Convenient Access to Archived Predictions

Canada's CaSPAr Platform

Adapted from “The Canadian Surface Prediction Archive (CaSPAr): A Platform to Enhance Environmental Modeling in Canada and Globally,” by **Juliane Mai** (University of Waterloo), **Kurt C. Kornelsen**, **Bryan A. Tolson**, **Vincent Fortin**, **Nicolas Gasset**, **Djamel Bouhemhem**, **David Schäfer**, **Michael Leahy**, **François Anctil**, and **Paulin Coulibaly**. Published online in *BAMS*, March 2020. For the full, citable article, see [DOI:10.1175/BAMS-D-19-0143.1](https://doi.org/10.1175/BAMS-D-19-0143.1).

Hydrologic, land surface, and other environmental models require meteorological input data—for example, precipitation and temperature to estimate discharge and soil moisture. In a hindcast these data are usually provided by ground observations, but in forecast mode, they require forecasted meteorological inputs from numerical weather prediction (NWP) models.

Archived numerical predictions are required to evaluate or improve such forecasting systems. However, archives of NWP data are rare. Furthermore, while data are usually disseminated by the agencies producing forecasts—including the European Centre for Medium-Range Weather Forecasts (ECMWF), NOAA, and Environment and Climate Change Canada (ECCC)—users of existing archives often need to download the full dataset since a spatial, temporal, and/or variable selection is not possible. There are exceptions, such as NOAA's National Operational Model Archive and Distribution System (NOMADS). In this repository of weather model outputs and inputs, and a limited subset of climate model datasets generated by NOAA, the user can select variables and specific time periods. But the download options, FTP and HTTP(S), are limited to very short archives of a few

days and only allow users to download the full data product. NOMADS also offers HTTP adaptive streaming for several products, which gives access to longer archives but only allows the download of the complete domain, variable, and forecast. For NOMADS, the Network Common Data Form (netCDF) spatial subsetting is limited to rectangular bounding boxes. Moreover, the netCDF subsetting service lacks a graphical user interface, which limits utility and access.

The ECMWF disseminates not only their own data but also several additional datasets. The disadvantage is that the user has to manually specify the variable(s), time period, ensemble size, issue date(s), and vertical level, and a spatial selection is not possible. The data are then provided as a download link after processing. The ECMWF's Climate Data Store online allows users to browse available products and supports building applications using these data.

An archive of NWP products issued by ECCC is not yet publicly accessible. This, along with the limitations of existing archives, led to the development of the Canadian Surface Prediction Archive (CaSPAR). CaSPAR is accessible free of charge and archives eight operational NWP models, one preoperational product, and one reanalysis dataset, all generated by ECCC, at hourly to daily temporal resolution and 2.5–50-km spatial resolution. Every night, about 185 GB of new data are processed and added to the database. As of March 2020, the database contained 394 TB of NWP products in NetCDF format.

CaSPAR provides a web-based GIS platform that allows the user to select variables, time period, and forecast horizons of interest, as well as spatial subsetting beyond bounding boxes: select the domain of interest by drawing a polygon (by hand) or uploading a shapefile. The data in CaSPAR are not interpolated.

CaSPAR provides access to forecasts issued seven days prior or earlier, not current forecasts, because it is decoupled from the data provider and ensures that all data retrieved are checked for consistency. The time delays helps ensure there is no confusion with the source of operational data. The user is notified by email with a link to download the requested data. The download procedure allows for secure and fast parallel data transfers free of charge.

.....

A core philosophy of CaSPAR is user data sovereignty. This means that users ultimately determine how their data should be represented for their unique modeling and analysis needs.

.....

A core philosophy of CaSPAR is user data sovereignty. This means that users ultimately determine how their data should be represented for their unique modeling and analysis needs. Unlike other data providers, CaSPAR stores all data on the original grid used by the NWP model. This provides transparency and gives the user the power to choose transformations and algorithms with full knowledge of their strengths and limitations. Indeed, the degree of customization and transparency afforded by CaSPAR is unique among publicly accessible NWP archives, and it minimizes user download requirements and local processing time.

Products and processing

Currently, CaSPAR archives eight operational products: the Global Ensemble Prediction System (GEPS), the Global Deterministic Prediction System (GDPS), the Regional Ensemble Prediction System (REPS), the Regional Deterministic Prediction System (RDPS), the High-Resolution Deterministic Prediction System (HRDPS), the Canadian Land Data Assimilation System (CaLDAS), the Canadian Precipitation Analysis on a coarse 10-km grid (CaPA coarse), and CaPA on a finer 2.5-km grid (CaPA fine). CaSPAR also hosts the finer CaPA product in its preoperational version (CaPA fine) and a reanalysis of RDPS called the Regional Deterministic Reanalysis System (RDRS). The archive starts for most products in May 2017.

GEPS and GDPS cover the whole globe, while REPS, RDPS, RDRS, and CaPA (coarse) are available for North America; HRDPS, CaLDAS, and CaPA fine (all 2.5-km resolution) cover Canada and the northern United States. The user can select a product and will receive a string that contains all the product-specific details such as available variables and forecast horizons.

To benchmark the run times and storage requirements of potential user-specific data requests of different complexity, we processed 216 test requests for the 9 available products, 6 domains of interest, and 4 different scenarios. The 6 watershed domains used for the benchmark range in area over several orders of magnitude. The 6 domains are submitted as shapefiles to CaSPAR, leading to the smallest possible file size—areas outside the domains are filled with no-data values.

Based on this test, data processing times can range from minutes to multiple days and

*** An example of the capabilities of CaSPAr, showing NWP products for the record flood starting at noon on 23 Jun 2017 (UTC) in the Grand River in Ontario, Canada (watershed mapped in white outline; domain in gray tiles). Colored lines show total precipitation rate for different forecasts and analyses. The CaPA precipitation analysis (gray bars) at 2.5-km resolution has been added for reference. All but one of the forecast products (HRDPS) predicted a heavy rainfall event but of lower volume than actually occurred.**

storage requirements can be 80 GB (or much more in other examples), depending on the user request. Users can estimate and plan for waits and storage requirements associated with their specific requests. Note that storage is limited of 1 TB and processing to 28 days. A larger request generates an error message for the user.

An analysis example

On 23 June 2017, some rain gauges in Ontario's Grand River watershed had their highest 1-day total rainfall since record-keeping began in 1950. The Grand River Conservation Authority (GRCA, the local flood forecasting entity) notes that the public ECCC weather forecast site predicted 5–10 mm of rainfall over the area on 23 June 2017. But over a 2-day period, precipitation exceeded 100 mm—and in several cases 130 mm—in the watershed.

The CaSPAr archive was already in place for one month in June 2017 (though the automatic user request system did not go online until June 2018). The 7-day delay of making data accessible would not have allowed the GRCA to draw the current forecasts. But the archive can show which data would have been available from ECCC, and allows for a detailed post-event analysis comparing, for example, the various ECCC precipitation products.

At the moment, users can download operational, real-time weather forecast data via ECCC's Datamart portal. The data are available in grib2 and are interpolated to a grid the grib2 format supports. This sort of postevent analysis using easily-accessed CaSPAr data helps people evaluate whether they should invest the effort to regularly access the operational data via Datamart. Spatial subsetting and constrained forecast horizons are not possible in Datamart. Subsetting through CaSPAr leads to significantly smaller downloads. For example, spatial subsetting for the Grand River watershed reduces the data needs by more than 99% for most of the products.

The data show that indeed, forecasts issued on 22 June 2017 all correctly predicted rainfall, but less than observed. GDPS forecasted 40 mm until 0000 UTC 24 June in its midnight issue on 22 June and only 20 mm in the noon (UTC) issue on 22 June. The same holds for all forecasts issued on 22 June for GEPS, RDPS, and REPS. With the exception of the high-resolution product HRDPS, they all forecasted at most 40 mm of rain over the entire watershed until midnight (UTC) 24 June. None of the forecasted amounts was as high as the 100–130 mm reported by the Conservation Authority.

CaSPAr provides an opportunity for forecasting agencies such as the GRCA to retrospectively test ECCC weather forecast products, understand the strengths and limitations of their current forecast systems, and develop new forecast products. CaSPAr aims to improve the understanding and dissemination of ECCC's available numerical products. With CaSPAr, the similar file structure makes it straightforward to compare several products.

The data are all available at the same location. The user only needs to download and process the data of interest.

Future work

CaSPAr data can be employed to analyze extreme rainfall events, but may also assist with atmospheric modeling experiments over limited regions. Moving the system onto ECCC's servers would be beneficial, allowing dissemination of current forecasts without the 7-day delay. We hope that CaSPAr will help practitioners such as conservation authorities improve flood forecasting and warnings and help researchers test and benchmark data assimilation techniques and/or postprocessing algorithms.

CaSPAr is working to establish an application interface so that advanced users can avoid the web interface and directly request data. We are also investigating the feasibility of implementing an OPeNDAP (www.opendap.org) access to the CaSPAr database. ☛

≡ METADATA

BAMS: How did you become interested in the topic of this article?

Juilane Mai: *My initial project was to work on data assimilation. But soon I came to know that none of the data I was supposed to use were available in a long and consistent record. I couldn't test any of the data assimilation methods with only 2–3 months of data. This led to starting to automatically download and archive the data that are now available under CaSPAr. The initial intent was only one script that downloads the data. It grew much bigger than we ever imagined, with now having all data in constant formats with augmented metadata and an automatic tool processing user requests and getting the users the data.*

BAMS: What surprises/ surprised you the most about the work you document in this article?

JM: *I think what surprised me the most is that archives of weather predictions are so rare. And that not such an open-access, free-of-charge archive existed for Canadian weather predictions. The past four years we worked on CaSPAr showed that lots of researchers and practitioners are interested in archived forecasts and strongly benefit from CaSPAr.*

BAMS: What was the biggest challenge you encountered while doing this work?

JM: *The biggest challenge is certainly to get funding to maintain the archive and support the challenges we*

face in the operational setup (downtimes of servers, inconsistent or missing data, changes of data, user interaction and Q&A). I think my appreciation for researchers that make data freely available grew significantly in the past years.

BAMS: How will you follow up?

JM: *We just got another three years of computational resources from Compute Canada approved. We will continue archiving the current datasets and will add new reanalysis products currently produced at ECCC. We also plan to improve our web-based frontend and include OPeNDAP functionality such that CaSPAr data can be accessed easier through platforms like Deltares' Flood Early Warning System (FEWS).*