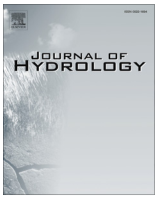
Journal of Hydrology 545 (2017) 120–131

Contents lists available at ScienceDirect

Journal of Hydrology

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



Research papers

Impacts of combining reanalyses and weather station data on the accuracy of discharge modelling

Gilles R.C. Essou [a,*], François Brissette [a], Philippe Lucas-Picher [a,b]

[a] *Department of Construction Engineering, École de technologie supérieure (ÉTS), Université du Québec, Montreal, Quebec, Canada*
[b] *Centre pour l'Étude et la Simulation du Climat à l'Échelle Régionale (ESCER), Department of Earth and Atmospheric Sciences, Université du Québec à Montréal (UQAM), Montreal, Quebec, Canada*

ARTICLE INFO

Article history:
Received 16 July 2016
Received in revised form 2 November 2016
Accepted 12 December 2016
Available online 15 December 2016
This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Jianzhong Yang, Associate Editor

Keywords:
Reanalyses
Observations
Hydrological modelling
Calibration
Data combination

ABSTRACT

Reanalyses are important sources of meteorological data. Recent studies have shown that precipitation and temperature data from reanalysis present a strong potential for hydrological modelling, especially in regions with a sparse observational network. The objective of this study is to evaluate the impacts of the combination of three global atmospheric reanalyses – ERA-Interim, CFSR and MERRA – and one gridded observation dataset on the accuracy of hydrological model discharge simulations. Two combination approaches were used. The first one combined reanalyses and the observational database using a weighted average of the precipitation and temperature inputs. The second one consisted in using all meteorological inputs separately and combining the simulated hydrographs. The combinations were performed over 460 Canadian watersheds (representing regions with a low density of weather stations) and 370 US watersheds (representing regions with a higher density of weather stations). Results showed significant improvements in the simulated discharges for 68% and 92% of the Canadian watersheds for the input combinations and output combinations, respectively. Moreover, both approaches led to significant improvements in the simulated discharges for 72% of the US watersheds studied. For all watersheds where simulated discharges using observational data had a Nash Sutcliffe efficiency (NSE) lower than 0.5, the combination with reanalyses resulted in a median NSE increase of 0.3. This indicates that reanalysis can successfully compensate for deficiencies in the surface observation record and provide significantly better hydrological modelling performance.

1. Introduction

It is well known that the quality of weather data used as input for hydrological models has a strong influence on the accuracy of river flow predictions (Duncan et al., 1993; Fekete et al., 2004). However, for many regions, such as Northern Canada, available surface weather stations are sparsely distributed, and the quality of historical measurements is often questionable. Therefore, finding adequate data for hydrological modelling is a real challenge in such areas.

In recent decades, significant effort has been dedicated to producing global datasets for climate monitoring and research using weather forecasting models and a complex assimilation of observations called "reanalyses". Reanalyses use a constant data assimilation scheme and a numerical forecasting model, which for their part use several observations every 6–12 h, over a given period (Dee et al., 2011; Saha et al., 2010). Available observations include radiosondes, satellites, buoys, aircraft and ship reports. While the assimilation scheme is constant, the observational network changes constantly. Nonetheless, reanalyses provide a physically consistent estimate of the climate state at each time step. In addition to global coverage, reanalyses typically cover several decades, and provide a large array of climate variables (Mesinger et al., 2006; Rienecker et al., 2011). Despite the spatial and temporal consistency of reanalyses, the observational database, which changes constantly over the duration of each reanalysis, can produce spurious trends and artificial variability. Reanalyses outputs are often biased, and especially so for surface fields, but have steadily improved in time. Part of the biases involved are due to the relatively coarse grid resolution, as well as to the parameterization of many physical processes, such as convective storms. Recent reanalysis outputs have been gradually made available at higher spatial and temporal scales, thus potentially reducing biases.

---

[*] Corresponding author at: Department of Construction Engineering, École de technologie supérieure, 1100 Notre-Dame Street West, Montreal, QC H3C 1K3, Canada.

E-mail addresses: essougilles@yahoo.fr, gilles.essou.1@ens.etsmtl.ca (G.R.C. Essou).

http://dx.doi.org/10.1016/j.jhydrol.2016.12.021
0022-1694/© 2016 Elsevier B.V. All rights reserved.

Recent studies have assessed the usefulness of reanalysis data for climate monitoring, and have found them to be extremely useful if used with care (Bosilovich, 2013; Lorenz and Kunstmann, 2012; Manzanas et al., 2014; Nigam and Ruiz-Barradas, 2006; Rusticucci et al., 2014; Zhang et al., 2013). Moreover, reanalyses have demonstrated good potential to drive hydrological models (Choi et al., 2009; Fuka et al., 2014; Vu et al., 2012). More recently, Essou et al. (2016b) used precipitation and temperature series from the NARR, ERA-Interim, CFSR and MERRA reanalyses and from one gridded observation database (Santa Clara gridded dataset) to calibrate a lumped hydrological model and to simulate river flows over 370 watersheds in the continental USA. They found that the river flows obtained using NARR forcing were as good as when gridded observations were used. Moreover, the Nash-Sutcliffe values of the river flows simulated using the other three reanalyses were equal to or better than those from the gridded observations, with the exception of the Humid continental and subtropical regions, where the precipitation seasonality was not well reproduced. Reanalyses may thus be useful for hydrological modelling, especially in areas with a sparse weather station density. However, instead of using either reanalyses or surface weather stations alone, a more optimal scenario may consist in combining both data sources. Such a multi-model approach involves the combination of several hydrological models to simulate river flows more accurately than the models taken individually (Ajami et al., 2006; Arsenault et al., 2015; Cavadias and Morin, 1986; Diks and Vrugt, 2010; Shamseldin et al., 1997). Recently, Arsenault et al. (2016a) combined three hydrological models and four climate datasets to produce multi-input averaged flows and found that this approach provides better results than the classical multi-model averaging. In their work, all the datasets used came from the same data source type (gridded databases). The combination of different data sources has the potential to improve the accuracy of simulated river flows. For instance, Sun et al. (2000) evaluated flood estimation combining radar and rain gauge data for the Finniss River catchment in Darwin, Australia, and found that rainfall estimated by coKriging both data sources considerably improved flood estimates. They showed that an optimal combination of both databases improves the estimation of subcatchment rainfall. To our knowledge, the potential presented by combining reanalyses and weather stations for hydrological modelling has never been investigated.

This study will focus on the impacts of combining three global atmospheric reanalyses – ERA-Interim, CFSR and MERRA – and one gridded observation dataset on hydrological model simulations. More specifically, it aims to assess the impacts of such a combination on the ability to simulate river discharges: (1) in the presence of a high density of surface weather stations (US watersheds); (2) in the presence of a low density of surface weather stations (Canadian watersheds); and (3) over watersheds where hydrological models perform poorly, thus calling into question the quality of surface observations. The results of this study will determine whether the combination of reanalyses and observations in regions with a sparse density of weather stations, such as Northern Canada, is impactful for hydrological modelling.

2. Study area

The study area consists of 830 watersheds in North America, 370 of which are located in the United States, and 460 in Canada. The watersheds are located in various hydro-climatic regimes. The US watersheds were selected because of their relatively high density of weather stations compared to their Canadian counterparts. They were derived from the MOdel Parameter Estimation eXperiment database (MOPEX) (Duan et al., 2006), and their total areas ranged between 104 and 10,325 km^2 with respectively an average and a median values of 2925 km^2 and 2070 km^2. The Canadian watersheds were selected because of their relatively low density of weather stations. They were derived from the CANadian mOdel Parameter EXperiment (CANOPEX) database (Arsenault et al., 2016b). Their total areas ranged between 450 and 127,635 km^2 with respectively an average and a median values of 7770 km^2 and 2310 km^2. The Canadian watersheds tend to be larger since major rivers are the only ones typically gauged in remote areas.

Over the study area, the mean annual precipitation is between 0 and 5 mm/day (Fig. 1a). The highest precipitation (>4 mm/day) area is located in the Southeastern US, and the lowest precipitation (<1 mm/day) area is located in Northern Canada. The mean annual temperature varies between −5 °C and 20 °C (Fig. 1b). The temperature decreases from South to North in the study area. Consequently, watersheds in the Southeastern US are the warmest, while the colder ones are located in Northern Canada.

3. Datasets

3.1. Observational datasets

The observational datasets consisted of daily meteorological (minimum and maximum temperature and precipitation) and hydrometric datasets derived from the Santa Clara and MOPEX databases, for the US watersheds, and from the NRCan and HYDAT databases, for the Canadian watersheds.

- The Santa Clara dataset (SClara) is a gridded database based on the high-density network of weather cooperative stations of the National Oceanic and Atmospheric Administration (NOAA) (average of 1 station per 700 km^2) (Maurer et al., 2002). The SClara gridded database consists of daily precipitation and maximum and minimum air temperature at a 0.125° × 0.125° spatial resolution (about 12 km × 12 km) for the period of 1949–2010. The Synergraphic Mapping System (SYMAP) algorithm was used for the data interpolation (Shepard, 1984; Widmann and Bretherton, 2000). The SClara gridded precipitation was scaled to match the long-term average precipitation of the Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly et al., 1994, 1997).
- The MOPEX database contains daily mean hydrometric and meteorological data covering the period of 1949–2003 (Duan et al., 2006). This study used only the MOPEX hydrometric data.
- NRCan is a gridded database based on the low-density network of Environment Canada weather stations. The NRCan dataset consists of daily precipitation and 2 m temperature at a 10 km spatial resolution over the period of 1950–2010. The Interpolation was performed using the thin plate smoothing splines (ANUSPLIN) method (Hopkinson et al., 2011; Hutchinson, 1995, 2004; Hutchinson et al., 2009).
- The HYDAT database contains daily mean discharge data from about 7000 hydrometric stations across Canada (Coulibaly et al., 2013; Winkler, 1993). Both of the previous datasets were brought together within the watershed-based coherent CANOPEX database (Arsenault et al., 2016b).

3.2. Reanalysis datasets

The daily mean precipitation and 2 m temperature from the ERA-Interim, CFSR and MERRA reanalyses were used to force the hydrological model described later in this article.