

CaSPAr

Canadian Surface Prediction Archive

Juliane Mai, Kurt C. Kornelsen, Bryan Tolson,
Paulin Coulibaly, François Anctil, Vincent Fortin,
Michael Leahy, Brent Hall



UNIVERSITÉ
LAVAL



esri Canada



Environment and
Climate Change Canada
Environnement et
Changement climatique Canada



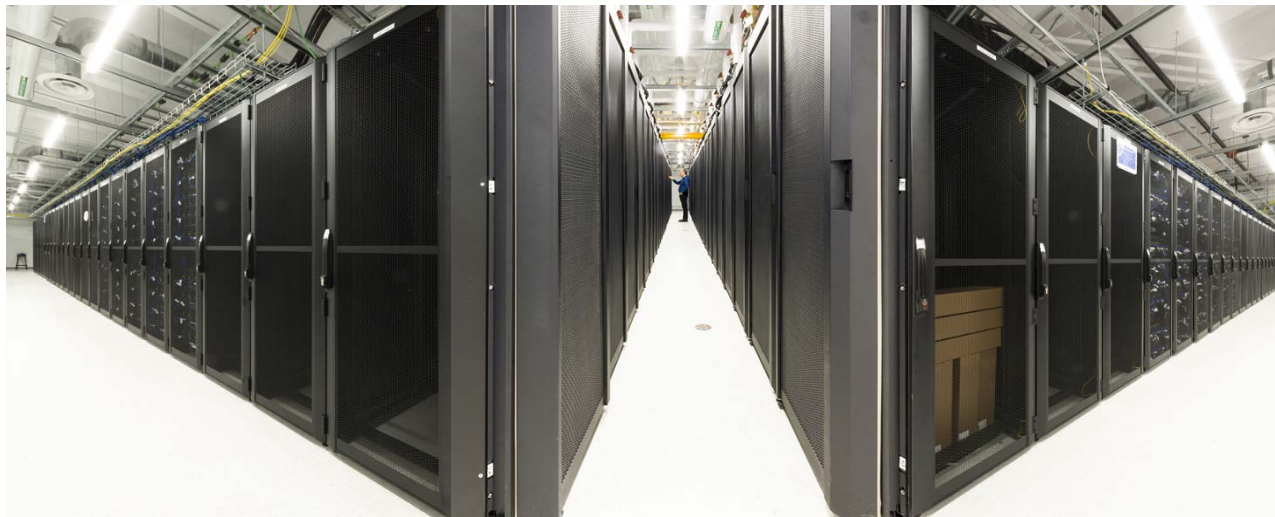
**NSERC
CRSNG**



**compute
canada** | **calcul
canada**

Overview

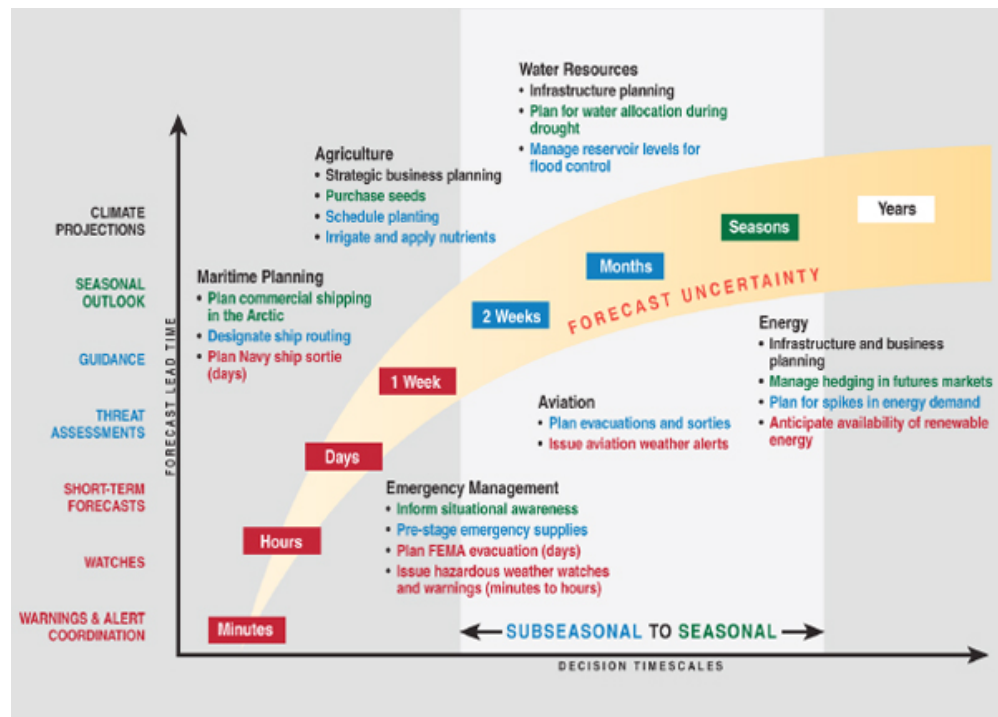
– New cluster “Graham” @ SHARCNET –



- Number of CPU cores: 32,168 (all other together: 22,580)
- Total memory (RAM): 149 TB (4.6 GB/core on average)
- Total storage: 5 PB = 5,120 TB

Overview

– RPP proposal –



CaSPAr will provide foundational **knowledge and data infrastructure** to link numerical weather prediction to environmental applications

Overview

– RPP proposal –

Year 1

Storage:

264 TB (**100% approved**)

- 116 TB (RDPS 2002-16)
- 55 TB (REPS 2013-16)
- 31 TB (CaLDAS, RE/DPS, GE/DPS Oct 2016-now)
- 62 TB (var. products 1 year)

Computational Resources:

275 core years (**49% approved**)

- 75 core yrs (User requests)
- 200 core yrs (Adding/formatting new data)

Year 2 & 3

Storage:

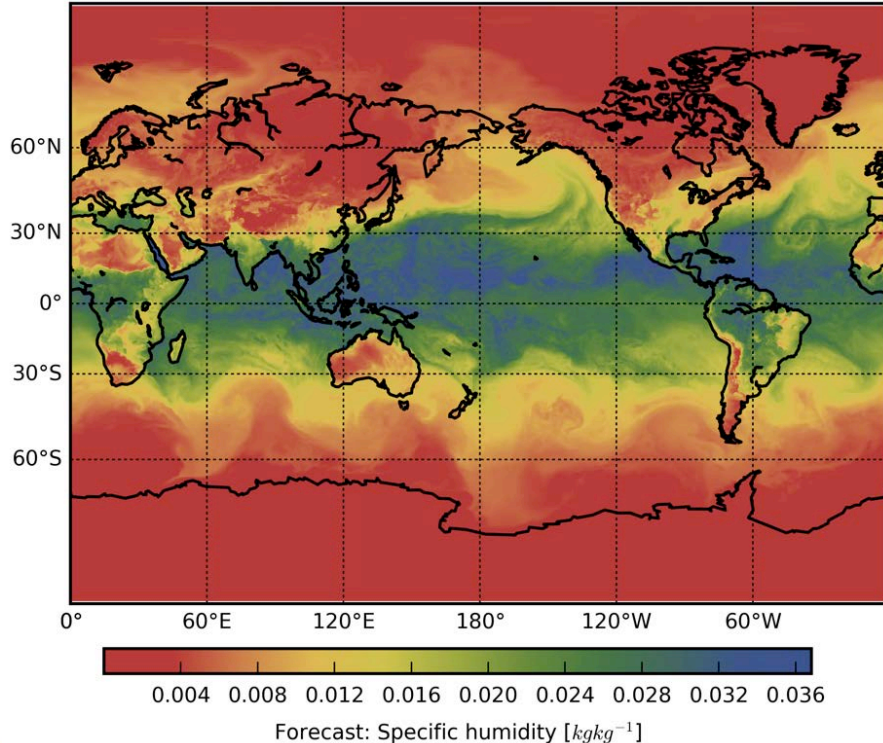
93 TB and 139 TB

Comp. Resources:

275 core yrs each

Products

– Example: Global Ensemble Prediction GEPS –



Forecasted time
steps [h]:

1,2,..., 72, 96, ... 384

Issues:

2/day

Ensemble size:

21

variables:

~21

Resolution:

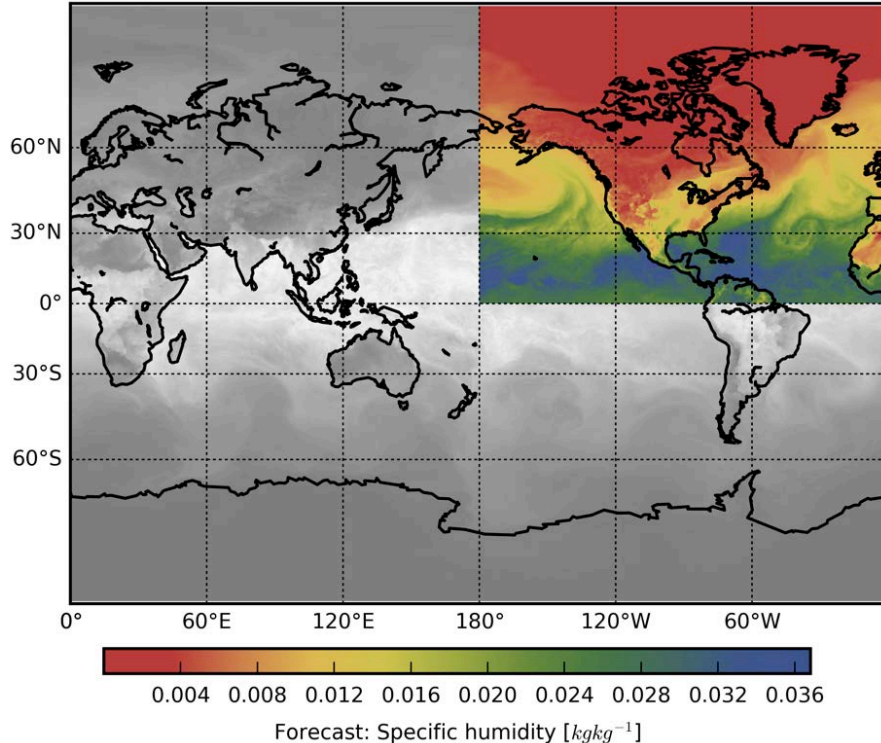
50 km

Domain:

global

Products

– Example: Global Ensemble Prediction GEPS –



Forecasted time
steps [h]:

1,2,..., 72, 96, ... 384

Issues:

2/day

Ensemble size:

21

variables:

~21

Resolution:

50 km

Domain:

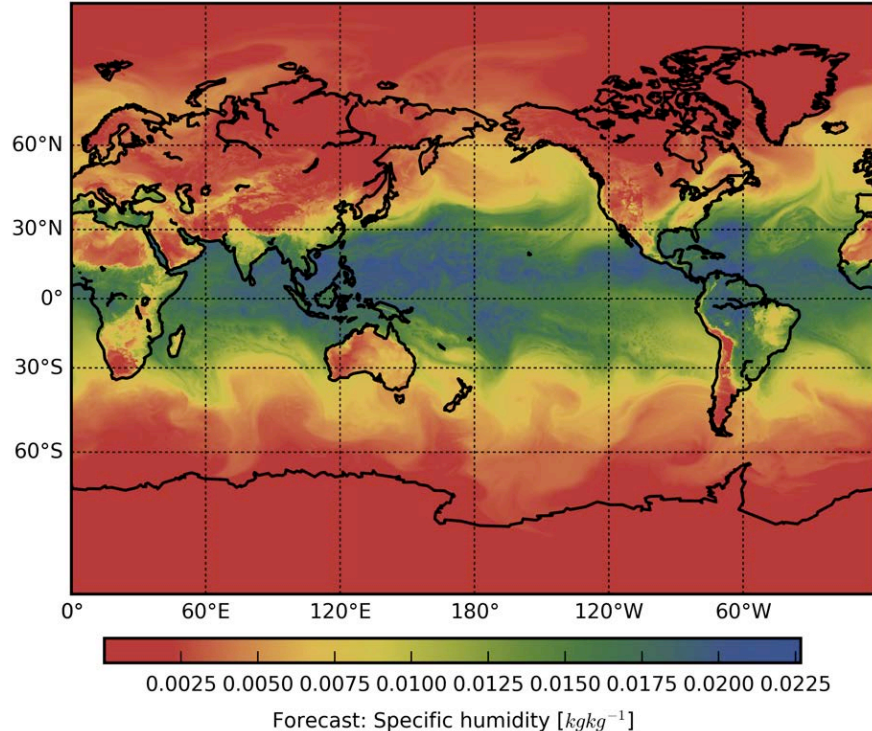
North America

Memory [GB/month]:

1245

Products

– Example: Global Deterministic Prediction GDPS –



Forecasted time
steps [h]:

1,2,...,144, 147,...,240

Issues:

2/day

Ensemble size:

1

variables:

~23

Resolution:

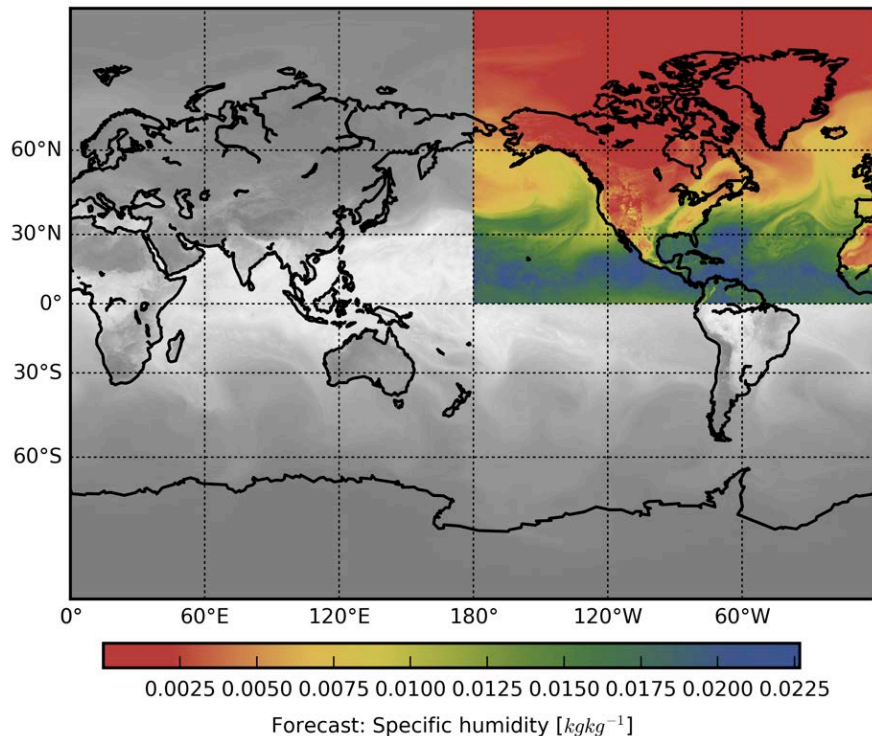
25 km

Domain:

global

Products

– Example: Global Deterministic Prediction GDPS –



Forecasted time
steps [h]:

1,2,...,144, 147,...,240

Issues:

2/day

Ensemble size:

1

variables:

~23

Resolution:

25 km

Domain:

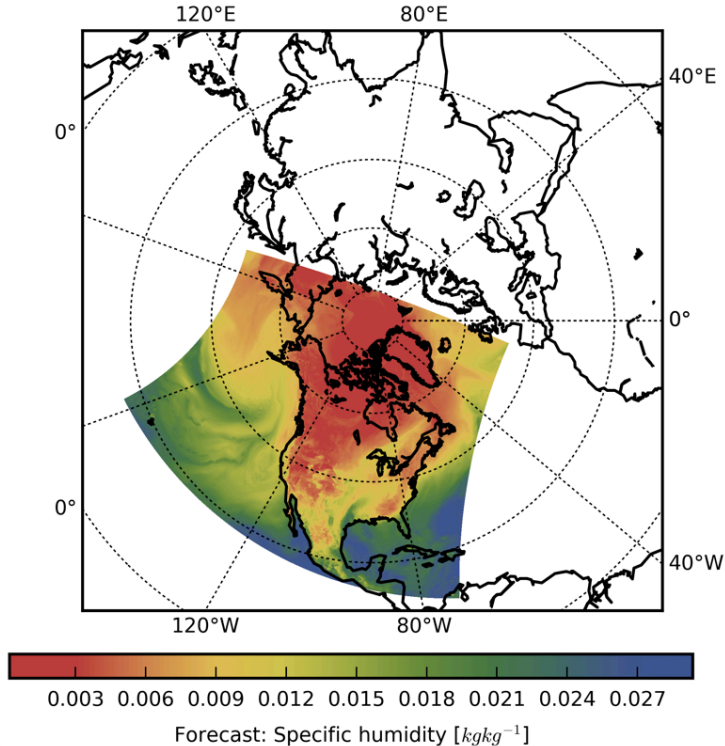
North America

Memory [GB/month]:

105

Products

– Example: Regional Ensemble Prediction REPS –



Forecasted time
steps [h]:

1,2,...,72

Issues:

2/day

Ensemble size:

21

variables:

~37

Resolution:

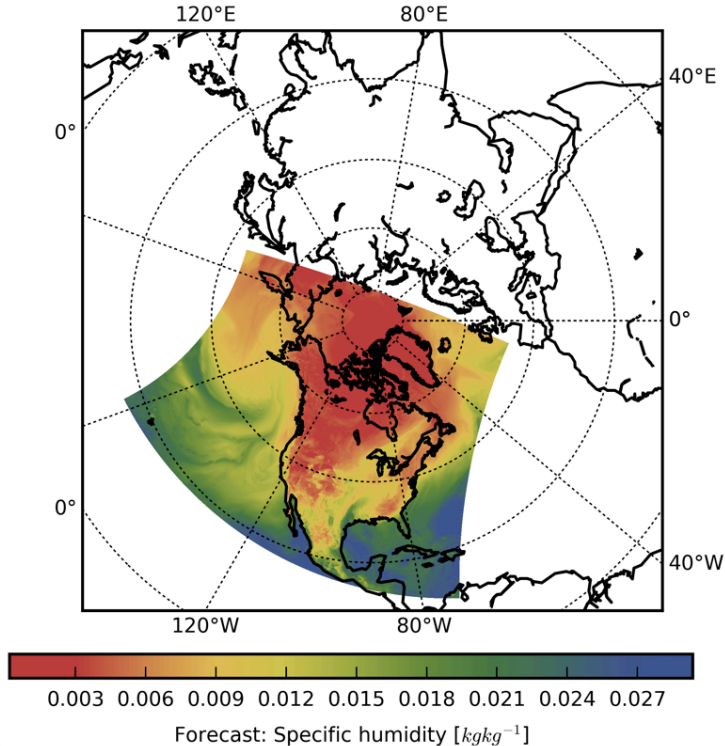
15 km

Domain:

North America

Products

– Example: Regional Ensemble Prediction REPS –



Forecasted time
steps [h]:

1,2,...,72

Issues:

2/day

Ensemble size:

21

variables:

~37

Resolution:

15 km

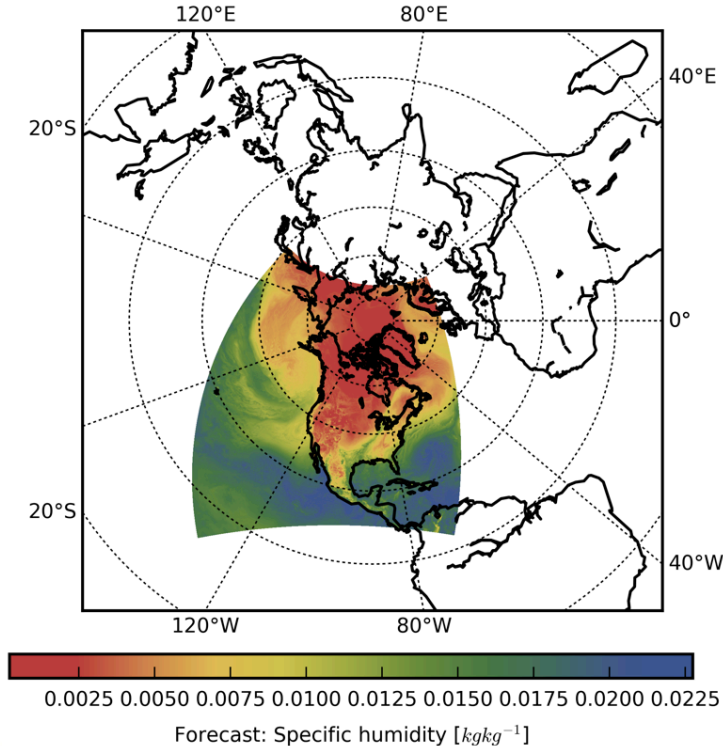
Domain:

North America

Memory [GB/month]: **1140**

Products

– Example: Regional Deterministic Prediction RDPS –



Forecasted time
steps [h]:

1,2,...,84

Issues:

4/day

Ensemble size:

1

variables:

~32

Resolution:

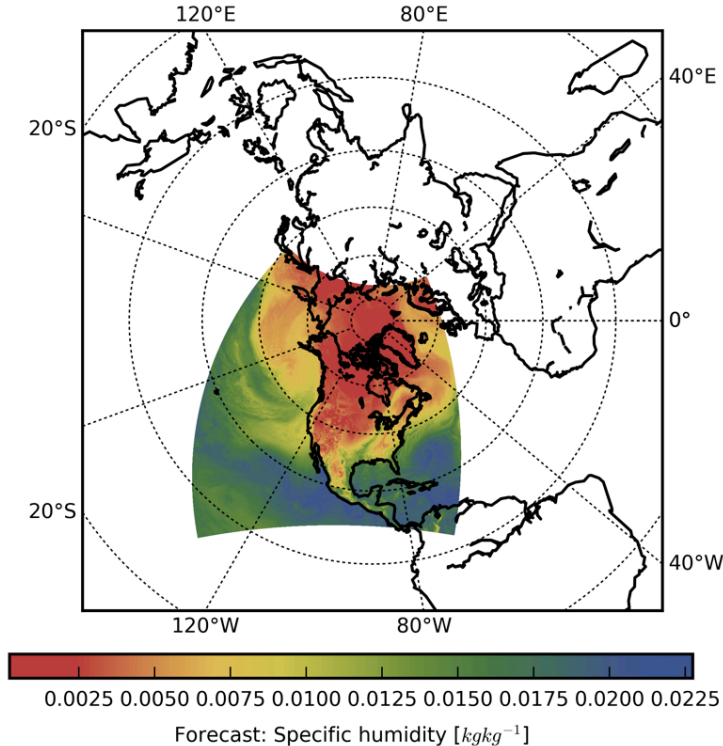
10 km

Domain:

North & Centr. America

Products

– Example: Regional Deterministic Prediction RDPS –



Forecasted time
steps [h]:

1,2,...,84

Issues:

4/day

Ensemble size:

1

variables:

~32

Resolution:

10 km

Domain:

North & Centr. America

Memory [GB/month]: **645**

Products

– all others –

- **CaLDAS: 756 GB/month**
Canada, 2.5 km resolution, 5 issues/day, 21 ens., ~6 variables
- **CaLDAS mean: 576 GB/month**
Canada, 2.5 km resolution, 24 issues/day, ~20 variables
- **HRDPS: 675 GB/month**
Canada, 2.5 km resolution, 4 issues/day, ~32 variables, $t = 1, 2, \dots, 84$ [h]
- **Reforecast: 6 GB/month**
North America, 50 km res., 1 issue/week, 4 ens., ~4 variables, $t = 24, 48, \dots, 768$ [h]
- **CaPA: 0.75 GB/month**
North America, 10 km resolution, 5 issues/day, 1 variable
- **CaPA: 6 GB/month**
Canada, 2.5 km resolution, 5 issues/day, 1 variable

Products

– all others –

- **CaLDAS: 756 GB/month**

Canada, 2.5 km resolution, 5 issues/day, 21 ens., ~6 variables

- **CaLDAS mean: 576 GB/month**

Canada, 2.5 km resolution

- **HRDPS: 675 GB/month**

Canada, 2.5 km resolution

- **Reforecast: 6 GB/month**

North America, 2.5 km resolution

- **CaPA: 0.75 GB/month**

North America, 2.5 km resolution

- **CaPA: 6 GB/month**

Canada, 2.5 km resolution, 5 issues/day, 1 variable

5,155 GB

new data added to CaSPAr
every month

...84 [h]

t = 24, 48, ...768 [h]

Architecture



Community



Front End Portal



esri

Portal for ArcGIS



PostgreSQL



openstack
CLOUD SOFTWARE

Back End Server



netCDF

CDO

Fortran

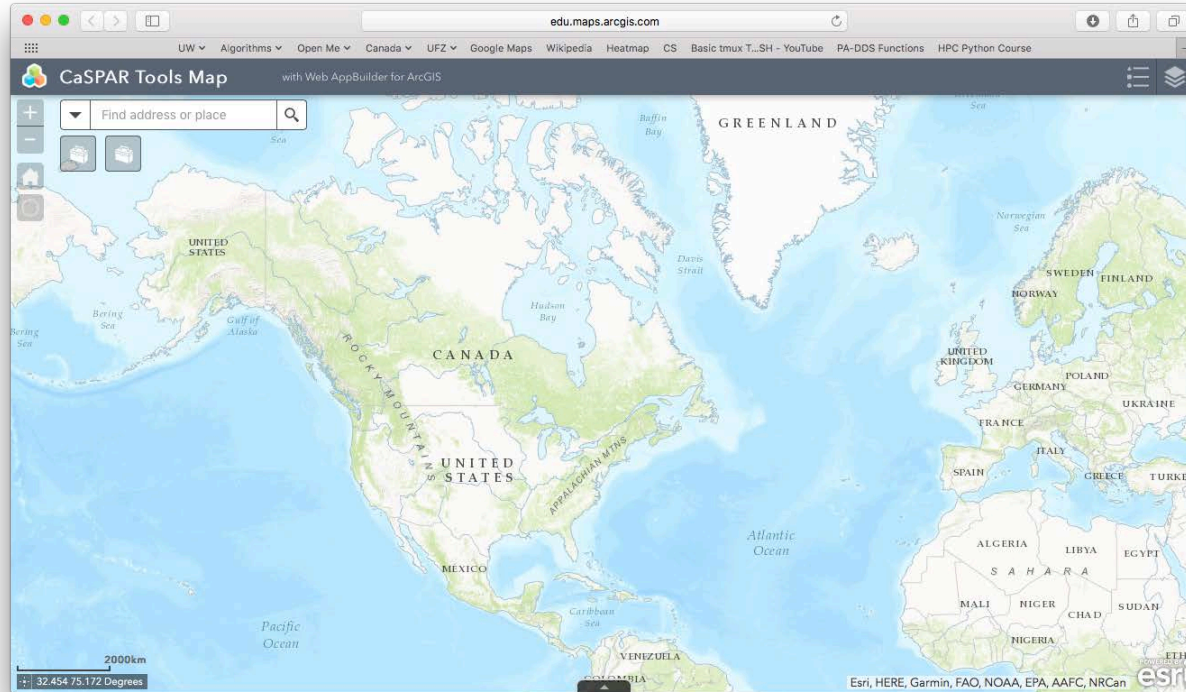


globus

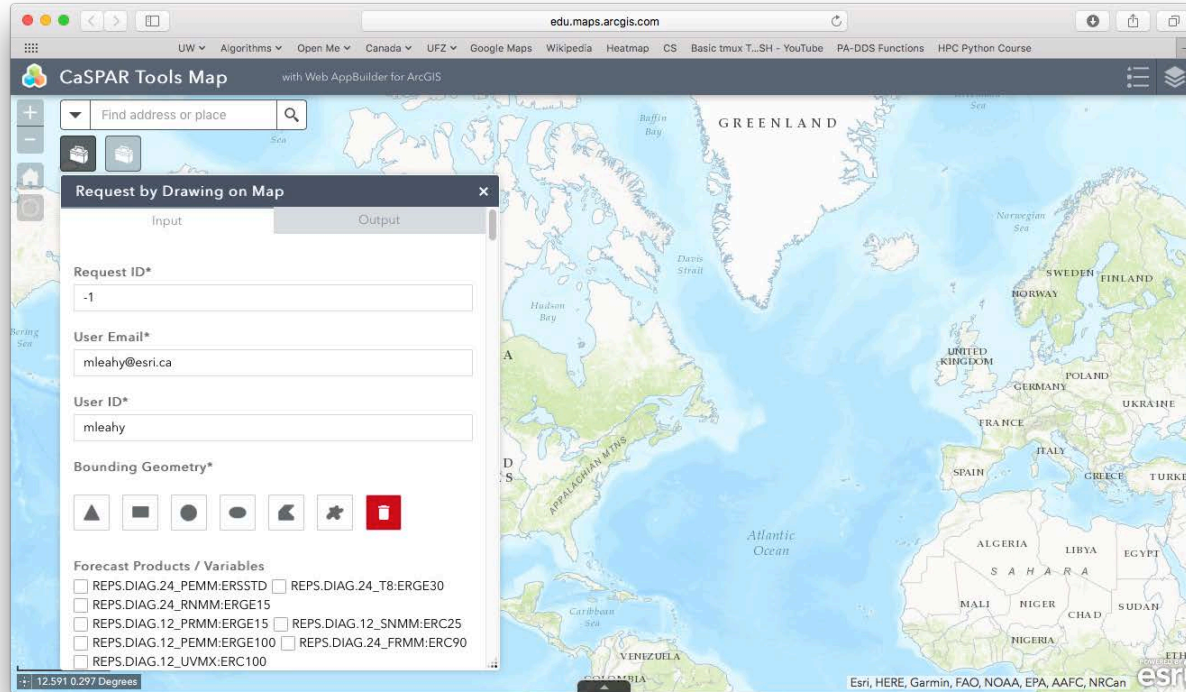
Front End View

REPS

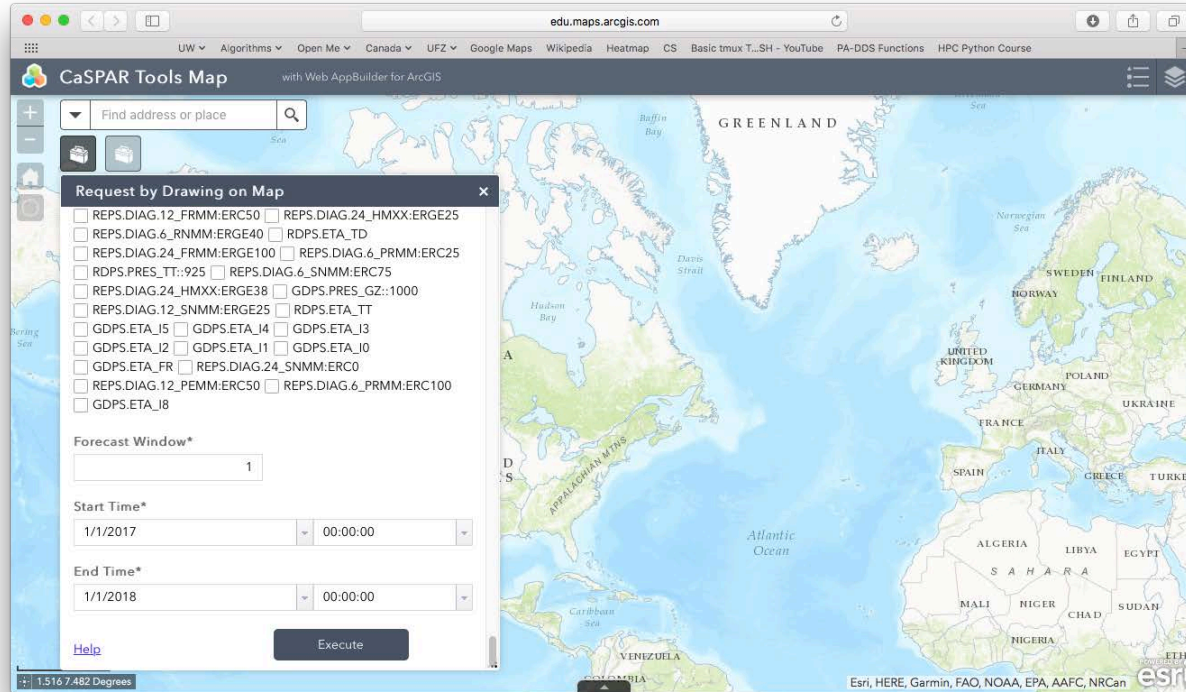
GDPS



Front End View



Front End View



Transfer



The screenshot shows the Globus web interface for managing data transfers. The top navigation bar includes "Manage Data", "Groups", "Support", and "tcrum". The main header shows "Transfer Files" and "Activity". The interface is divided into two panels, each showing a list of files for a specific endpoint and path.

Left Panel:

- Endpoint: rda#datashare
- Path: /ds084.4/2015/201502/
- Files list (partial):

File Name	Size
gdas1.sflux.2015022200.f00.grib2	160.70 MB
gdas1.sflux.2015022200.f03.grib2	164.71 MB
gdas1.sflux.2015022200.f06.grib2	165.58 MB
gdas1.sflux.2015022200.f09.grib2	165.58 MB
gdas1.sflux.2015022206.f00.grib2	162.98 MB
gdas1.sflux.2015022206.f03.grib2	167.60 MB
gdas1.sflux.2015022206.f06.grib2	167.60 MB
gdas1.sflux.2015022206.f09.grib2	167.60 MB
gdas1.sflux.2015022212.f00.grib2	163.55 MB
gdas1.sflux.2015022212.f03.grib2	168.36 MB
gdas1.sflux.2015022212.f06.grib2	168.36 MB
gdas1.sflux.2015022212.f09.grib2	168.36 MB
gdas1.sflux.2015022218.f00.grib2	162.47 MB
gdas1.sflux.2015022218.f03.grib2	166.04 MB
gdas1.sflux.2015022218.f06.grib2	166.04 MB
gdas1.sflux.2015022218.f09.grib2	166.04 MB
gdas1.sflux.2015022300.f00.grib2	161.19 MB
gdas1.sflux.2015022300.f03.grib2	165.23 MB

Right Panel:

- Endpoint: tcrum#cisl-toulon.scd.ucar.edu
- Path: /~/datasets/ds084.4/
- Files list (partial):

File Name	Size
gdas1.sflux.2015022200.f00.grib2	23.75 MB
gdas1.sflux.2015022200.f03.grib2	14.75 MB
gdas1.sflux.2015022200.f06.grib2	27.75 MB
gdas1.sflux.2015022200.f09.grib2	30.00 MB

Source: globus.org

Back End

– Workflow –

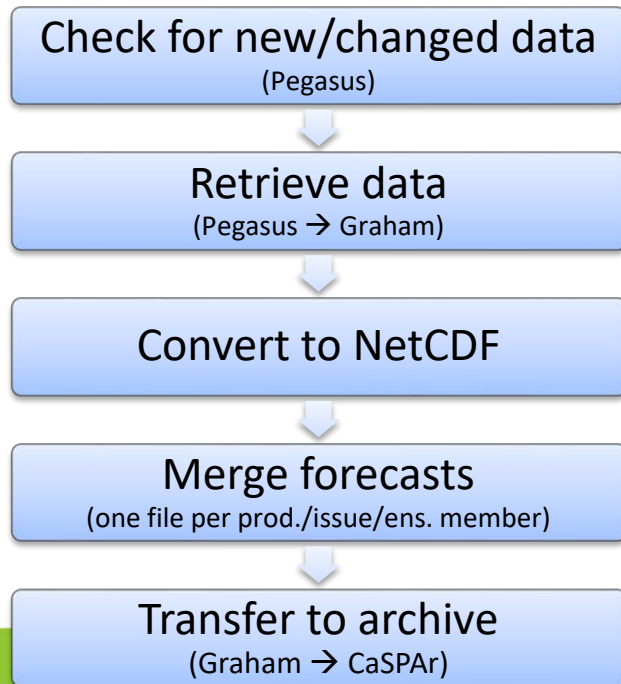
Task 1: Import new data to database

Task 2: Process data for user requests

Back End

– Workflow –

Task 1: Import new data to database

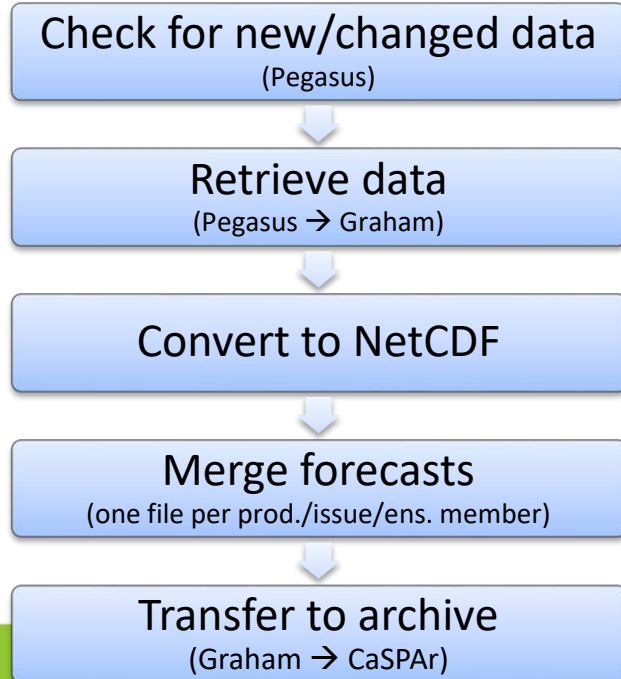


Task 2: Process data for user requests

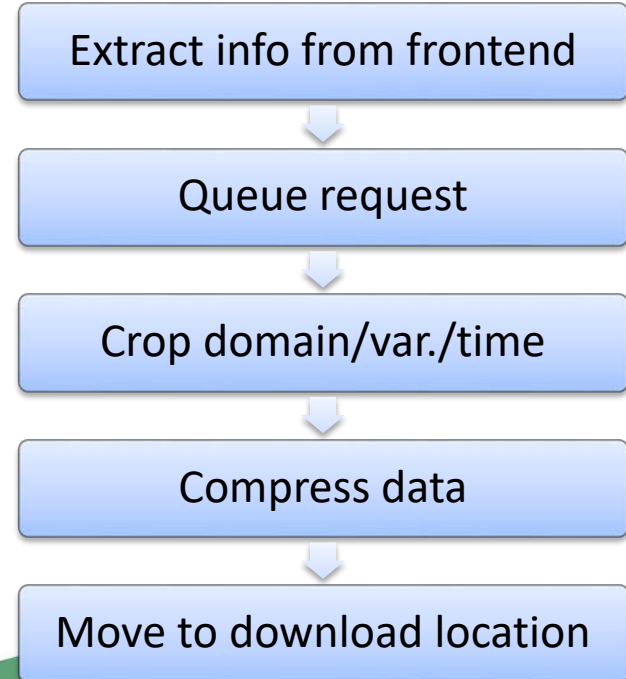
Back End

– Workflow –

Task 1: Import new data to database



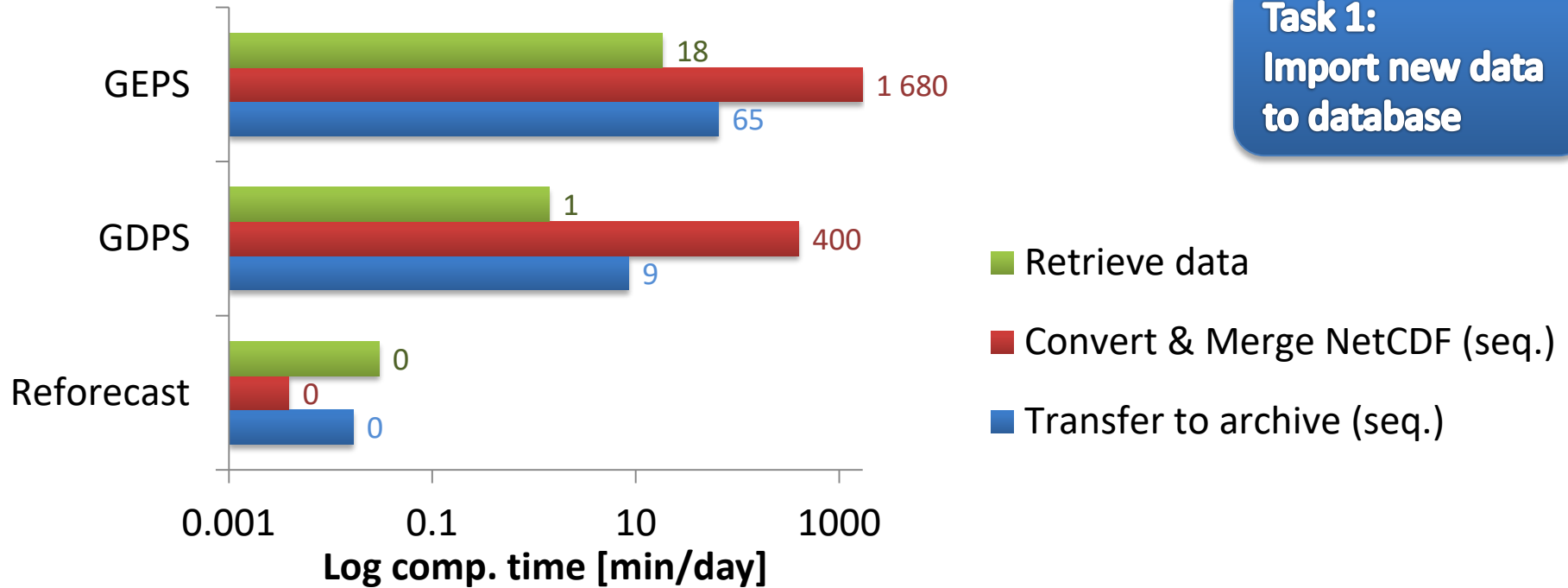
Task 2: Process data for user requests



Back End

– Some scenarios and their timing –

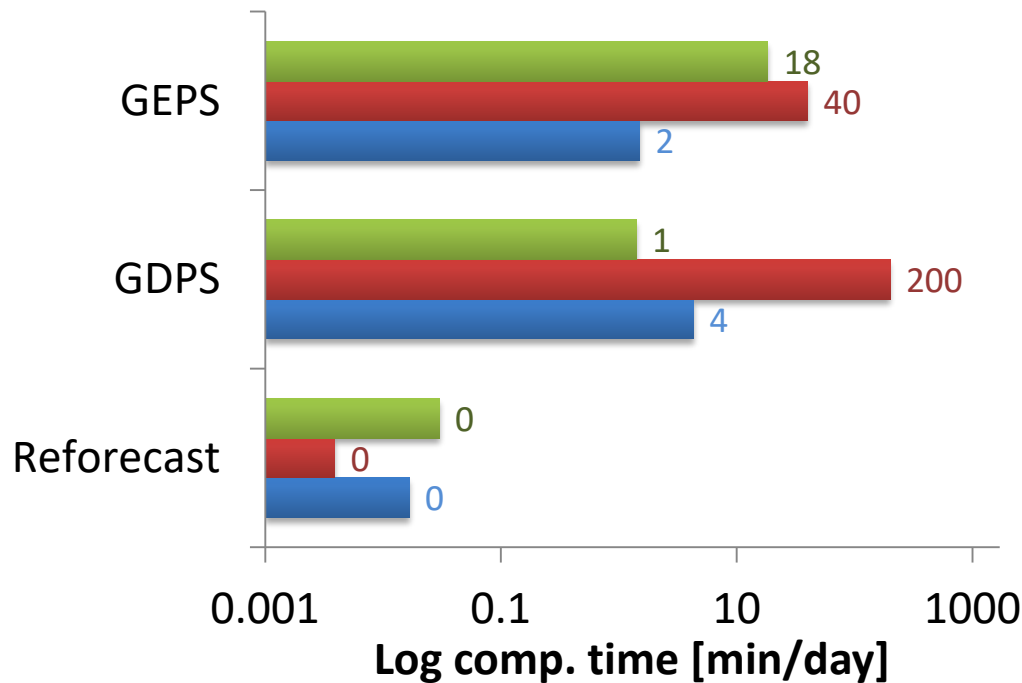
Task 1:
Import new data
to database



Back End

– Some scenarios and their timing –

Task 1:
Import new data
to database



Retrieve data

Convert & Merge NetCDF (par.)

Transfer to archive (par.)

Back End

– Some scenarios and their timing –

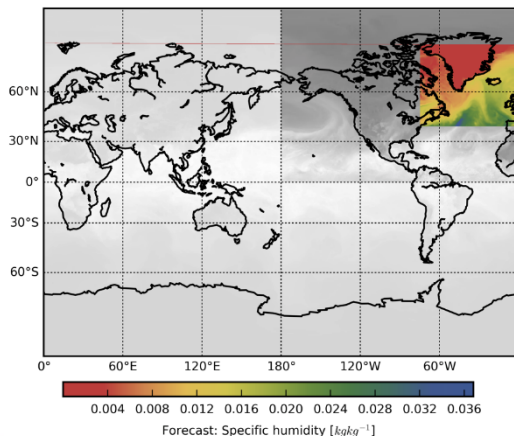
User request #1:

GEPS

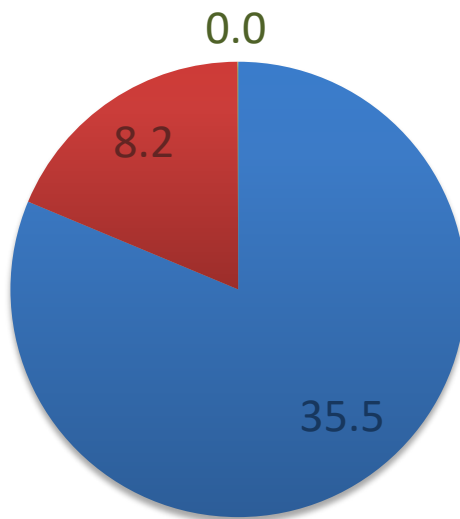
1 variable

1 day

21 ensemble members



Comp. time [min]



43.7 min

1 file (153 MB)

Task 2:
Process user
request

- Extract domain/var./time (seq.)
- Compress data
- Move to download dest.

Back End

– Some scenarios and their timing –

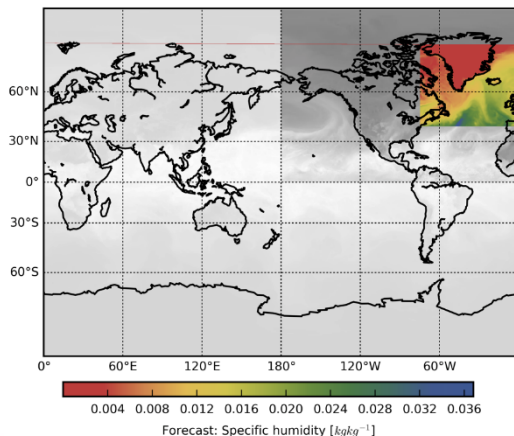
User request #1:

GEPS

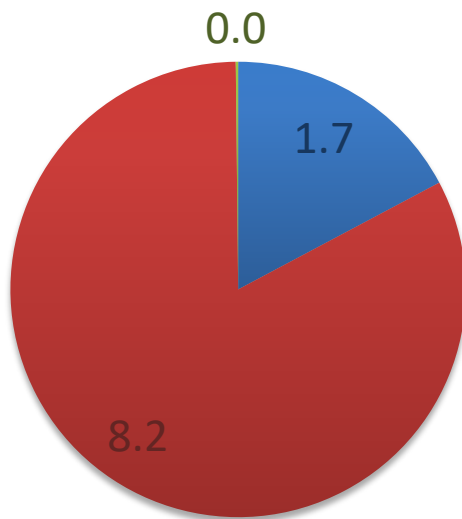
1 variable

1 day

21 ensemble members



Comp. time [min]



9.9 min

1 file (153 MB)

Task 2:
Process user
request

- Extract domain/var./time (par.)
- Compress data
- Move to download dest.

Back End

– Requirements –

- ✗ ARMNLIB to convert FST format
- ✓ NetCDF library
- ✓ Climate data operators (CDO) library
- ✓ Python
- ✓ Fortran compiler
- ✗ NetCDF operators (NCO) library
- ✗ NCAR common language (NCL)



CDO



Thank you!



Juliane Mai, Kurt C. Kornelsen, Bryan Tolson,
Paulin Coulibaly, François Anctil, Vincent Fortin,
Michael Leahy, Brent Hall



UNIVERSITÉ
LAVAL



esri Canada



Environment and
Climate Change Canada
Environnement et
Changement climatique Canada



NSERC
CRSNG



compute
canada | calcul
canada

Backup slides

Back End

– Some scenarios and their timing –

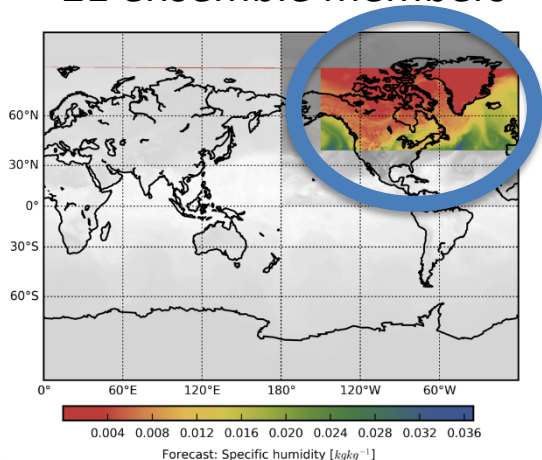
User request #2:

GEPS

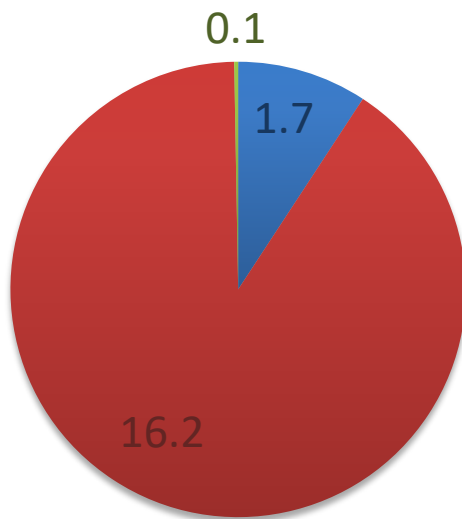
1 variable

1 day

21 ensemble members



Comp. time [min]



17.9 min

1 file (283 MB)

Task 2:
Process user request

- Extract domain/var./time (par.)
- Compress data
- Move to download dest.

Back End

– Some scenarios and their timing –

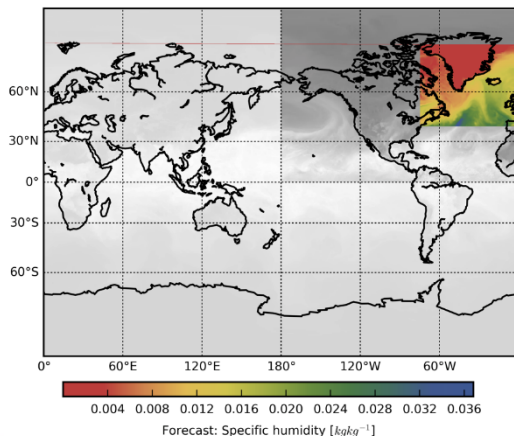
User request #3:

GEPS

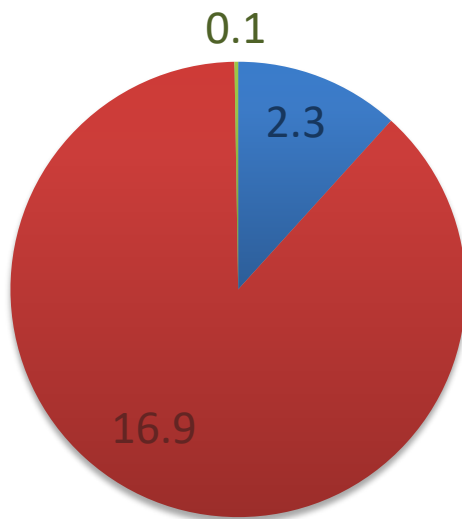
1 variable

2 days

21 ensemble members



Comp. time [min]



19.2 min

1 file (303 MB)

Task 2:
Process user request

- Extract domain/var./time (par.)
- Compress data
- Move to download dest.

Back End

– Some scenarios and their timing –

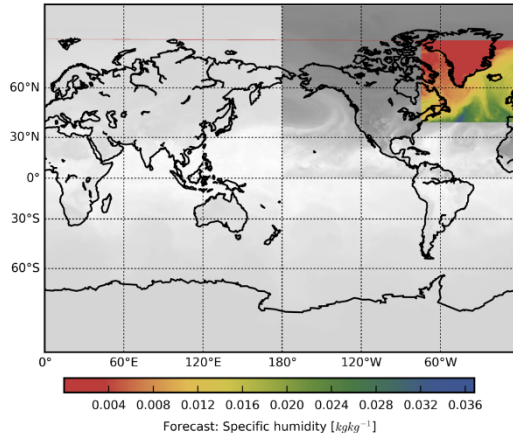
User request #4:

GEPS

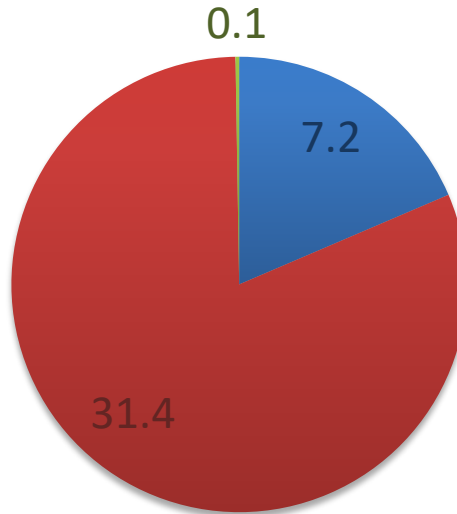
3 variables

1 day

21 ensemble members



Comp. time [min]



38.7 min

1 file (629 MB)

Task 2:
Process user request

- Extract domain/var./time (par.)
- Compress data
- Move to download dest.