

GWF Core Computer Science

Kevin Schneider

Department of Computer Science, University of Saskatchewan

Operations Meeting December 9, 2021







GWF Core Computer Science Team

- Co-Leads
 - Kevin Schneider (USask)
 - Jimmy Lin (UWaterloo)
- Faculty
 - Ray Spiteri (USask) Scientific Computing
 - Chanchal Roy (USask) Software Engineering
 - Jyoti Mondal (USask) Visualization
 - Carl Gutwin (USask) Human Computer Interaction
 - Banani Roy (USask) Software Engineering
- HQP
 - 10 Undergraduates; 12 MSc; 5 PhD; 4 PDFs; 2 Ras
- Complementary funding
 - NSERC CREATE for Software Analytics Research (SOAR) (awarded 2020)
 - Software Analytics Research Laboratory CFI \$350,000 (awarded 2021)



GWF Core Computer Science Sub-Projects

- Hydrological Modelling Software Infrastructure
 - BRoy Lead, Schneider, CRoy, Software Research Lab (SRLab)
 - Hydrological Modelling Software Development Environment
 - Hydrological Modelling Software Renovation (Next-gen CRHM)
- Visualization and Decision Support Platform
 - Mondal Lead, Mondal, Gutwin, Schneider, SRLab, HCILab, VGALab
 - Interactive Visualization Toolset and Framework
- Hydrological Model Acceleration
 - Spiteri Lead, Numerical Simulation Research Lab (SimLab)
 - Hydrological Model Acceleration Toolset and Framework
- Automatic Knowledge Extraction and Dataset Management
 - Lin Lead
 - Data Management Toolset and Framework



Hydrological Modelling Software Infrastructure

Banani Roy-Lead, Kevin Schneider, Chanchal Roy

GWF Core Computer Science

December 9, 2021







Nutrient App [D. Costa et al., JEM'21] (https://gwf.usask.ca/projects-facilities/nutrient-app.php#Overview)

- Monitors of dissolved nutrient concentrations (N, P) in aquatic freshwater systems.
- **Developed API for Nutrient App**
- Developed web-based Nutrient App.
- Featured in Starphoenix

(https://thestarphoenix.com/news/local-news /young-innovators-new-u-of-s-app-tracks-causes-of-algae-bloom)



Medicine

Costa D., Aziz U., Elliott J., Baulch, H., Roy, B., Schneider K. A. and Pomeroy, J. (2020), "The Nutrient App: Developing a smartphone application for on-site instantaneous community-based NO3 and PO4 monitoring", Environmental Modelling and Software. 10.1016/j.envsoft.2020.104829.133, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2020.104829

Kelowna



Regina

0 maNO₃-N/I

enStreetMan contributors CC-BY-SA

CRHM Renovation

- Migrated Borland CRHM to modern compilers and libraries
 - Removed dependence on Borland
 - Reduced dependence on external and proprietary libraries
 - Multi-platform support
 - Shared source code for both gcc and vcc
- Modern Development Environment
 - Distributed version control
 - Automated testing¹
 - Incremental build
- Additional features
 - 64-bit and double precision
 - Faster
 - Reduced memory use



Renovation Phases

- **Phase I: Exploration²** (identify migration issues)
- Phase II: Create Next-Gen DevOps Environment
- Phase III: Migrate
- Phase IV: Renovate (in progress)

2. Bhattacharjee A., Roy B and Schneider K. "Supporting Readability by Comprehending the Hierarchical Abstraction of a Software Project", in the Proceedings of the 15th Innovation in Software Engineering Conference, India, February 2022. (Accepted)



^{1.} Khodabandehloo H, Roy B, Roy CK, Schneider KA, Mondal M. "A Testing Approach While Re-engineering Legacy Systems: An Industrial Case Study", in the *Proceedings of the 28th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 2021, pp. 600-604



Next-Gen CRHM Recent Improvements

- Improve Maintainability
 - Borland TStringList structures replaced with type-safe library containers
 - New Source Code Model one class per file where possible and organized into folders
- Reduce Memory Use Output now uses FileStream
 - Memory footprint almost halved
 - Output produced while model runs and not at end of run
 - Current large projects do not crash Next-Gen CRHM gcc
- Improve Performance Macros as Modules
 - Can significantly improve performance
 - E.g., SmithCreek project take 1/3 the time when Grow_Crop is a module



Next-Gen CRHM Plans for 2022

- Release Next-Gen CRHM to community
 - GUI and Command Line Versions
- Develop CRHM Share
 - parameter and project editing tools
- Validate Water Quality modules
- Increase automated regression testing robustness
- Explore ways to make module development easier



Visualization and Decision Support Platform

Jyoti Mondal – Lead, Carl Gutwin, Kevin Schneider

GWF Core Computer Science

December 9, 2021







Visualization Project

- Management Team
 - Mondal, Gutwin, Schneider, Pietroniro
- Development Team
 - Sean Wang, Rakib Hasan
- Completed
 - Proof-of-concept, Spring 2021
 - Research Prototype, Fall 2021
- Planned
 - Beta Version, Winter/Spring 2022
 - Engagement and Training, Summer/Fall 2022
 - Production Release, Fall 2022



Visualization Platform



A MODEL AGNOSTIC APPROACH TO VISUALIZATION





Code



Visualization Framework Features

Prioritize added value beyond what is currently available

- Model Agnostic
 - Model output -> Programmable Interface -> Visualization
- Scalable
 - Functional, Computational, Administrative
- Interactive
 - Data Exploration
- Programmable
 - Customizable, Transparent, Incremental, Extensible, Dashboard
- Collaborative
 - Share visualizations, data, and programs
- Data Analytics
 - Visual Comparison, Derived Variables, Auto-Summarization, Scenario Analysis, Decision Making

https://www.youtube.com/watch?v=-mhe-lh7j34



References – Peer-Reviewed, International, Archival

- 1. Zonayed Ahmed, Michael Beyene, Debajyoti Mondal, Chanchal K. Roy, Christopher Dutchyn, and Kevin A. Schneider: ContourDiff: Revealing Differential Trends in Spatiotemporal Data. In Proceedings of the 25th International Conference Information Visualization (IV), 2021: 35-41
- 2. Gazi MH Zahan, Debajyoti Mondal, and Carl Gutwin: Contour Line Stylization to Visualize Multivariate Information Graphics Interface, In Proceedings of the 46th Graphics Interface Conference (GI), 2021: 198 207.
- 3. Mohammad Rakib Hasan, Debajyoti Mondal, Jarin Tasnim and Kevin A. Schneider: Putting Table Cartograms into Practice. In Proceedings of the 16th International Symposium on Visual Computing (ISVC), 2021. Accepted, In press.
- 4. Jared Espenant and Debajyoti Mondal StreamTable: An Area Proportional Visualization for Tables with Flowing Streams. In Proceedings of the 16th International Conference and Workshop on Algorithms and Computation (WALCOM), 2022. Accepted, In press.
- 5. Jarin Tasnim and Debajyoti Mondal: Data Reduction and Deep-Learning Based Recovery for Geospatial Visualization and Satellite Imagery. IEEE BigData 2020: 5276-5285
- 6. Debajyoti Mondal, Manishankar Mondal, Chanchal K. Roy, Kevin A. Schneider, Shisong Wang, and Yukun Li: Clone-World: A Visual Analytic System for Large Scale Software Clones. Journal of Visual Informatics, 3(1): 18-26 (2019).











Hydrological Model Acceleration

Ray Spiteri – Lead GWF Core Computer Science December 9, 2021







Hydrological model acceleration

Progress to date:

- CHM performance improvements:
 - Faster multi-threaded solvers
 - Linear Algebra / Preconditioning
 with Trilinos
 - Move to distributed (hybrid MPI/OpenMP) architecture to run on clusters
 - Large-scale scaling
 - Lookup table model evaluations
- Bug fixes in HPC libraries:
 - scipy linear iterative solvers
 - Trilinos Kokkos / OpenMP interaction in multithreaded datastructure construction on version update





Hydrological model acceleration

Current activities:

- SUMMA/SUNDIALS
 - Better management of timestepping methods in SUMMA
- Incorporating SUMMA's hydrological components into CHM



- Snowcast v2
 - Blowing snow
 - Running nightly on copernicus
- Setup for ever-larger simulations
 - issues with consistency of physics/model setup and solve on very large meshes





Hydrological model acceleration

Goals for 2022:

- More generic coupler implementation complete
 - state variables on different meshes
 - spatial resolutions and representations
 - temporal resolutions
- Further enhancement of distributed solvers (Linear Algebra / Preconditioning)





Automatic Knowledge Extraction and Dataset Management

Jim Lin – Lead

GWF Core Computer Science

December 9, 2021



UNIVERSITY OF SASKATCHEWAN Global Water Futures



GWF Midterm Review

It is not clear whether the data strategy incorporates the machine learning and deep learning/Artificial Intelligence (AI) that is necessary in order to ensure the prediction capability that is crucial in industrial and policy contexts...

An open data platform where AI tools can be leveraged by third parties would be of value to several stakeholders in the water research community and would help accelerate the emergence of an economic cluster in water-related technology.

Our response...



AI at the CS/DM Intersection

- Apply natural language processing (NLP) and information retrieval (IR) techniques based on artificial intelligence (AI) to the management of textual metadata in GWF.
- Automate (some) analysis of textual metadata associated with:
 - GWF publications
 - Metadata records
 - •



Covidex

- Created in March 2020 as a tool to help scientists search the rapidly growing scientific literature on COVID-19.
- Exploits the latest AI-based neural networks for text analysis.
- Capabilities:
 - Semantic search
 - Faceted browsing
 - Platform for additional analyses

Covidex	× +	~
igstarrow igstarro		🖈 🖈 🋞 :
Covidex		, С
Search CORD-19	v for something about COVID-19 Q Search	

Covidex applies state-of-the-art neural network models and artificial intelligence (AI) techniques to answer questions using the COVID-19 Open Research Dataset (CORD-19) provided by the Allen Institute for AI, which is curated dataset of scientific articles about COVID-19 and coronavirus-related research drawn from a variety of sources including PubMed, a curated list of articles from the WHO, as well as preprints from bioRxiv and medRxiv. We hope that our technologies can contribute to the fight against this global pandemic by helping policy makers and clinicians make better-informed decisions and by helping researchers generate new insights.

Our system has been evaluated in the TREC-COVID challenge organized by the National Institute of Standards and Technology (NIST). Submissions based on Covidex components were the highest-scoring "automatic" runs in rounds 4 and 5 of the evaluation. We provide here as a working prototype a simplified version of complete architecture described in the following paper:

Edwin Zhang, Nikhil Gupta, Raphael Tang, Xiao Han, Ronak Pradeep, Kuang Lu, Yue Zhang, Rodrigo Nogueira, Kyunghyun Cho, Hui Fang, and Jimmy Lin. Covidex: Neural Ranking Models and Keyword Search Infrastructure for the COVID-19 Open Research Dataset. *Proceedings of the 1st Workshop on Scholarly Document Processing*, pages 31-41, November 2020.

Since the conclusion of the TREC-COVID challenge in August 2020, our team has not had sufficient resources to fully keep pace with releases of the CORD-19 dataset. The prototype here still uses the CORD-19 version used in round 5 of the TREC-COVID challenge, dating from July 16, 2020 (and we have no intention of updating it). As a result, the system returns results that are outdated and thus should **not** be used for making medical, public health, and other decisions.

This project is led by Jimmy Lin from the University of Waterloo and Kyunghyun Cho from NYU, with a small team of wonderful collaborators and students: Edwin Zhang, Johnson Han, Nikhil Gupta, Ralph Tang, Rodrigo Nogueira, and Ronak Pradeep. Special thanks to Colin Raffel for his help in pretraining T5 models for the biomedical domain. We are grateful for support from CIFAR for an AI and COVID-19 Catalyst Grants, Microsoft for an AI for Good COVID-19 Grant, and Compute Canada for computational resources that are sustaining this prototype.

https://covidex.ai/



Covidex × -	+
\leftarrow \rightarrow C \triangleq https://covidex.ai/?query=What	t%20is%20the%20prognostic%20value%20of%20IL-6%20levels%20in%20COVID-19? 🖈 🔅 :
Covidex	Ģ
Search CORD-19 v fo	What is the prognostic value of IL-6 levels in COVID-19?
Filter your search	1. Interleukin-6 as prognosticator in patients with COVID-19: IL-6 and Covid-19
Publish Time	Elisa Grifoni, Alice Valoriani, Francesco Cei, Roberta Lamanna, Anna Maria Grazia Gelli, Benedetta Ciambotti, Vieri Vannucchi, Federico Moroni, Lorenzo Pelagatti, Roberto Tarquini, Giancarlo Landini, Simone Vanni, Luca Masotti. <i>J Infect</i> (2020-06-08)
2016 2020	In conclusion, in our COVID-19 population, IL-6 levels at hospital admission seem to be a good prognostic ator for the combined endpoint progression to severe disease and/or in-hospital mortality, and it seems to be the best prognostic ator for negative outcome
Authors	Show more ×
Find a filter	
	SARS-CoV-2 infection: the role of cytokines in COVID-19 disease
A. Charney A. Elgaml	Víctor J. Costela-Ruiz, Rebeca Illescas-Montes, Jose M. Puerta-Puerta, Concepción Ruiz, Lucia Melguizo- Rodríguez. <i>Cytokine Growth Factor Rev</i> (2020-06-02)
A. Firpo-Betancourt	detected elevated IL-6 levels in one-third of patients with mild symptoms and three-quarters of those with severe symptoms,
More filters	concluding that IL-6 , alongside IL- 10, may be of prognostic value in patients with COVID-19 [54] Various authors have detected this interleukin in patients with COVID-19 and related its levels to disease severity and progression, as in the case of other cytokines [17,43,57, 6 4,82,83,8 6 ,88,89] and it has been reported to have possible prognostic value [54]
Source	Show more ×
 Elsevier MedRxiv Medline PMC 	3 . Position Paper for the State-of-the-Art Application of Respiratory Support in Patients with COVID-19
🗌 who	Michael Pfeifer, Santiago Ewig, Thomas Voshaar, Winfried Johannes Randerath, Torsten Bauer, Jens Geiseler, Dominic Dellweg, Michael Westhoff, Wolfram Windisch, Bernd Schönhofer, Stefan Kluge, Philipp M. Lepper. <i>Respiration</i> (2020-06- 19)
Journal	• <)





Covidex ---> GWF

- Bring the capabilities of the Covidex to GWF
- Semantic search and facet browsing of GWF publications in progress
- Here's an early analysis...







Global Water Futures Core Computer Science Project