Data Notes

HP TODAY

HYDROLOGICAL PROCESSES
Hydrol. Process. **30,** 2734–2736 (2016)
Published online 6 June 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hyp.10880

CANOPEX: A Canadian hydrometeorological watershed database

Richard Arsenault,[1]*
Rachel Bazile,[1]
Camille Ouellet Dallaire[2] and
François Brissette[1]

[1] *Department of Construction Engineering, École de technologie supérieure, Université du Québec, Montreal, Québec, H3C 1K3, Canada*
[2] *Department of Geography, University McGill, Montreal, Québec, H3A 2K6, Canada*

*Correspondence to: Richard Arsenault, Department of Construction Engineering, École de technologie supérieure, Université du Québec, 1100 Notre-Dame Street West, Montreal, Québec H3C 1 K3, Canada.
E-mail: richard.arsenault.1@ens.etsmtl.ca

Received 1 February 2016
Accepted 8 April 2016

Abstract

Over the past few years, many international initiatives and collaborations were launched to improve and share knowledge in hydrology research. Large databases allowed finding patterns and relationships across regions and scales. This paper introduces the Canadian model parameter experiment (CANOPEX) database, which is adapted from the US MOPEX project data and methods. The CANOPEX database includes meteorological and hydrometric data as well as watershed boundaries for 698 basins. Two sets of basin-averaged meteorological data (Maximum and minimum temperature and precipitation) are provided. The first dataset is directly taken from Environment Canada's weather stations whereas the second is extracted from the Natural Resources Canada gridded climate data product. Data are provided in MOPEX and MATLAB formats. CANOPEX watersheds are well distributed over Canada, which allows investigating a variety of physiological and climatological conditions. The CANOPEX database can be used in a variety of hydrologic research projects such as climate change impact studies, model comparisons, multi-modelling, ensemble streamflow prediction and model parameter estimation. CANOPEX could be used to generalize findings to other cold climate catchments as well as assess the robustness of research methodologies and procedures. Copyright © 2016 John Wiley & Sons, Ltd.

Key words hydrological database; CANOPEX; MOPEX; large dataset; Canadian hydrology

Methods

Selection and filtering of streamflow gauges

The methods used to select suitable catchments for the Canadian model parameter experiment (CANOPEX) database were inspired by the seminal paper introducing the MOPEX project catchment selection process (Schaake *et al.,* 2006). The first step was to collect streamflow records from Environment Canada's hydrometric data portal (HYDAT). HYDAT is a database containing the daily observed hydrometric data for publicly funded gauges in Canada. Also available in HYDAT are metadata about the hydrometric stations, such as latitude and longitude, catchment area, record length and information regarding the flow conditions (current status, regulated or natural regime). It currently contains data for approximately 7000 hydrometric stations across Canada. The raw data themselves as well as an extractor executable are publicly available from Environment Canada's FTP website at ftp://arccf10.tor.ec.gc.ca/wsc/software/HYDAT. However, HYDAT lacks catchment boundaries for the stations on record, which is a key piece of information for hydrological research and applications. The database is updated four times per year. CANOPEX is based on HYDAT Release 19 (version 20150126). Potential sites were filtered according to the following criteria, mirroring Schaake *et al.* (2000):

1. Length of observational record must be equal or superior to 20 years
2. River regimes must be unregulated (we only keep naturally flowing basins)
3. Catchment area must be larger than 300 km^2 due to delineation uncertainty and daily time step.

At this point, 1024 candidate basins remained out of the approximately 7000 original catchments.

Watersheds delineation

Hydrometric data portal does not contain catchment boundaries for its hydrometric stations. This is one of the key problems in large-scale hydrological studies in Canada because the hydrometric and weather data are publicly available. This project relied on Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales (HydroSHEDS) (Lehner *et al.*, 2008) to provide the necessary hydrological data. More information is provided through Lehner *et al.* (2006) and data are available on the HydroSHEDS project website: http://www.hydrosheds.org/. Watershed boundaries were extracted and contributing areas compared with values registered in HYDAT. Values with a difference greater than 10% were manually inspected and their outlet was moved to a better location when possible. If this step proved inconclusive, they were ultimately discarded. After this step, 698 streamflow gauges remained in the CANOPEX database.

Meteorological data

The CANOPEX database contains two sources of basin-averaged meteorological data. The first is derived from weather station observations directly, and the second is based on a gridded climate data product. Observed daily minimum and maximum temperatures as well as precipitation data were collected from Environment Canada's weather station database, which contains almost 8000 stations. For each basin, stations inside or near the basin boundaries were selected as possible contributors to the basin-averaged meteorological time series. Stations within a 0.5° buffer around the basin boundaries were also considered. Thiessen polygons were used to compute the basin-averaged data for each day to account for missing data events on one or multiple stations on given days. An alternative dataset provided by Natural Resources Canada (NRCan) uses an average of a daily 10 km gridded climate dataset for the given catchments, resulting in no missing data for the period 1950–2010. A comprehensive analysis of the NRCan gridded dataset can be found in Hutchinson *et al.*, 2009.

Results

The 698 basins that compose the CANOPEX database are shown in Figure 1, which also shows the catchments' annual precipitation based on the NRCan dataset.

The CANOPEX watersheds are well distributed over the territory, except for the prairies area due to the difficulty of delineating basins in flat topography and lack of measurements as well as in areas where rivers are regulated. CANOPEX catchment areas vary from small drainage areas of 292 km^2 (smaller than the 300 km^2 threshold, but within the 10% difference allowance with HYDAT) to immense regions covering more than 120 000 km^2. The database also covers and identifies 11 kinds of different climates according to the 1901–2010 Koppen–Geigger classification (Chen and Chen, 2013).

Format and Availability of CANOPEX Datasets

For each of the 698 CANOPEX database watersheds, daily precipitation and minimal and maximal temperatures were extracted from both observations and NRCan gridded data. Daily flow values, watershed boundaries and hydrometric station metadata are also made

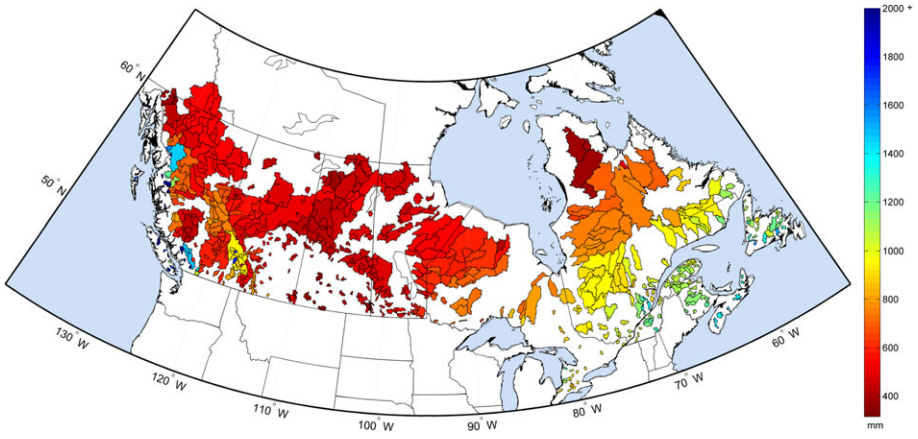

Figure 1. Spatial distribution of the 698 CANOPEX catchments and their mean total annual precipitation from the NRCan database (mm)

available. The CANOPEX data can be downloaded in MATLAB or ASCII formats on the project website: http://canopex.etsmtl.net. CANOPEX data are provided freely for non-commercial use. More information about terms and conditions are available on the website.

Possible Uses

Canadian model parameter experiment is a large-scale hydrometeorological database designed to facilitate hydrologic studies over Canada. Its 698 basins, for instance, could be used in model structure comparisons, namely testing how different hydrological models react according to the type of climate and characteristics of tested watersheds. Some of the catchments are part of the Reference Hydrometric Basins Network and could be used to study climate change impacts (Whitfield *et al.*, 2012). Moreover, ensemble streamflow prediction could be processed over a large number of basins, which help to assess robustness of models and procedures. Hydrologists are encouraged to use this free database to help improve current knowledge and methods in northern climates where flow regimes are dominated by snowmelt. Studies on the impacts of objective function selection and calibration algorithm due to the strong peak floods, snow-model performance and model parameter regionalization in this particular climate would be most welcome.

Acknowledgements

We are grateful to Bernhard Lehner for his advice concerning the watershed boundaries delineation, to M. Pierre Brigode for his help with the basin shapefiles and to two anonymous reviewers whose comments improved the dataset. Streamflow data and hydrometric station metadata (except for watershed areas) provided in the CANOPEX database are extracted from the HYDAT database. Weather station data was supplied by Environment Canada's Climate archives. Both products are official works that are published by Environment Canada. This reproduction has not been produced in affiliation with the Government of Canada but respects the terms of use of both products. This project is partly funded by the Natural Science and Engineering Research Council (NSERC) of Canada.

References

Chen D, Chen HW. 2013. Using the Köppen classification to quantify climate variation and change: An example for 1901–2010. *Environmental Development* 6: 69–79. DOI:10.1016/j.envdev.2013.03.007

Hutchinson MF, McKenney DW, Lawrence K, Pedlar JH, Hopkinson RF, Milewska E, Papadopol P. 2009. Development and testing of Canada-wide interpolated spatial models of daily minimum-maximum temperature and precipitation for 1961–2003. *Journal of Applied Meteorology and Climatology* 48: 725–741.

Lehner B, Verdin K, Jarvis A. 2006. *Hydrological data and maps based on Shuttle elevation derivatives at multiple scales (HydroSHEDS)-Technical Documentation*, World Wildlife Fund US, Washington, DC, Available at http://hydrosheds.cr.usgs.gov.

Lehner B, Verdin K, Jarvis A. 2008. New global hydrography derived from spaceborne elevation data. *Eos, Transactions AGU* 89(10): 93–94.

Schaake JC, Duan Q, Smith M, Koren V. 2000. Criteria to select basins for hydrologic model development and testing. *Preprints, 15th Conference on Hydrology*, 10–14.

Schaake J, Cong S, Duan Q. 2006. The US MOPEX data set, *IAHS publication*, *307*, 9.

Whitfield PH, Burn DH, Hannaford J, Higgins H, Hodgkins GA, Marsh T, Looser U. 2012. Reference hydrologic networks I. The status and potential future directions of national hydrologic networks for detecting trends. *Hydrological Sciences Journal* 57(8): 1562–1579.