Core Modelling Team

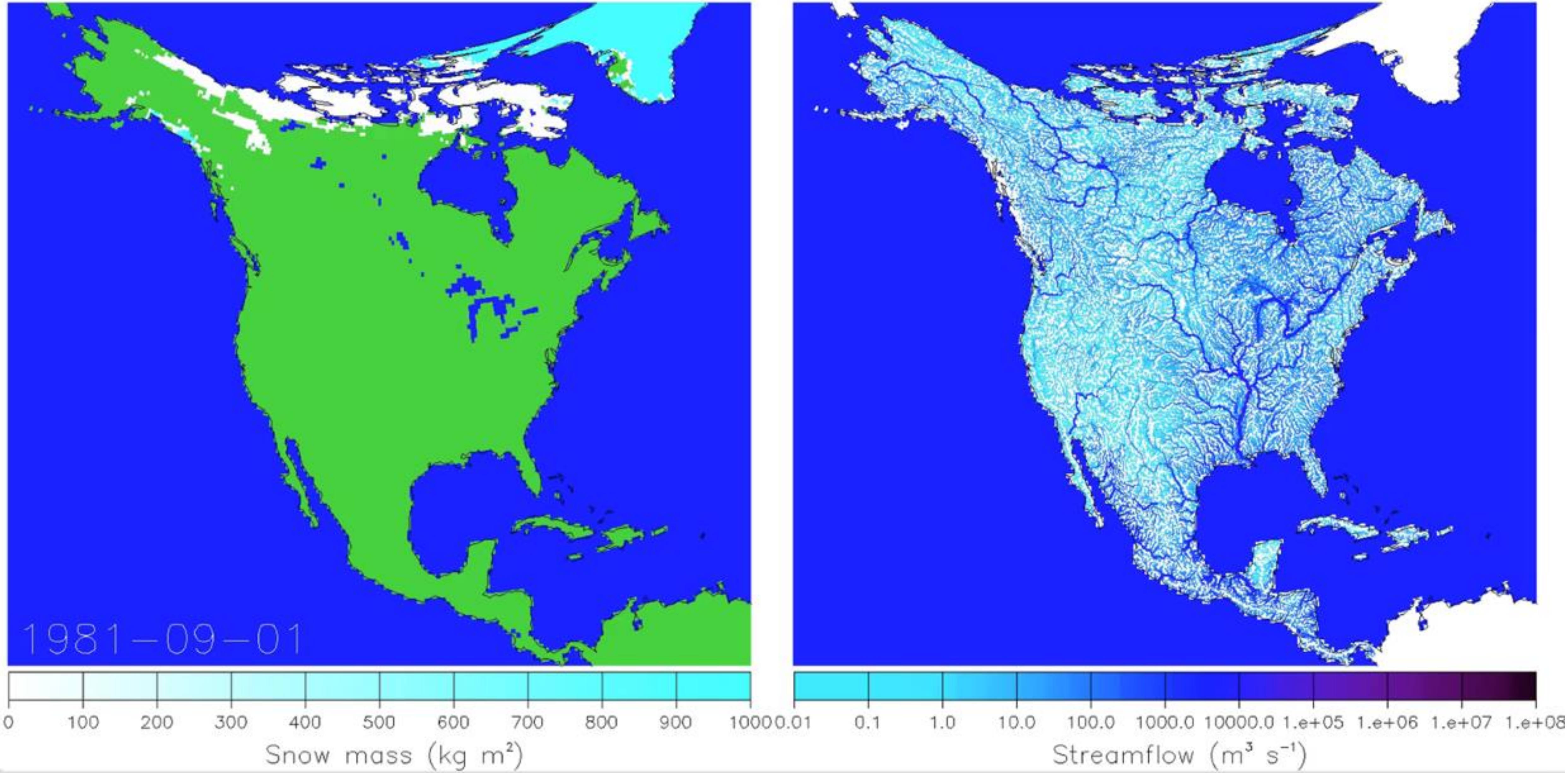
Global datasets to support hydrological modelling in alpine regions

Martyn Clark, Alain Pietroniro, John Pomeroy, and the GWF core modelling team

INARCH meeting, 19 October, *2021*

The North American domain

Core Modelling Team



GWF Planetary Water Prediction Initiative

Core Modelling Team

- Exploit GWF science advances to improve our understanding and predictions of the global hydrological cycle

- Build computational infrastructure (models, data) to enable state-of-the-art hydrological simulations anywhere on the planet.
  - ➢ *Improve global datasets on climate forcing, DEMs, veg, soil etc.*
  - ➢ *Use these datasets to configure hydrological models for the global domain.*
  - ➢ *Address challenging questions of global hydrological change*

- Develop regional models in key areas where GWF has strong collaborations (Arctic, Himalaya, Andes, central Asia) to address pressing societal needs.
  - ➢ *The regional models will be constructed as cut-outs from the global-scale fabric*
  - ➢ *Address challenging questions of regional change*

Topics for today

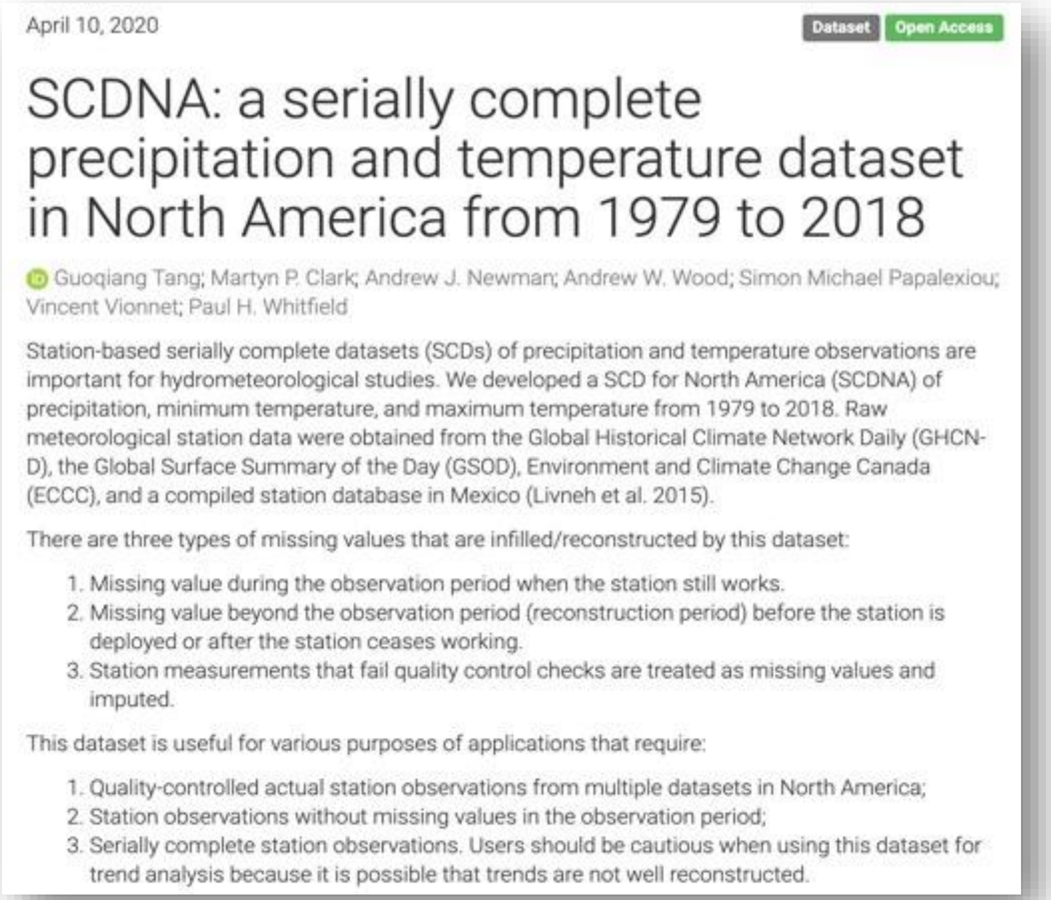
Core Modelling Team

- Spatial meteorological forcing data
- Geospatial intelligence
- Summary and outlook

SCDNA: Serially Complete Dataset for North America

Core Modelling Team

40-year gapless precipitation and temperature data from~27,000 stations.



April 10, 2020

Dataset Open Access

SCDNA: a serially complete precipitation and temperature dataset in North America from 1979 to 2018

Guoqiang Tang; Martyn P. Clark; Andrew J. Newman; Andrew W. Wood; Simon Michael Papalexiou; Vincent Vionnet; Paul H. Whitfield

Station-based serially complete datasets (SCDs) of precipitation and temperature observations are important for hydrometeorological studies. We developed a SCD for North America (SCDNA) of precipitation, minimum temperature, and maximum temperature from 1979 to 2018. Raw meteorological station data were obtained from the Global Historical Climate Network Daily (GHCN-D), the Global Surface Summary of the Day (GSOD), Environment and Climate Change Canada (ECCC), and a compiled station database in Mexico (Livneh et al. 2015).

There are three types of missing values that are infilled/reconstructed by this dataset:

1. Missing value during the observation period when the station still works.
2. Missing value beyond the observation period (reconstruction period) before the station is deployed or after the station ceases working.
3. Station measurements that fail quality control checks are treated as missing values and imputed.

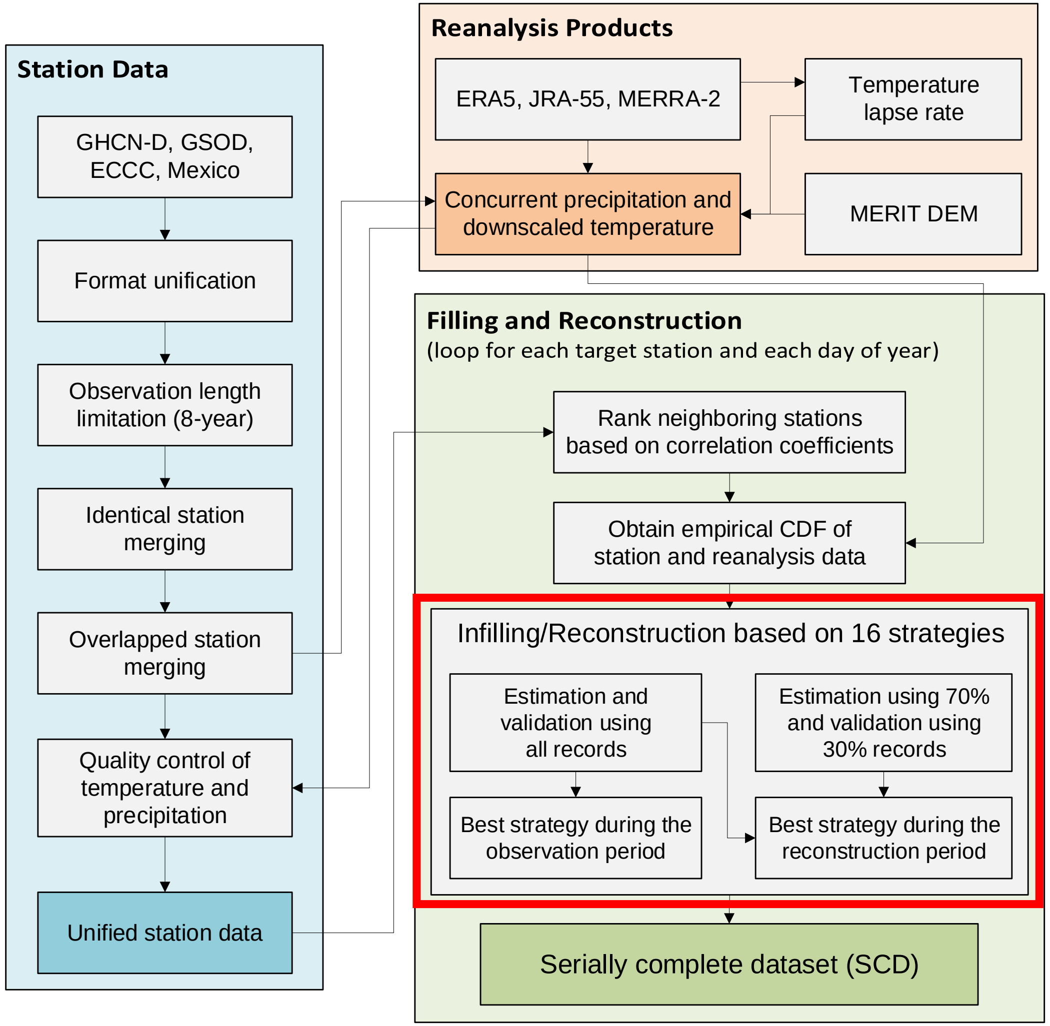
This dataset is useful for various purposes of applications that require:

1. Quality-controlled actual station observations from multiple datasets in North America;
2. Station observations without missing values in the observation period;
3. Serially complete station observations. Users should be cautious when using this dataset for trend analysis because it is possible that trends are not well reconstructed.

Tang and Clark et al. (2020), ESSD

Dataset link: https://doi.org/10.5281/zenodo.3735533

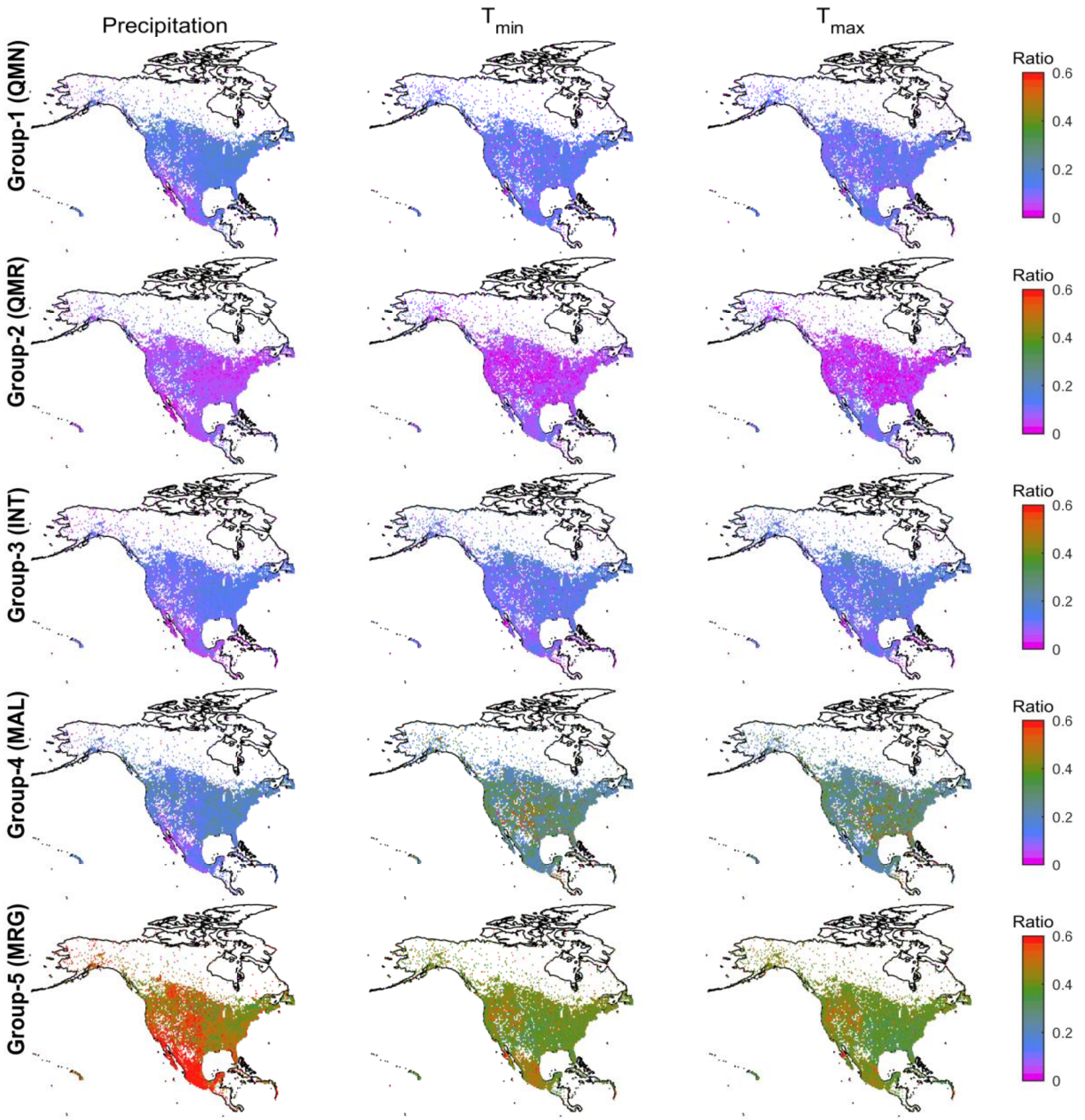
- Multi-source: Raw stations + Reanalysis
- Multi-strategy: 16 gap filling strategies



Serially-complete station data

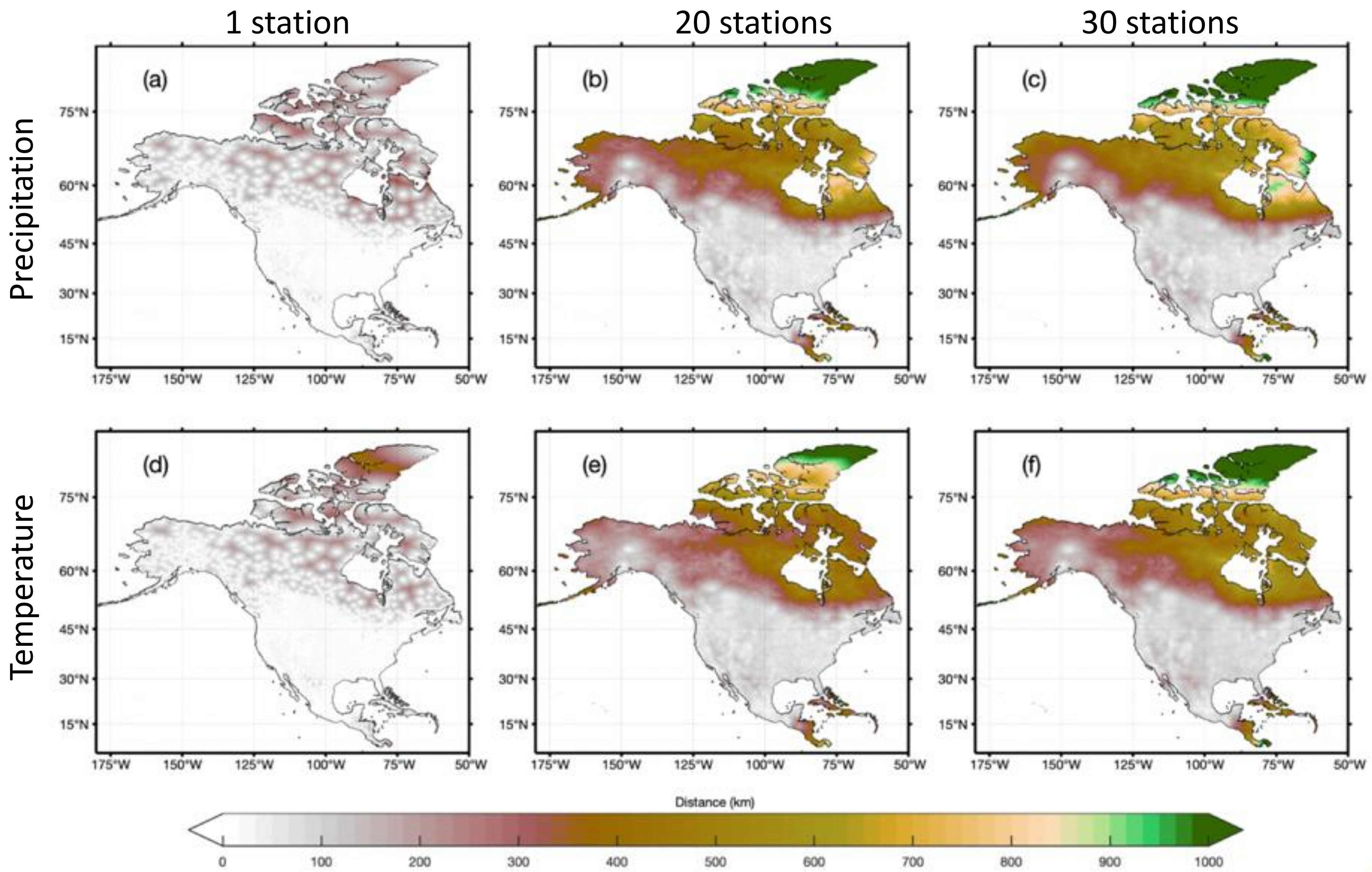
Methods:

1. **Quality control**: five checks for temperature and six checks for precipitation (Durre et al. 2010; Hamada et al. 2011; Beck et al. 2019);
2. **16 infilling/reconstruction strategies**: quantile mapping with neighboring stations (QMN), quantile mapping with concurrent reanalysis estimates (QMR), machine learning methods (MAL), & multi-strategy merging methods (MRG);
3. **Independent validation** based on 30% station records;
4. **Correct climatological biases** (quantile mapping and mean-value correction);
5. **Compare to four benchmark datasets** (ERA5, MERRA-2, JRA-55, and MSWEP).



Tang et al., ESSD (2020)

EMDNA: Ensemble Meteorological Dataset for North America



- For most regions in the CONUS, the search radius is smaller than 100 km.
- For the regions northern to 50°N or southern to 20°N, the search radius is much larger and even exceeds 1000 km in the Arctic Archipelago.
- The **sparse station network** at higher latitudes motivates the decision to optimally combine station data with reanalysis products.

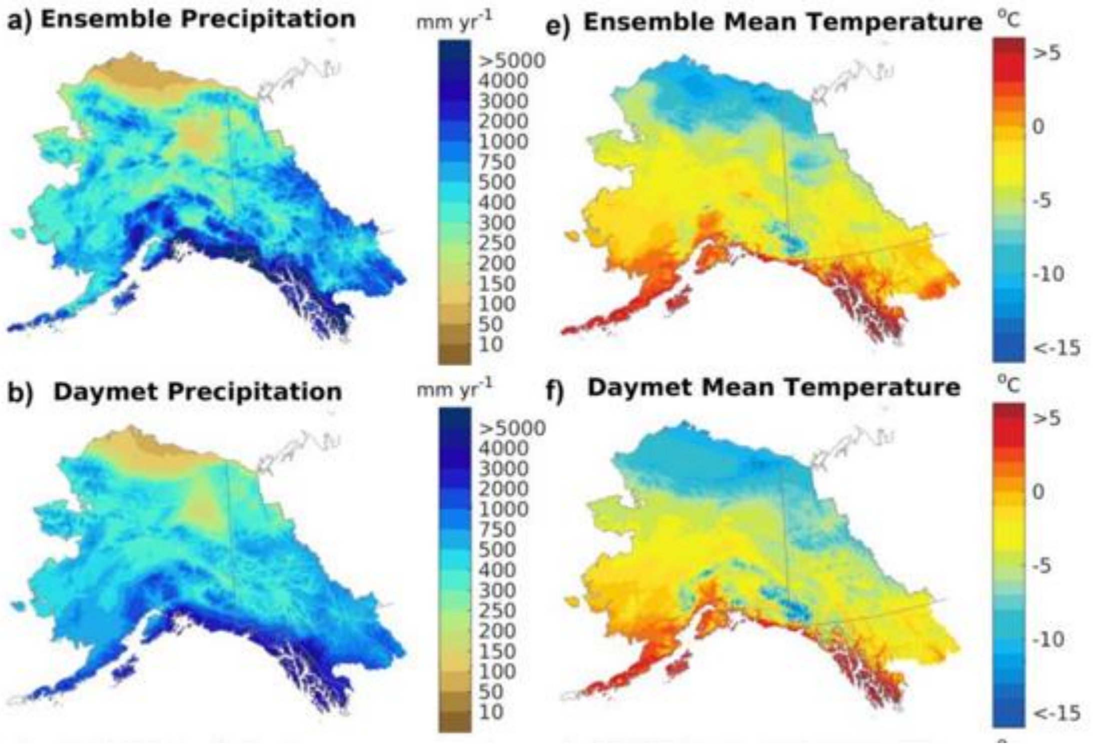
Tang et al., ESSD (2020)

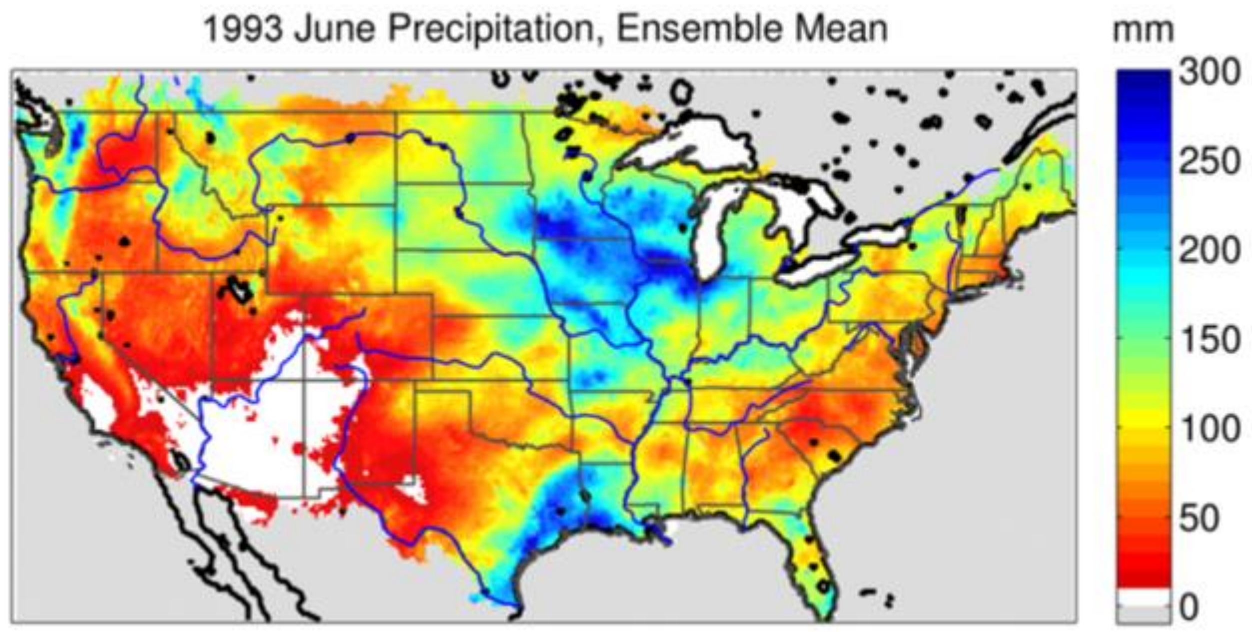
EMDNA: Ensemble Meteorological Dataset for North America

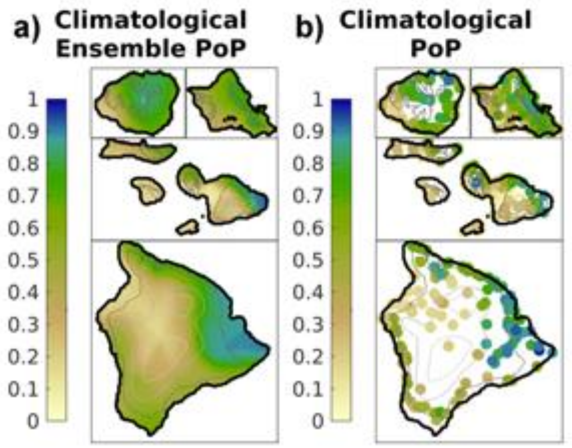
Core Modelling Team

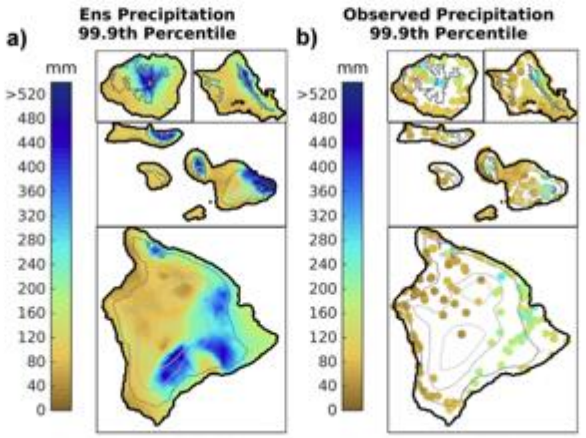
Previous work: station-based estimates

- ➢Newman 2015 (CONUS/ S. Canada, 1980-2012, 12-km)
- ➢Newman 2019 (Hawaii, 1990-2014, 1-km)
- ➢Newman 2020 (Alaska and Yukon, 1980-2012, 12-km)





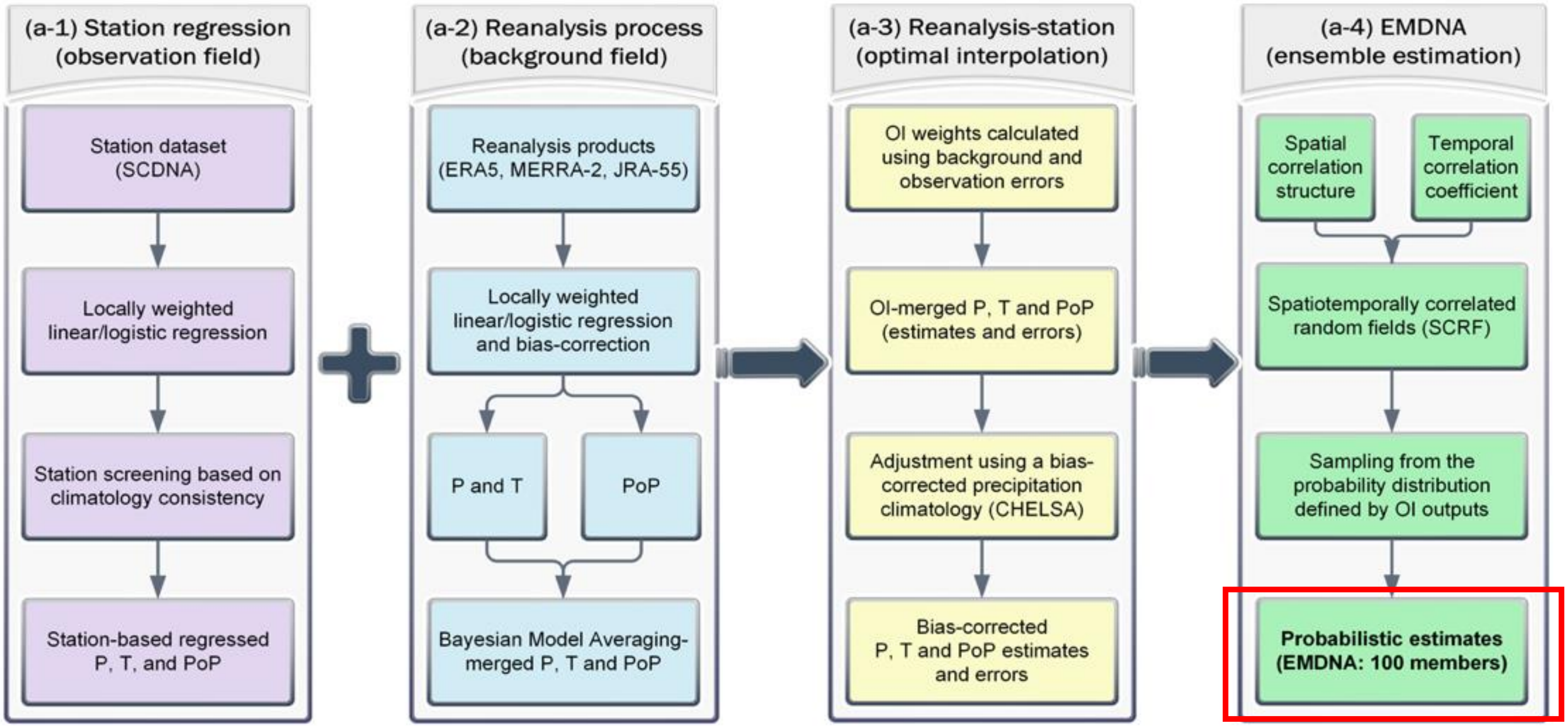




EMDNA: Ensemble Meteorological Dataset for North America

Core Modelling Team

Station	Reanalysis	Reanalysis + Station	EMDNA
(SCDNA)	(ERA5, MERRA-2, JRA-55)	(Optimal Interpolation)	(Probabilistic estimate)

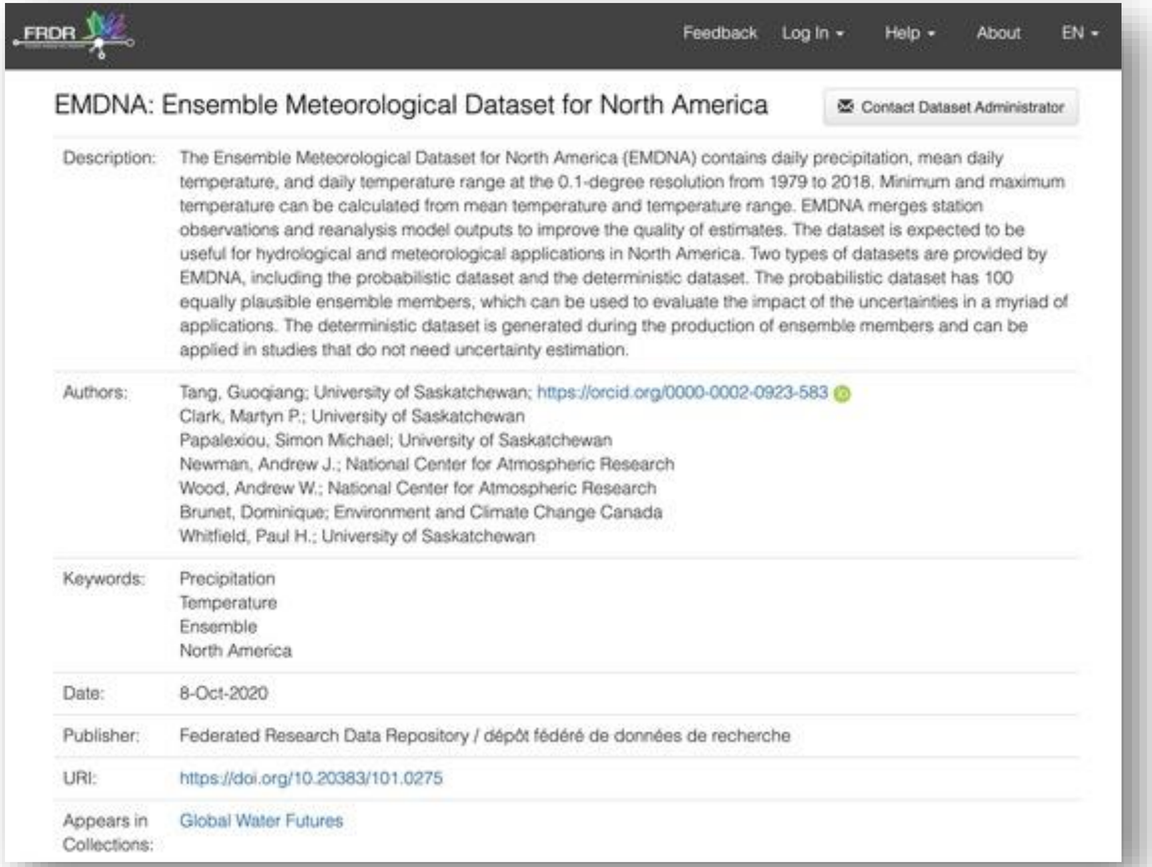


EMDNA: Ensemble Meteorological Dataset for North America

Core Modelling Team

Precipitation and temperature estimates:
0.1°×0.1°; daily;100 members; 1979-2018.

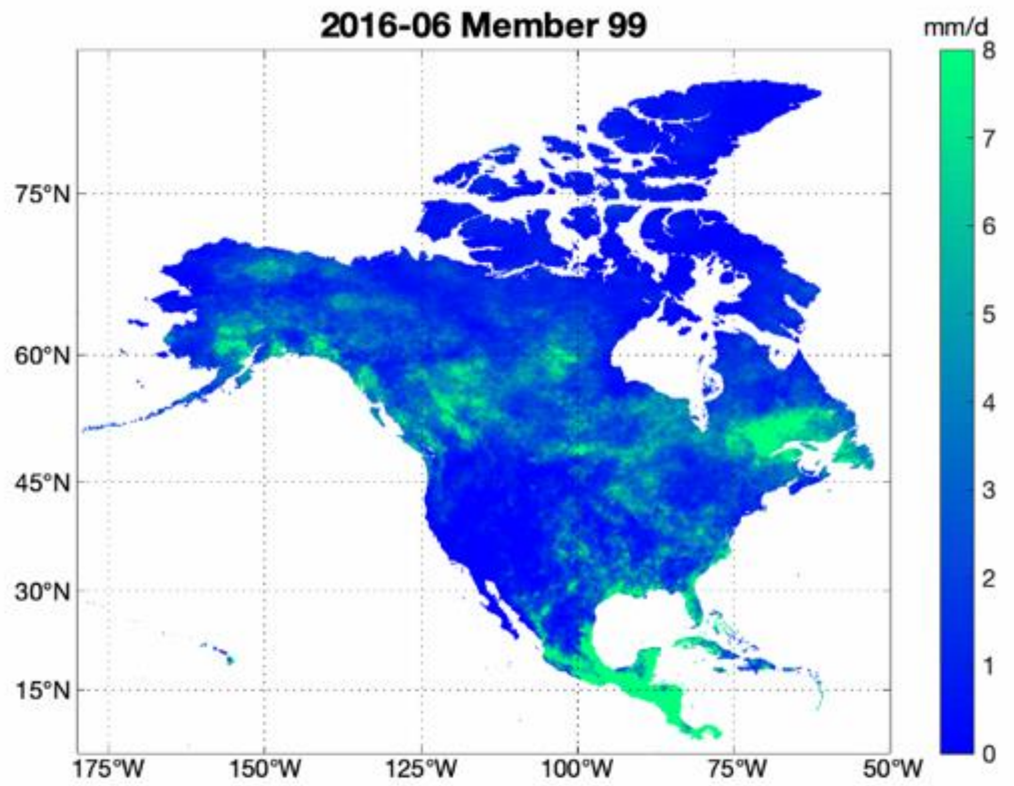
Precipitation from 100 members


FRDR Feedback Log In Help About EN

EMDNA: Ensemble Meteorological Dataset for North America

Contact Dataset Administrator

Description:	The Ensemble Meteorological Dataset for North America (EMDNA) contains daily precipitation, mean daily temperature, and daily temperature range at the 0.1-degree resolution from 1979 to 2018. Minimum and maximum temperature can be calculated from mean temperature and temperature range. EMDNA merges station observations and reanalysis model outputs to improve the quality of estimates. The dataset is expected to be useful for hydrological and meteorological applications in North America. Two types of datasets are provided by EMDNA, including the probabilistic dataset and the deterministic dataset. The probabilistic dataset has 100 equally plausible ensemble members, which can be used to evaluate the impact of the uncertainties in a myriad of applications. The deterministic dataset is generated during the production of ensemble members and can be applied in studies that do not need uncertainty estimation.
Authors:	Tang, Guoqiang; University of Saskatchewan; https://orcid.org/0000-0002-0923-583 Clark, Martyn P.; University of Saskatchewan Papalexiou, Simon Michael; University of Saskatchewan Newman, Andrew J.; National Center for Atmospheric Research Wood, Andrew W.; National Center for Atmospheric Research Brunet, Dominique; Environment and Climate Change Canada Whitfield, Paul H.; University of Saskatchewan
Keywords:	Precipitation Temperature Ensemble North America
Date:	8-Oct-2020
Publisher:	Federated Research Data Repository / dépôt fédéré de données de recherche
URI:	https://doi.org/10.20383/101.0275
Appears in Collections:	Global Water Futures

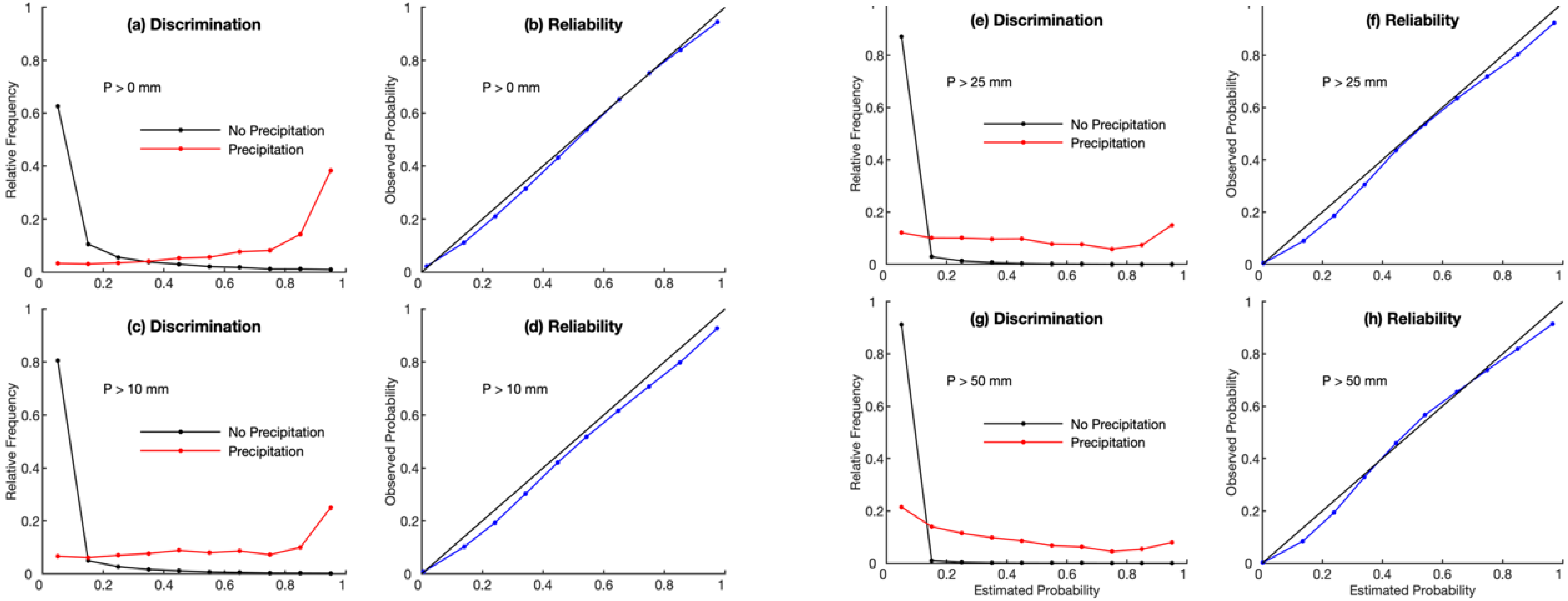


Tang and Clark et al. (2021), ESSD
Dataset link: https://doi.org/10.20383/101.0275

EMDNA: Ensemble Meteorological Dataset for North America

Core Modelling Team

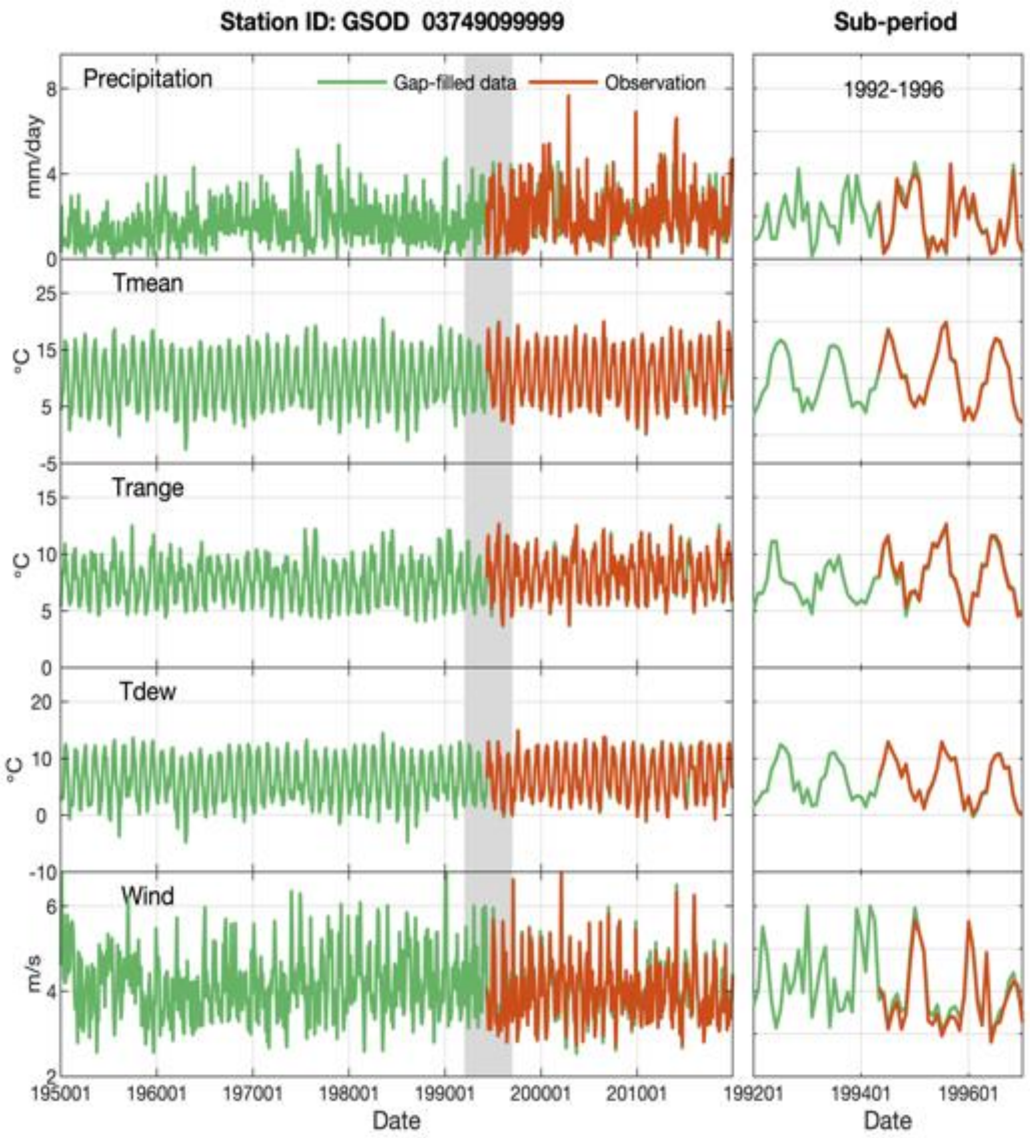
Evaluation of EMDNA precipitation estimates is based on independent stations that are not used in neither SCDNA or EMDNA.



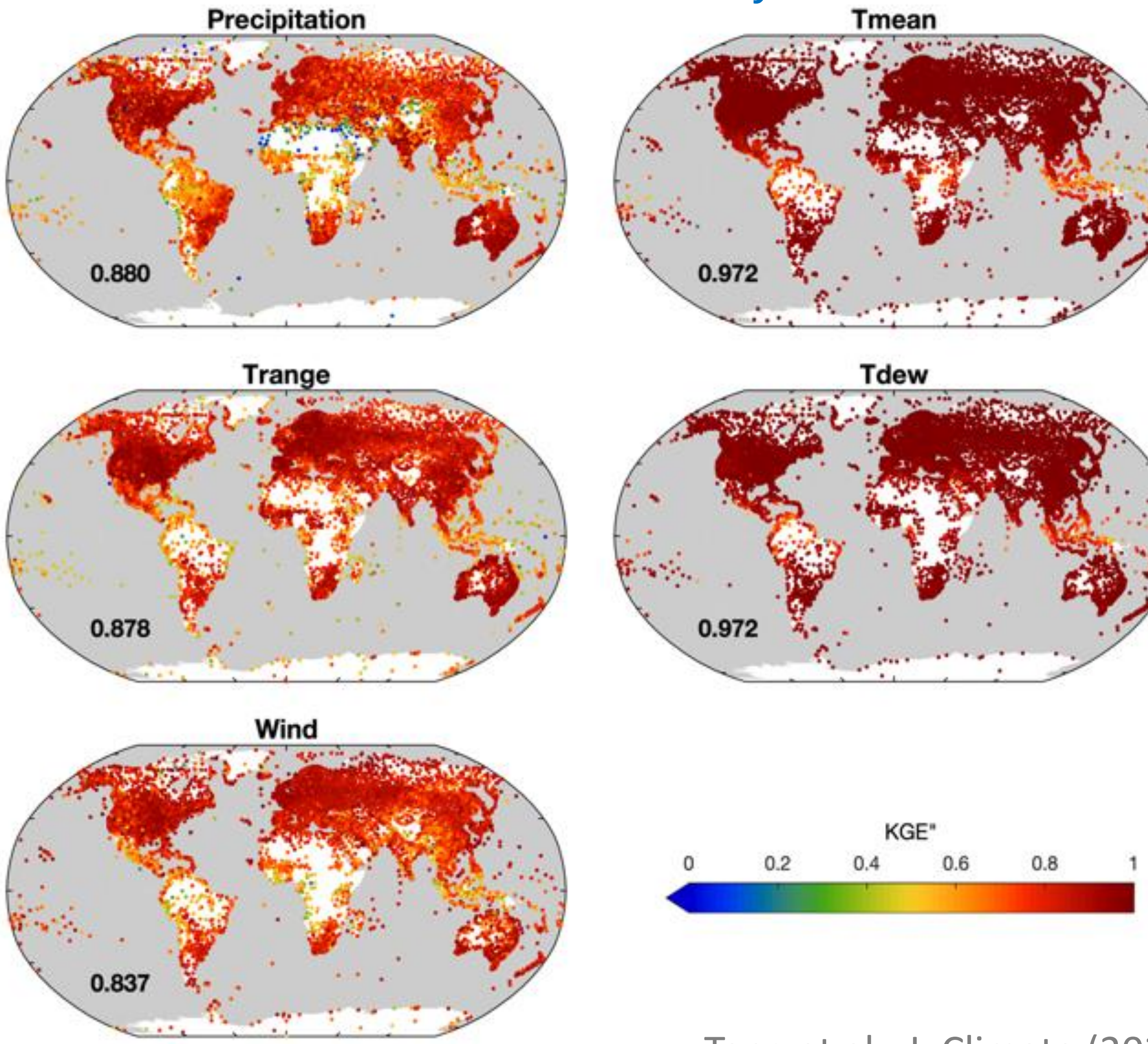
SC-Earth: A serially complete dataset for a small planet

Core Modelling Team

Example Station



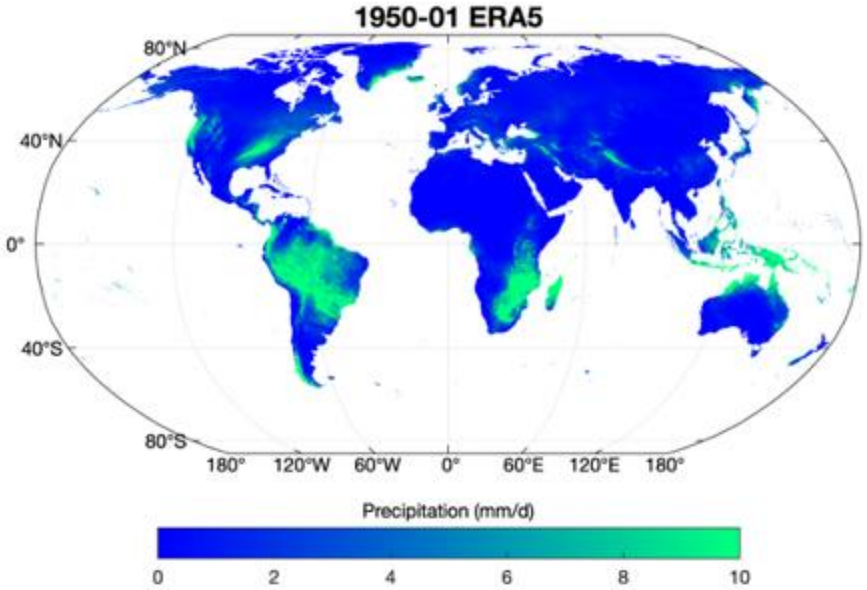
Global Accuracy

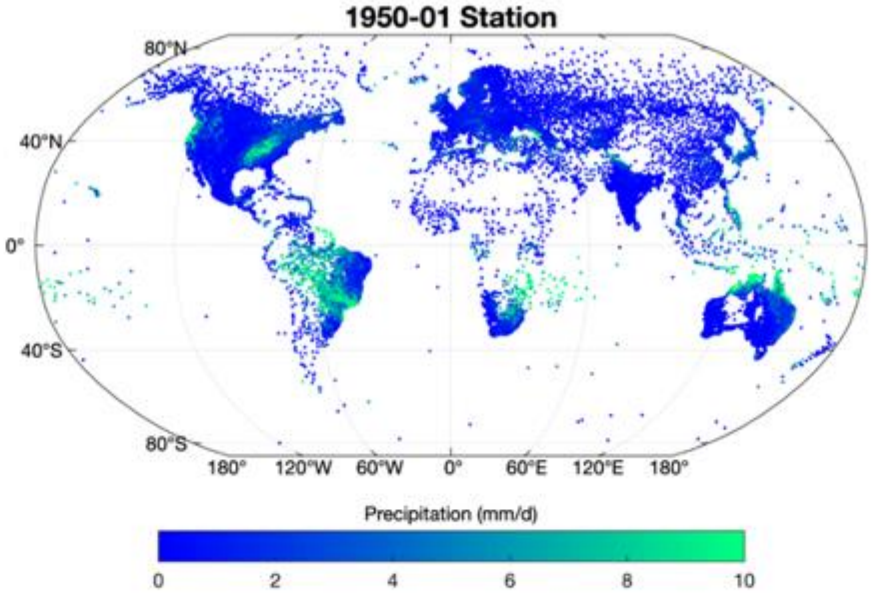


Tang et al., J. Climate (2021)

EM-Earth: Ensemble meteorological data

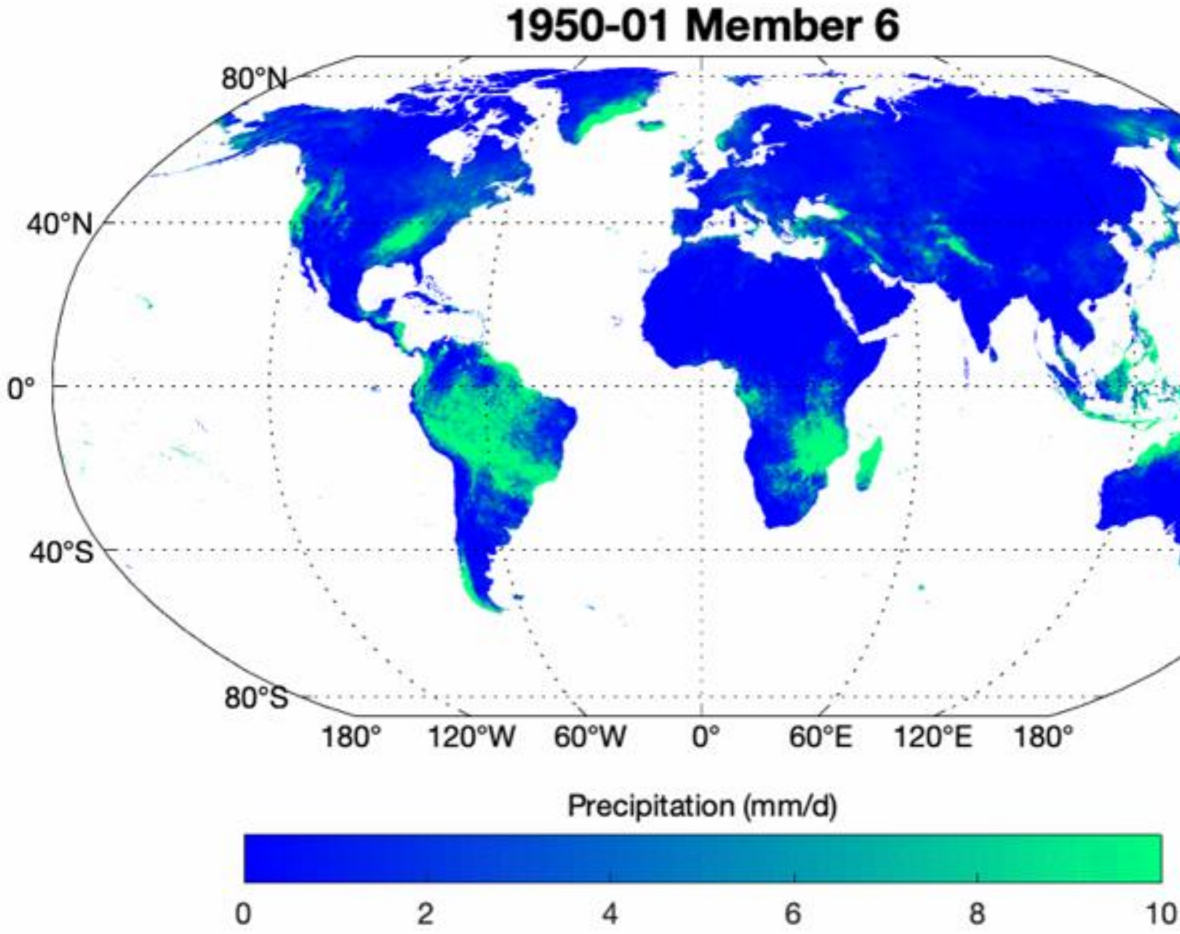
Core Modelling Team





EM-Earth

Monthly Precipitation in 1950-01



Tang et al., BAMS (under review)

Topics for today

- Spatial meteorological forcing data
- Geospatial intelligence
- Summary and outlook

Develop model-agnostic workflows for large-domain applications

Core Modelling Team

GWF Global Water Futures

While models have similar data requirements, they are configured in an individualistic and ad-hoc way

AGU PUBLICATIONS

Water Resources Research

COMMENTARY

10.1002/2016WR019285

Most computational hydrology is not reproducible, so is it really science?

Christopher Hutton[1], Thorsten Wagener[1,2], Jim Freer[2,3], Dawei Han[1], Chris Duffy[4], and Berit Arheimer[5]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK, [2]Cabot Institute, Royal Fort House, University of Bristol, Bristol, BS8 1UJ, Bristol, UK, [3]School of Geographical Sciences, University of Bristol, Bristol, UK, [4]Department of Civil Engineering, Pennsylvania State University, State College, Pennsylvania, USA, [5]Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Key Points:

- Articles that rely on computational work do not provide sufficient information to allow published scientific findings to be reproduced
- We argue for open reuseable code, data, and formal workflows, allowing published findings to be verified
- Reproducible computational hydrology will provide a more robust foundation for scientific advancement and policy support

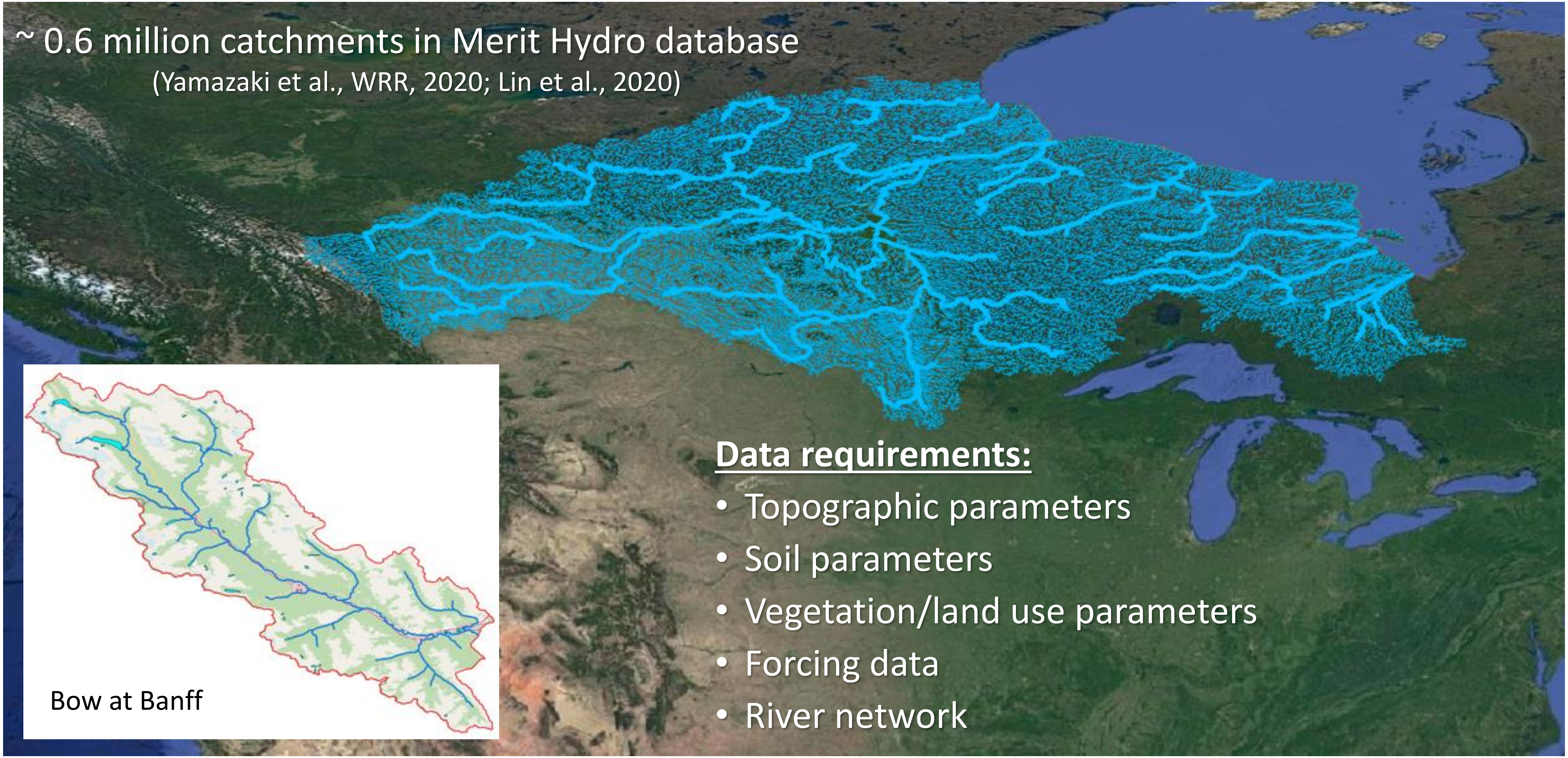
Correspondence to:
C. Hutton,
chutton294@gmail.com

Citation:
Hutton, C., T. Wagener, J. Freer, D. Han, C. Duffy, and B. Arheimer (2016), Most computational hydrology is not reproducible, so is it really science?, *Water Resour. Res., 52*, 7548–7555, doi:10.1002/2016WR019285.

Abstract Reproducibility is a foundational principle in scientific research. Yet in computational hydrology the code and data that actually produces published results are not regularly made available, inhibiting the ability of the community to reproduce and verify previous findings. In order to overcome this problem we recommend that reuseable code and formal workflows, which unambiguously reproduce published scientific results, are made available for the community alongside data, so that we can verify previous findings, and build directly from previous work. In cases where reproducing large-scale hydrologic studies is computationally very expensive and time-consuming, new processes are required to ensure scientific rigor. Such changes will strongly improve the transparency of hydrological research, and thus provide a more credible foundation for scientific advancement and policy support.

Example: North America setup for SUMMA

Core Modelling Team



~ 0.6 million catchments in Merit Hydro database

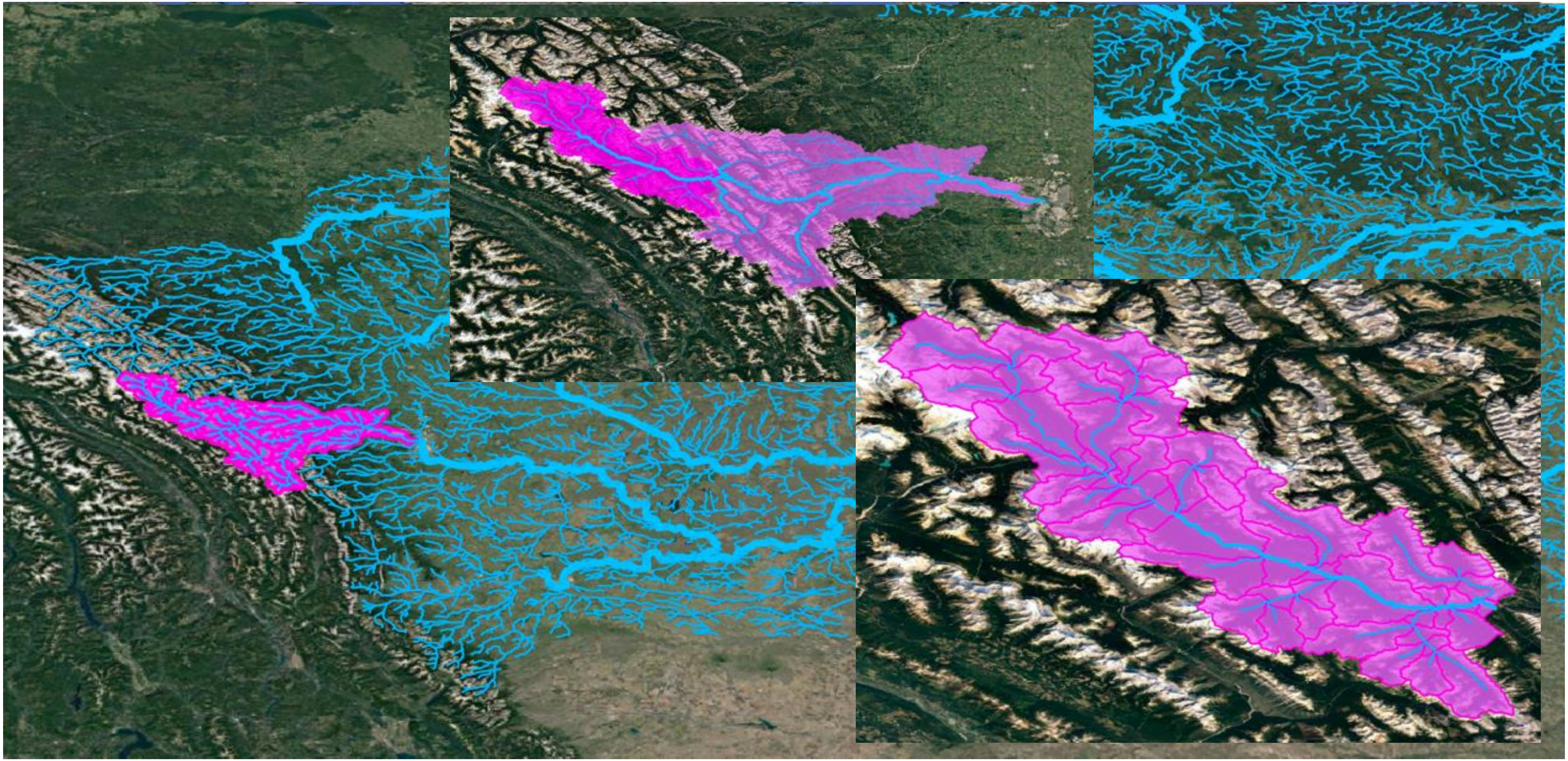
(Yamazaki et al., WRR, 2020; Lin et al., 2020)

Bow at Banff

Data requirements:

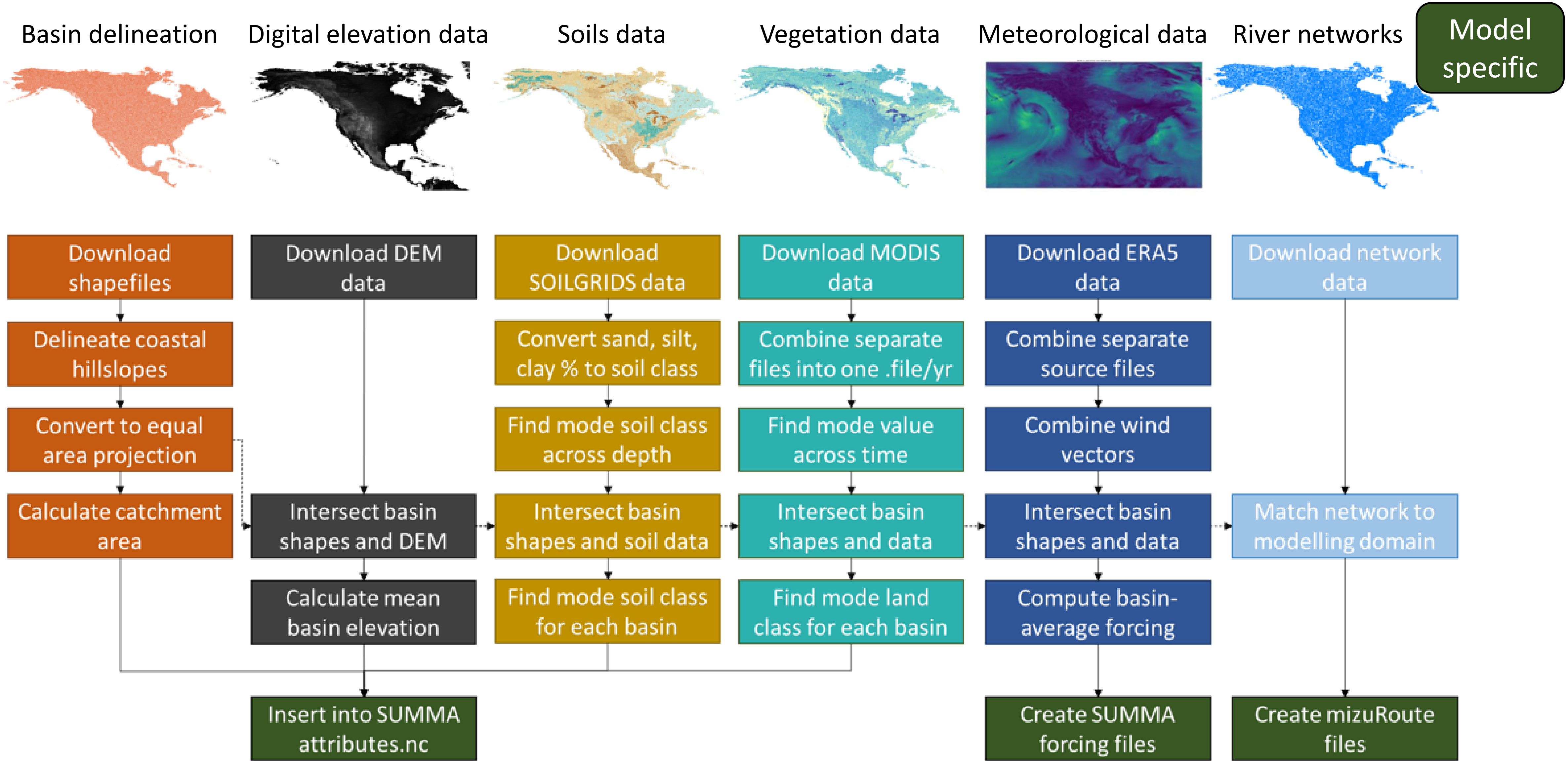
- Topographic parameters
- Soil parameters
- Vegetation/land use parameters
- Forcing data
- River network

Global network datasets: MERIT-Hydro



Developing model-agnostic workflows

Core Modelling Team

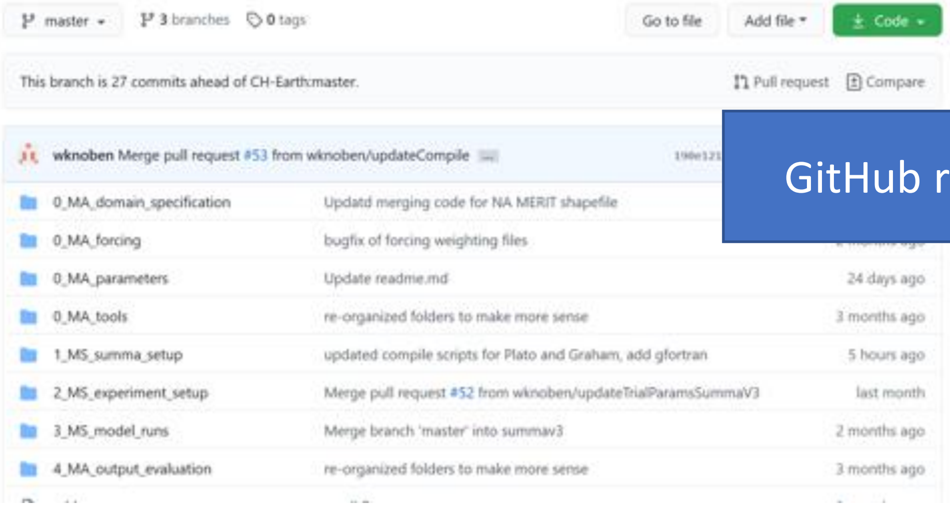


Wouter Knoben

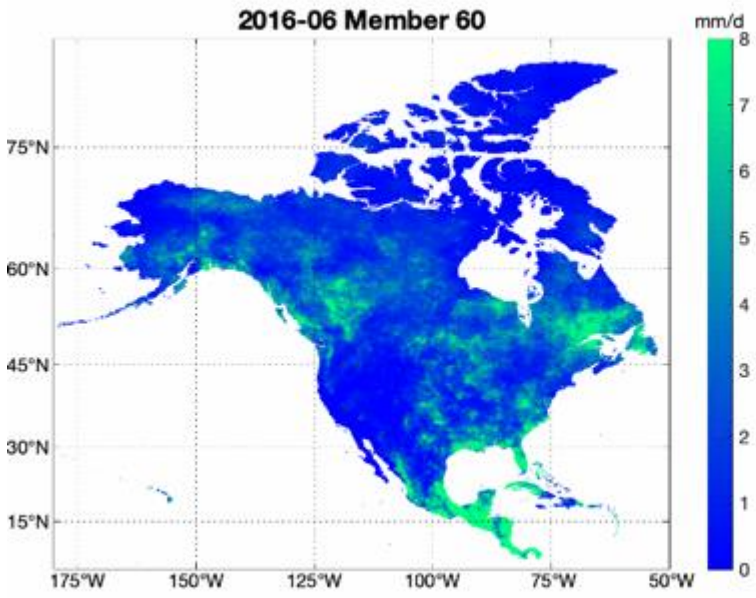
Developing model-agnostic workflows

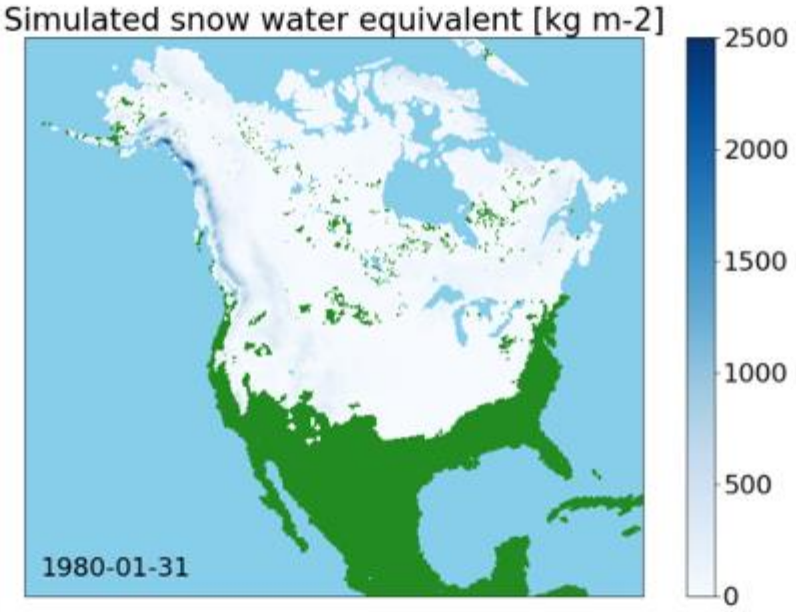
Core Modelling Team

- Goal: Improve the efficiency of continental-domain model implementation tasks
- Easier to collaborate; easier to keep track of work for reporting and paper reviews
- Increase transparency, reproducibility, and code re-use
- Advance ***community hydrological modelling***, rather than a *community hydrological model*



GitHub repository



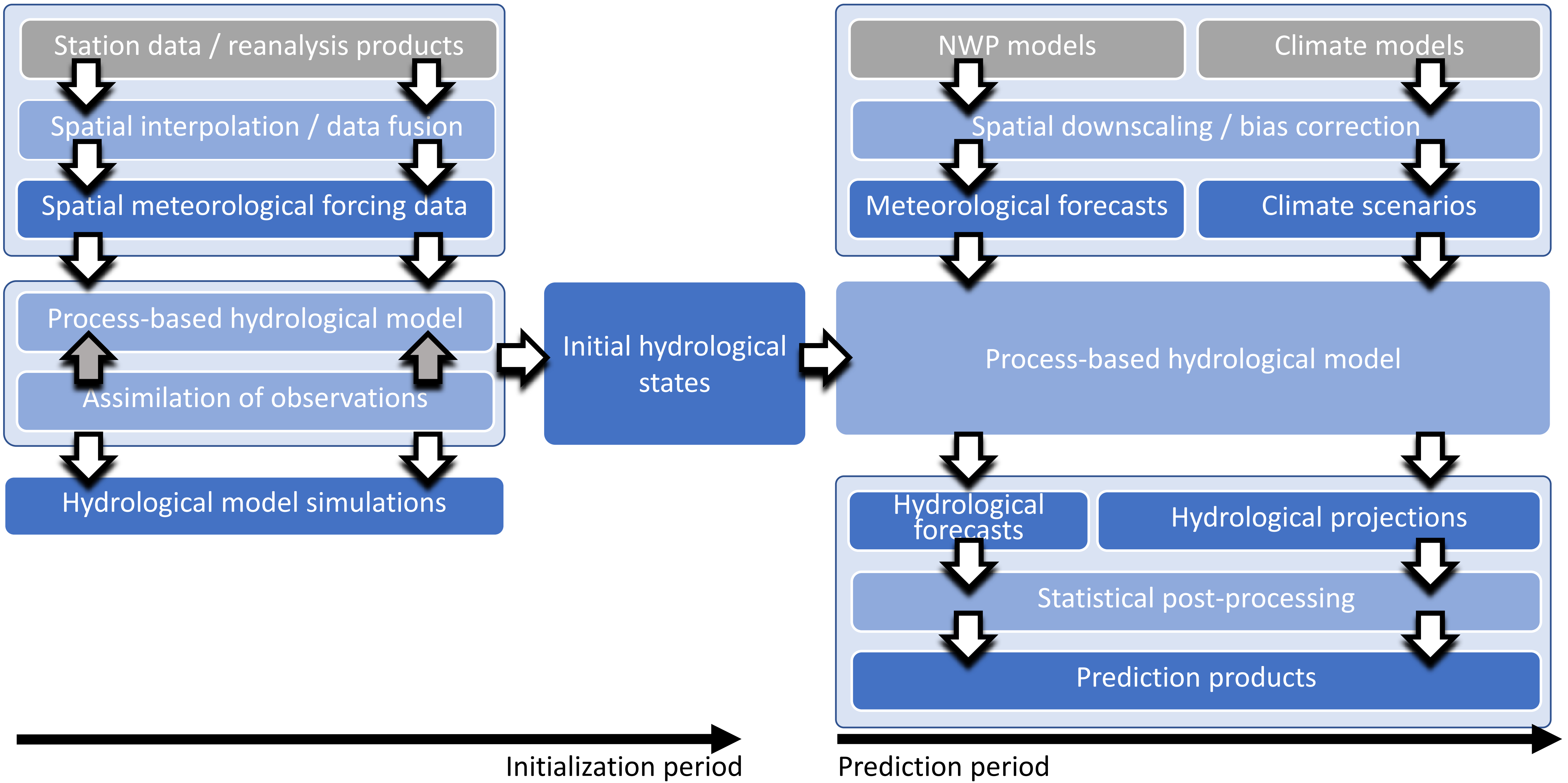


Streamflow and lake level

Topics for today

- Spatial meteorological forcing data
- Geospatial intelligence
- Summary and outlook

Hydrological prediction applications

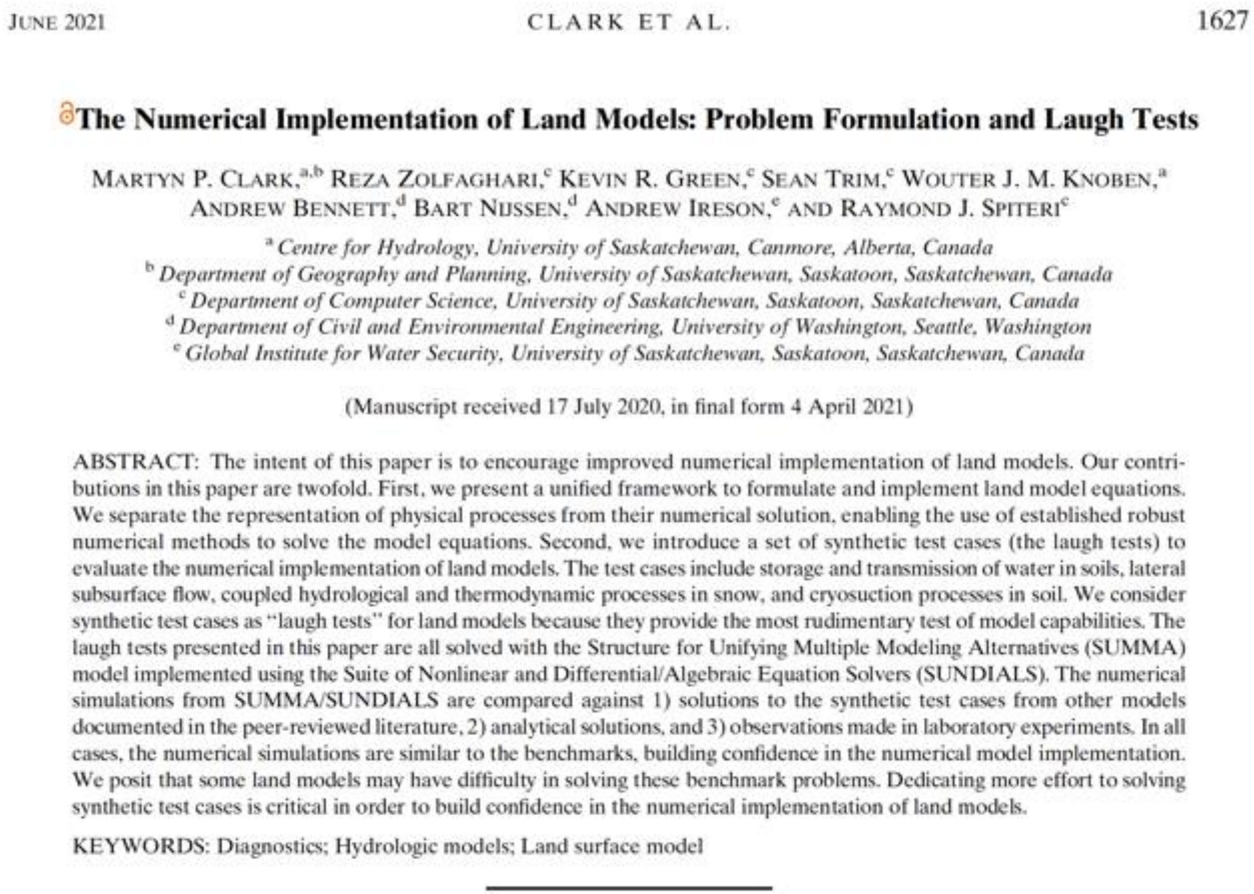


Summary and outlook

- Developing computational infrastructure to enable large-domain hydrological modelling
  - Deterministic and probabilistic forecasting capabilities
- Model agnostic philosophy
  - Focus on *community hydrological modelling* rather than developing a community hydrological model
  - Advance a community of practice to more effectively share code and concepts across different model development groups
- Interested in developing state-of-the-art capabilities in predicting terrestrial water, energy, and biogeochemical cycles
  - Looking forward to collaborations

Fun fact: Most land models do not conserve energy

Core Modelling Team



The Numerical Implementation of Land Models: Problem Formulation and Laugh Tests

MARTYN P. CLARK,[a,b] REZA ZOLFAGHARI,[c] KEVIN R. GREEN,[c] SEAN TRIM,[c] WOUTER J. M. KNOBEN,[a] ANDREW BENNETT,[d] BART NIJSSEN,[d] ANDREW IRESON,[e] AND RAYMOND J. SPITERI[c]

[a] *Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada*
[b] *Department of Geography and Planning, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*
[c] *Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*
[d] *Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington*
[e] *Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*

(Manuscript received 17 July 2020, in final form 4 April 2021)

ABSTRACT: The intent of this paper is to encourage improved numerical implementation of land models. Our contributions in this paper are twofold. First, we present a unified framework to formulate and implement land model equations. We separate the representation of physical processes from their numerical solution, enabling the use of established robust numerical methods to solve the model equations. Second, we introduce a set of synthetic test cases (the laugh tests) to evaluate the numerical implementation of land models. The test cases include storage and transmission of water in soils, lateral subsurface flow, coupled hydrological and thermodynamic processes in snow, and cryosuction processes in soil. We consider synthetic test cases as "laugh tests" for land models because they provide the most rudimentary test of model capabilities. The laugh tests presented in this paper are all solved with the Structure for Unifying Multiple Modeling Alternatives (SUMMA) model implemented using the Suite of Nonlinear and Differential/Algebraic Equation Solvers (SUNDIALS). The numerical simulations from SUMMA/SUNDIALS are compared against 1) solutions to the synthetic test cases from other models documented in the peer-reviewed literature, 2) analytical solutions, and 3) observations made in laboratory experiments. In all cases, the numerical simulations are similar to the benchmarks, building confidence in the numerical model implementation. We posit that some land models may have difficulty in solving these benchmark problems. Dedicating more effort to solving synthetic test cases is critical in order to build confidence in the numerical implementation of land models.

KEYWORDS: Diagnostics; Hydrologic models; Land surface model

- Enthalpy formulation
- The apparent heat capacity is not representative of the change over the time step
- See Equation 44 in Clark et al. 2021