-GLOBWB by comparing results from both models to independent measurements. We obtained a regression model explaining 89% of the variance in MAF based on catchment area and catchment averaged mean annual precipitation and air temperature, slope and elevation. The regression model performed better than PCR-GLOBWB for the prediction of MAF, as root-mean-square error (RMSE) values were lower (0.29–0.38 compared to 0.49–0.57) and the modified index of agreement (d) was higher (0.80–0.83 compared to 0.72–0.75). Our regression model can be applied globally to estimate MAF at any point of the river network, thus providing a feasible alternative to spatially explicit process-based global hydrological models.

1. Introduction

Mean annual discharge or flow of rivers (hereafter abbreviated as MAF) is an important indicator of global water supply, with applications in irrigation supply assessment, climate change vulnerability assessment (Chang, 2003; Santini and di Paola, 2015), hydropower assessment (Hall et al., 2004), water footprinting (Hanafiah et al., 2011; Hoekstra et al., 2011; Jefferies et al., 2012; Pfister et al., 2009; Tendall et al., 2014), and for quantifying sediment fluxes (Syvitski et al., 2003). It also represents one of the most important factors determining the integrity of freshwater biodiversity (Oberdorff et al., 1995, 2011; Poff and Zimmerman, 2010; Xenopoulos and Lodge, 2006; Xenopoulos et al., 2005). Despite its importance, streamflow data availability is limited, and monitoring is in rapid decline since the mid-1980s (Shiklomanov et al., 2002). Modelling approaches have long been used to estimate MAF at ungauged sites and are generally divided into two categories: spatially explicit process-based models and regression-based empirical models.

State-of-the-art spatially explicit numerical models for global-scale calculations of streamflow are Macroscale Hydrological Models (MHM) or Global Hydrology and Water Resources Models (GHWM) (Alcamo et al., 2003; Gosling and Arnell, 2011; Hanasaki et al., 2008; van Beek and Bierkens, 2008; Van Der Knijff et al., 2010; Widén-Nilsson et al., 2007; Wisser et al., 2010). As these models account for the spatial variability of the physical processes involved within catchment hydrology and are capable of predicting streamflow even at the daily time scale, they are computationally and data intensive.

Regression-based approaches to calculate MAF are less time-consuming and computationally less intensive. Moreover, regres-

* Corresponding author.
E-mail address: vbarbarossa@science.ru.nl (V. Barbarossa).

http://dx.doi.org/10.1016/j.jhydrol.2016.11.053
0022-1694/

sion equations relating streamflow to explanatory catchment characteristics like upstream drainage area, precipitation and temperature may help to better understand general hydrological patterns and processes across different scales (Burgers et al., 2013; Farmer et al., 2015). However, to date, regression-based approaches relating mean annual streamflow to catchment characteristics have been mainly applied at a regional scale (Hortness and Berenbrock, 2001; Stuckey, 2006; Tran et al., 2015; Verdin and Worstell, 2008; Vogel et al., 1999) or to specific climate zones (Syvitski et al., 2003), and the extent to which these models can be extrapolated to other regions is not known. Regression relationships at the global scale have hardly been established so far. An exception is Burgers et al. (2013), who derived MAF relationships at a global scale using precipitation and catchment area as predictors. However, their model explained only 56% of the variance in MAF, which is low compared to the range of 77–99% achieved by regional regression models (e.g. Verdin and Worstell (2008)). Yet, the regional studies typically included a larger number of predictors, which suggests that the explanatory power of a global-scale regression model may increase if relevant predictors are added. In addition, the applicability of global regression relationships for the prediction of mean annual streamflow has not yet been tested. Therefore, the aim of this study was twofold: (1) to establish an empirical regression model relating MAF to easily retrievable catchment characteristics at the global scale; (2) to test the predictive ability of the regression model in a backcasting analysis and compare its performance with the predictive performance of PCR-GLOBWB, a spatially explicit MHM (van Beek et al., 2011). To our knowledge, our study is the first to make an explicit comparison of the predictive abilities of a process-based and a regression-based global-scale model for MAF.

We based our regression model on measured long-term average MAF from 1885 catchments worldwide, ranging from 2 km^2 to 10^6 km^2 in size. We used five predictor variables, including two climatic variables – mean annual precipitation and air temperature – and three geomorphologic variables – area, mean slope and mean elevation of the catchment. Drainage area, mean annual precipitation and mean annual temperature are often used as predictors of MAF in regional regression modelling studies (Verdin and Worstell, 2008; Vogel and Sankarasubramanian, 2000; Vogel et al., 1999). The dependence of MAF on drainage area is a well-accepted power relationship reflecting the self-similarity of river systems (Rodríguez-Iturbe and Rinaldo, 2001). Mean annual precipitation represents the potential runoff of the catchment, as it equals the amount of water supplied to the catchment (Thomas and Benson, 1970). We selected the mean annual temperature as a proxy for the potential evapotranspiration (PET), because temperature is a major determinant of evapotranspiration (Hamon, 1963; Lu et al., 2005; Thornthwaite, 1948). Furthermore, previous regression analyses of MAF have shown an increased explained variance when additional geomorphologic parameters were considered (Hortness and Berenbrock, 2001; Stuckey, 2006; Vogel et al., 1999). Therefore, we included average slope and elevation of the catchment as additional predictors in our study. Although elevation and slope alone may not directly influence MAF, they may serve as proxies for other factors causing inter-basin streamflow variation which are difficult to measure, e.g. radiation, wind, vegetation and basin ruggedness (Thomas and Benson, 1970).

2. Materials and methods

2.1. Mean annual discharge data

We retrieved worldwide MAF data from the Global Runoff Data Centre (GRDC) database, which provides daily or monthly observations of 9213 gauging stations monitored from 1806 to 2015, with variable record length (GRDC, 2015). The GRDC has spent more than 25 years gathering river discharge data from the National Hydrological Services of all the World Meteorological Organization (WMO) state members, which has resulted in a discharge dataset unprecedented in size. For example, the SAGE Global River

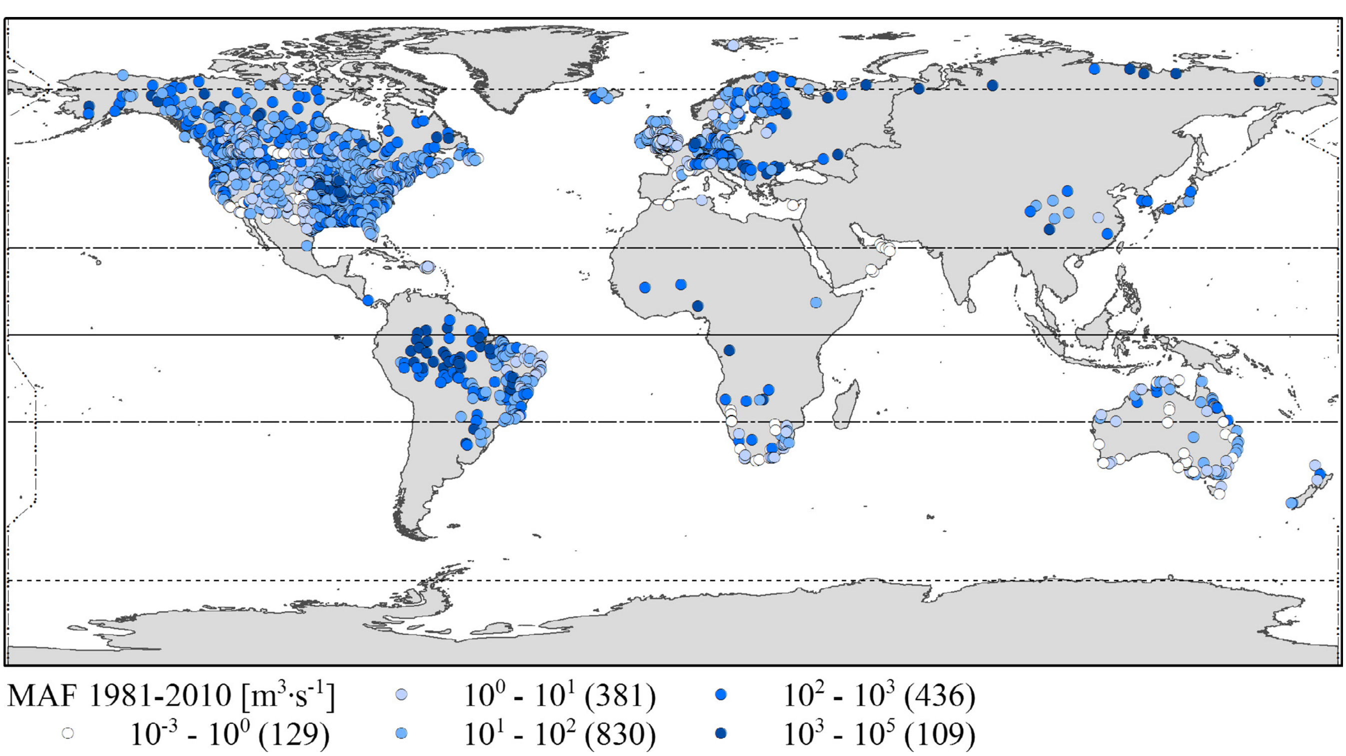

Fig. 1. Distribution of the 1885 GRDC gauging stations monitored for at least 15 years in the 1981–2010 period. The stations are grouped based on the mean annual flow (MAF) recorded at each station. Next to each MAF category, the number of observations is provided in brackets.