

This is an author produced version of a paper published in Hydrological Processes. This paper is the final submitted version before typesetting and journal pagination.

An edited version of this paper was published by Wiley. Copyright (2016) Wiley.

The published paper can be found at DOI: 10.1002/hyp.10880.

Full citation for the published paper :

Arsenault, R., Bazile, R., Dallaire-Ouellet, C. and Brissette, F. (2016). CANOPEX: A Canadian hydrometeorological watershed database. Hydrological Processes, 30(15), 2734-2736. doi:10.1002/hyp.10880.

The CANOPEX dataset can be downloaded at: <http://canopex.etsmtl.net/>

CANOPEX : A Canadian hydrometeorological watershed database

Richard Arsenault⁻¹, richard.arsenault.1@ens.etsmtl.ca
Rachel Bazile⁻¹, rachel.bazile@gadz.org
Camille Ouellet Dallaire⁻², camille.o.dallaire@gmail.com
François Brissette⁻¹, françois.brissette@etsmtl.ca

⁻¹Department of Construction Engineering, École de technologie supérieure, Université du Québec, 1100 Notre-Dame Street West, Montreal, Québec H3C 1K3, Canada

⁻²Department of Geography, University McGill, 805 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6

Running Head: “CANOPEX : A Canadian hydrometeorological watershed database”

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 NRCan 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-modeling, 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.

Methods

Selection and filtering of streamflow gauges

The methods used to select suitable catchments for the 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 itself 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 Shaacke 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 300km² due to delineation uncertainty and daily time step.

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

2.2 Watersheds delineation

HYDAT does not contain catchment boundaries for its hydrometric stations. This is one of the key problems in large-scale hydrological studies in Canada since the hydrometric and weather data are publicly available. This project relied on HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales) [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 to 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.

2.3 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 10km 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 Hutchison et al. 2009.

3. 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.

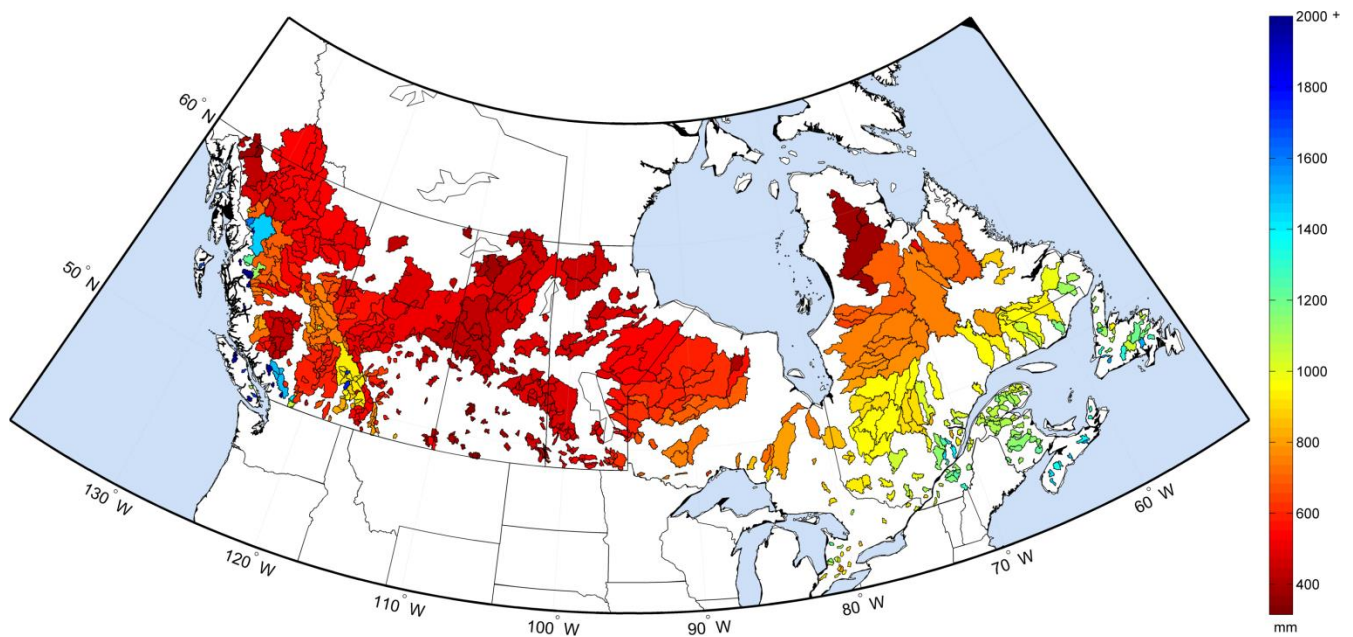


Figure 1: CANOPEX database basin locations and average annual precipitation according to the NRCan dataset.

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 km² (smaller than the 300km² threshold, but within the 10% difference allowance with HYDAT) to immense regions covering more than 120000 km². The database also covers and identifies 11 kinds of different climates according to the 1901-2010 Koppen-Geiger classification (Chen and Chen, 2013).

4. Format and availability of CANOPEX datasets

For each of the 698 CANOPEX database watersheds, daily precipitation, 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 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.

5. Possible uses

CANOPEX 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 (RHBN) and could be used to study climate change impacts (Whitfield et al. 2012). Moreover, ensemble streamflow prediction could be processed over a large amount 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 2 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. and H. W. Chen, 2013: Using the Köppen classification to quantify climate variation and change: An example for 1901–2010. *Environmental Development*, 6, 69-79, 10.1016/j.envdev.2013.03.007.
- Hutchinson, M. F., McKenney, D.W., Lawrence, K., Pedlar, J.H., Hopkinson, R.F., Milewska, E., Papadopol, P. (2009), Development and testing of Canada-Wide Interpolated Spatial Models of Daily Minimum-Maximum Temperature and Precipitation for 1961-2003, *American Meteorological Society*,(April): 725-741.
- Lehner, B., Verdin, K., Jarvis, A. (2008), New global hydrography derived from spaceborne elevation data, *Eos, Transactions, AGU*, 89(10): 93-94.
- 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>.
- Nash, J., & Sutcliffe, J. V. (1970), River flow forecasting through conceptual models part I—A discussion of principles, *Journal of hydrology*, 10(3), 282-290.
- Schaake, J. C., 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, P.H., Burn, D.H., Hannaford, J., Higgins, H., Hodgkins, G.A., 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.