A 35-YEAR NORTH AMERICAN PRECIPITATION AND SURFACE RETROSPECTIVE ANALYSIS

EXECUTIVE SUMMARY AND PROGRESS REPORT

Dr. Nicolas Gasset, Post-doctorate fellow; Dr. Vincent Fortin, Research scientist

Meteorological Research Division, Environment and Climate Change Canada (ECCC)

Environment and Climate Change Canada (ECCC) has designed a system which can provide accurate and consistent estimates of precipitation and surface state (including all major components and fluxes of the surface water cycle). This system is based on components of ECCC's operational weather and environmental prediction system, but it has been modified in order to be able to simulate any period in time between 1980 and 2015, and to cover all of North America as well as the Arctic Basin in the same 15-km domain. This report provides a brief description of the method and a status update on the project.

KEY MESSAGES

- This IWI project was initiated in July 2015, with the primary objective of producing a climatology for the Canadian Precipitation Analysis (CaPA), but also for ECCC's component Net Basin Supply (NBS) estimates.
- In a March 2016 update, ECCC reported having been able to design and test such a precipitation and surface reanalysis system, but results showed a slight deterioration of atmospheric variables near the surface, especially in summer.
- During the first half of FY 2016-2017, the system was improved by coupling it to the Canadian Land Data Assimilation System (CaLDAS), solving this issue to our satisfaction.
- In order to reduce overall computing time for the production of a 35-year hindcast, we proposed to run the system in parallel for seven periods of five years.
- The production cycle requires that the GEM atmospheric model be run twice for every 12-hour period: once at 50-km resolution over the whole globe, then at 15-km resolution over North America.
- During the second half of FY 2016-2017, production of a 5-year hindcast started in order to demonstrate the feasibility of the approach. We were able to complete the 50-km runs, but due to limited access to the supercomputer only half of the 15-km runs were completed. We currently expect to complete the 5-year hindcast during the spring of 2017, and to present an evaluation of this hindcast at the annual conference of the International Association for Great Lakes Research (May 2017) and at the FloodNet Annual General Meeting (June 2017).

CONTENTS

THE NEED FOR A PRECIPITATION AND SURFACE REANALYSIS SYSTEM

Near real-time high-resolution estimates of precipitation and surface variables (including component Net Basin Supplies) are available from operational and experimental systems developed and operated by Environment and Climate Change Canada (ECCC). ECCC, however, does not currently produce either a reanalysis or climatology of these estimates. This requires running atmospheric forecast and analysis systems back in time for many decades, which has never been done by ECCC in part because of the very significant cost in both human and computer resources of running not only the atmospheric model but also the data assimilation system.

Yet, a reanalysis would allow bias correction of current estimates and forecasts, and help decision maker understand and communicate by how much the current and forecasted state of the system differs from the recent past. Although other reanalysis products exist, they have many limitations:

- They often show strong discontinuities at the Canada-US border (for example NARR and NLDAS)
- They are provided at fairly low horizontal resolution (such as ERA-Interim/Land and MERRA Land), thus not adequately resolving important local features such as the Great Lakes
- They generally do not incorporate information from most climate stations (in particular Canadian stations).

There is therefore a need for a higher resolution precipitation and surface reanalysis product that relies on more accurate climate station data and that is seamless at the Canada/US border. This has clear applications for adaptive management of transboundary watersheds.

AN OPPORTUNITY FOR A PRECIPITATION AND SURFACE REANALYSIS

In November 2014, however, ECCC started to produce operational global ensemble reforecast based on the ERA-Interim reanalysis for the 1995-2012 period at 50-km resolution (Gagnon et al. 2014). This reforecast system demonstrated the ability to initialize the Global Environmental Multiscale (GEM) model with non-ECCC data, thus allowing running it back in time without the need for an associated atmospheric assimilation system. A dynamical downscaling to finer resolution such as 15-km would then be relatively straightforward, as this is the current horizontal resolution of the Regional Ensemble Prediction System for North America.

In the Spring of 2015, ECCC informed the IJC that this presented an opportunity for developing a precipitation and surface reanalysis system which could be used for adaptive management of transboundary watersheds.

DESIGNING A PRECIPITATION AND SURFACE REANALYSIS SYSTEM

In July 2015, ECCC was tasked by IWI to develop a prototype of a system that would allow retrospective analysis of precipitation and surface variables over the 1979-2015 period. Considering the complexity of the ECCC`s forecast and assimilation systems, it was decided to rely

essentially on systems that are designed and tested for operational analysis and forecasting. Key components from ECCC systems for the project are listed below.

- The Global Environmental Multiscale (GEM) model is the atmospheric forecast model operated at ECCC in various configurations, ranging from low resolution seasonal forecasts to very high-resolution hourly forecasts.
- The Canadian Precipitation Analysis (CaPA) system provides real-time gridded precipitation analyses over North America at a 10-km resolution and a 6-h time step. It relies on the GEM atmospheric model to obtain a background field of precipitation that is modified based on ground, radar and satellite observations of precipitation.
- The Canadian Land Data Assimilation System (CaLDAS) provides near real-time gridded surface analyses at 25-km globally and at 2.5-km over Canada, both at 3-h time resolution. It relies on CaPA to obtain precipitation and on GEM to get background fields of snow depth and atmospheric surface layer temperature and humidity. These fields are then modified based on ground and satellite observations.

During this design phase, a system based on ECCC's ensemble reforecast system, but relying on a single deterministic non-perturbed member has been developed in order to run GEM back to 1979. The system requires that GEM be run twice: once at 50-km resolution over the whole globe, then a second time at 15-km over North America.

The CaLDAS system was not used at first, as it complicated the design substantially. It was demonstrated that 15-km gridded precipitation forecasts of quality similar to operational 10-km GEM forecasts could be obtained in winter, while less biased but less skillful summer precipitation forecasts were obtained (Gasset et al., 2016). In addition, a number of limitations of the existing ensemble reforecast approach were identified that needed to be resolved in order to use it for the development of a retrospective analysis of precipitation and surface variables.

The computing cost and time for creating a 35-year retrospective analysis of precipitation and surface fields has also been evaluated, which led to the conclusion that it would not be feasible to perform a continuous 35-year simulation in a timely manner. Based on similar studies used for other atmospheric reanalysis experiments, it was proposed to split the 35-year simulation in seven production periods of five years that could be done in parallel. Each five-year production period would be preceded by a one year spin-up period.

Despite this strategy, computing time was still judged excessive for the existing operational supercomputer used by ECCC to run GEM. Fortunately, the effective computing time is expected to go down significantly once ECCC takes possession of a new computing infrastructure (now expected to occur in the first half of FY 2017-2018). In the meantime, it was proposed to focus on improving the design and running the system for one period of five years.

USING CALDAS TO IMPROVE THE QUALITY OF THE REANALYSIS

During the second year of the project (FY 2016-2017), the focus was put on coupling the GEM model to the CaLDAS system, in order to obtain better initial surface conditions. The surface

analysis of soil temperature, soil moisture and snowpack produced by CaLDAS can indeed be provided to the GEM atmospheric model, together with an upper-air analysis, in order to initialize a new forecast cycle. Such an approach is currently operational only for the most advanced high-resolution forecast system from ECCC. Other systems rely on a much simpler approach to provide them with surface initial conditions. The more sophisticated approach to initializing GEM which rely on CaLDAS showed clear benefits for surface variables such as temperature and humidity, but had significant impacts for the boundary layer of GEM as well, with positive impacts observed up to the middle of the troposphere (approximately 5-km above ground). It was thus decided to include CaLDAS in the reanalysis system, despite the added complexity and computational cost. It was further shown that coupling of CaLDAS and GEM was only required for the 15-km configuration of GEM.

The addition of CaLDAS created specific challenges for the production of the 5-year hindcast period, as the CaLDAS system is inefficient on the current supercomputer operated by ECCC. A high-end desktop computer was acquired for the project and dedicated to running CaLDAS, but despite these efforts it was not possible to complete in time the production of the 5-year hindcast. Because CaLDAS was not necessary for the 50-km GEM runs, the lower resolution GEM runs were completed in time. About half of the 15-km runs were completed by March 2017. The remaining half should be completed during the spring of 2017.

DATASETS USED BY THE PRECIPITATION AND SURFACE REANALYSIS

In order to produce the 35-year hindcast, two main sources of input data required are required:

- Upper-air analysis: the Interim version of the European atmospheric Reanalysis (ERA-Interim) from the European Centre for Medium Range Forecast (ECMWF) serves as initial upper-air condition for GEM (instead of an ECCC analysis).
- Surface stations: high-quality surface station data is needed in order to be able to deliver on this project. Absolute temperature, humidity, precipitation and snow depth observations are the main input (along model outputs) of the precipitation and surface variable analysis systems, i.e. CaPA and CaLDAS. These observations are already available in ECCC databases but they cover only a half of the period of interest. Contacts have been made with NOAA/ESRL (Dr. Tom Hamill), NOAA/GLERL (Dr. Andrew Gronewold), Mesonet-Québec (Alexandre Vanasse) and ECCC's Climate Research Division (Dr. Xiaolang Wang) in order to obtain the most accurate surface station data currently available. The following surface stations datasets will be used: ISD, TDL, AHCCD, MesoNet Quebec and ECCC surface observation database.

CaPA and CaLDAS can also take advantage of radars and dedicated satellites observations. However, considering the short period of availability of these observations, and in order to retain a constant quality of the final product across the period of interest, it was decided to use the latter for evaluation purpose only.

STATUS UPDATE ON TECHNICAL ASPECTS OF THE PROJECT

The following technical aspects need to be underlined:

- Design of the downscaling to 15-km over North America and of the coupled system (i.e. CaPA/CaLDAS coupled with GEM) is now completed and optimised. Various configurations have been evaluated.
- The approach considered as the most promising has been identified, and is being used to produce a 5-year sample (2010-2015) which is required in order to complete a strong proof of concept.
- All of the results are obtained using ECCC's current – but soon to be replaced – supercomputer. Due to limited access to this computer in the middle of the migration efforts, coupling of CaPA/CaLDAS and GEM is only achieved at the regional level. Fortunately, this has been shown to have limited impact for the region of interest.
- ERA-Interim dataset is fully converted in a format usable by GEM for the whole period of interest (1979-2015). Sea surface temperature and ice cover coming from ERA-Interim are also used.
- To produce the current 5-year evaluation period, archived surface observations from ECCC's operational systems are used. This dataset only starts in the mid-nineties and has not been subject to offline quality control procedures. For the final product, a dataset of higher quality that is furthermore consistent in time is being prepared.
- Surface observation processing for the period 1979-2015 is almost completed: converters for all input surface observation datasets (ISD/TDL/AHCCD/MesoNet Quebec) are now tested and fully functional, and the latter can now be directly used as input of CaLDAS and CaPA.
- Migration of individual components of the system (i.e. GEM, CaPA and CaLDAS) to the new ECCC super-computer and environment is almost completed: they are currently being evaluated on the latter.
- Last but not least, discussions with the ECCC section in charge of running operational systems are ongoing in order for them to take care of running the full 35-year period. This will allow ECCC's research scientists to focus on in-depth evaluation of the product in peer-reviewed publications.

Evaluation of the results has started, using independent surface, satellite and upper-air observations. In order to evaluate the skill of the system, official ECCC operational system outputs for the same period are used as the reference. The following preliminary conclusions have been reached:

- Results of the coupled approach are inline with the reference operational approach that was operated from 2012 to 2016 (GEM 3.3, 10-km resolution) for both summer and winter seasons: a little degradation is generally observed in the upper-air (mostly in the stratosphere), but results in the lower portion of the atmosphere are very similar and sometimes better than the operational approach.
- A clear benefit of the coupled approach is observed compared to a simpler reforecast (without data assimilation), notably in the summer with regard to surface variables, but also upper-air

fields. As an example, surface layer dew point and absolute temperature errors are reduced by more than a couple of degrees during summer months in the eastern half of North America. These improvements are mainly associated with a more realistic representation of root-depth soil moisture (which as less impact in winter).

- Absolute values as well as spatial and temporal variability of soil moisture from the coupled approach are generally in good agreement with observation from SMOS satellite (when and where this data is available and valid).
- Improvements in the western part of North America are less significant due to the complex topography and the associate challenge to assimilate surface observations that are also scarcer in these regions.
- Precipitation from the coupled system is almost unchanged (before being analysed) compared to a simple reforecast approach.

As a summary of the tasks that need to be completed, the following aspects are to be underlined:

- A 5-year evaluation period (2010-2015, including one year of spin-up) is nearing completion:
    - Global simulations (50-km) are completed.
    - Half of the period of interest is completed by the coupled regional approach (15-km).
    - The other half should be completed by June 2017.
- The final retrospective analysis will be based on the various observation datasets previously introduced, not just ECCC databases. However, a strong redundancy exists between the various datasets. We are planning to take advantage of this redundancy as part of an off-line quality control through an aggregation procedure of redundant stations. The aggregation procedure is already implemented in CaPA for precipitation, and a similar approach now needs to be applied to the other variable of interest in order to produce the final input observation datasets.
- The effort of evaluation and validation need to continue, based the 5-year evaluation period. Various additional independent datasets will be used in order to have a better picture of the retrospective analysis and forecast quality. Some of these diagnostics may be implemented directly within the system so that they are routinely and automatically produced along with the computations. We are also planning to achieve additional evaluation of the results with:
    - other surface analysis datasets such as ERA-Interim/Land, MERRA/Land.
    - satellite observations for other variables such as surface temperature retrieval from GOES satellite.
    - soil moisture and temperature, sensible and latent heat fluxes from surface station Network as FLUXNET and SCAN.
    - surface run-off and rivers flows for various North American watershed allowing for an integral evaluation.
- Upon completion of the migration and evaluation of independent components of the system on the new ECCC super-computer, we will be able to start migrating the current coupled system.
- An official proposal to the “Comité des Passes Opérationnelles et Parallèles” (CPOP) will follow, together with the production of a technical guide to operate the system.
- Peer-reviewed papers will be prepared to document the method and evalute the results.
- Results will be communicated at various conferences and workshops.

- Outputs of interest to scientists and users will be made available on the pegasus cluster, which can be accessed by external users.
- Selected fields will also be made available through the publicly accessible RPN-WMS web mapping service.

TIMINGS

To complete one month of reforecast and retrospective analysis, it typically takes:

- 50-km global reforecast (without coupling): 1.2 days
- 50-km global reforecast with CaPA/CaLDAS coupling above North America: 2 days (Linux super computer for CaPA/CaLDAS + IBM supercomputer for GEM)
- 15-km regional reforecast with CaPA/CaLDAS coupling: 1.6 days (dedicated workstation for CaPA/CaLDAS+ IBM supercomputer for GEM)
- 15-km regional reforecast with CaPA/CaLDAS coupling: 5 days (operational front-end Linux supercomputer for CaPA/CaLDAS + IBM back-end supercomputer for GEM)

Thus, based on the current configuration (i.e. simple global reforecast without CaPA/CaLDAS coupling + regional approach coupled to CaPA/CaLDAS for North America) and considering that seven time windows of 5+1 years are run in parallel, it would take about 1.6 x 12 x 6 = 115-days of computation on 2016 CPU cores to complete the 35-years of regional reforecast and retrospective analysis of precipitation and surface fields on the old ECCC super-computer. If the seven time windows cannot be done in parallel, then doing a single 35-year run would take 1.6 x 12 x 35 = 672 days of computation on 288 cores.

CONCLUSIONS

This project aims to develop a system allowing the production of a retrospective analysis of precipitation and surface fields over North America at 15-km resolution for the 1979-2015 period, using three components of ECCC's numerical weather prediction infrastructure: the atmospheric model GEM, the land-data assimilation system CaLDAS and the precipitation analysis system CaPA. These systems are initialized by ERA-Interim upper-air analyses and constrained at the surface by data from climate stations. This report marks the end of the second year of the project.

During the first year, a system based on GEM and CaPA was designed and thoroughly evaluated, leading to the conclusion that CaLDAS needed to be used in order to improve estimations of the land-surface state, as well as near-surface predictions of the state of the atmosphere. This work was completed during the second year, and the production of a 5-year reanalysis was initiated in order to refine the design and better assess the quality of the product in a near-operational framework.

The production of the 5-year precipitation and surface reanalysis has progressed, and it should be completed by the end of June 2017. This product will need to be further evaluated, in particular in its ability to simulate net basin supplies. Nonetheless, the technology transfer from research to operations can start soon after the commissioning of ECCC's new supercomputer in the first half of

FY 2017-2018. Once the technology transfer is completed, it would take from four months to two years to produce 35 years of data, depending on the amount of CPU resources dedicated to the project.