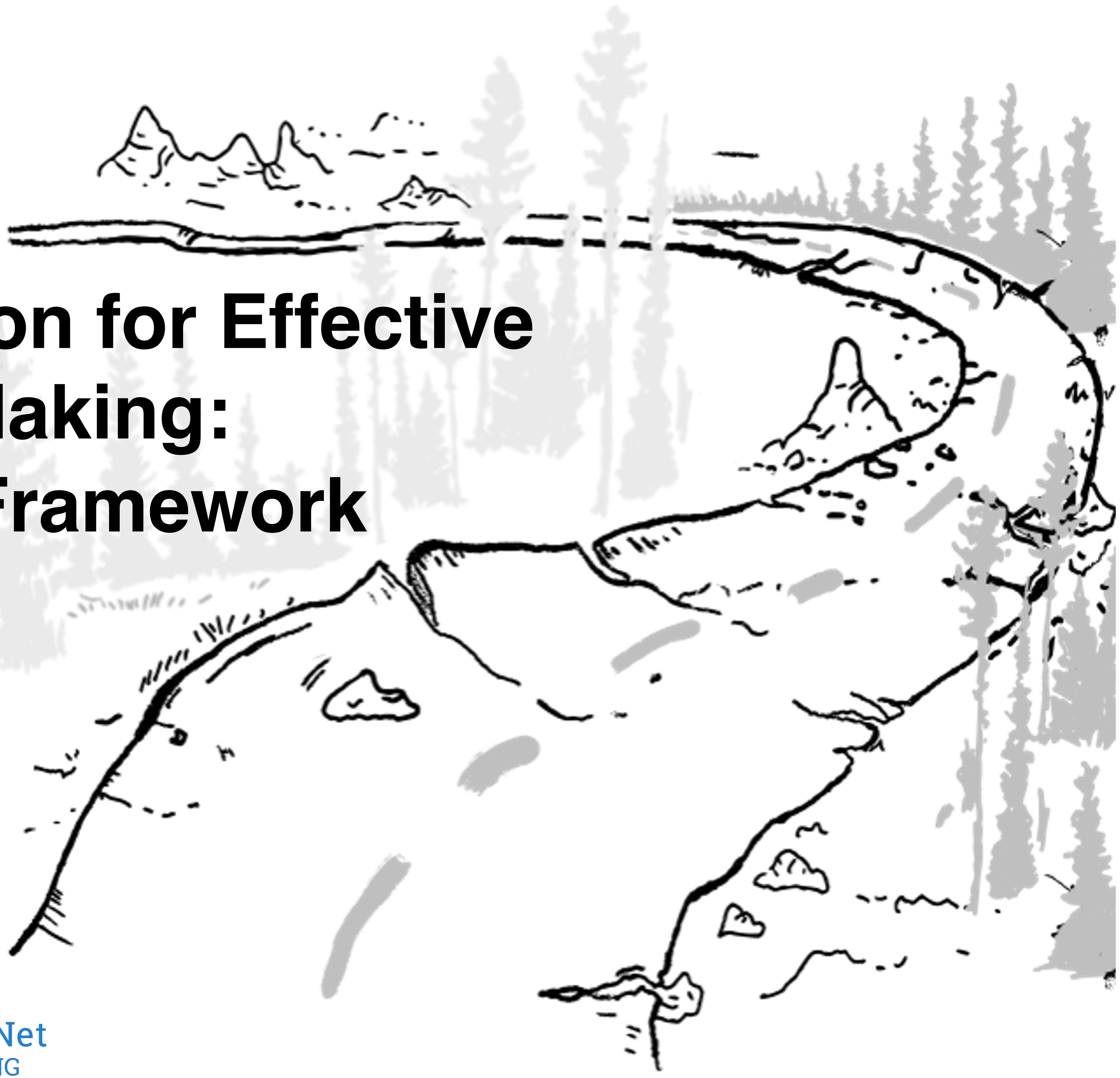
Guiding Model Selection for Effective Adaptation Decision Making: A Statistical Ranking Framework

T3-PhD4: Quantifying Confidence in Ground Temperature Simulations



Hannah Macdonell NOV 2023

Carleton University

Northwest Territories

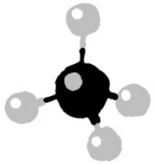
PermafrostNet
NSERC | CRSNG

How can modelling help?

Making useful predictions for current and future:

Ground ice content

Active layer thickness



Carbon storage

Ground temperatures

Modelling Evaluation Obstacles

Statistics

Data Availability

Modelling Evaluation Obstacles

Statistics

Lack of statistical consensus

Data Availability

Modelling Evaluation Obstacles

Statistics

- Lack of statistical consensus
- Interpretation of statistical values

Data Availability

Modelling Evaluation Obstacles

Statistics

- Lack of statistical consensus
- Interpretation of statistical values

Data Availability

- Limited spatial coverage

Modelling Evaluation Obstacles

Statistics

- Lack of statistical consensus
- Interpretation of statistical values

Data Availability

- Limited spatial coverage
- Incomplete observational datasets

Modelling Evaluation Obstacles

Statistics

- Lack of statistical consensus
- Interpretation of statistical values

Data Availability

- Limited spatial coverage
- Incomplete observational datasets
- Observations ≠ variables of interest

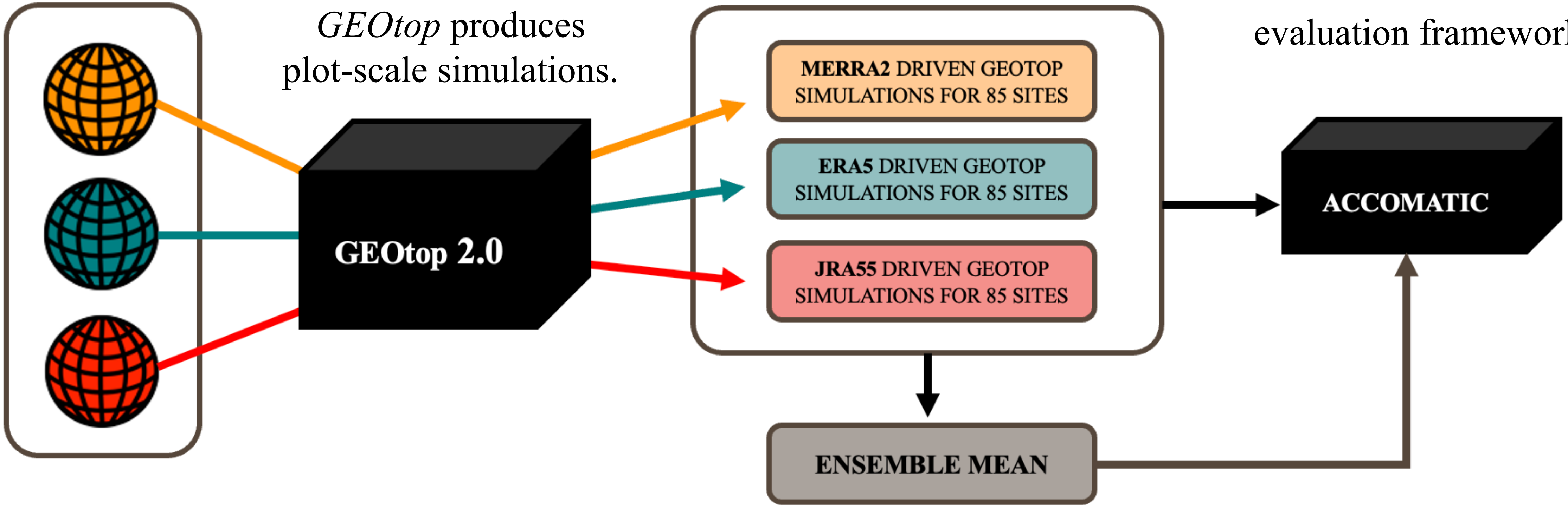
Carleton University
Northwest Territories
PermafrostNet
NSERC | CRSNG

Producing GST Simulations

Three reanalysis data products are used as driving data.

GEOtop produces plot-scale simulations.

Four ***models*** **(simulation output)** are fed into the model evaluation framework.



Producing GST Simulations

~ 10 cm below the ground surface

Mini loggers that measure GST.

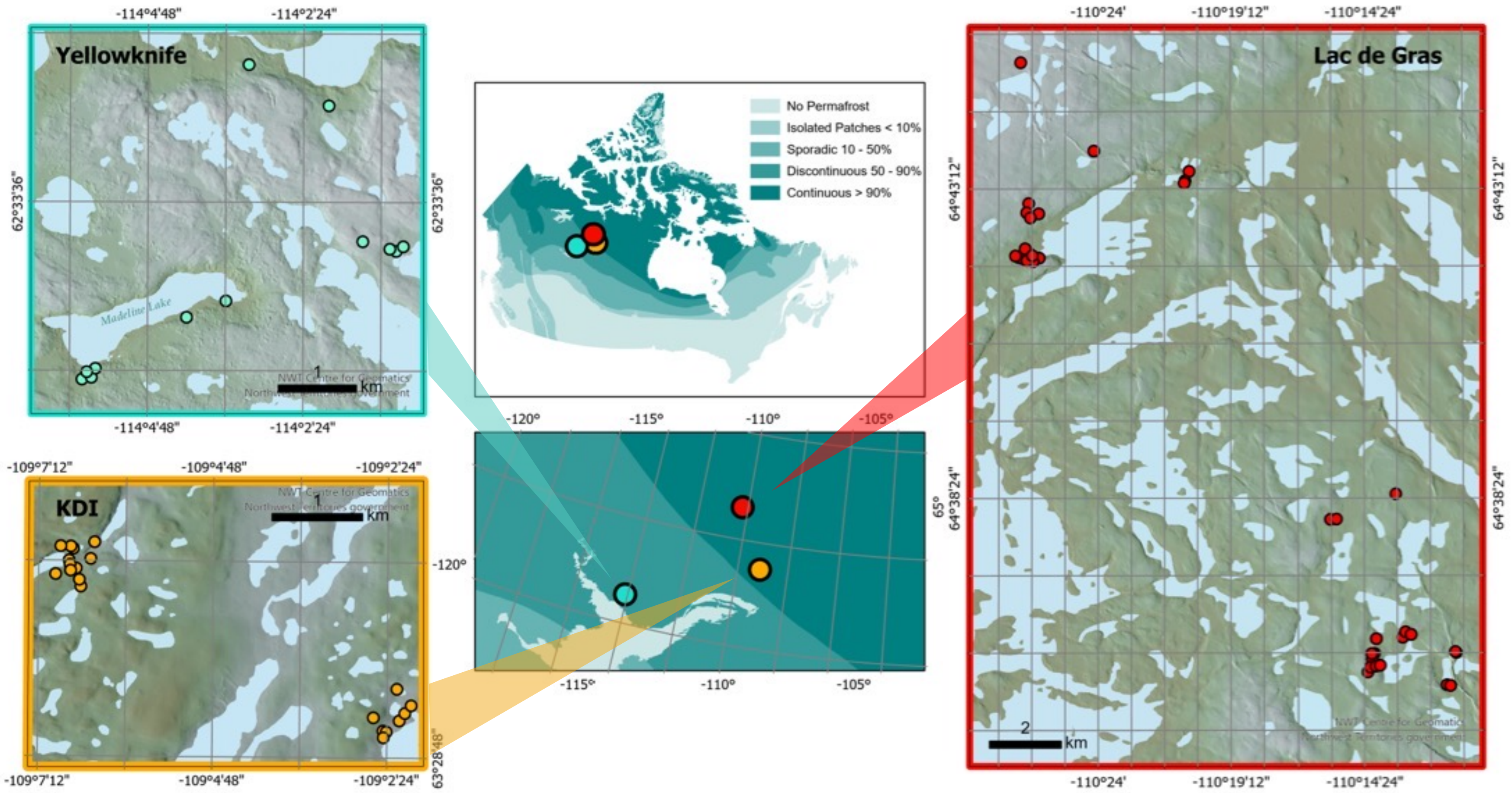


Fig. 1 Map of GST site clusters in Canada.

Producing GST Simulations

Describing site characteristics

- Vegetation
- Snow collection
- Self-shading
- Terrain wetness
- Ground material

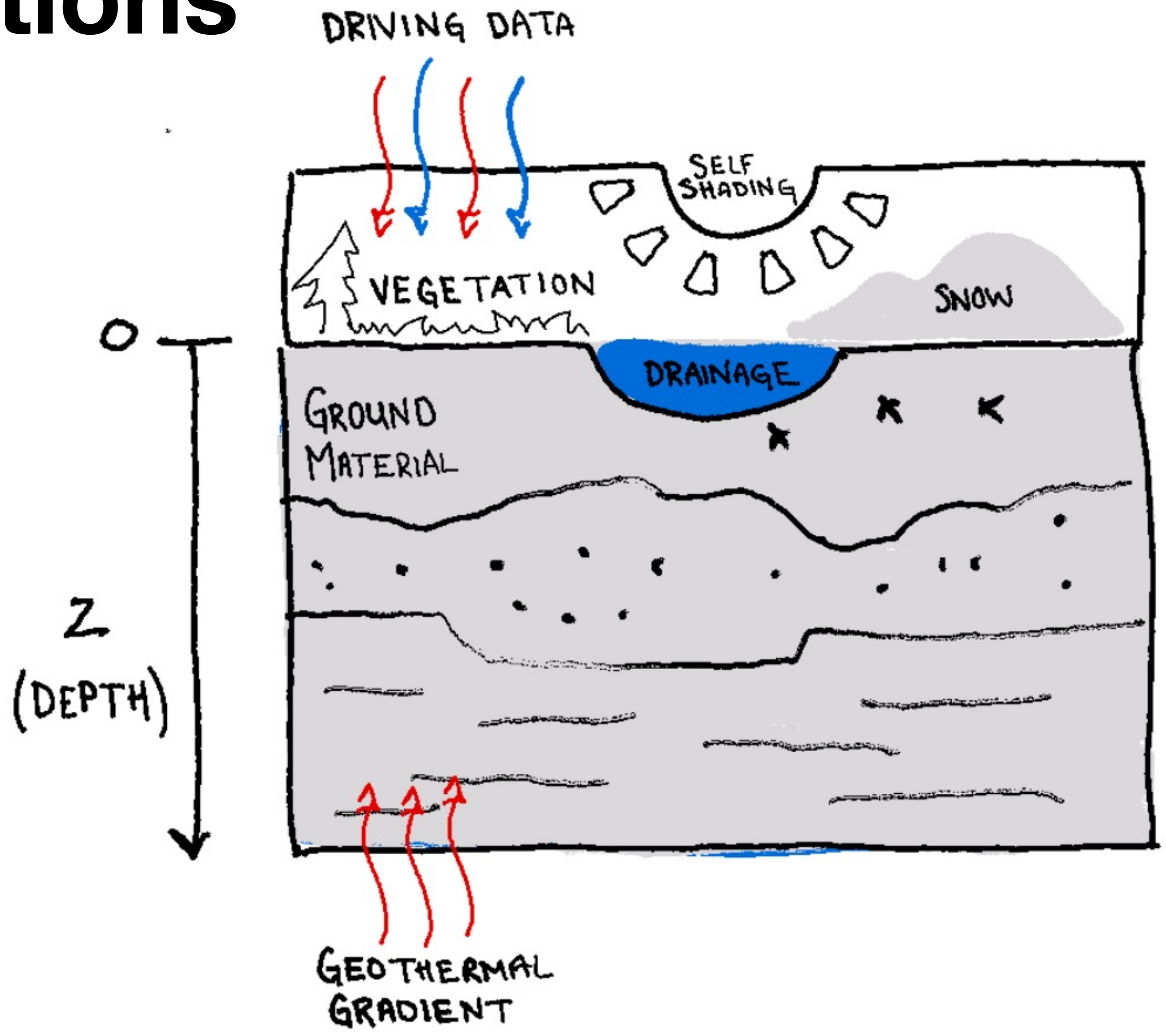


Fig. 2 Rough diagram of components used to predict GST.

Producing GST Simulations for Evaluation

Describing surface characteristics of each site

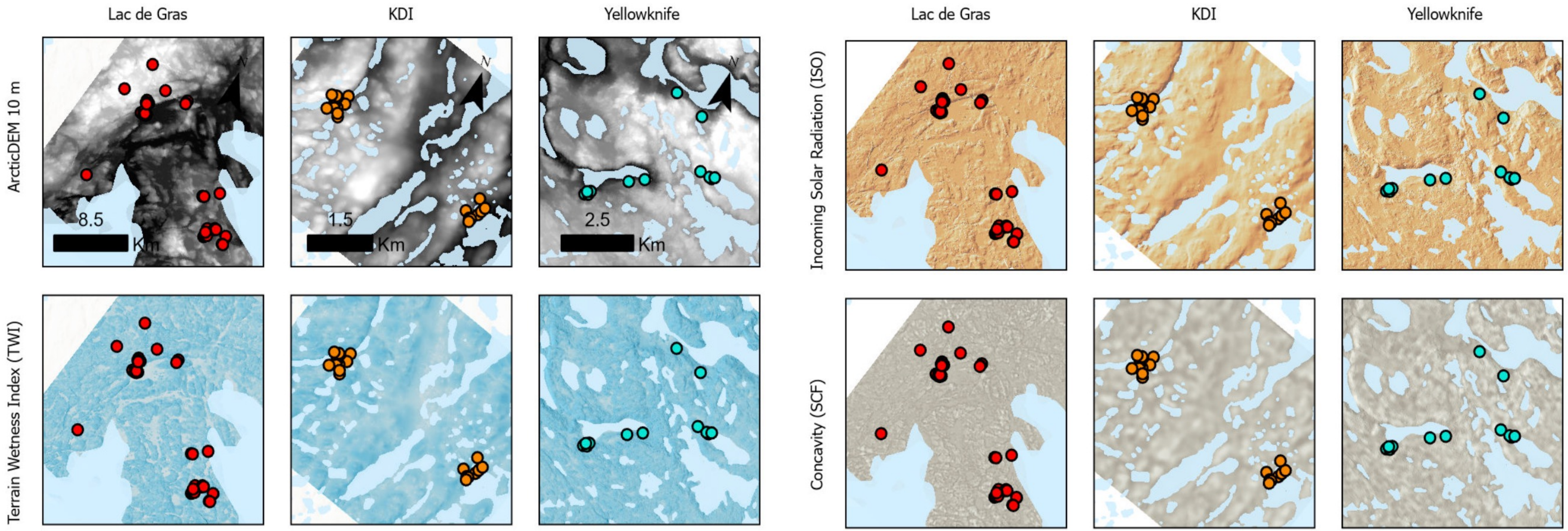


Fig. 3 Maps of three GST clusters in NWT showing elevation, drainage, self-shading and concavity.

Producing GST Simulations for Evaluation

Describing surface characteristics of each site

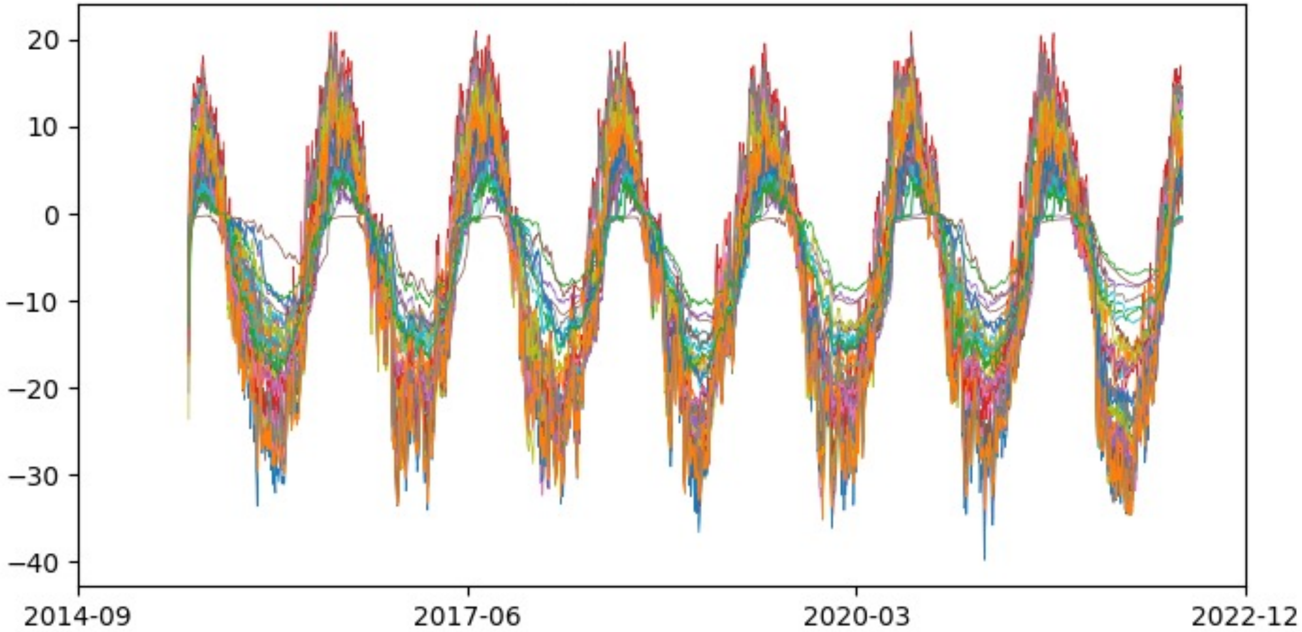


Fig. 4 *Visualization* of timeseries GST output from GEOtop for multiple sites

Producing GST Simulations for Evaluation

Describing surface characteristics of each site

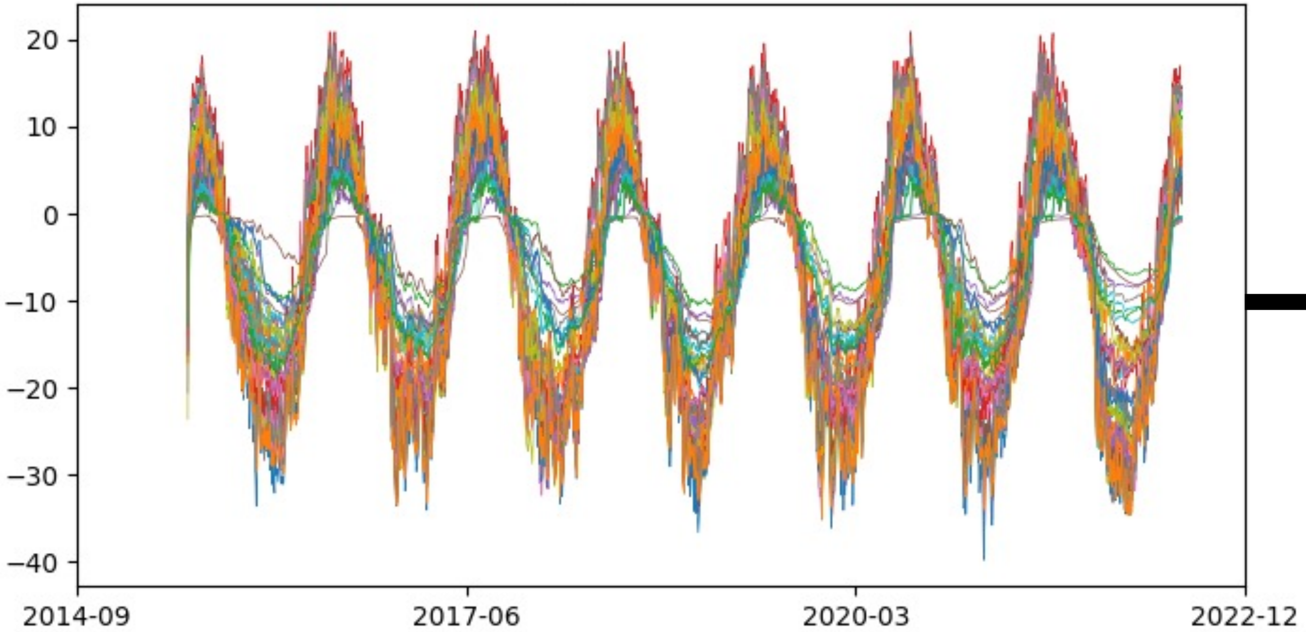


Fig. 4 *Visualization* of timeseries GST output from GEOtop for multiple sites

Accomatic: A ranking Framework

Lack of statistical consensus

Incomplete observational datasets

Interpretation of statistical values

Limited spatial coverage

Observations not always var Of interest

Lack of statistical consensus

Model Evaluation Anarchy

Models cannot be compared due to the lack of consensus over which statistics to use.

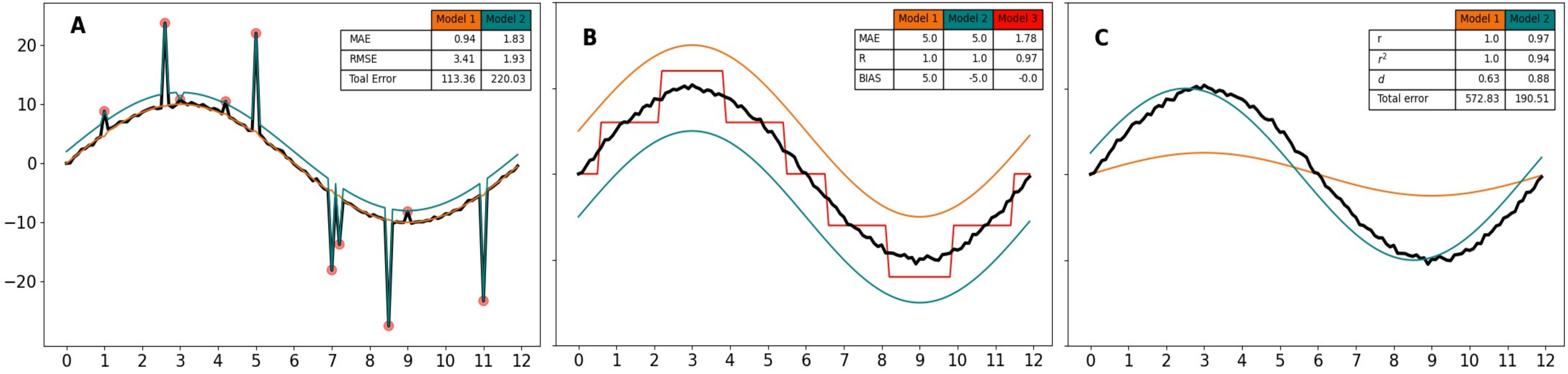


Fig. 5 Synthetic data shows the problem with (A) RMSE, (B) Bias and (C) r statistics.

Interpretability of statistics

Solution: A Ranking Framework

"Statistics are the grammar of science."

- Karl Pearson

Most statistical values are **intangible in reality**, and often **mathematically unrelated** to one another**.** Many domains rely on **rankings** to establish "the best".

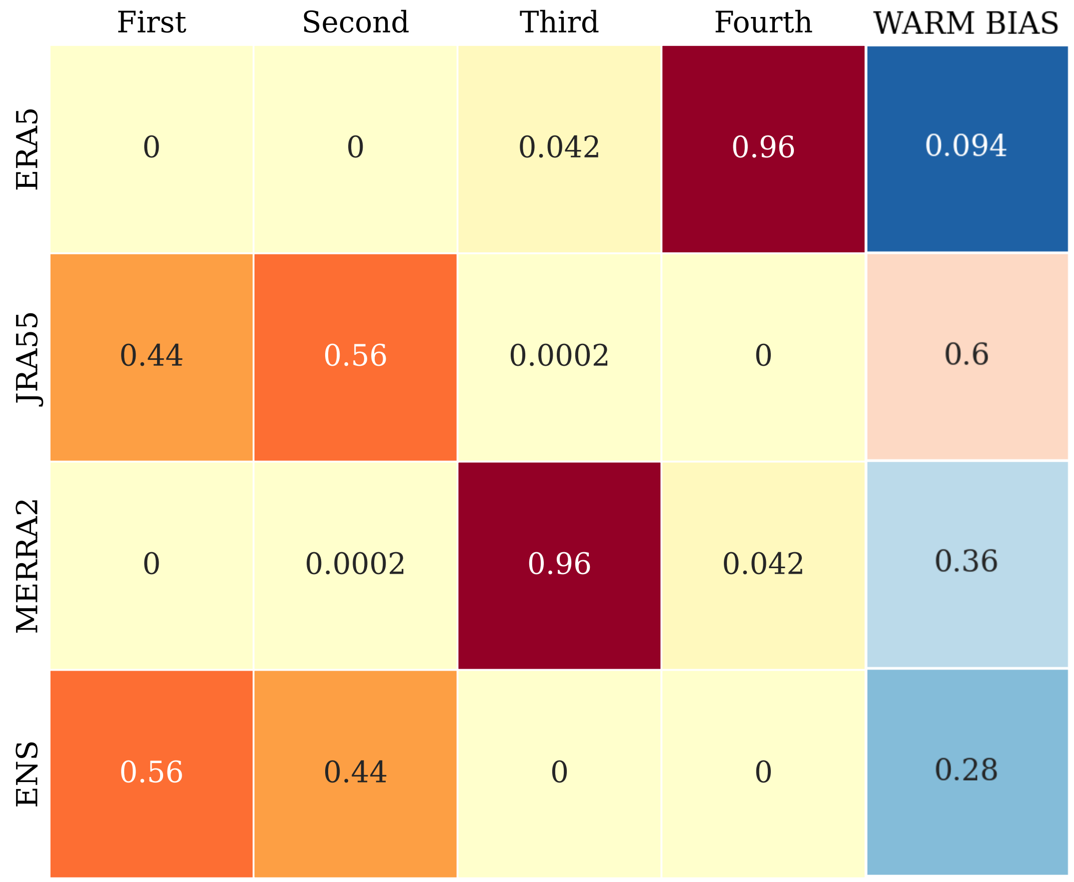


Fig. 6 Rank distribution for four models and their biases.

[1] Drew, L. J., & Schuenemeyer, J. H. (2011). *Statistics for earth and environmental scientists*. John Wiley & Sons.

Incomplete observational datasets

Bootstrapping timeseries observations

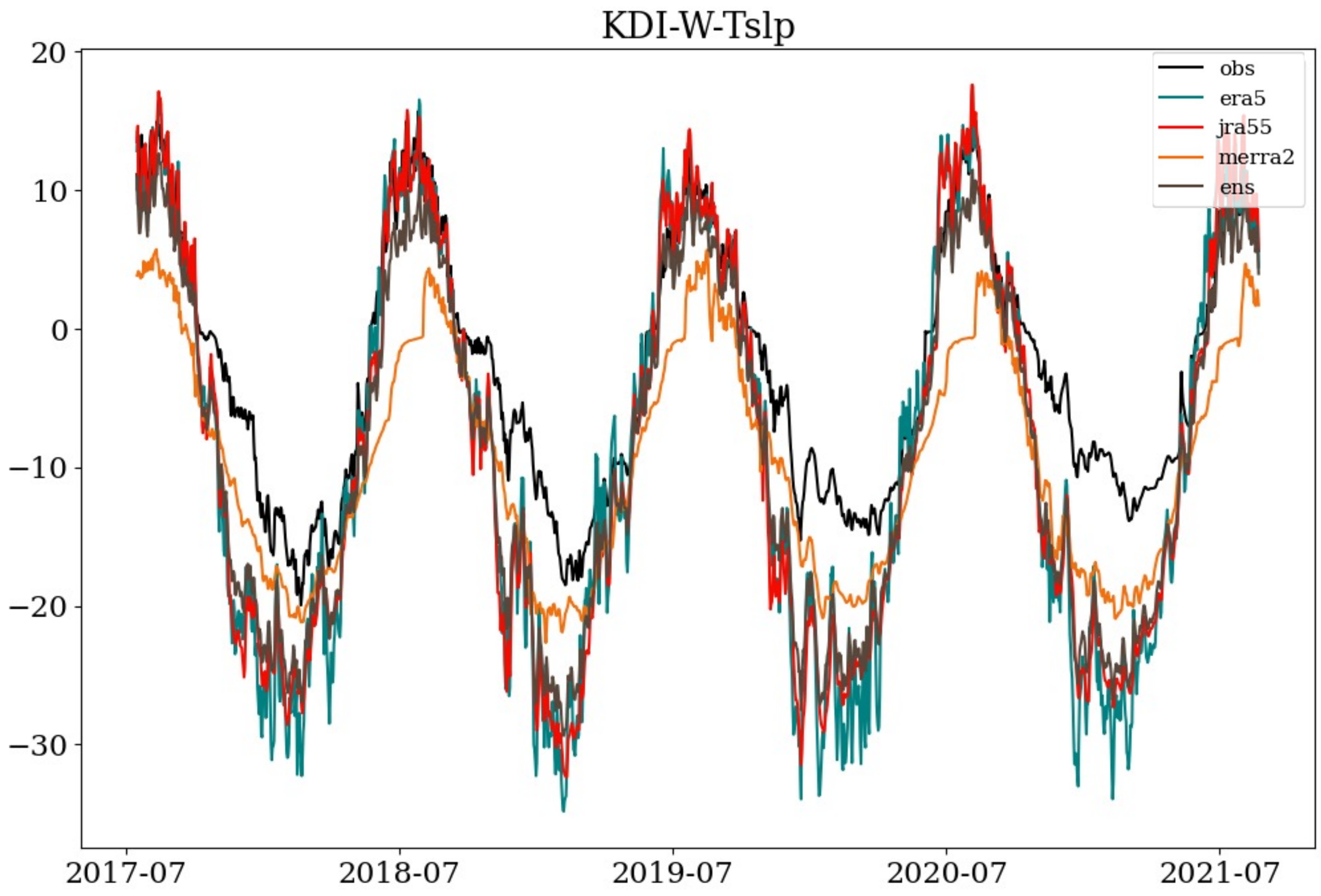


Fig. 9 Timeseries GST data modelled and observed (black).

To avoid introducing seasonal bias into model results, **complete years** of data are favoured for evaluation. This means lots of **data is lost** from model evaluation.

Incomplete observational datasets

Bootstrapping timeseries observations

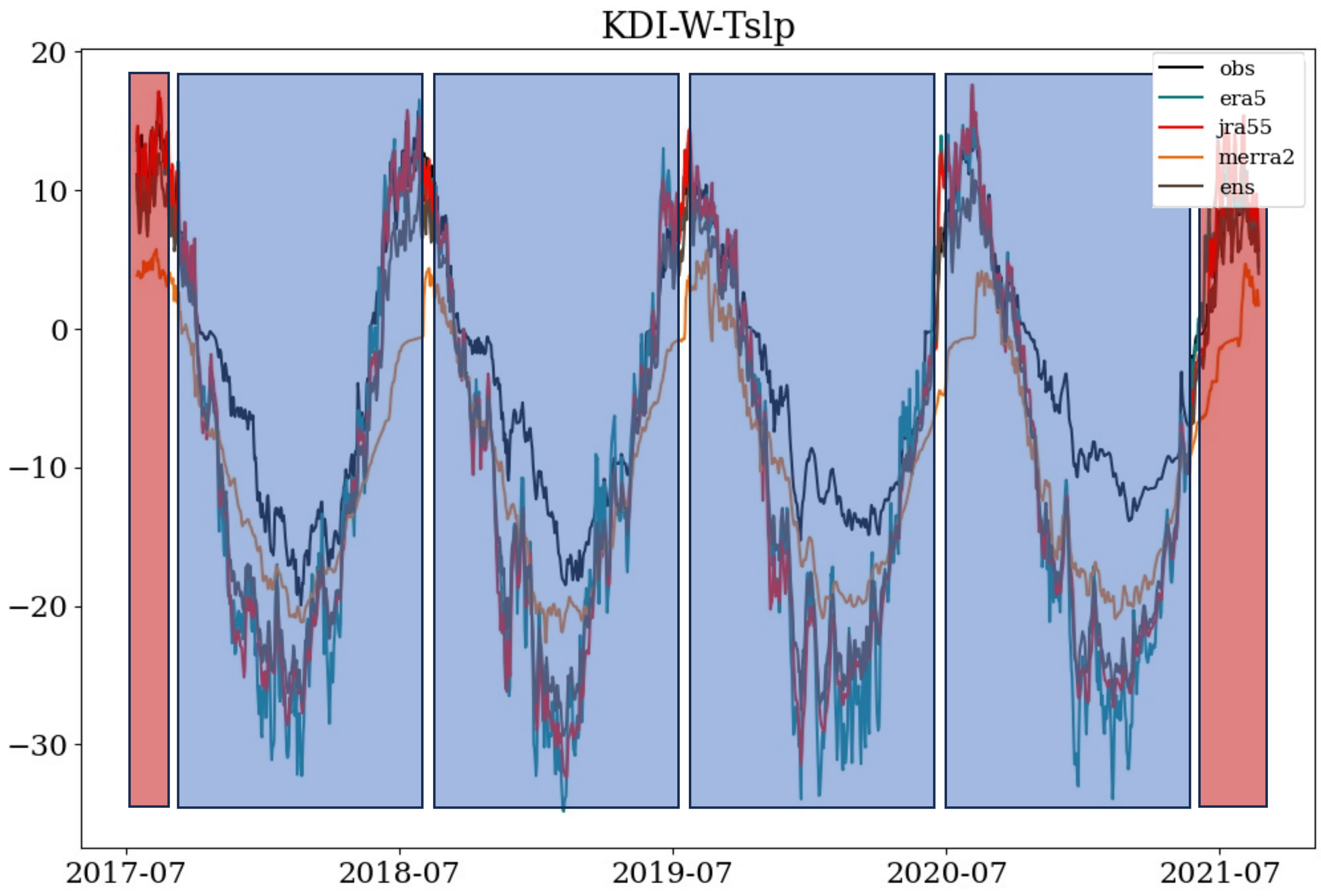


Fig. 9 Timeseries GST data modelled and observed (black).

To avoid introducing seasonal bias into model results, **complete years** of data are favoured for evaluation. This means lots of **data is lost** from model evaluation.

Incomplete observational datasets

Bootstrapping timeseries observations

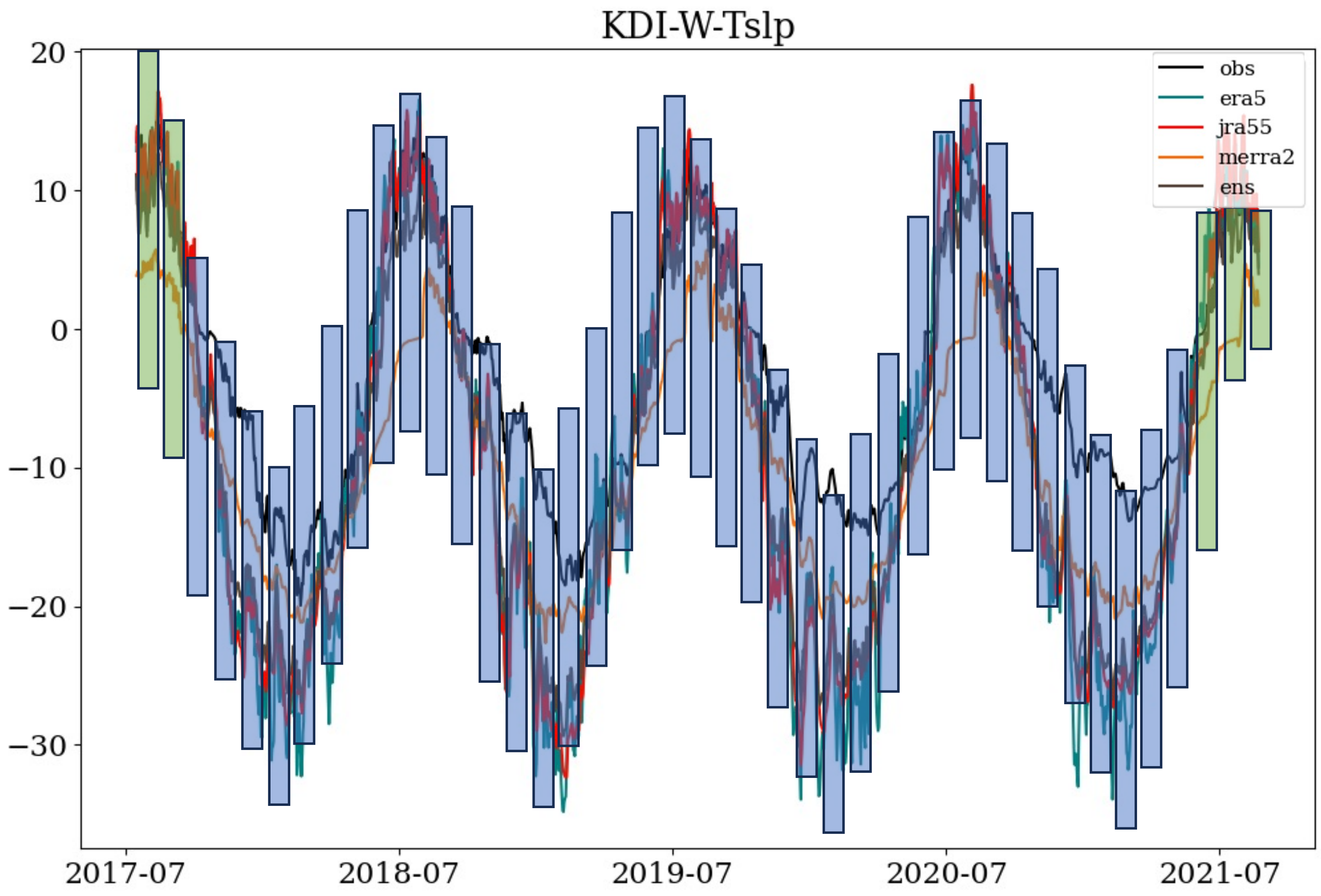


Fig. 9 Timeseries GST data modelled and observed (black).

Subsetting model evaluation by terrain type can **mitigate** any **potential bias** towards terrains with more observations.

Incomplete observational datasets

Bootstrapping timeseries observations

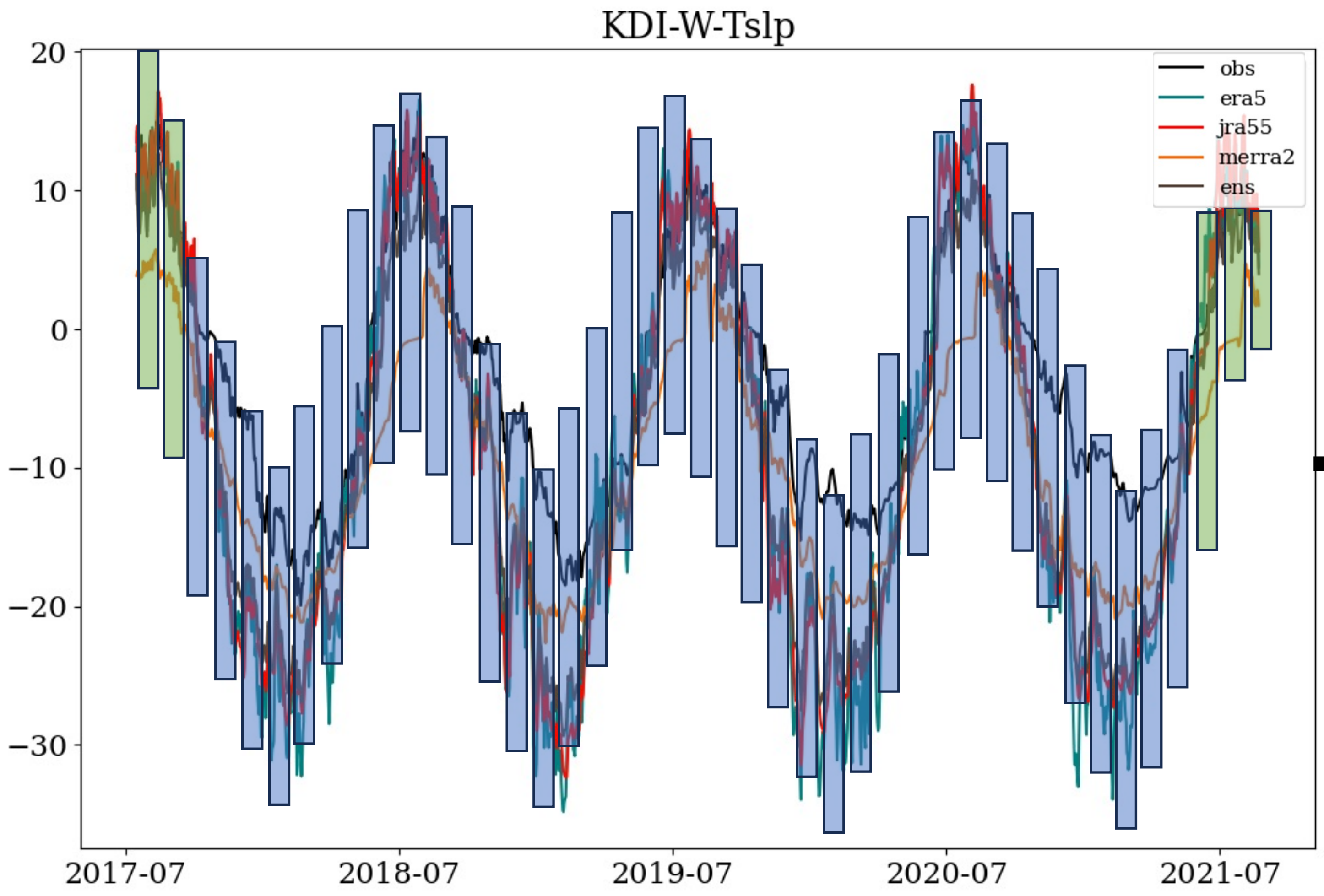


Fig. 9 Timeseries GST data modelled and observed (black).

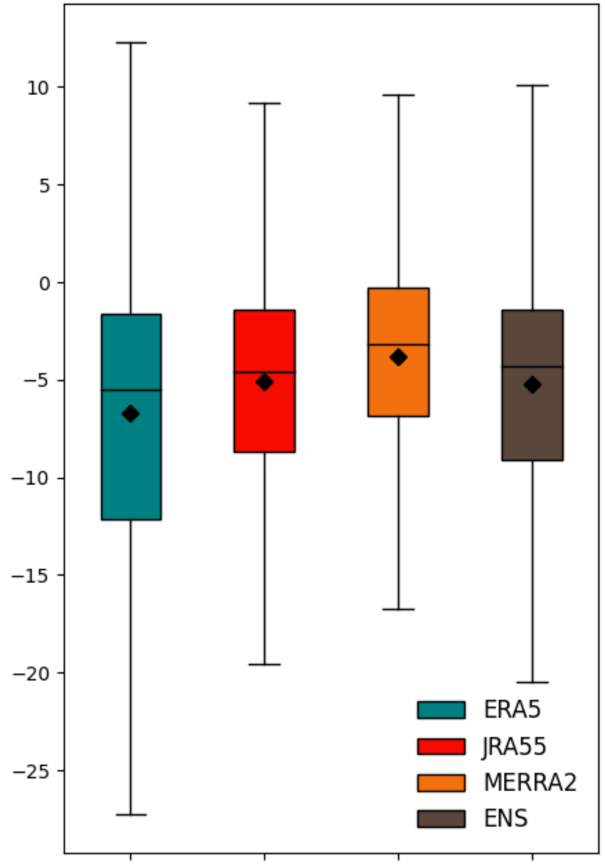


Fig. 10 Bootstrap results for a BIAS metric in Nov.

Limited spatial coverage of observations

Specifying biogeoclimatic zones

Analysing performance **across different terrains** leads to a better understanding of model **strengths** and **weaknesses**.

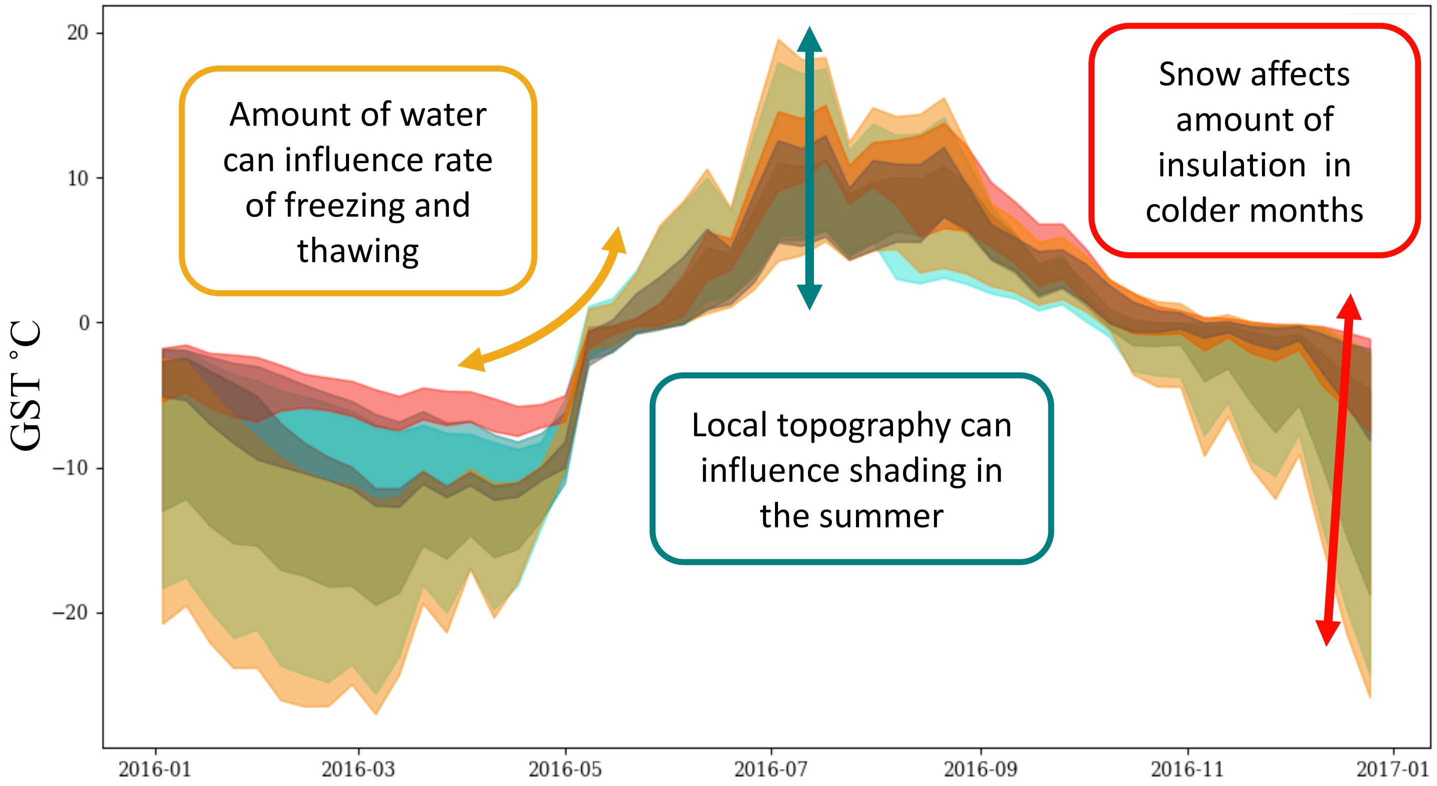


Fig. 11 Range of ground surface temperatures observed across terrain types.

≠ Observations ≠ variables of interest

Extension of simulations to greater depths

- Essentially: are our “best” simulations able to be “best” elsewhere
- How can we measure our ability to predict deeper temperatures?

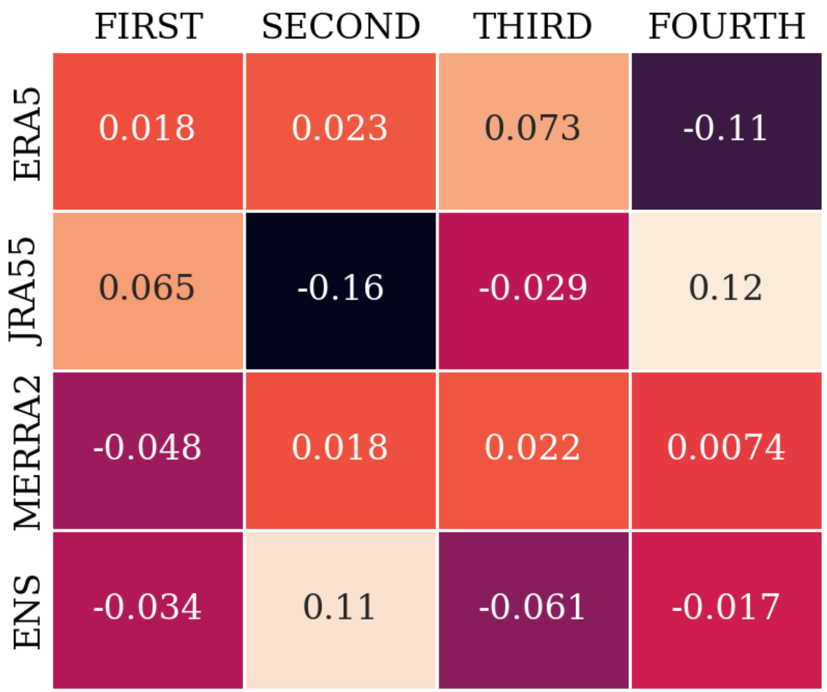


Fig. 12 Heatmap of differences in rank distribution with depth.

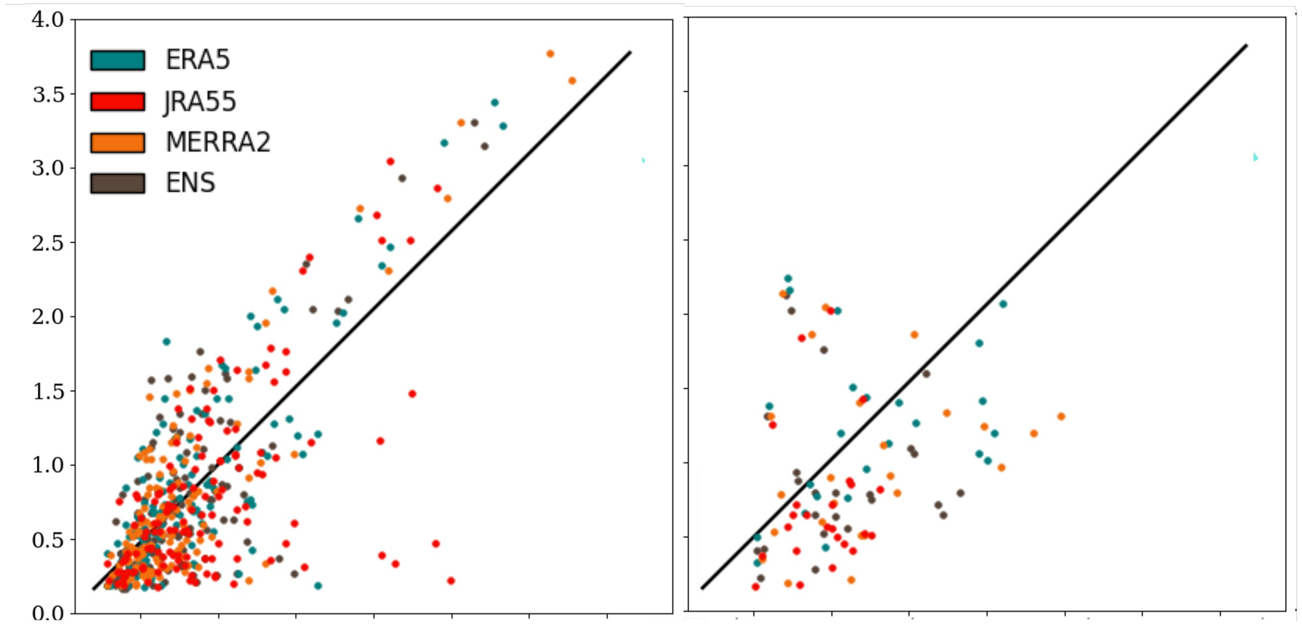


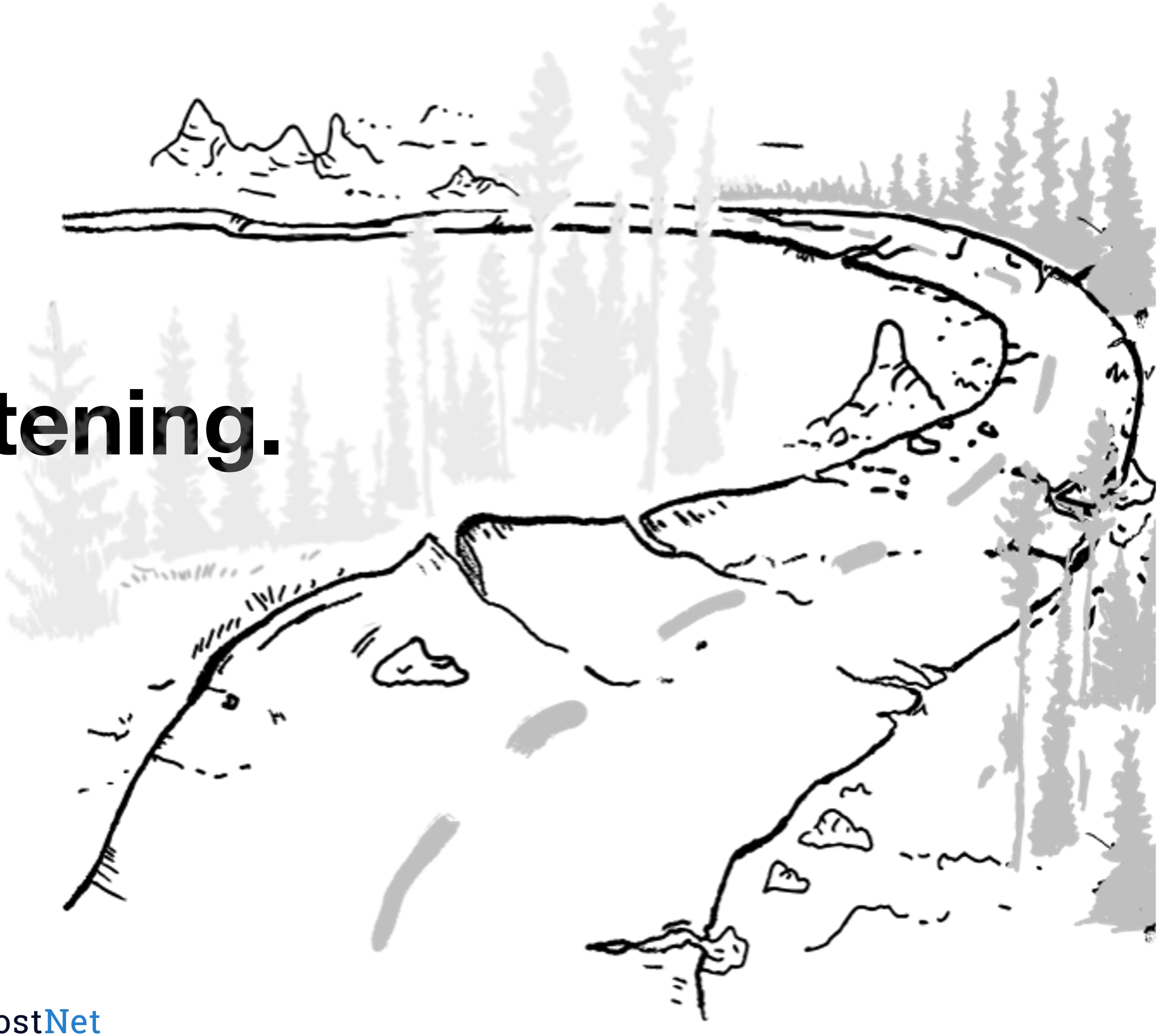
Fig. 13 Correlation of model performance at 0.1 and 0.5 m depth.

AGM 23 Victoria
Hannah Macdonell

Recap: Modelling and evaluation challenges… and their solutions

Challenge	Solution
Limited spatial coverage	Sub-setting and weight model performance by terrain type
Incomplete datasets	Bootstrapping
Lack of statistical consensus	Fit statistics to your variable of interest
Interpretability of statistics	Rank models
Observed ≠ Interesting	Do model results extend to greater depths?

Thank you for listening.



Hannah Macdonell NOV 2023

Carleton University

Northwest Territories

PermafrostNet
NSERC | CRSNG