www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

A practitioner-oriented regional hydrology data product for use in site-specific hydraulic applications

Kevin R. Shook[1], Zhihua He[1], John W. Pomeroy[1], Chris Spence[2] & Colin J. Whitfield[3,4] ✉

In the 465,000 km^2 Canadian Prairies ecozone, robust hydrological input data for hydraulic model applications are uncommon because of the sparse monitoring network and the intermittently connected stream network. New hydrological datasets can offer a valuable advancement for making water management decisions and designing infrastructure in this water stressed region. The Prairie Hydrology Design and Analysis Product (PHyDAP) was created to address existing limitations, and provides a comprehensive regional dataset for use in hydraulic modelling applications. PHyDAP is a collection of outputs from a physically based hydrological modelling framework, run for periods ranging from 38 to 150 years, according to three climate forcing datasets. The dataset includes vertical and lateral fluxes (rainfall, snowmelt, upland runoff, and open water evaporation) and basin streamflow, at hourly or 3-hourly intervals for the >4000 small basins of approximately 100 km^2 that span the region. This contribution describes the motivation for this work and methodology used to derive the data product, summarizes the data and its accessibility, and provides an overview of potential use cases.

Background & Summary

Predicting the state of water resources in water insecure regions is of critical importance. Water resource, hydrological, and hydraulic models can all be effective tools in addressing this problem, however limitations of these models can be an impediment to producing high quality information to support practitioner decision-making. For example, some model structures may not include important processes or behaviours predominant in some regions (e.g., snow redistribution, radiation driven snowmelt, infiltration to frozen ground, or intermittent streamflow applicable in cold semi-arid regions), with the result that different model types are appropriate for distinctive applications and model selection is important. Hydrological models cannot be expected to provide strong representation of streamflow within highly channelized and engineered drainage networks (e.g., with culverts, roads, gates or pumps), or reliably simulate how modifications to the network will impact behaviour. Conversely, hydraulic models are not well suited for representing landscapescale hydrological behaviour. New approaches that provide flexibility while better representing both hydrological and hydraulic processes to improve water resource prediction are needed.

The Canadian Prairie of western Canada is a region with well-known issues of water insecurity[1] where these modelling challenges are apparent. The region has undergone widespread environmental change associated with proliferation of agricultural development and management during the 20th century; tillage, fallowing, and cropping practices have all changed substantively during and since this time, with consequences for water storage and hydrological regimes[2–4]. Widespread wetland drainage to expand the agricultural land base is also common to the region, with hydrological and other environmental impacts of this practice well documented[5]. Coupled with ongoing land management impacts to water resources are the effects of climate change. The region is experiencing strong climate change and is projected to undergo increases in mean annual temperature of 1.9–6.5 °C by 2100, while future changes in annual precipitation depths remain less certain[6]. Associated changes to precipitation phase[7], and the potential for more multiple-day rainfall events[8] are anticipated. Most climate model outputs for the region suggest greater precipitation, especially in winter, and with a greater fraction of

[1]Centre for Hydrology, University of Saskatchewan, Saskatoon, SK, Canada. [2]Environment and Climate Change Canada, Saskatoon, SK, Canada. [3]School of Environment and Sustainability, University of Saskatchewan, Saskatoon, SK, Canada. [4]Global Institute for Water Security, University of Saskatchewan, Saskatoon, SK, Canada. ✉e-mail: colin.whitfield@usask.ca

this precipitation falling as rain[9]. Collectively, these changes add uncertainty to water management decisions, including decisions around infrastructure design.

Added to this challenge is the complex hydrography of the region that makes hydrological modeling of its basins difficult[10–12]. Depressional storage and large non-effective runoff contributing fractions, blowing snow and snow redistribution, snowmelt dominated runoff, frozen soil dynamics, and sparse meteorological data all contribute to this complexity[13]. While physically based hydrological modelling approaches have been successfully employed in some small basins[14,15], deploying physically based hydrological models over the entire Prairie ecozone is challenging[12,16].

In the absence of robust methods to support hydraulic modelling, runoff frequency and magnitude estimates used to design Canadian Prairie infrastructure are commonly estimated from historical streamflow gauge records. Applying this method has four key disadvantages. First, large and well-gauged rivers such as the Saskatchewan River and its tributaries derive the vast majority of their flow from the Canadian Rockies and foothills, and so those flow records cannot be used to estimate local prairie runoff. Second, the very low density of gauges measuring streamflow from prairie basins reduces the availability of representative data from comparable basins (i.e., area, effective drainage area, wetland distribution, soils, land use)[4]. Third, assuming historical flows can be used to design infrastructure for the future neglects that neither climate nor widely used agricultural practices are constant[8,17,18]. Finally, the Rational Method and Soil Conservation Service Curve numbers often used to calculate runoff values require assumptions that are typically violated in Canadian Prairie basins (e.g., due to predominance of snowmelt runoff, and a sizeable portion of water being captured by topographic depressions).

Herein, we present a novel dataset that addresses the limitations of current methods, and supports hydraulic investigations of engineered infrastructure. The Prairie Hydrology Design and Analysis Product (PHyDAP) is an ensemble of hydrological outputs generated via long-term simulations of more than 4000 small basins across the Canadian Prairie region that can be used as input to hydraulic models. These hydrological outputs, including rainfall, snowmelt, upland runoff, open water evaporation and basin stream discharge, were generated using a physically-based hydrological modelling platform via long-term model runs, as a means of producing robust data for forcing hydraulic models. Importantly, because many regulations for the design and operation of hydraulic infrastructure are based on return-period flows, the multi-decadal simulation period, produced using three long-term climate forcing datasets from historical and future climate periods, supports the use of stochastic rather than deterministic approaches. The outputs can be used as inputs to hydraulic models to estimate changes in runoff and flooding at specific locations and under a variety of design scenarios, offering improvement over existing techniques which suffer from the limitations described above.

Methods

Study domain. This study focused on the Canadian Prairie region (Canadian Prairies ecozone), spanning 465,000 km^2 across southern parts of the provinces of Alberta, Saskatchewan, and Manitoba (Fig. 1). Once an extensive grassland biome, since the late 19th century much of the region has been converted to support agricultural uses[19]. The climate is variable and alternates between dry and wet phases, ranging from semi-arid to sub-humid conditions, respectively, with mean annual (1970–2000) precipitation ranging from 350 to 650 mm, generally increasing to the east[20]. Potential evapotranspiration of 600–800 mm yr^{-1} exceeds precipitation in most years[21]. Mean annual temperatures vary from 1–6 °C, with the southwest being warmer[22,23] (http://climate.weather.gc.ca/climate_normals/index_e.html).

Relief across the region is low, and except for several large uplands, topography can vary from flat (glaciolacustrine deposits) to hummocky (moraine and till plains). The region deglaciated approximately 10,000 years before present, establishing a landscape with numerous small depressions over a low permeability glacial clay till substrate. Depressions on the landscape are important for local hydrology[24] and often feature wetlands which can be sites of depression focussed groundwater recharge[25]. Owing to a short period since glaciation, a relatively dry climate and the hummocky terrain, drainage networks remain poorly established, and vast expanses of the region are often hydrologically disconnected from higher order streams[4,26,27].

Surface hydrology is strongly influenced by long winters. This allows snow to be redistributed across the landscape, accumulating in depressions[28,29]. The snowmelt period, when infiltration into frozen soils is largely determined by moisture content at freeze-up and subsequent ice layer formation, tends to promote upland runoff at the expense of infiltration[30]. This makes it the key hydrological period during which depressional wetlands fill, merge and spill, and surface runoff networks can be activated[31,32]. During summer when monthly precipitation is typically highest (e.g., June and July), infrequent rainfall coupled with deep soils, high evapotranspiration and consequent large infiltration capacities contribute to vastly reduced upland runoff.

Drainage basin classification and virtual basin modelling framework. In this analysis, we use a classification-based virtual basin modeling approach, as it permits modelling of the variables of interest: open-water evaporation (wetland ponds and stream channels, or lakes), upland runoff (above wetlands, ponds and lakes), rainfall, and snowmelt, with sufficient resolution to produce data for site-specific investigations. Basin streamflow (discharge) is also included, albeit with caveats discussed below. Within the Prairie ecozone, topographic delineation of basins (watersheds) is available from the HydroSHEDS global dataset[33], with boundaries of ~100 km^2 basins derived using the Shuttle Radar Topographic Mission (SRTM) digital elevation model (DEM) calculated at a 15 arc-second resolution. These basins ($n = 4175$) lying entirely within the Prairies ecozone domain, and excluding urbanized areas are the basis for the analysis.

The foundation of the modelling approach is a classification of these headwater basins. This classification used a suite of biophysical parameters (e.g., elevation, surficial geology, wetland distribution, land cover, soil type) in a hierarchical clustering of principal components[34] which identified seven basin classes for the region

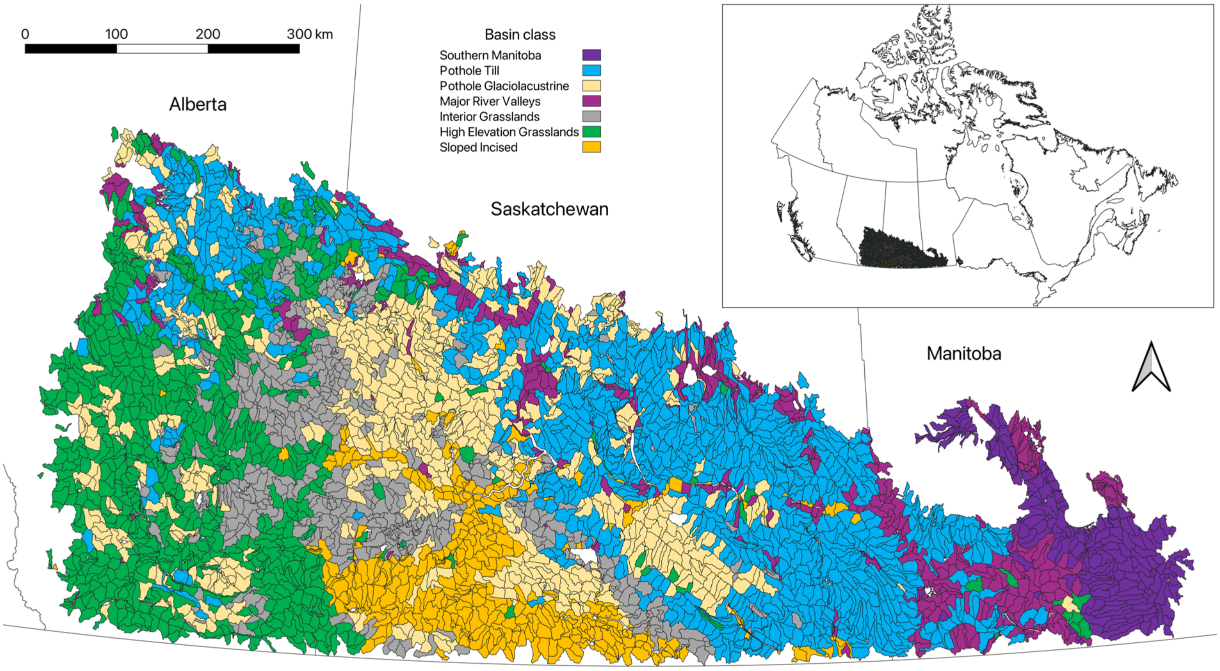

Fig. 1 Map of the >4000 small HydroSHEDS basins[33] in the Canadian provinces of Alberta, Saskatchewan and Manitoba used in the modelling framework to create PHyDAP. Individual basins are shown, with colours representing seven different classes of basin. The inset map shows the PHyDAP domain in Canada.

(Fig. 1). The classification described in ref. [13] was used to build PHyDAP, as it best serves as the basis for hydrological modelling. In addition, HydroSHEDS basins representing 346 lakes (105 of which are named) having areas between 2 and 520 km^2 were also selected for modelling of vertical fluxes (rainfall, snowmelt, evaporation) on these water bodies across the region.

A virtual representation of the typical basin in each of the seven basin classes was constructed using the median characteristics identified during the classification exercise. The basin classes feature characteristic differences[34]. Southern Manitoba features predominantly black soils, high cropland and low wetland coverage with mild slopes. The two pothole classes (Pothole Till and Pothole Glaciolacustrine) have high non-effective areas linked to high coverage of depressional wetlands, with high cropland cover and low areas of unmanaged grasslands. These two classes differed in surficial geology, with glaciolacustrine deposits with higher clay and silt content soils being an important feature of Pothole Glaciolacustrine basins. Major River Valleys basins are generally found along the banks of major rivers in the region, and have relatively high slopes and higher incidence of black soils. Three of the classes (Interior Grassland, High Elevation Grassland, and Sloped Incised) are considered grassland dominated, with Interior Grassland commonly featuring brown soils, and having a large non-effective fraction with large areas of lower elevation than the basin outlet. High Elevation Grassland basins feature higher slopes and smaller non-effective areas than those of the Interior Grassland class, while Sloped Incised Basins are relatively steep, with low wetland density and higher effective areas. The basin characteristics contribute to different hydrological behaviour among the classes[35].

There is prior precedent for use of virtual approaches in the region, with earlier work used successfully to examine spatial variability of evapotranspiration over the Prairies[36]. The virtual modelling foundation for PHyDAP has been described in detail before[13,35,37]. Briefly, the Cold Regions Modelling Platform (CRHM[28,38]) was used to develop virtual basin models of each of the seven basin classes (Fig. 1). CRHM is modular, process-based, spatially semi-distributed hydrological model, featuring key cold region and warm season hydrological processes of importance in this region. These processes include wind redistribution and sublimation of snow, energy balance snowmelt, infiltration to frozen and unfrozen soils, groundwater dynamics, crop growth and evapotranspiration, among others. Virtual basins for each class were configured using multiple hydrological response units (HRUs), discrete areas each having a defined land use and soil type that contribute to a characteristic hydrological behaviour of the HRU. HRUs used in each virtual basin model included fallow fields, cultivated fields, grasslands, shrublands, woodlands, wetlands, and stream channels, with their areas established using median land covers observed across all basins of that class[35]. For lakes, similar CRHM models were developed but limited to the lake body itself (excluding terrestrial HRUs), which allowed simulation of the evaporation, rainfall, and snowmelt across the lake surface only. Modellers wishing to simulate the basin of a lake should use PHyDAP outputs for the upland basins surrounding the lake.

Model forcings and implementation. Simulations of CRHM virtual basins for 4175 basins and 346 lakes were run using three long-term forcing datasets available for the region (Table 1) using the virtual basin structures in ref. [35]. Use of long-term (>30 years) climate forcing datasets here is important to enable modellers to

Name	Acronym	Nature of data	Period	Data interval	Grid resolution	Ensembles
Regional deterministic reforecast system[46]	RDRS	Gridded weather stations	1981–2018	Hourly	10 km	NA
CanRCM4_Cor_WFDEI-GEM-CaPA[9,47]	CWGC	Combined reanalysis and downscaled and bias adjusted future climate simulations	1951–2100	3-hourly	0.125°	15
Fifth generation ECMWF atmospheric reanalysis of the global climate[48,49]	ERA5	Gridded reanalysis	1950–2020	Hourly	31 km	NA

Table 1. Details of the three long-term climate forcing datasets available for the study domain and used to produce PHyDAP.

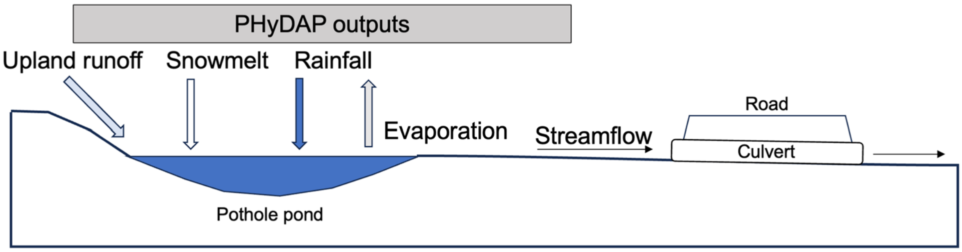

Fig. 2 Conceptual diagram illustrating the PHyDAP outputs including vertical fluxes (block arrows: upland runoff, snowmelt, rainfall, evaporation) and streamflow that can be used in hydraulic modelling applications (line arrows), such as transportation infrastructure design.

find return-period values. Multiple sets of forcings best equip practitioners with PHyDAP outputs matching the climate forcings that are of most utility for their targeted application. Future forcings of the CWGC dataset[9] follow the RCP 8.5 high emission scenario as it best matches observed increases in atmospheric greenhouse gas concentrations.

Basin-specific mean fluxes (precipitation, air temperature, vapour pressure, wind speed and incoming shortwave radiation) were derived from the gridded forcings in two steps. In the first, the global gridded values were clipped to the extent of the Canadian Prairies using the Python program datatool (https://github.com/kasra-keshavarz/datatool). In the second step, the gridded data were interpolated and averaged over each ~100 km^2 basin, using the Python program EASYMORE (https://github.com/ShervanGharari/EASYMORE/tree/main/env). The horizontal grid resolution of these climate forcings (Table 1) is comparable to the average basin size, such that each basin, with few exceptions for smaller adjacent basins using ERA5 data, will have unique climate forcings following interpolation over the basin areas. Using these basin-specific climate forcings, 4521 CRHM simulations for each of the three climate forcings (Table 1) were batch-run on Compute Canada's high performance computing system to generate three sets of PHyDAP time-series outputs, one for each of the RDRS, CWGC and ERA5 climate forcings. Each time-series consists of hourly or 3-hourly (see Table 1) rainfall (computed by CRHM from RDRS, CWGC and ERA5 precipitation data) specific to individual basins, and CRHM virtual basin simulations of snowmelt, evaporation, and upland runoff (Fig. 2). The PHyDAP dataset includes HYBAS IDs and basin classifications for each upland basin, HYBAS IDs for each lake, and output files for each forcing dataset (and ensemble member) and upland basin flux, as well as lake fluxes for each forcing dataset (Table 2). Snowmelt is reported as the flux for an average sized wetland HRU, while evaporation fluxes from open water are the same for all depressions. The term "runoff" is used by many hydrologists to refer to stream discharge expressed as a uniform basin depth. The PHyDAP "upland runoff" refers to the computed depth of water running off an upland surface in a time interval. It is produced by rainfall and snowmelt on the surface, and accounts for infiltration loss to the soils, which are often deeply frozen. The PHyDAP upland runoff is taken from HRUs representing cropped fields and pastures (the majority of the area of each basin). Thus, in a hydraulic model, one need only multiply the upland runoff depth by the area contributing flow to get the volume of water contributed to a water body (wetland pond, stream channel) where alteration by storage attenuation can occur. No upland runoff is simulated for lake basins. The PHyDAP dataset also includes streamflow estimates for each of the 4175 classified basins (see below).

Data Records

PHyDAP is available at the Federated Research Data Repository (FRDR), which is custom built for preservation and sharing of large Canadian research data sets. The data are available as three analogous sets corresponding to the three forcings described above; a netCDF file for each climate forcing and output variable was created, with file sizes for RDRS, CWGC and ERA5 climate forcings totalling approximately 180, 913, and 73 GB, respectively. Metadata providing Hydrosheds and basin class information are also provided. The data are available at: https://doi.org/10.20383/102.0694 (ref. [39]).

Rainfall, snowmelt, evaporation, upland runoff, discharge. PHyDAP provides, for the first time in this region, estimates of vertical fluxes using local (small basin-specific) climate forcings. Basin rainfall is computed by CRHM using basin-interpolated precipitation for each climate dataset (data not shown) and a physically

Data type	Cases	Variables	File name[a]
Upland basins	4175	HYBAS ID, Basin class number, Basin class name	PHyDAP_basin_data.csv
Lakes	346	HYBAS ID, Lake name	hydrosheds_prairie_named_lakes.csv
CWGC evaporation	438280	HYBAS ID, time[b], evaporation	PHyDAP_CWGC_rxxxxxxx_3hourly_evaporation.nc
CWGC rainfall	438280	HYBAS ID, time, rainfall	PHyDAP_CWGC_rxxxxxxx_3hourly_rainfall.nc
CWGC upland runoff	438280	HYBAS ID, time, upland runoff	PHyDAP_CWGC_rxxxxxxx_3hourly_upland_runoff.nc
CWGC snowmelt	438280	HYBAS ID, time, snowmelt	PHyDAP_CWGC_rxxxxxxx_3hourly_snowmelt.nc
CWGC streamflow	438280	HYBAS ID, time, streamflow	PHyDAP_CWGC_rxxxxxxx_3hourly_streamflow.nc
CWGC lake fluxes	438280	HYBAS ID, time, evaporation, rainfall, snowmelt	PHyDAP_CWGC_rxxxxxxx_3hourly_lake_fluxes.nc
RDRS evaporation	341760	HYBAS ID, time, evaporation	PHyDAP_RDRS21_hourly_evaporation.nc
RDRS rainfall	341760	HYBAS ID, time, rainfall	PHyDAP_RDRS21_hourly_rainfall.nc
RDRS upland runoff	341760	HYBAS ID, time, upland runoff	PHyDAP_RDRS21_hourly_upland_runoff.nc
RDRS snowmelt	341760	HYBAS ID, time, snowmelt	PHyDAP_RDRS21_hourly_snowmelt.nc
RDRS streamflow	341760	HYBAS ID, time, streamflow	PHyDAP_RDRS21_hourly_streamflow.nc
RDRS lake fluxes	341760	HYBAS ID, time, evaporation, rainfall, snowmelt	PHyDAP_RDRS21_hourly_lake_fluxes.nc
ERA5 evaporation	622344	HYBAS ID, time, evaporation	PHyDAP_era5_hourly_evaporation.nc
ERA5 rainfall	622344	HYBAS ID, time, rainfall	PHyDAP_era5_hourly_rainfall.nc
ERA5 upland runoff	622344	HYBAS ID, time, upland runoff	PHyDAP_era5_hourly_upland_runoff.nc
ERA5 snowmelt	622344	HYBAS ID, time, snowmelt	PHyDAP_era5_hourly_snowmelt.nc
ERA5 streamflow	622344	HYBAS ID, time, streamflow	PHyDAP_era5_hourly_streamflow.nc
ERA5 lake fluxes	622344	HYBAS ID, time, evaporation, rainfall, snowmelt	PHyDAP_era5_hourly_lake_fluxes.nc

Table 2. Metadata summary showing key information for each of the data components making up the PHyDAP dataset. [a]Note that for CWGC, which has 15 ensemble members, there is a corresponding file for each member and type of data (rxxxxxxx serves as a placeholder for individual ensemble member numbers). [b]Note that time is represented as hours since 1900-01-01 00:00, and readily converted to datetime when reading the files (e.g. in R, using ch_read_PHyDAP).

based psychrometric energy balance calculation of precipitation phase[40]. Other vertical fluxes illustrate the nature of the PHyDAP outputs from the virtual basin modelling framework.

Snowmelt in the study domain is an important component of the water balance, as blowing snow transports snow from more exposed or short vegetation areas, with accumulation in vegetated or wind-sheltered depressions[41]. Snowmelt is the key hydrological period, as the influx of water can lead to activation of surface flow networks. With simulations spanning up to 150 years (e.g., CWGC), it is possible to contrast current or historic with predicted future fluxes using the PHyDAP outputs. In this region experiencing notable climate warming, potential changes in future snowmelt amount are evident at some basins (e.g., Fig. 3).

PHyDAP also includes the first regional estimates of open water evaporation rates. For example, PHyDAP open water evaporation from small waterbodies varies from less than 400 to more than 1000 mm yr^{-1} (under RDRS climate forcings; Fig. 4). This spatial variability demonstrates some regional climate influences. For example, lower evaporation rates were simulated for the higher elevation and cooler Cypress Hills area in the southwest of the study domain. These rates contrast the high evaporation rates in surrounding basins, and especially the drier and warmer areas to the north. It is important to note that while one open water evaporation estimate is discussed here (for RDRS), this and the other vertical fluxes will vary depending on the climate forcing dataset used.

One of the important findings from earlier virtual basin modelling of hydrological behaviour in the region has been the importance of place[35]. This is clearly apparent again in PHyDAP for upland runoff. The dataset highlights characteristic differences in upland runoff amongst basin classes (Fig. 5). Because of key differences in runoff contributing areas, wetland coverage, stream network density and other parameters, some basin classes are more efficient at generating upland runoff to the wetland complex. The different basin classes provide a useful tool for typifying the hydrological behaviour of basins across the region. Furthermore, PHyDAP has proven successful in illustrating how local climate contributes to variability in hydrological fluxes within a basin class. For example, while the same virtual basin is used for all basins in each class, long-term upland runoff amounts vary by <50 mm yr^{-1} for Sloped Incised basins, by ~75 mm yr^{-1} for both Southern Manitoba and Pothole Till basins, and by 100 or more mm yr^{-1} for (Interior and High Elevation) Grassland basins (Fig. 5). These differences are due to hydrological responses to different (RDRS) climate forcings across basins within each class. By capturing this variability, PHyDAP represents a major advance in providing foundational data for local scale hydraulic modelling applications.

Streamflow. In addition to the vertical fluxes, PHyDAP also includes basin streamflow expressed as a depth over the basin. Very importantly, basin streamflow is a function of the interactions among all the HRUs and includes effects of the contributing fraction of the basin class. Because of the virtual basin approach, however, these simulations cannot be expected to accurately represent any specific local basin streamflow, except in cases where the basin (or basins) of interest have effective area fractions and wetland areas that closely align with those

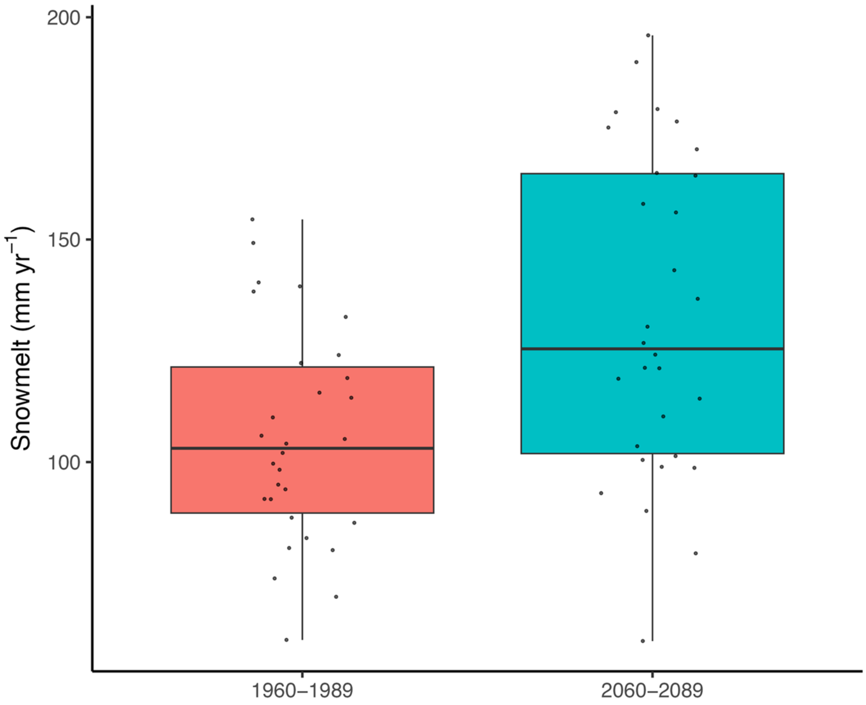

Fig. 3 One example of PHyDAP simulated snowmelt for two climate windows, for a single basin of the Southern Manitoba class, according to CWGC climate forcings from a single ensemble member. Each point in the boxplots represents a different year from the time window.

for the median of the class (see technical validation below). While streamflow data have been included in this data product, these estimates should be used with an abundance of caution.

Technical Validation

Previous evaluations have demonstrated the ability of CRHM to represent hydrological processes in the region, including snowmelt generated streamflow[42]. It has also been demonstrated that the basin classification–virtual basin approach, extended here to a regional analysis, is robust in capturing the major hydrological state and fluxes of each basin class, including annual and monthly streamflows[35], peak snow water equivalent[13], snow depth[35], and observed wetland pond depth[37]. Additionally, CRHM has been shown to provide reasonable simulation of evapotranspiration at multiple locations in the region under both wet and dry conditions[36]. A comprehensive summary of CRHM performance for a broad selection of variables in cold regions basins in the Prairie and internationally highlights the model's robustness[38].

Parameters used in developing PHyDAP were selected from the outcomes of over 60 years of hydrological research observations in Prairie basins by the Division of Hydrology and later the Centre for Hydrology, University of Saskatchewan. The parameter sets used within the CRHM representation of each basin class were not calibrated for two reasons. First, a virtual basin has no geographic location, so it cannot be calibrated against any specific gauged basin. Second, there are almost no unregulated gauged basins of ~100 km^2 in the ecozone, so the ability to simulate hydrological fluxes without calibration is advantageous. To quantify the ability of CRHM to simulate hydrological fluxes refs. [13,35] compared mean monthly simulated streamflow depths to Water Survey of Canada streamflow data from gauges in proximity to climate stations from which data were used to force virtual basin models. Mean bias ranged from −0.71 to 0.54, with the best performance in the Pothole Till basin class (0.14), and the worst performance in the High Elevation Grasslands class (−0.71). Mean bias for basins of the Southern Manitoba, Pothole Glaciolacustrine, Major River Valleys, Interior Grasslands, and Sloped Incised classes were 0.54, −0.37, −0.45, −0.42, and −0.58, respectively[35]. Differences can be attributed to the fact that virtual basin models were structured and parameterized using the median characteristics of each basin class. These basin characteristics are different to those in the specific gauged basins used in evaluating the virtual basins, so differences are expected. It is also important to note that the climate station data used in the virtual basin models are different to the climate conditions experienced at the gauged basins, which will contribute to differences in streamflow.

This highlights an important limitation of this approach. The model outputs are expected to represent the hydrological behaviour of a typical basin within a class. This also means the outputs are unlikely to explicitly represent any specific basin within a class. They do not capture the physiographic variability in basin structure within a class, only among classes. For example, a major controlling factor on runoff generation in the Prairie ecozone is the fraction of non-effective area contributing to the basin outlet. As discussed earlier, depression storage capacity can intercept much hillslope runoff before it reaches the basin outlet, and the location and distribution of this storage capacity controls the size of the area contributing water to the basin outlet. The size of

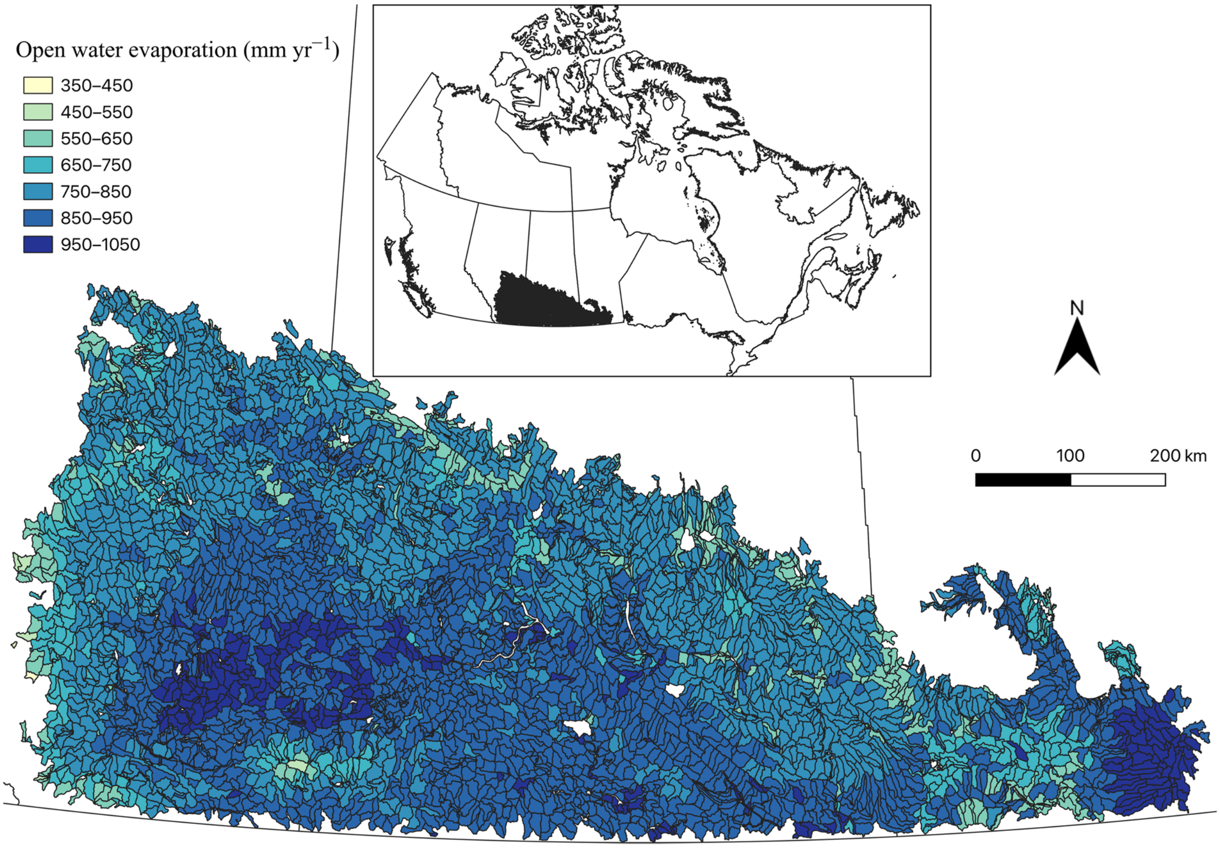

Fig. 4 PHyDAP long-term (1980–2018) mean annual open water evaporation for all 4175 upland basins, using RDRS climate forcings. The location of the PHyDAP domain in Canada is shown in the inset image.

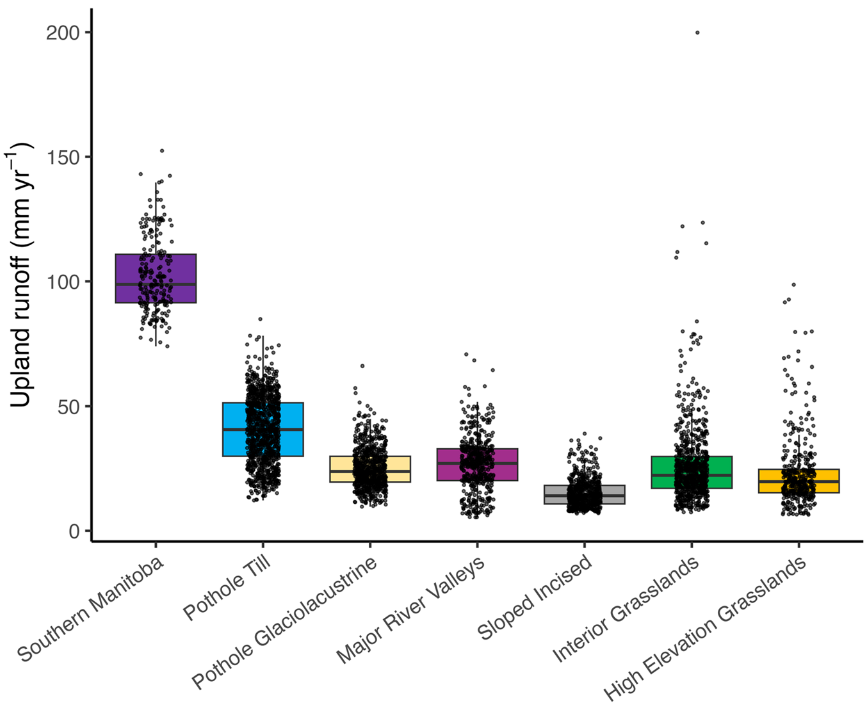

Fig. 5 Long-term mean annual upland runoff for all basins in each of the seven classes, according to RDRS (1980–2018) climate forcings. Each point in the boxplot for a particular watershed class represents a different basin subject to its own local climate.

this area has been referred to as the effective basin area. Conversely, the area that is disconnected from the basin outlet is the non-effective area, and the portion of that area in the gross drainage area is the non-effective fraction; of which there is considerable variability within each basin class. The difference between the virtual basin non-effective fraction and the actual fraction in any basin of interest will indicate whether the virtual basins will tend to over or underestimate streamflow.

Model outputs capture a major source of variability, climate, by applying distributed forcing data across all 4521 basins. The original scientific articles (Table 1) discuss the strengths and weaknesses of each climate forcing dataset. We encourage practitioners using PHyDAP to consider carefully which of the datasets derived from the three sets of climate forcings are best suited to their application.

Usage Notes

Given the challenge of working with datasets of this size, and in accessing complex netCDF files, we have implemented a purpose-built R function (ch_read_PHyDAP) to assist practitioners with accessing PHyDAP outputs. This allows users to extract data for one or more basin locations simultaneously from a netCDF file. It is worth cautioning, however, that given the large netCDF file sizes, users are likely to encounter insufficient memory if attempting to read and process these data for all basins simultaneously (e.g., aiming to extract one output parameter for all basins). After reading data, these can be efficiently summarized from sub-daily to longer timesteps, or otherwise manipulated as needed by the user for their intended purposes. We warn users of this dataset that the framework which allows regionalization of this modelling approach also means that basin-specific vertical and lateral fluxes are subject to some uncertainty; differences among outputs from different climate forcings have been identified, and streamflow predictions should be used with caution.

Types of applications. As described above, the intent in developing PHyDAP is to provide data sets for forcing hydraulic models of small regions (basins, sub-basins, or even individual fields) within the Canadian Prairies. Because the PHyDAP dataset is very new, it has not yet been used extensively. CRHM model outputs, however, have previously been coupled unidirectionally to hydraulic models in the Canadian Prairies, with some success. CRHM outputs have been used to force a small-scale hydraulic model[43]. In this application in Smith Creek basin near Langenburg, SK, a Water Survey of Canada gauge that defines the research basin outlet was in a stream valley immediately upstream of a culvert through a road. CRHM alone was unable to simulate the complex hydraulics. To overcome this limitation, a Storm Water Management Model (SWMM[44]) of the stream valley and culvert site ingesting CRHM runoff outputs was developed to capture the system hydraulic behaviour. This allowed successful simulation of gauged flows[43]. Also, CRHM outputs from a virtual basin model, although not one of the PHyDAP datasets, have been used to force an application of SWMM to a small (~3 km^2) area consisting of two fields on separate sides of a road (K. Altraide, unpublished data). The fields were connected by a culvert through the roadbed, which was frequently damaged by flooding. By running SWMM iteratively, it was possible to optimise the design of the road culverts. Notably, PHyDAP was recently used as input to a hydraulic model to successfully predict the spatial extent of a flooding event that occurred in a rural area featuring a major highway with numerous culverts running underneath it[45].

Code availability

An R script was created for practitioners who use PHyDAP to facilitate working with this very large dataset. The function ch_read_PHyDAP is available at https://cshs.cwra.org/en/hydrology/ and can be used for extraction of basin-specific data from netCDF files. Use of this tool will assist users in efficiently extracting data specific to one or more basins for use in local applications for which PHyDAP was intended. The data are formatted for ease of use with several common hydraulic modelling platforms. CRHM parameter files for the seven virtual basin models (e.g., ref. [35]) used herein are available from the authors on request.

Received: 14 March 2024; Accepted: 1 October 2024;
Published online: 14 October 2024

References

1. Gober, P. & Wheater, H. S. Socio-hydrology and the science–policy interface: a case study of the Saskatchewan River basin. *Hydrology and Earth System Sciences* **18**, 1413–1422 (2014).
2. Awada, L., Lindwall, C. W. & Sonntag, B. The development and adoption of conservation tillage systems on the Canadian Prairies. *International Soil and Water Conservation Research* **2**, 47–65 (2014).
3. van der Kamp, G., Hayashi, M. & Gallén, D. Comparing the hydrology of grassed and cultivated catchments in the semi-arid Canadian prairies. *Hydrological Processes* **17**, 559–575 (2003).
4. Shook, K., Pomeroy, J. & van der Kamp, G. The transformation of frequency distributions of winter precipitation to spring streamflow probabilities in cold regions; case studies from the Canadian Prairies. *Journal of Hydrology* **521**, 395–409 (2015).
5. Baulch, H. *et al*. Synthesis of science: findings on Canadian Prairie wetland drainage. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques* **46**, 229–241 (2021).
6. Bush, E. & Lemmen, D. S. *Canada's Changing Climate Report*. (Government of Canada= Gouvernement du Canada, 2019).
7. Vincent, L. A. *et al*. Observed Trends in Canada's Climate and Influence of Low-Frequency Variability Modes. *Journal of Climate* **28**, 4545–4560 (2015).
8. Shook, K. & Pomeroy, J. Changes in the hydrological character of rainfall on the Canadian prairies. *Hydrological Processes* **26**, 1752–1766 (2012).
9. Asong, Z. E. *et al*. High-Resolution Meteorological Forcing Data for Hydrological Modelling and Climate Change Impact Analysis in Mackenzie River Basin (2019).
10. Gray, D. M. Snow hydrology of the prairie environment. *Snow hydrology* 21–34 (1970).
11. Pomeroy, J. W. *et al*. *Prairie Hydrological Model Study Final Report*. 126 (2009).
12. Unduche, F., Tolossa, H., Senbeta, D. & Zhu, E. Evaluation of four hydrological models for operational flood forecasting in a Canadian Prairie watershed. *Hydrological Sciences Journal* **63**, 1133–1149 (2018).

13. Spence, C. *et al.* Assessing hydrological sensitivity of grassland basins in the Canadian Prairies to climate using a basin classification-based virtual modelling approach. *Hydrology and Earth System Sciences* **26**, 1801–1819 (2022).
14. Mengistu, S. G. & Spence, C. Testing the ability of a semidistributed hydrological model to simulate contributing area. *Water Resources Research* **52**, 4399–4415 (2016).
15. Shook, K., Pomeroy, J. W., Spence, C. & Boychuk, L. Storage dynamics simulations in prairie wetland hydrology models: evaluation and parameterization. *Hydrological Processes* **27**, 1875–1889 (2013).
16. Pomeroy, J. W. *et al. Improving and Testing the Prairie Hydrological Model at Smith Creek Research Basin*. 102 https://research-groups.usask.ca/hydrology/documents/reports/chrpt14_phm_scrb.pdf (2014).
17. Dumanski, S., Pomeroy, J. W. & Westbrook, C. J. Hydrological regime changes in a Canadian Prairie basin. *Hydrological Processes* **29**, 3893–3904 (2015).
18. Whitfield, P. H., Shook, K. R. & Pomeroy, J. W. Spatial patterns of temporal changes in Canadian Prairie streamflow using an alternative trend assessment approach. *Journal of Hydrology* **582**, 124541 (2020).
19. Marchildon, G. P. The Prairie Farm Rehabilitation Administration: climate crisis and federal–provincial relations during the Great Depression. *Canadian Historical Review* **90**, 275–301 (2009).
20. Gray, D. M. & Landine, P. G. An energy-budget snowmelt model for the Canadian Prairies. *Canadian Journal of Earth Sciences* **25**, 1292–1303 (1988).
21. Morton, F. I. Operational estimates of lake evaporation. *Journal of Hydrology* **66**, 77–100 (1983).
22. Mekis, É. & Vincent, L. A. An Overview of the Second Generation Adjusted Daily Precipitation Dataset for Trend Analysis in Canada. *Atmosphere-Ocean* **49**, 163–177 (2011).
23. Vincent, L. A. *et al.* A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *Journal of Geophysical Research: Atmospheres* **117**, (2012).
24. Hayashi, M., van der Kamp, G. & Rosenberry, D. O. Hydrology of Prairie Wetlands: Understanding the Integrated Surface-Water and Groundwater Processes. *Wetlands* **36**, 237–254 (2016).
25. van der Kamp, G. & Hayashi, M. Groundwater-wetland ecosystem interaction in the semiarid glaciated plains of North America. *Hydrogeology Journal* **17**, 203–214 (2009).
26. Godwin, R. B. & Martin, F. R. J. Calculation of gross and effective drainage areas for the Prairie Provinces. in *Canadian Hydrology Symposium-1975 Proceedings, 11–14 August 1975, Winnipeg, Manitoba* 219–223 (1975).
27. Stichling, W. & Blackwell, S. R. *Drainage Area as a Hydrologic Factor on the Glaciated Canadian Prairies*. (Prairie Farm Rehabilitation Administration Regina, SK, Canada, 1957).
28. Pomeroy, J. W. *et al.* The cold regions hydrological model: a platform for basing process representation and model structure on physical evidence. *Hydrological Processes: An International Journal* **21**, 2650–2667 (2007).
29. Fang, X. & Pomeroy, J. W. Modelling blowing snow redistribution to prairie wetlands. *Hydrological Processes: An International Journal* **23**, 2557–2569 (2009).
30. Gray, D. M., Granger, R. J. & Dyck, G. E. Overwinter soil moisture changes. *Transactions of the ASAE* **28**, 442–0447 (1985).
31. Shaw, D. A., Vanderkamp, G., Conly, F. M., Pietroniro, A. & Martz, L. The fill–spill hydrology of prairie wetland complexes during drought and deluge. *Hydrological Processes* **26**, 3147–3156 (2012).
32. Leibowitz, S. G., Mushet, D. M. & Newton, W. E. Intermittent surface water connectivity: fill and spill vs. fill and merge dynamics. *Wetlands* **36**, 323–342 (2016).
33. Lehner, B., Verdin, K. L. & Jarvis, A. New global hydrography derived from spaceborne elevation data. *Eos, Transactions, American Geophysical Union* **89**, 93–94 (2008).
34. Wolfe, J. D., Shook, K. R., Spence, C. & Whitfield, C. J. A watershed classification approach that looks beyond hydrology: application to a semi-arid, agricultural region in Canada. *Hydrology and Earth System Sciences* **23**, 3945–3967 (2019).
35. He, Z., Shook, K., Spence, C., Pomeroy, J. W. & Whitfield, C. J. Modelling the regional sensitivity of snowmelt, soil moisture, and streamflow generation to climate over the Canadian Prairies using a basin classification approach. *Hydrology and Earth System Sciences* **27**, 3525–3546 (2023).
36. Armstrong, R. N., Pomeroy, J. W. & Martz, L. W. Variability in evaporation across the Canadian Prairie region during drought and non-drought periods. *Journal of Hydrology* **521**, 182–195 (2015).
37. Spence, C. *et al.* Assessing runoff sensitivity of North American Prairie Pothole Region basins to wetland drainage using a basin classification-based virtual modelling approach. *Hydrology and Earth System Sciences* **26**, 5555–5575 (2022).
38. Pomeroy, J. W. *et al.* The Cold Regions Hydrological Modelling Platform for hydrological diagnosis and prediction based on process understanding. *Journal of Hydrology* **615**, 128711 (2022).
39. Shook, K., He, Z., Spence, C., Whitfield, C., & Pomeroy, J. PHyDAP - Prairie Hydrology Design and Analysis Product, *Federated Research Data Repository.*, https://doi.org/10.20383/102.0694 (2024).
40. Harder, P. & Pomeroy, J. Estimating precipitation phase using a psychrometric energy balance method. *Hydrological Processes* **27**, 1901–1914 (2013).
41. Pomeroy, J. W. & Li, L. Prairie and arctic areal snow cover mass balance using a blowing snow model. *Journal of Geophysical Research: Atmospheres* **105**, 26619–26634 (2000).
42. Fang, X. *et al.* Prediction of snowmelt derived streamflow in a wetland dominated prairie basin. *Hydrology and Earth System Sciences* **14**, 991–1006 (2010).
43. Annand, H. The influence of climate change and wetland management on prairie hydrology — insights from Smith Creek, Saskatchewan. PhD Thesis (University of Saskatchewan, Saskatoon, Canada) (2022).
44. Gironás, J., Roesner, L. A., Rossman, L. A. & Davis, J. A new applications manual for the Storm Water Management Model (SWMM). *Environmental Modelling & Software* **25**, 813–814 (2010).
45. Khatibi, A. Coupling a hydraulic model with the cold regions hydrological modelling platform to predict the spatial extent of localized flooding. MSc Thesis (University of Saskatchewan, Saskatoon, Canada). https://hdl.handle.net/10388/15923 (2024).
46. Gasset, N. *et al.* A 10 km North American precipitation and land-surface reanalysis based on the GEM atmospheric model. *Hydrology and Earth System Sciences* **25**, 4917–4945, https://doi.org/10.5194/hess-25-4917-2021 (2021).
47. Asong, Z. E. *et al.* High-resolution meteorological forcing data for hydrological modelling and climate change impact analysis in the Mackenzie River Basin. *Earth System Science Data* **12**, 629–645, https://doi.org/10.5194/essd-12-629-2020 (2020).
48. Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049, https://doi.org/10.1002/qj.3803 (2020).
49. Hersbach, H. *et al.* ERA5 hourly data on single levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)* https://doi.org/10.24381/cds.adbb2d47 (2023).

Acknowledgements

This work was completed as part of the Prairie Water project of the Global Water Futures program funded by the Canada First Research Excellence Fund. We are grateful to the Prairie Water Community Advisory Committee for valuable feedback on this work, and to all participants who joined in the PHyDAP development workshops that guided this work. This research was enabled in part by support provided by the Digital Research Alliance of Canada (alliancecan.ca).

Author contributions

K.R.S. — conceptualization, methods, analysis, data management, manuscript writing; Z.H. — methods, analysis, manuscript editing; J.W.P. — conceptualization, methods, manuscript editing; C.S. — conceptualization, methods, manuscript writing; C.J.W. — conceptualization, methods, manuscript writing.

Competing interests

The authors have no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.J.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.