



# GWF Computer Science Team Update

Kevin A. Schneider

Global Water Futures Operations Team Meeting  
January 21-23, 2019





# GWF Core Computer Science Team

- Kevin Schneider  
(co-lead, USask)
- Jimmy Lin  
(co-lead, UWaterloo)
- Ray Spiteri
- Chanchal Roy
- Jyoti Mondal
- Carl Gutwin
- Kevin Green
- Wenxian Guo
- Ian McConkey
- Binben Wang
- Tuan Nguyen
- Rayhan Ferdous
- Mahjabin Alam
- Hamid  
Khodabandehloo
- Manishankar Mondal
- Banani Roy
- Julianne Mai
- Zhenhua Li
- Homa Kheyrollah Pour
- Martin Gauch
- Yixin (Ethan) Wang
- Luchen Tan
- Yanping Li
- Gazi Hasnat
- Jarin Tasnim
- Zonayed Ahmed
- Sean Wang
- Yukun Li
- Ilia Rudnisky
- ...



# Goal

- Create a GWF platform and toolset
  - To support integration of hydrological, water quality and economic models
  - To provide infrastructure for migrating existing modelling tools to the common GWF platform
  - To provide new algorithms and approaches for accelerating modelling tool performance
  - To support interactive visualization and end-user mobile applications
  - To provide a next-generation data management environment



# Inception Strategy

- Objectives
  - Improved **evolution** of hydrological modelling tools
  - Improved **scalability** of hydrological modelling
  - Improved **delivery** of data and knowledge
  - Improved **management** of sensor data
- Deliverables
  - Software reference architecture and platform
  - Toolsets and frameworks
- 5 Projects
- Computing Infrastructure
- Radiam – CANARIE project



# Radium – CANARIE project

- Data management
  - for finalized datasets ready for publication is getting easier: many repository solutions
  - Data management for data actively collected, processed and analysed is equally important
- Increasingly, large research projects are carried out at a scale where placing all the data in one storage device or one institution is infeasible
  - *“GWF has identified the need to have a centrally accessible platform and catalogue to link together and manage current researcher data and metadata over the many locations where that data is stored. Ideally, this platform would provide a scalable integration of our resources in Compute Canada, at the University of Saskatchewan, and other storage spaces within and outside of the USask domain.”*
- Need tools (Radium) to manage active research data:
  - Through the entire data lifecycle
  - While it is being actively collected, processed and analysed
  - When the data is geographically distributed
  - To facilitate publication into repositories once datasets are finalized





# GWF Computing Infrastructure Phases

- Phase 0 – 15%
  - basic infrastructure
    - 40gbps, cluster management, login node, rack, UPS, network, gateway
  - individual lab workstations / dedicated compute nodes
- Phase 1 – 55%
  - HPC File System
    - 240-400TB, 1200MB/s
  - Compute/GPU Nodes
    - 384 Core, 2-4 V100s, 1920GB RAM
- Phase 2 – 30% (if cost effective to delay purchase)
  - based on experience with Phase 1
    - expand storage, compute and GPUs and necessary



# Projects & Leads

1. Software Reference Architecture (i.e., platform architecture)
  - Kevin Schneider, Chanchal Roy
2. Software Co-evolution (i.e., migration and renovation)
  - Kevin Schneider, Chanchal Roy
3. Hydrological Model Acceleration
  - Ray Spiteri
4. Interactive Visualization
  - Jyoti Mondal, Carl Gutwin
5. Managing Environmental Monitoring Data (Data Cusinart)
  - Juliane Mai, Jimmy Lin





# Hydrological Model Acceleration

- Lead: Ray Spiteri
- Purpose: Extend capability or accelerate the performance of new or existing GWF hydrological modelling frameworks
- Personnel
  - Kevin R. Green, Research Scientist, September 1, 2018 – current
  - Wenxian Guo, part-time Research Assistant, May 1-September 30, 2018
  - Ian McConkey, BSc student, May 1-August 31, 2018
  - Binben Wang, MSc student, September 1, 2018-current
  - Tuan Nguyen, MSc student, co-supervised, September 1, 2018-current



# Hydrological Model Acceleration - Completed

- **CHM updates**
  - Module organization
    - Factories for modules and stations.
    - Build system update to allow easier development of external tools that use CHM components.
  - Mesh input with ability to reorganize mesh elements.
- **Mesher update**
  - Allow orderings of mesh elements to be changed to optimize properties of nearest-neighbour connectivity matrices.
- **Knowledge base**
  - CHM tutorial/setup documentation (github)
  - Captured baseline performance of CHM prior to making changes.
  - Tested various linear algebra libraries (need to work well with raw MPI and OpenMP)
  - Reviewed Kokkos and Tpetra (within the Trilinos software stack) to ensure they can be used for individual modules without destroying the possibility of other modules using raw MPI/OpenMP.
    - Kokkos: Performance portability programming model for manycore devices
    - Tpetra: Implements parallel sparse linear algebra objects
    - NOTE: the Kokkos + Tpetra combination abstracts a hybrid MPI+X programming model for linear algebra, where X can be either OpenMP or CUDA



# Hydrological Model Acceleration – In progress

- Current
  - Organizing distance-dependent mesh connectivities to determine how to distribute a mesh in such a way that CGAL methods will still work with their local portions of the mesh.
- Near term
  - Replace existing linear algebra with Tpetra stack (in current shared memory code)
  - Design ghosted distributed mesh around preserving CGAL's methods.
  - Bring the previous two points together.
- Other
  - A patch has been submitted to scipy (<https://github.com/scipy/scipy/pull/9525>) regarding its interface to SuperLU (a C library for direct solvers for linear algebraic systems), and is currently being reworked based on feedback. The patch fixes an issue reported over a year ago at (<https://github.com/scipy/scipy/issues/7700>).



# CRHM Migration & Restructuring Phases

- Exploration
  - Identify migration issues using source code analysis and prototypes
- Develop Next-Gen CRHM Environment and Architecture
  - Based on requirements identified during exploration
- Migrate and Restructure
  - Convert current CRHM to new environment



# CRHM Migration & Restructuring - Exploration

- Analyzed different versions of CRHM
- Migrated CRHM 2012 Console version to Microsoft Visual C++ (MSVC++) 2017
- Prototyped CRHM GUI in MSVC++ 2017
- Created t-charts using observation files in MSVC++ 2017
- Partially migrated CRHM 2018 version to MSVC++ 2017 core version.
- Connected CRHM 2018 GUI MSVC++ prototype to partially migrated CRHM 2018 MSVC++ version
  - Open and run a project file
  - Integrated T-chart ActiveX control to show simulation results
  - Save and Save as functionality
  - Close project file.
- Integrated migrated CRHM with Google Test Framework
- Setup a Github repository for version control





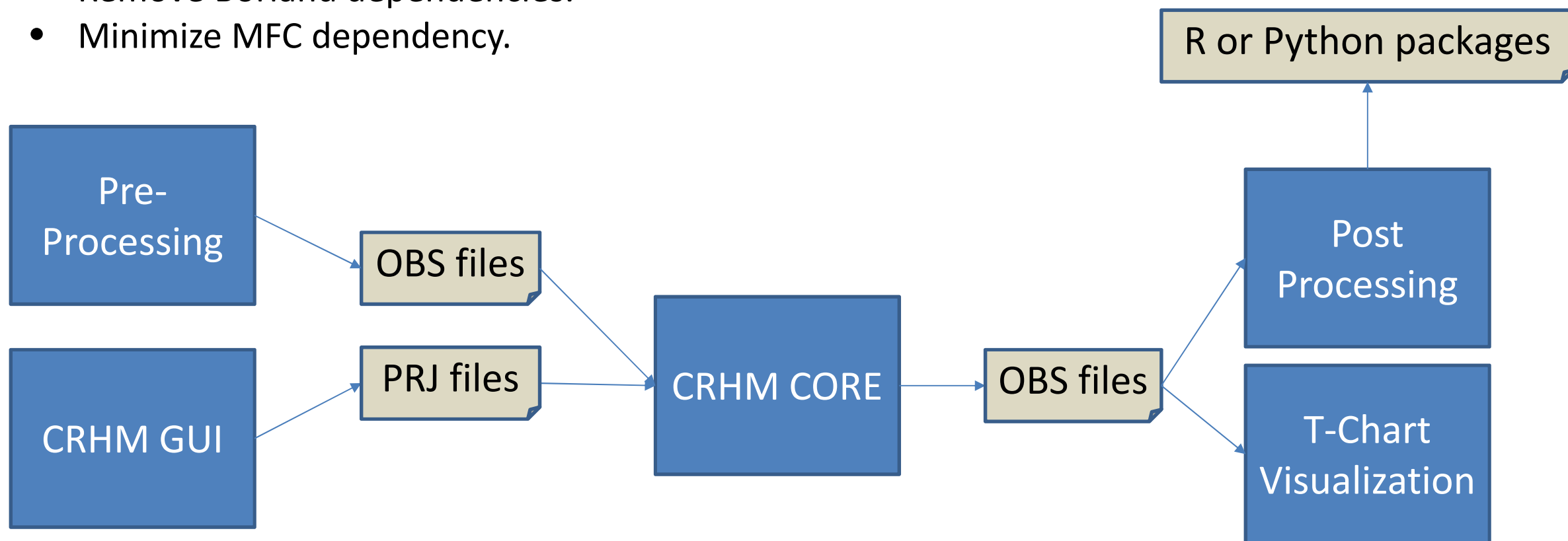
# Next-Gen CRHM Development Environment

- **A modern compiler and libraries**
  - E.g., MS Visual C++ (VC++) 2017, GCC
- A Distributed **Version Control System** to support multi-version experimentation
  - E.g., Git
- An **Automated testing** framework to support rapid deployment of new features
  - E.g., test scripts, Travis-CI, Google Test
- **Virtual Containers** to support packaging, deployment & portability
  - E.g., Docker – Build, Ship and Run



# Next-Gen CRHM

- Code Separation: Core and GUI components.
- Develop a console version of CRHM that runs in a standard C++ environment.
- Create APIs to access CRHM core data structures.
- Use the API to populate components for GUI applications.
- Remove Borland dependencies.
- Minimize MFC dependency.



# CRHM Restructuring – Long term

- Plug-in architecture
  - identify and isolate dependencies; modularize
- Scientific Workflow Management System
  - Support logging and provenance tracking
  - Monitor software usage and performance
- Next-Gen GUI for PRJ files and to view T-Charts
- CRHM as a “family of products” (i.e., Product Family Engineering)
  - Web/Mobile/Desktop App
  - Cloud Computing
  - Different uses (e.g., expert, training)



# CRHM Migration - Personnel

- Kevin Schneider, Faculty, Computer Science
- Chanchal Roy, Faculty, Computer Science
- Rayhan Ferdous, M.Sc. Student
- Mahjabin Alam, M.Sc. Student
- Hamid Khodabandehloo, M.Sc. Student
- Manishankar Mandol, Post Doc
- Banani Roy, Research Associate
- 2 undergraduate summer students



# Publications in preparation

- Banani Roy, Kevin Schneider, Chanchal Roy, Rayhan Ferdous and Mahjabin Alam. An Approach for effectively Understanding a Legacy Software System for Migration: A Case Study using a Borland Application. In *Proceedings of International Conference on Software Architecture (ICSA) 2019*. In review.
- Rayhan Ferdous, Banani Roy, Chanchal Roy and Kevin Schneider. ProvMod: A Workflow Programming Model to Offer Automated Provenance and Log Analysis. In *Proceedings of International Conference on Web Services (ICWS), 2019*. In review.





# Visualization Project Goals



- **Browser-based visualization** technologies for complex geospatial queries in interaction time
- A unified platform where the researchers can share the findings and work in a **collaborative environment**
- **Machine learning** (Deep Learning) models that leverage the big-data stored in Compute Canada that can be used to create summarized interactive visualizations for typical laptops
- **Innovative visualizations** for CRHM and WRF simulation datasets

# Analysis Phase: Existing Datasets and Software

A collection of existing visualizations and web platforms investigated by the research group

<https://d3js.org/>  
<https://bl.ocks.org/syntagmatic/05a5b0897a48890133beb59c815bd953>  
<http://bl.ocks.org/syntagmatic/raw/3150059/>  
<https://bl.ocks.org/mbostock/4163057>  
<https://riccardoscalco.github.io/textures/>  
<https://dabrownstein.com/tag/d3/>  
<https://gist.github.com/curran/0ad2eef56811e04f3aa6>  
<http://bl.ocks.org/syntagmatic/fc62d1de6ada1e06b803>  
<https://bl.ocks.org/mbostock/9539958>

....

<https://reactjs.org/>  
<https://angular.io/>  
<https://www.ventusky.com>  
<https://www.youtube.com/watch?v=igihU5x6jhw>  
<https://www.youtube.com/watch?v=y3o-uFjd4Eo>  
<https://www.youtube.com/watch?v=79n6SXGPp7U>  
<https://www.youtube.com/watch?v=jr6Ly7RBZg>  
<https://www.youtube.com/watch?v=m0VM4dp640Y>  
<https://earth.nullschool.net/#current/wind/surface/level/orthographic=-38.78,26.97,398>  
<https://www.windy.com/52.140/-106.608?51.365,-106.608,7>  
<http://www.meteoearth.com/#/,122.23,29.11,2.50,10,1,1,0,0,0,time=1521147270>  
<http://www.maartenlambrechts.com/2014/07/29/introducing-the-weather-browser.html>  
[https://soilgrids.org/#!/?layer=TAXNWRB\\_250m&vector=1](https://soilgrids.org/#!/?layer=TAXNWRB_250m&vector=1)  
<https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=stelprdb1254424>  
<http://www.climex-project.org/en/module-hydrology>  
<http://sunsp.net/demo/GeogTreeMaps/>

....

# Analysis Phase: Existing Datasets and Software

Remote site: /project/6008034/Model

Weather Research and Forecasting (WRF) data at graham.computecanada.ca (in TBs)

Filename	Filetype	Last modified	Permissions	Owner/Group
..				
CanRCM4_Corrected	File folder	6/11/2018 11:3...	drwxrwsrwx	bzdravko rpp-hwheater
GEM-CaPA_interpolated_0.125	File folder	7/27/2018 5:54:...	drwxr-s---	melshamy rpp-hwheater
WRF	File folder	9/5/2018 3:04:2...	drwxrwsr--	bzdravko rpp-hwheater
CCRN	File folder	9/21/2018 11:3...	drwxr-s---	aez849 rpp-hwheater
CanRCM4_Raw	File folder	10/9/2018 9:04:...	drwxrwsr-x	bzdravko rpp-hwheater
CanRCM4-22	File folder	10/9/2018 9:04:...	drwxrwsr-x	bzdravko rpp-hwheater
AEP_MESH	File folder	12/4/2018 12:5...	drwxrwsr-x	zelalem rpp-hwheater
Land_Cover_2005v2_TIFF.zip	Compressed...	9/14/2018 2:32:...	-rw-r-----	melshamy rpp-hwheater

Selected 1 directory.

Each day corresponds to about 1GB data in a netCDF file with about 100 weather parameters

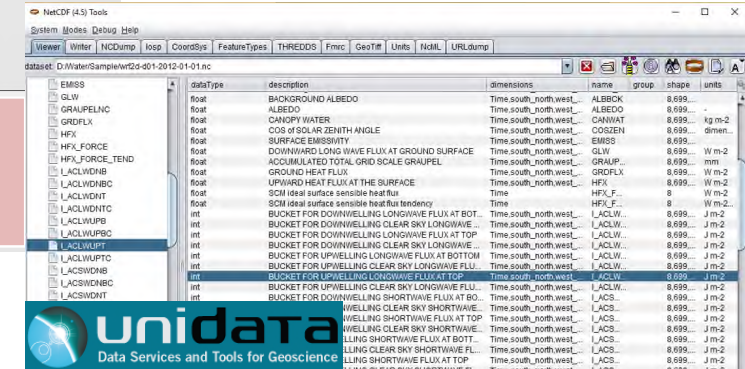
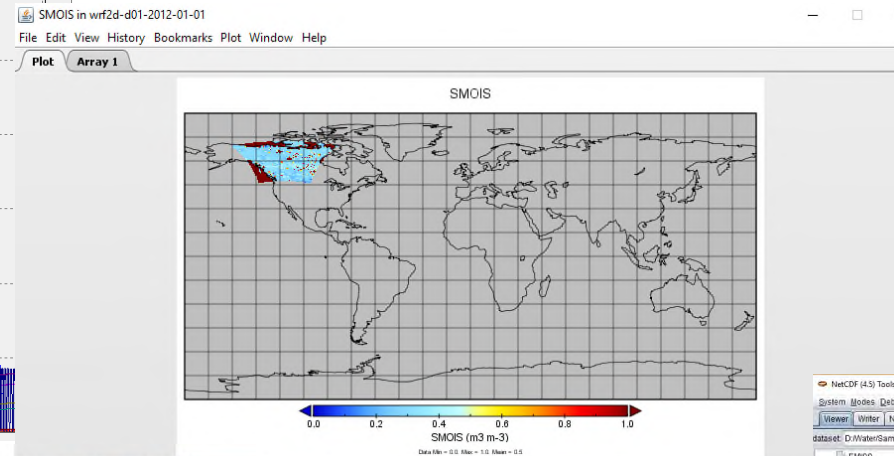
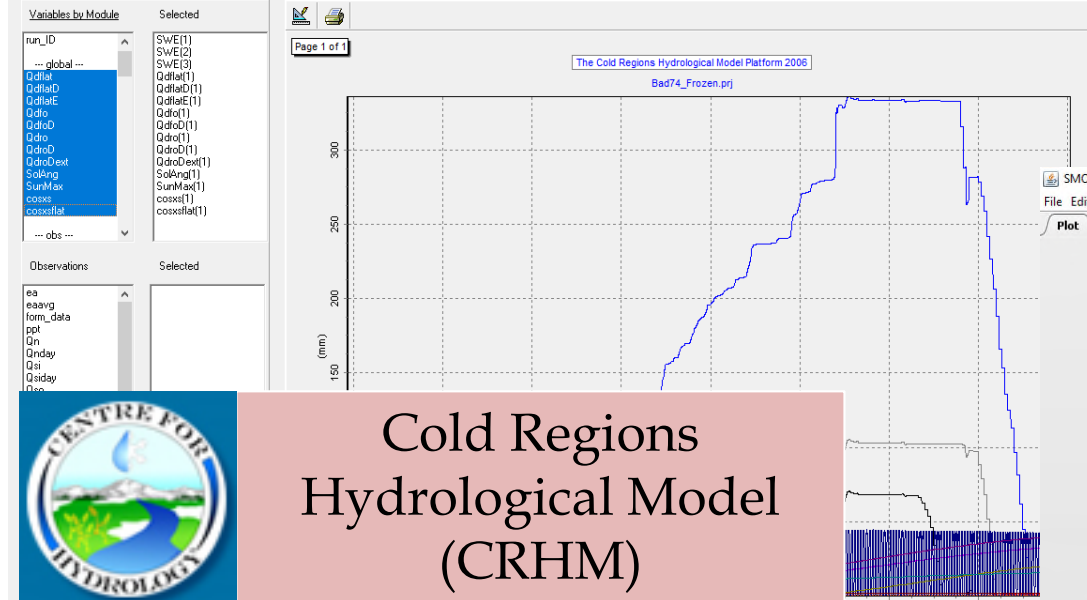
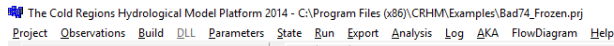
NetCDF file viewer showing metadata for 'wrf2d-d01-2012-01-01.nc'.

File type: Hierarchical Data Format, version 5

```
netcdf file:/D:/Water/Sample/wrf2d-d01-2012-01-01.nc {
  dimensions:
    Time = UNLIMITED; // (8 currently)
    south_north = 699;
    west_east = 639;
    soil_layers_stag = 4;
    DateStrLen = 19;
    west_east_stag = 640;
    south_north_stag = 700;
  variables:
    float ACGRDFLX(Time=8, south_north=699, west_east=639);
      :FieldType = 104; // int
      :MemoryOrder = "XY ";
      :description = "ACCUMULATED GROUND HEAT FLUX";
      :units = "J m-2";
      :stagger = "";
      :coordinates = "XLONG XLAT";
      :_ChunkSize = 4, 515, 471; // int

    float ACHFX(Time=8, south_north=699, west_east=639);
      :MemoryOrder = "XY ";
      :description = "ACCUMULATED UPWARD HEAT FLUX AT THE SURFACE";
      :units = "J m-2";
      :stagger = "";
      :coordinates = "XLONG XLAT";
      :FieldType = 104; // int
```

# Analysis Phase: Existing Datasets and Software



# Panoply netCDF, HDF and GRIB Data Viewer

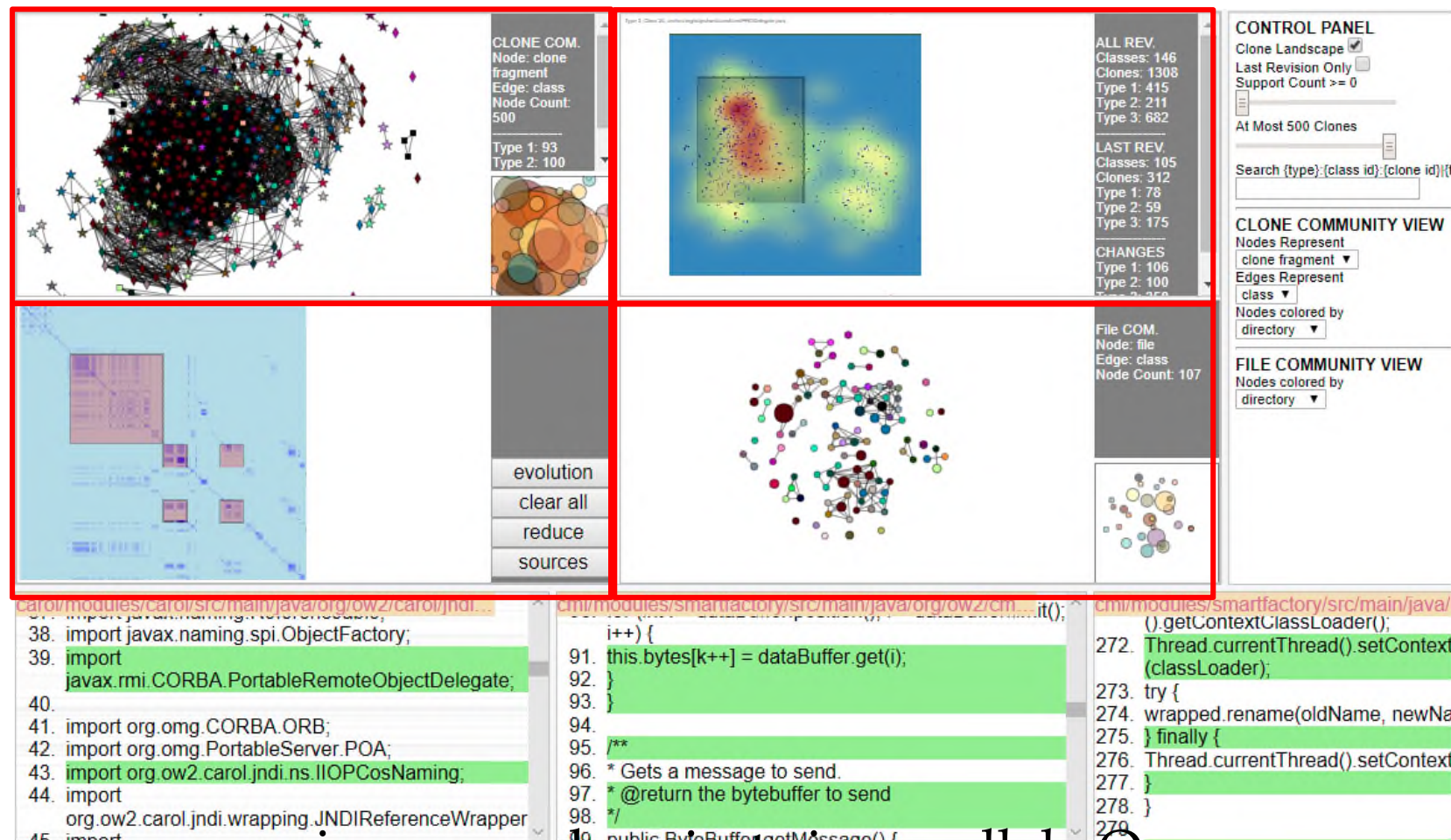
UniDATA netCDF Viewer



# Browser-based visualizations 1/5

A web-platform that supports 4 inter-linked visual displays.

Interaction in one display updates the others.



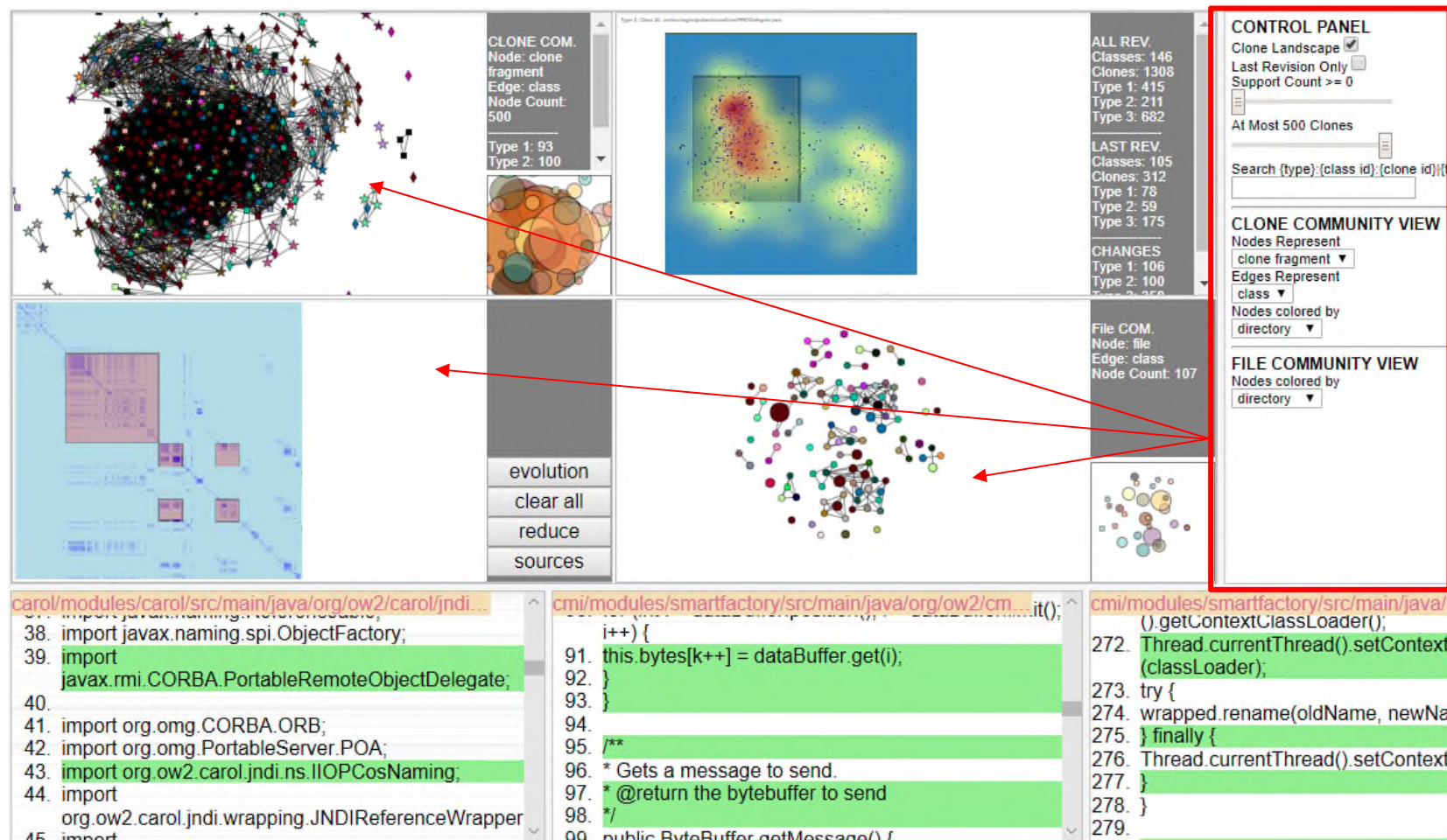
Note: For faster development, we are running several projects in parallel. Our group investigated netCDF file parsers to visualize WRF datasets, and meanwhile, used existing lab dataset to develop this web platform.

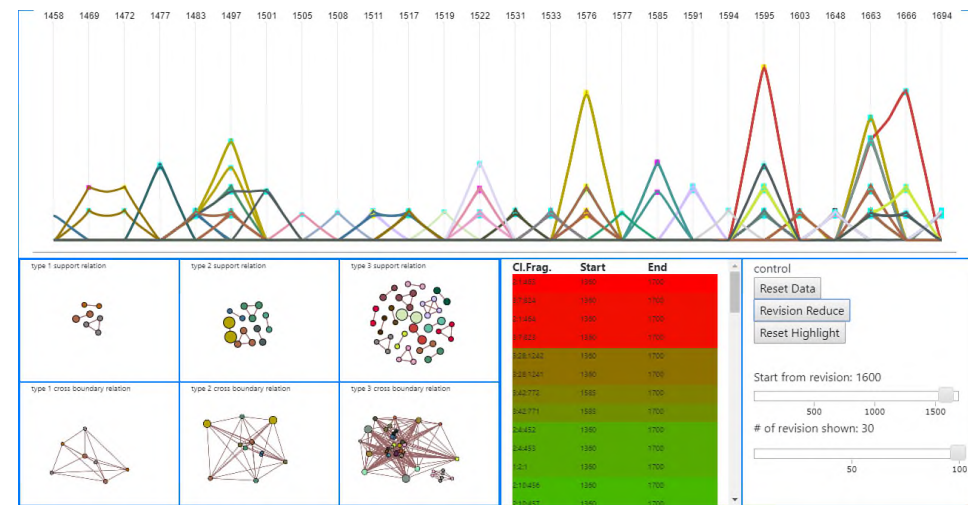


# Browser-based visualizations 2/5

An interactive control panel that filters the information in all displays in real time.

The current system is using d3-based rendering technology.

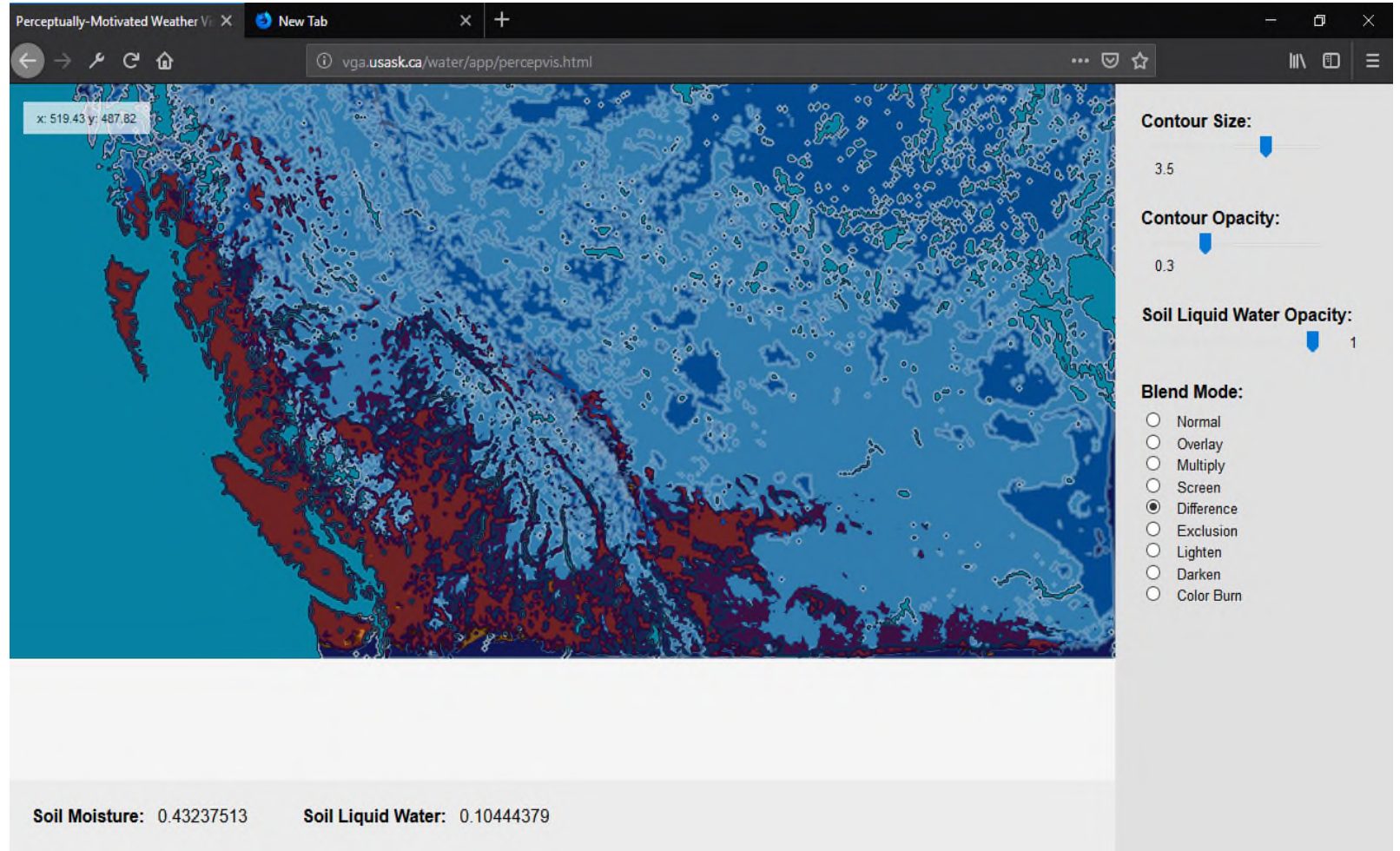






# Browser-based visualizations 4/5

We are analyzing how the correlation between multiple parameters of WRF datasets can be analyzed using color blending in a web platform.



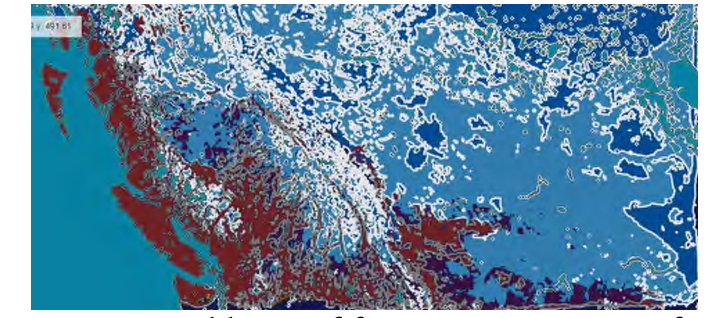
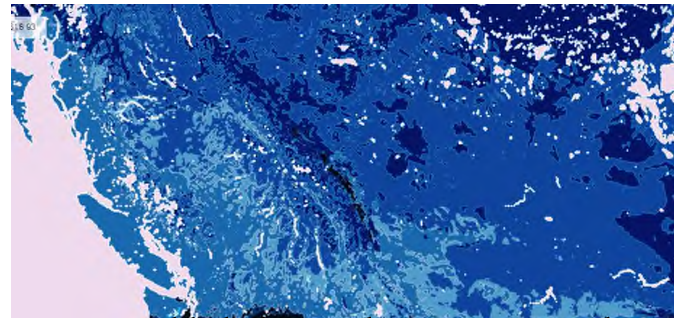
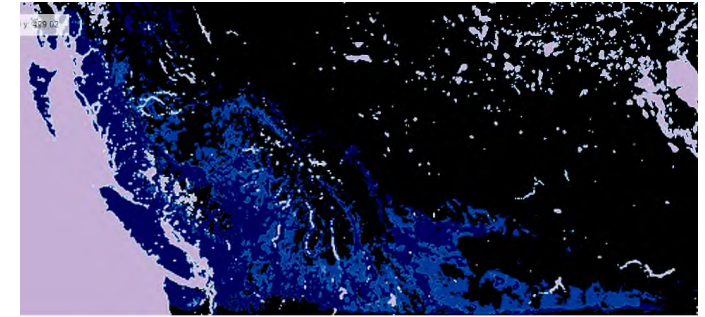
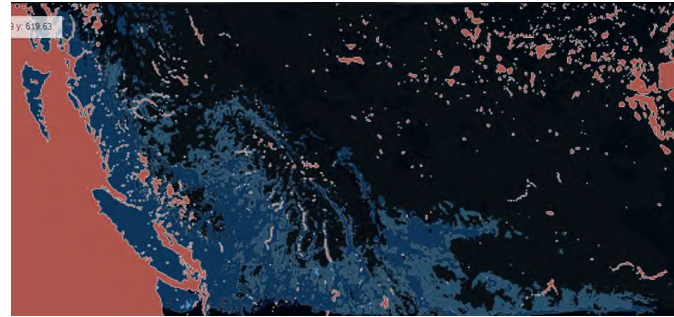
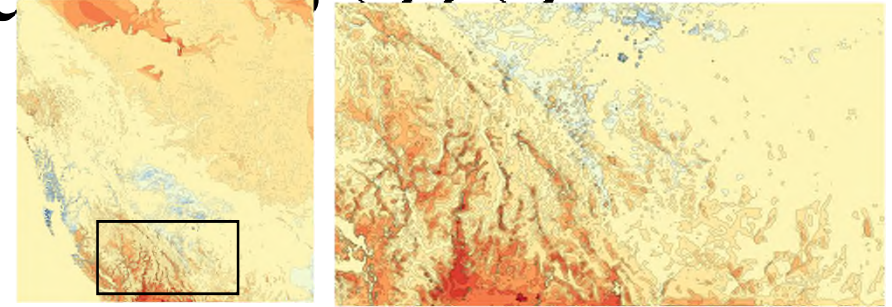
Here a 'difference' blending mode has revealed the correlation between the soil moisture and soil liquid

# Browser-based visualizations 5/5

Users can zoom/pan, select different blending modes, change the contour threshold, opacity and thickness.

<http://vga.usask.ca/water/app/percepvis.html>

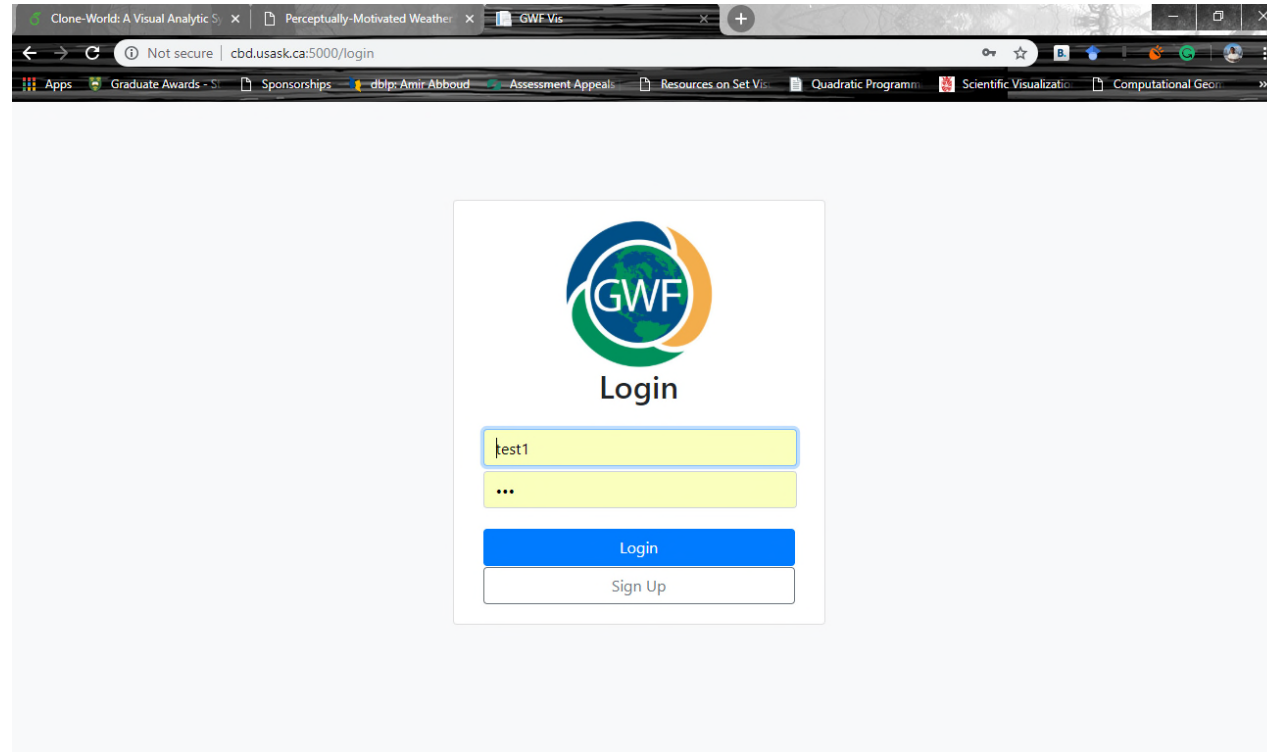
Note: use Chrome, zoom is disabled in this particular release



Not all blending mode are equally effective, we found the 'difference' mode to reveal the correlation the best for some pairs of weather parameters.



# Collaborative Environment



GWF login interface

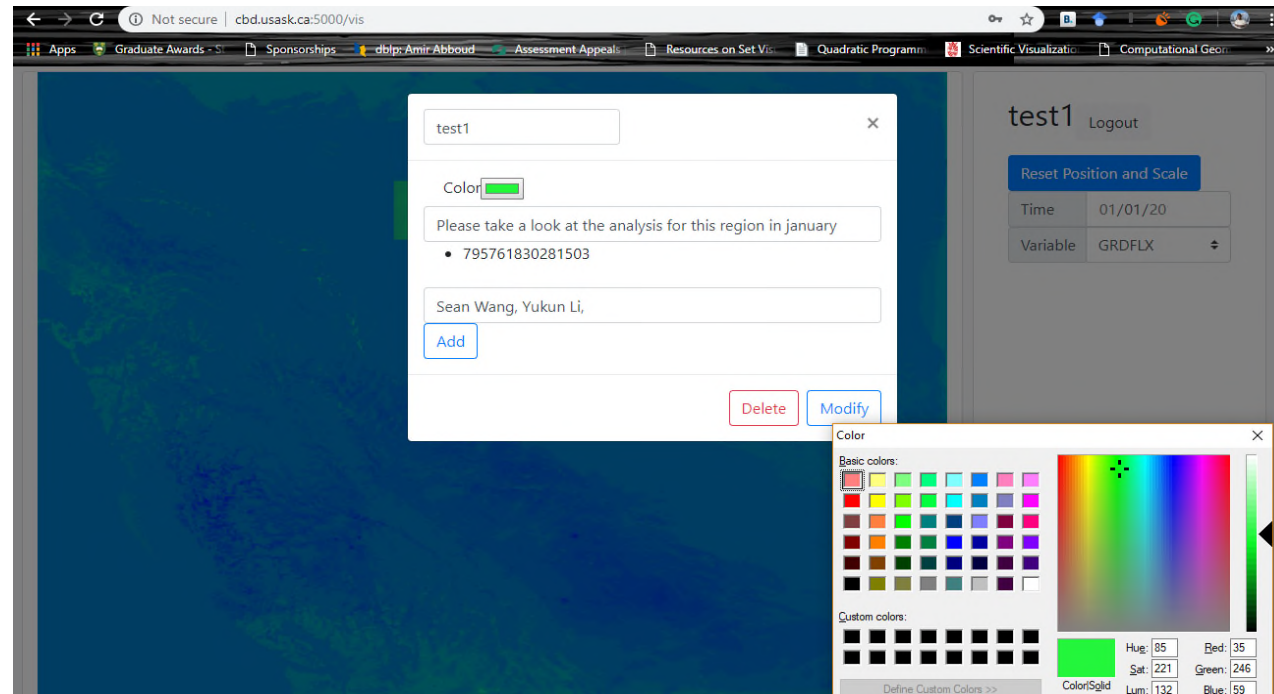
Users will be able to enter their NSID and password to access (under development)





# Collaborative Environment

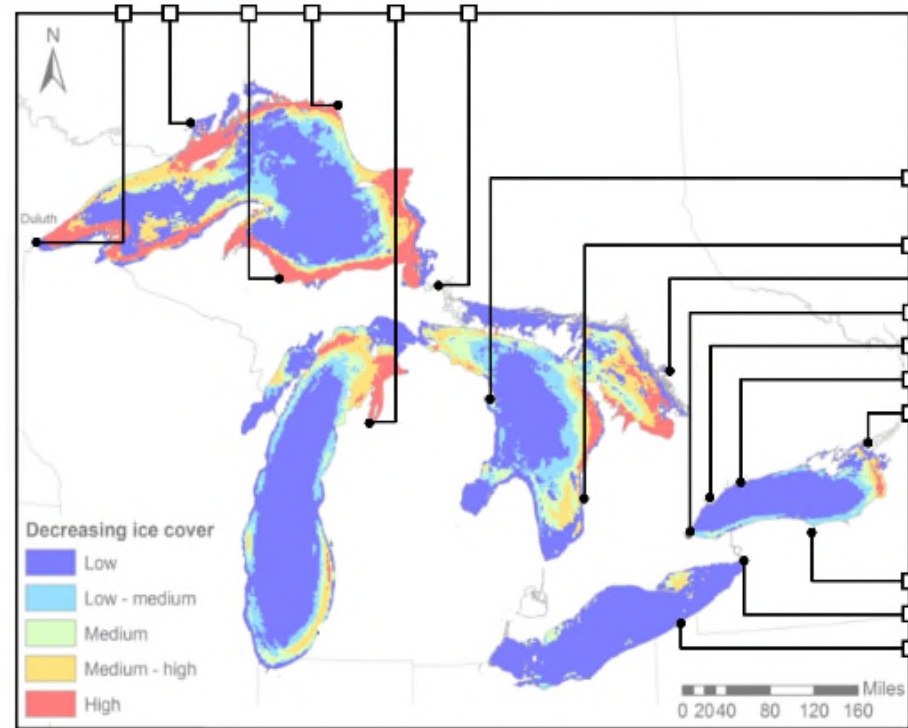
Users will be able to comment, tag, and share visualizations with a Facebook-like interface.



A user is sharing comments  
to a list of other researchers  
(Work in Progress)



# Collaborative Environment

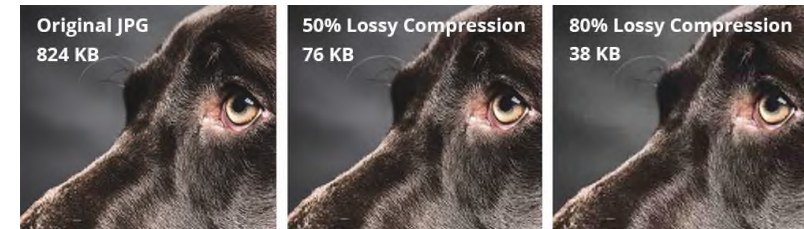
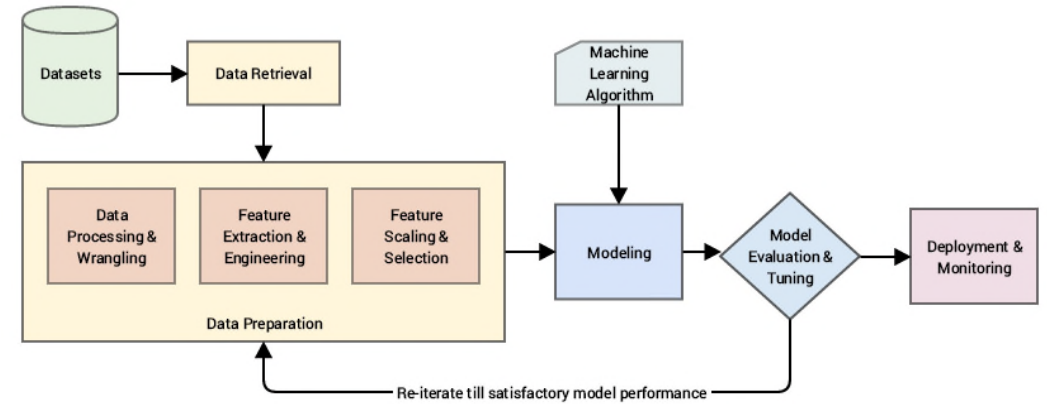


An example of labelling different  
user comments in a diagram  
(Work in Progress)



# Machine Learning: Summarized Visualizations

- Can WRF datasets be used to create machine learning models to (a) compress the dataset and (b) simulate user changes?
- Can we predict a large amount of the dataset accurately from a small part of the dataset, if so we can compress the dataset to explore on typical laptops and workstations. We are currently exploring dimensionality reduction and other feature engineering techniques in contour plots.

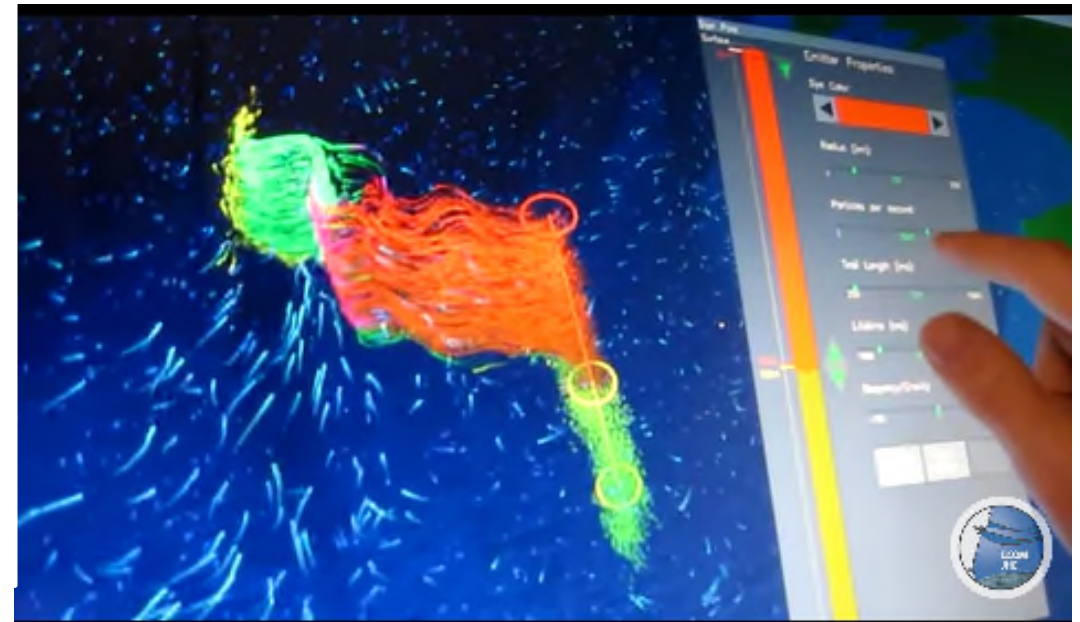
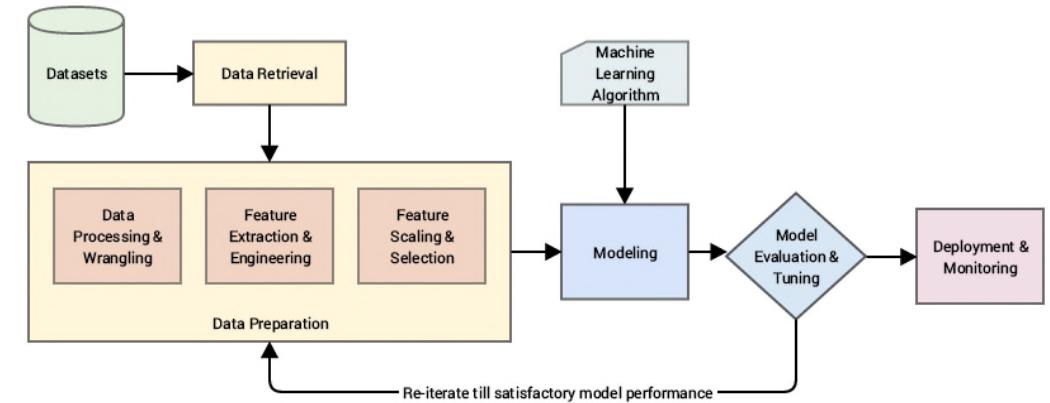


The idea is inspired by image compression using deep learning, whereas we are examining the feasibility of compressing years of WRF data (TBs) in a few Gigabytes through the compression of visual outputs.



# Machine Learning: Simulated User Interactions

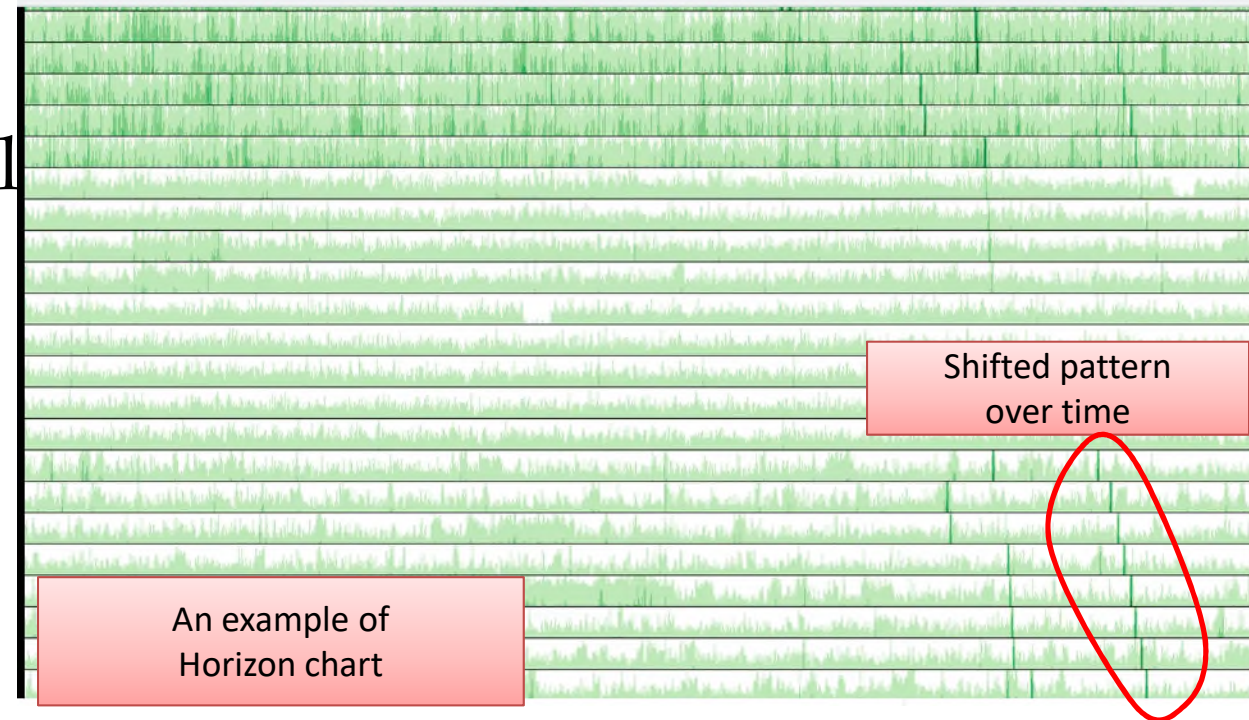
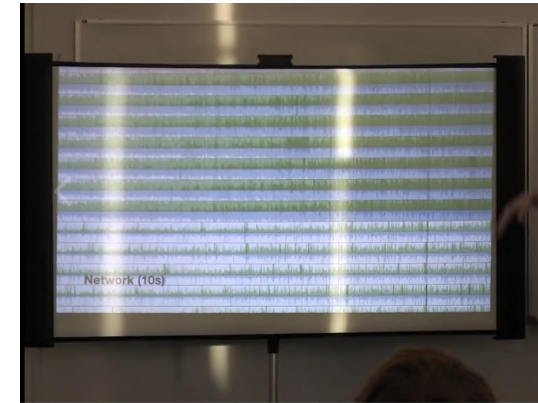
- Using machine learning models to simulate user interactions, e.g., a user may want to see how the increase of temperature/precipitation might impact the contour plots of other weather parameters, or what might happen to the weather if we create canals for irrigation.
- While this is a bold approach to having an accurate model based on just historical data, it would be interesting to explore its potential and see whether it may help improve existing modeling software.





# Innovative Visualizations

- In both CRHM and WRF, several parameters are overlaid on a single chart.
- A cutting edge approach is to use Horizon chart visualization to deal with high-dimensional time series data.
- Many dimensions are visualized in parallel on a large display, a Horizon chart can reveal patterns and anomalies.



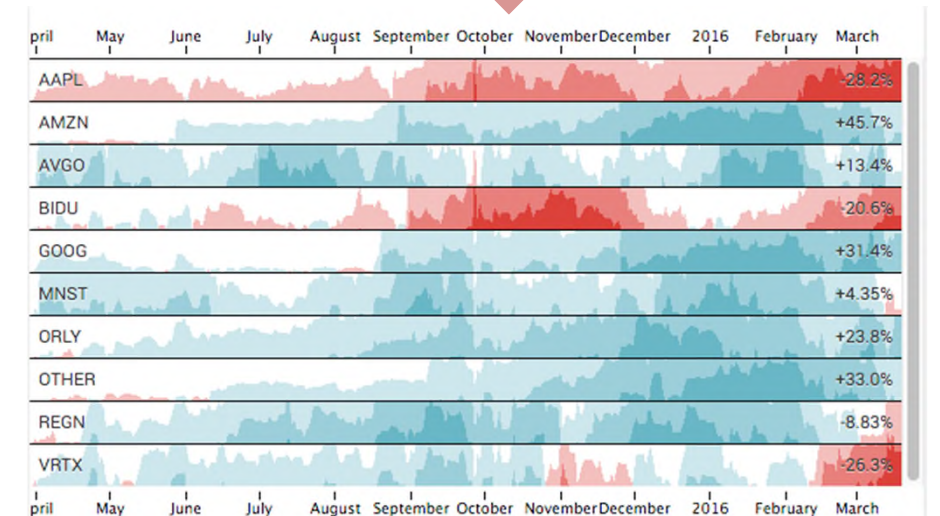
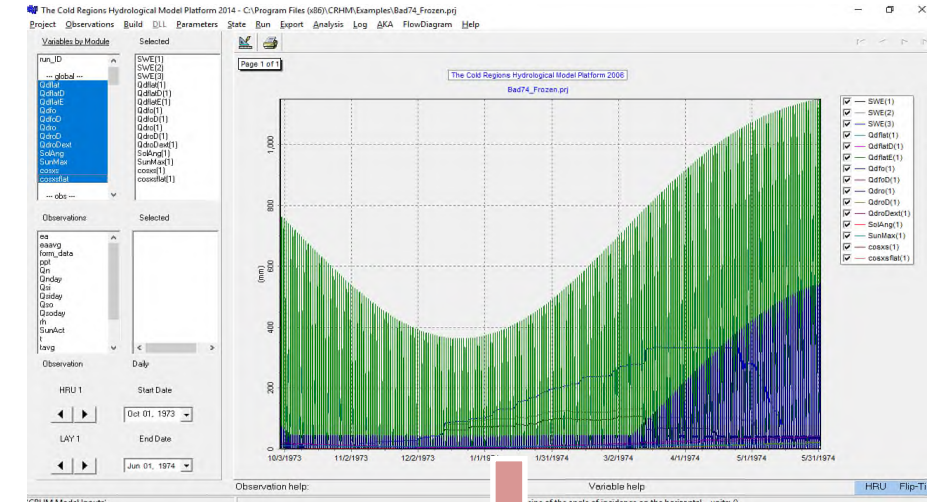
Horizon chart can visualize large information in small space

<https://www.youtube.com/watch?v=M53dYUcQbCs>



# Innovative Visualizations

Exploring the benefits of visualizing CRHM data with an Horizon chart.

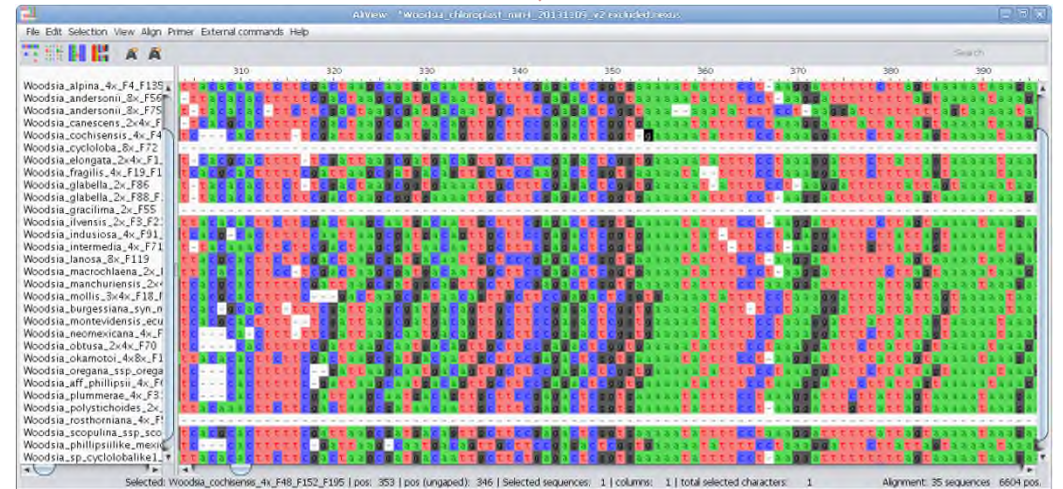
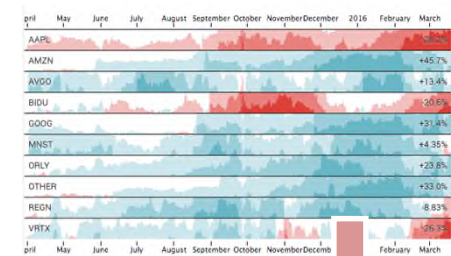
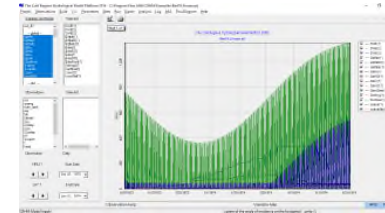




# Innovative Visualizations

We also want the process to be semi-automatic, i.e., rather than the researcher looking for a pattern, the system will automatically suggest important patterns.

The idea is to encode the peaks, valley and other interesting changes by symbols and then using multiple sequence alignment algorithms to visualize potentially interesting patterns to the researchers.





# GWF Visualization - Personnel

Researcher/Developer	Position	Projects
Kevin Schneider	Faculty, CS	
Debajyoti Mondal	Faculty, CS	
Chanchal Roy	Faculty, CS	
Gazi Hasnat	Graduate Student	2, 3
Jarin Tasnim	Graduate Student	3, 4
Zonayed Ahmed	Graduate Student	3, 4
Sean Wang	Research Assistant	1, 3
Yukun Li	Research Assistant	1, 2
Ilia Rudnisky	Research Assistant	2, 3





# GWF Visualization – Related Publications

Prosenjit Bose, Paz Carmi, Saeed Mehrabi, J. Mark Keil, and Debajyoti Mondal. *Boundary Labeling for Rectangular Diagrams*. Scandinavian Symposium and Workshops on Algorithm, Malmo, Sweden, 2018

Ilia Rudnitskiy and Debajyoti Mondal. *Interactive Visualization of Geospatial Water Datasets*. The Annual Science Meeting of Global Water Futures, Ontario, Canada 2018.

Debajyoti Mondal, Manishankar Mondal, Chanchal Roy, Kevin Schneider, Sean Wang, Yukun Li, *Clone-World: A Visual Analytic System for Large Scale Software Clones*. Computer Science Research Fest, Saskatoon, 2018 (submitted to ICSE).

Debajyoti Mondal, *Open Problems on Interactive Geo-spatial Visualization*, in the 21st Korean Workshop on Computational Geometry, Slovenia, 2018

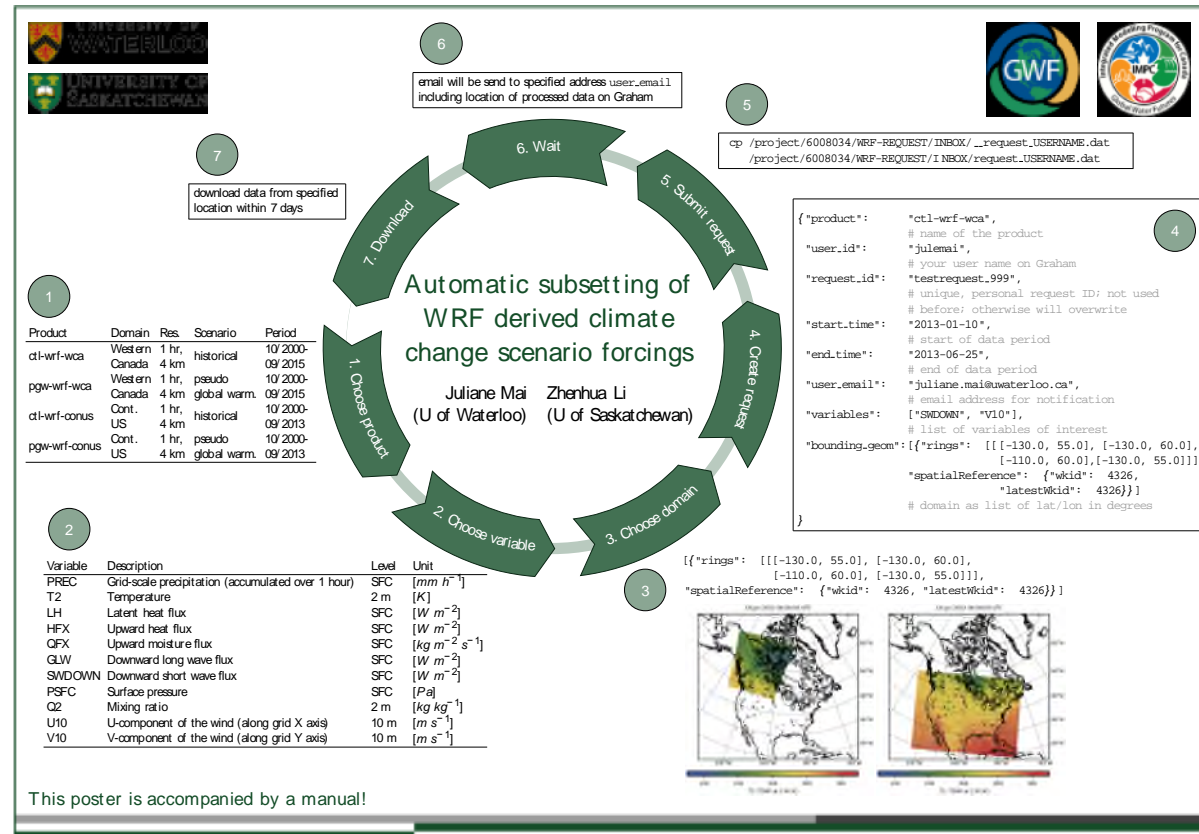
William Evans, Stefan Felsner, Michael Kaufmann, Stephen G. Kobourov, Debajyoti Mondal, Rahnuma Islam Nishat and Kevin Verbeek. *Table Cartograms*. Computational Geometry: Theory and Application. 68: 174-185, 2018.



# GWF data Cuizinart

## – Dissemination of Large Gridded Datasets Within and Beyond GWF –

- Team: Juliane Mai, Jimmy Lin, Zhenhua Li, Al Pietroniro, +
- CaSPAR-like tool to archive, distribute and publish large gridded datasets
- Data available at the moment are 4 different version of WRF runs
- Backend presented July 2018 (no data)







# GWF Computer Science Team Update

Kevin A. Schneider

Global Water Futures Operations Team Meeting  
January 21-23, 2019

