Five Reasons Why You Should Know CaSPAr and Cuizinart

Juliane Mai, Bryan A Tolson, Kurt C Kornelsen, Paulin Coulibaly, David Schäfer, Nicolas Gasset, Vincent Fortin, Djamel Bouhemhem, Michael Leahy, François Anctil, Brent Hall

Juliane Mai, Martin Gauch, Jimmy Lin, Bhaleka Persaud, Bryan Tolson, Ethan Wang, Alex Weatherhead, Kaisong Huang

Kitchener – October 14, 2020

Portage and Global Water Futures webinar series on Research Data Management

Motivation for CaSPAr

- Environmental models used for various purposes (drought and flood simulations, climate predictions, hydro-power applications)
- Models are all different in their conceptualization and processes
- But all models require inputs (e.g., precipitation, temperature)
- Observations available when model is run in prior period
- If models run in forecast mode, these data are Numerical Weather Predictions (NWP)

www.caspar-data.ca
caspar.data@uwaterloo.ca

Motivation for CaSPAr

For forecast setup of models you might want to access:

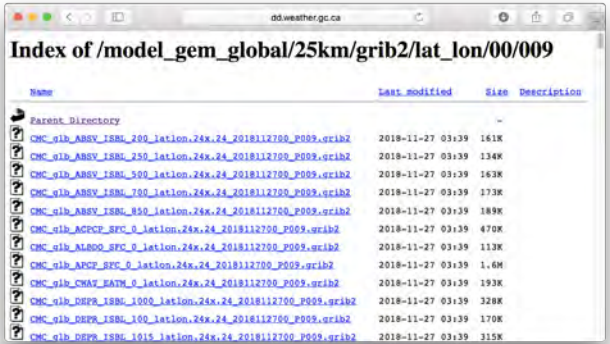
- Different **variables** (e.g., temperature, precipitation, wind speed)
- Different **forecast horizons** (e.g., 24h lead time, 1 week lead time)
- Different **time period** (e.g., July 2019 and July 2020)
- Different **resolution** (e.g., 2.5km and 10km product)
- Different **domain** (e.g., Ontario and BC)

www.caspar-data.ca
caspar.data@uwaterloo.ca

ECCC's Current NWP Data Distribution System

DataMart

http://dd.meteo.gc.ca/about_dd_apropos.txt



- no archive of forecasts
- data interpolated to grib2 supported grid
- no spatial cropping
- heavy post-processing

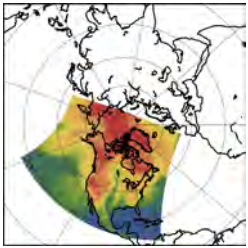
www.caspar-data.ca
caspar.data@uwaterloo.ca

Aim of CaSPAr

archive
forecasts & analyses
produced by

Environment and Climate Change Canada
Environnement et Changement climatique Canada

convert
raw FST data to
standardized NetCDF



provide
gis-based web interface
to make data available

www.caspar-data.ca
caspar.data@uwaterloo.ca

#1

Individual Subsetting of Data

www.caspar-data.ca
caspar.data@uwaterloo.ca

Frontend

– Submit Your Request –

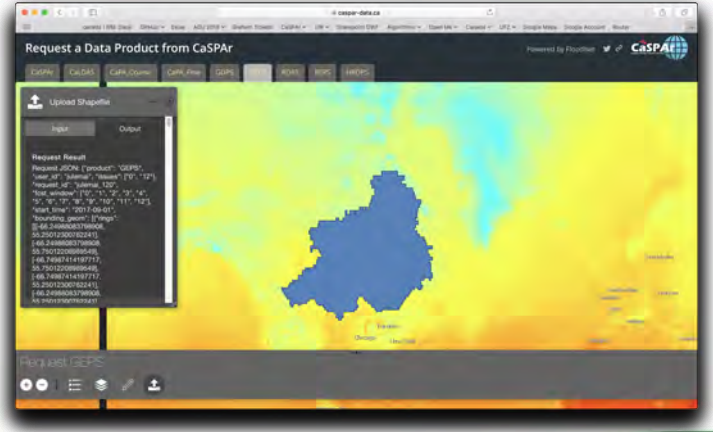
- ☐ Select product
- ☐ Credentials
- ☐ Domain
- ☐ Variables
- ☐ Horizons
- ☐ Issues
- ☐ Time period
- ☐ Submitted!

www.caspar-data.ca
caspar.data@uwaterloo.ca

Frontend

– Submit Your Request –

- ✓ Select product
- ✓ Credentials
- ✓ Domain
- ✓ Variables
- ✓ Horizons
- ✓ Issues
- ✓ Time period
- ✓ Submitted!



www.caspar-data.ca
caspar.data@uwaterloo.ca

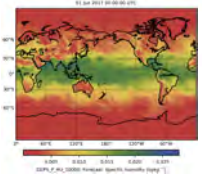
#2

Easy Comparison of NWP Products

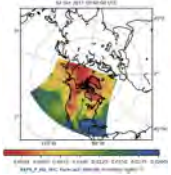
www.caspar-data.ca
caspar.data@uwaterloo.ca

Easy Comparison of NWP Products

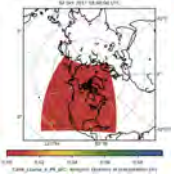
– Products Available (7 days after issue) –



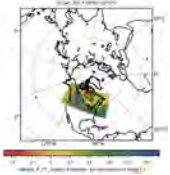
GEPS



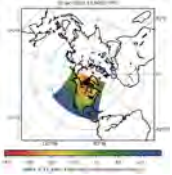
REPS



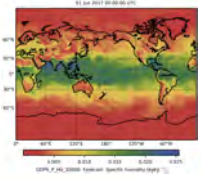
CaPA 10.0k



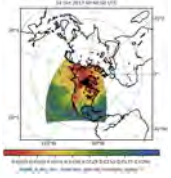
HRDPS



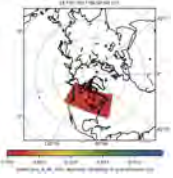
RDRS v1 (15.0k)



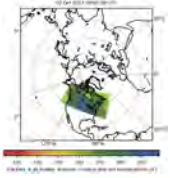
GDPS



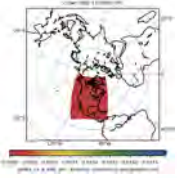
RDPS



CaPA 2.5k



CaLDAS



RDRS v2 (10.0k)

www.caspar-data.ca
caspar.data@uwaterloo.ca

Easy Comparison of NWP Products

– Products Available (7 days after issue) –

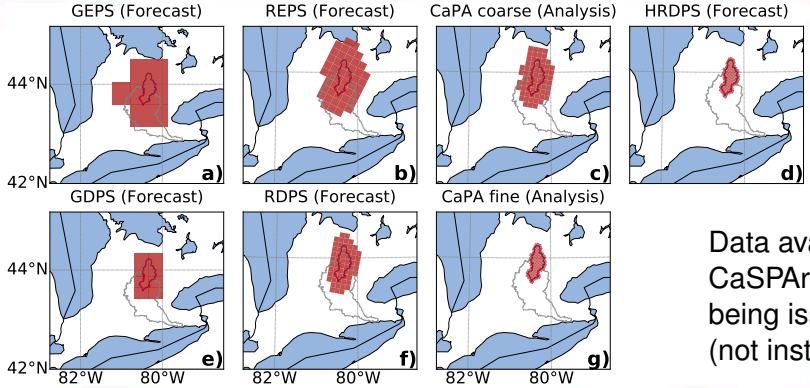
11+ TB
new data
per month

(40+ months are archived to date)

www.caspar-data.ca
caspar.data@uwaterloo.ca

Easy Comparison of NWP Products

– Grand River Flood Event June 2017 –



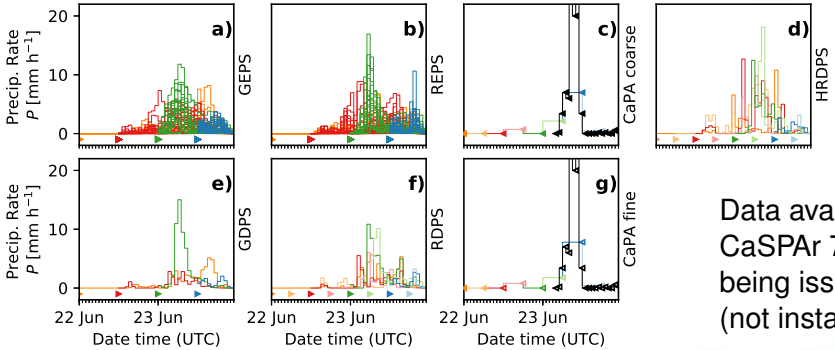
Data available in CaSPAr 7 days after being issued (not instantaneous)

Mai et al. (2020), BAMS.
https://doi.org/10.1175/BAMS-D-19-0143.1

Easy Comparison of NWP Products

– Grand River Flood Event June 2017 –

Forecasted precipitation – Grand River watershed upstream of West Montrose



Data available in CaSPAr 7 days after being issued (not instantaneous)

Mai et al. (2020), BAMS.
https://doi.org/10.1175/BAMS-D-19-0143.1

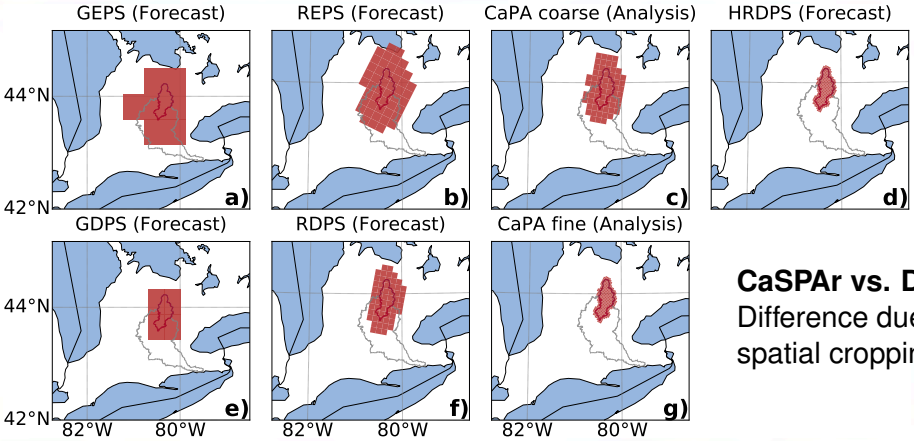
#3

Drastic Reduction of Data Amount to Download

www.caspar-data.ca
caspar.data@uwaterloo.ca

Drastic Reduction of Data Amount to Download

– Grand River Flood Event June 2017 –

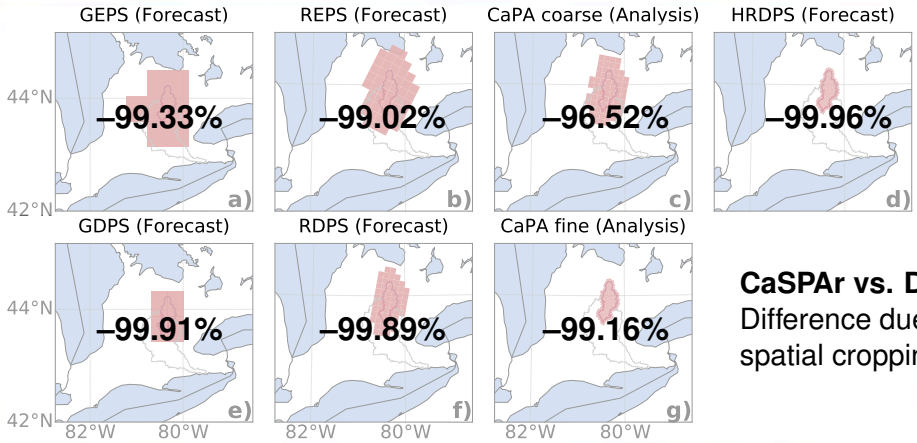


CaSPAr vs. DataMart
Difference due to spatial cropping

www.caspar-data.ca
caspar.data@uwaterloo.ca

Drastic Reduction of Data Amount to Download

– Grand River Flood Event June 2017 –



CaSPAr vs. DataMart
Difference due to spatial cropping

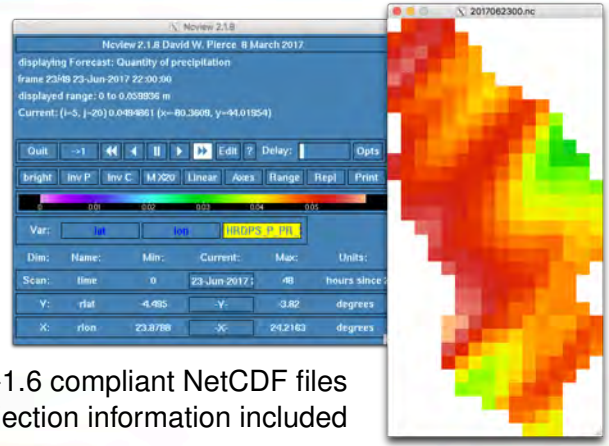
www.caspar-data.ca
caspar.data@uwaterloo.ca

#4

Standardized File Format

www.caspar-data.ca
caspar.data@uwaterloo.ca

Standardized File Format



- ✓ CF-1.6 compliant NetCDF files
- ✓ projection information included

www.caspar-data.ca
caspar.data@uwaterloo.ca

#5

Seamless Utilization of Input Data

www.caspar-data.ca
caspar.data@uwaterloo.ca

Seamless Utilization of Input Data

– Models running with NetCDF inputs directly –

- Hydrologic Modeling Framework RAVEN
- Hydrologic Modeling Framework SUMMA
- Variable Infiltration Capacity Model VIC v5.0
- Mesoscale Hydrologic Model – mHM
- MEC-Surface & Hydrology model – MESH
- Weather Research and Forecasting Model – WRF-Hydro
- Community Land Model – CLM
- Multi-parametrization Land Surface Model – Noah-MP

www.caspar-data.ca
caspar.data@uwaterloo.ca

Summary of CaSPAr Highlights

1. **Individual subsetting** of data via web-interface and (soon) API
2. **Easy comparison** of NWP products
3. **Drastic reduction** of data amount to download
4. **Standardized** NetCDF file format (CF-1.6 compliant)
5. **Seamless utilization** of input data for model runs

6. **Fast** data download via Globus
7. **GitHub** library of data processing scripts
8. Modular and decoupled **Frontend and Backend**

CaSPAr
Canadian Surface Prediction Archive

@CaSPArData

www.caspar-data.ca
caspar.data@uwaterloo.ca

Global Water Futures

– Motivation Data Portal –

The world's largest university-led freshwater research program.

http://cuizinart.io
gwf.cuizinart@uwaterloo.ca

Cuizinart

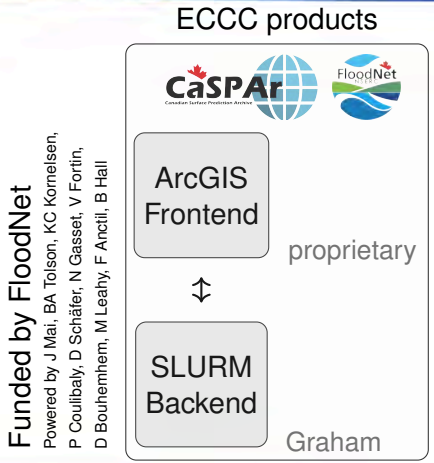
– The Analogy –

Slice-and-dice
large, gridded
NetCDF datasets

http://cuizinart.io
gwf.cuizinart@uwaterloo.ca

Cuizinart

– CaSPAr vs. Cuizinart –



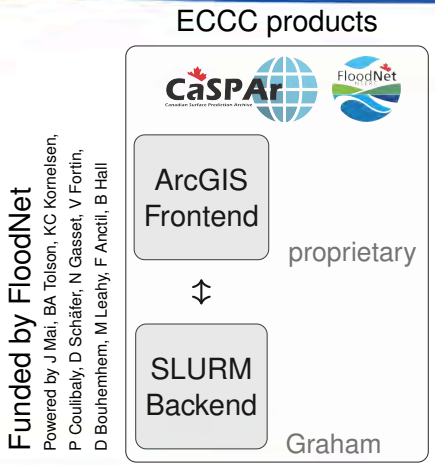
Funded by FloodNet

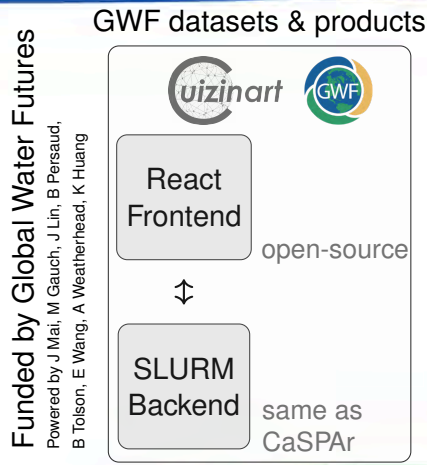
Powered by J Mai, BA Tolson, KC Kornelsen, P Coulibaly, D Schäfer, N Gasset, V Fortin, D Bouhemhem, M Leahy, F Anctil, B Hall

http://cuizinart.io
gwf.cuizinart@uwaterloo.ca

Cuizinart

– CaSPAr vs. Cuizinart –

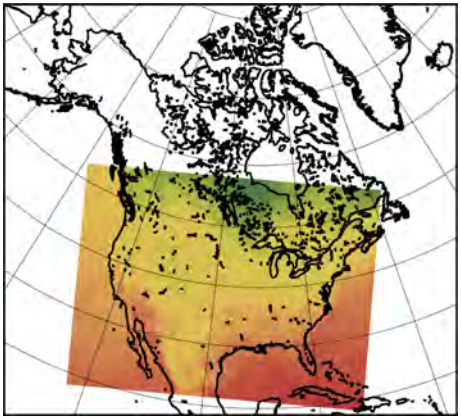




http://cuizinart.io
gwf.cuizinart@uwaterloo.ca

Cuizinart

– Dissemination of Large Gridded Datasets Within and Beyond GWF –



WRF control run over CONUS

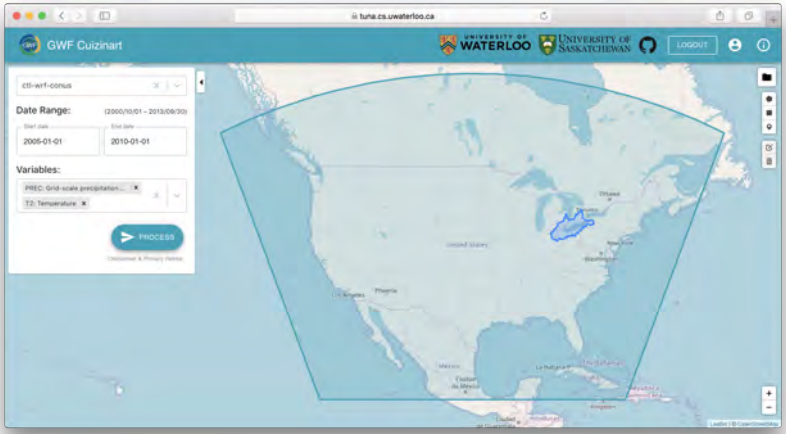
(Zhenhua Li, Yanping Li, et al.)

resolution:	4 km, hourly
# of variables:	11
time period:	13 years
storage:	**6.4 TB**

http://cuizinart.io
gwf.cuizinart@uwaterloo.ca

Cuizinart

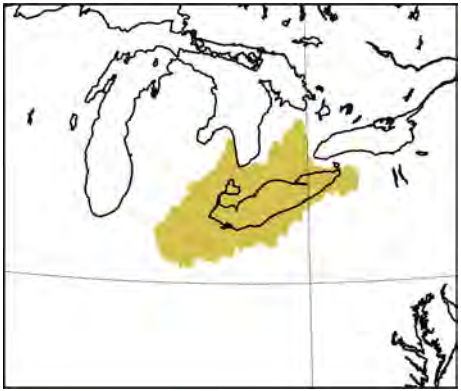
– Dissemination of Large Gridded Datasets Within and Beyond GWF –



http://cuizinart.io
gwf.cuizinart@uwaterloo.ca

Cuizinart

– Dissemination of Large Gridded Datasets Within and Beyond GWF –



WRF control run over **Lake Erie**

resolution:	4 km, hourly
# of variables:	4 (of 11)
time period:	5 (of 13) years
storage:	**11 GB** (0.2% of 6.4 TB)
processing time:	64 min (serial)

http://cuizinart.io
gwf.cuizinart@uwaterloo.ca

Cuizinart

– Dissemination of Large Gridded Datasets Within and Beyond GWF –

- Cuizinart provides framework to **disseminate large, gridded datasets** used and produced by GWF including data forcings, static inputs, and model results
- Data need to be provided in standard CF-1.6 compliant NetCDF format
  - Attributes `long_name` and `units` provided for each variable
  - 2D fields of latitudes and longitudes (in degrees) need to be available
  - Unique name for product need to be provided
  - ...
- Checklist for specifications available:
  `http://www.civil.uwaterloo.ca/jmai/Checklist_NetCDF_Cuizinart.pdf`

http://cuizinart.io
gwf.cuizinart@uwaterloo.ca

Summary

1. **Individual subsetting** of data via web-interface
2. **Easy comparison** of different products
3. **Drastic reduction** of data to handle
4. **Standardized** NetCDF file format
5. **Seamless utilization** of input data for model runs

@CaSPArData

Thank you!

Juliane Mai, Bryan A Tolson, Kurt C Kornelsen, Paulin Coulibaly, David Schäfer, Nicolas Gasset, Vincent Fortin, Djamel Bouhemhem, Michael Leahy, François Anctil, Brent Hall

Juliane Mai, Martin Gauch, Jimmy Lin, Bhaleka Persaud, Bryan Tolson, Ethan Wang, Alex Weatherhead, Kaisong Huang

UNIVERSITY OF WATERLOO

McMaster University

Environment and Climate Change Canada
Environnement et Changement climatique Canada

UNIVERSITÉ LAVAL

compute canada

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

esri

NSERC CRSNG

juliane.mai@uwaterloo.ca

www.caspar-data.ca
http://cuizinart.io