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

PermafrostNet
NSERC | CRSNG

HANNAH MACDONELL
M.Sc. D.S. CARLETON UNIVERSITY

Carleton University

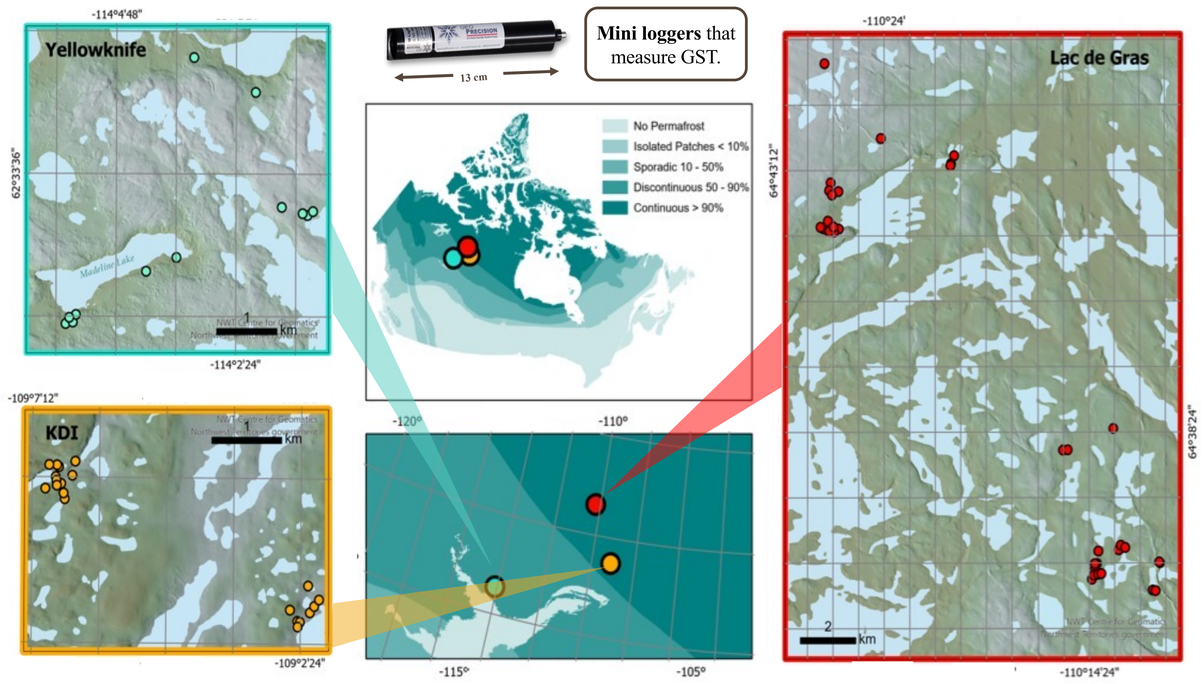


Figure 1: Clusters of ground surface temperature observations used for this study.

Introduction / Background

Permafrost modelling can contribute to **informing adaptation** in permafrost regions by characterizing the subsurface thermal. However, **models vary** in their performance. We need to be able to make justifiable comparisons between simulation products to improve the representation of permafrost processes in modelling software.

Consistency in metrics for model evaluation provides an opportunity to better compare the relative strengths of multiple models. In this study, we evaluate models under a range of accordance measures, for differing terrain types, and temporal subsets. Through review and experimental testing, we aim to develop **a ranking of simulation quality** that accounts for the specific characteristics of ground surface temperatures (GST) in permafrost areas.

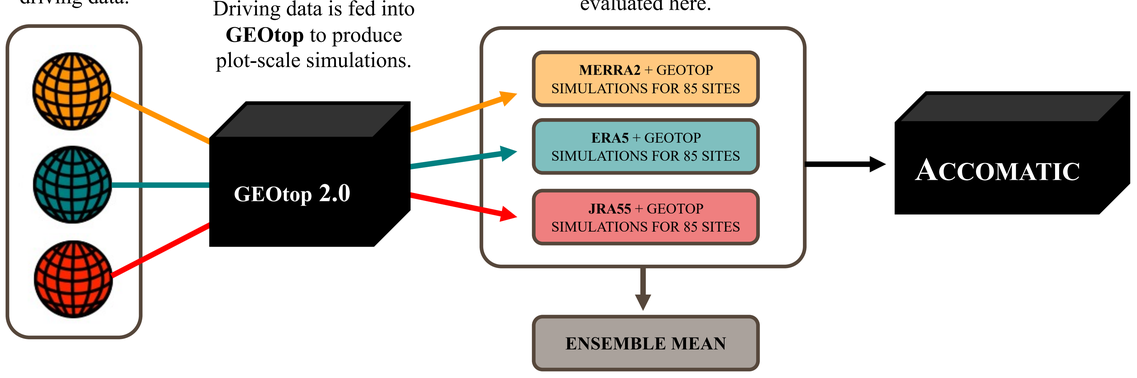
Methodology

SIMULATIONS We are comparing different *models* or *simulation outputs* which are a combination of **driving data** (MERRA2, JRA55, ERA5), **modeling software** (GEOtop), and **parameters.**

ACCOMATIC The python package used produce a suite of summary statistics and generate model rankings. Each simulation will be tested against a range of accordance measures, then split by season and terrain type.

APPROACH On the right are **five model evaluation** obstacles, and the solutions implemented by this framework. They are categorized into data availability and statistics.

Producing GST Simulations



ACCOMATIC: A RANKING FRAMEWORK

Obstacles & Solutions

Lack Of Statistical Consensus

PROBLEM: Models cannot be compared due to the **lack of consensus** over which statistics to use. There are problems with commonly used statistics: *RMSE*, *bias*, and *r*. In **ACCOMATIC**, three statistics are carefully selected to evaluate temperature simulations: *BIAS*, *MAE*, and Willmott's index of agreement d_r.

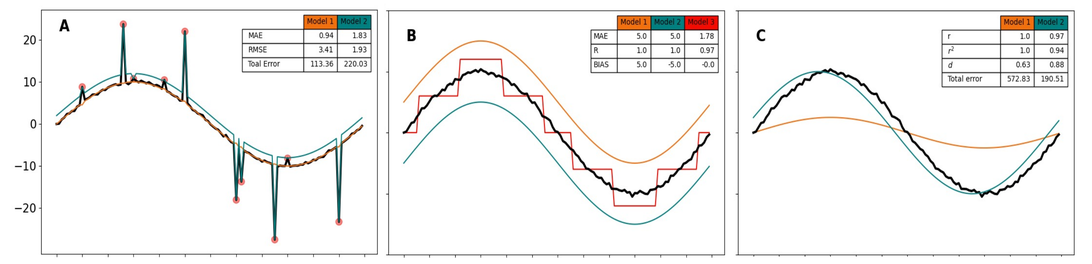


Figure 2: Demonstrating how statistical measure selection influences our interpretation of performance.

Incomplete Observational Datasets

PROBLEM: 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.

The bootstrap procedure implemented by **ACCOMATIC** segments modelled and observed timeseries into month-long sections, then evaluates **random samples** from this set, getting a distribution of model performance.

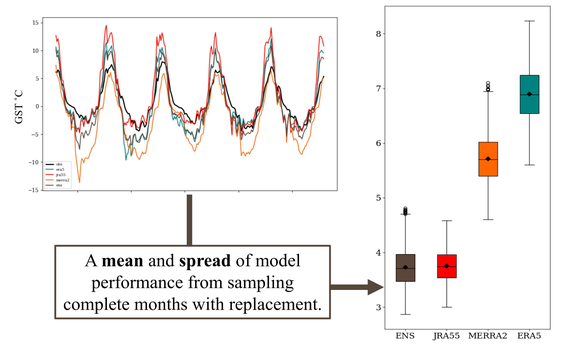


Figure 3: Interpreting model performance using a bootstrap procedure.

Limited Spatial Coverage

PROBLEM: Permafrost environments exhibit remarkable heterogeneity and model evaluation can be biased towards areas for which we have more data.

In **ACCOMATIC**, model evaluation is **subset** by **shared surface terrain features.** This allows for a better understanding of how the model performs in different environments, **mitigating any potential bias** towards terrains with abundant observations

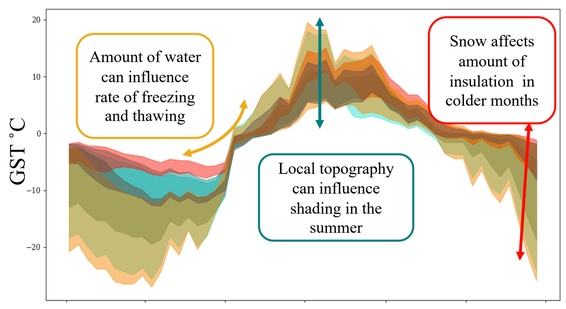


Figure 4: Range of GST observed across terrain types.

Interpretation Of Statistical Values

PROBLEM: Most statistical values are **intangible** and often **mathematically unrelated** to one another, making them difficult to interpret.

ACCOMATIC implements a ranking procedure that is sufficiently **transparent** to allow for reproducibility and enable critical evaluation. Here we see a visualization of ranking with confidence. For example, the ERA5 model ranks last **96%** of the time (don't pick this one).

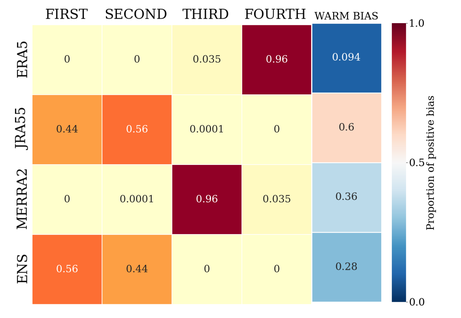


Figure 5: Heatmap showing the proportion of instances each model occupied a certain rank, and bias.

Observations Not Always Variables Of Interest

PROBLEM: Variables that have lots of observations (e.g. GST), are not what we're most interested in (Active layer thickness, ground ice content) when it comes to future permafrost change.

Using **ACCOMATIC**, the extent to which model performance at the surface (GST) **differs** from **performance at deeper depths** is explored. For example, we see to what extent the distribution of model rankings shift with depth, either for the better (+) or worse (-)

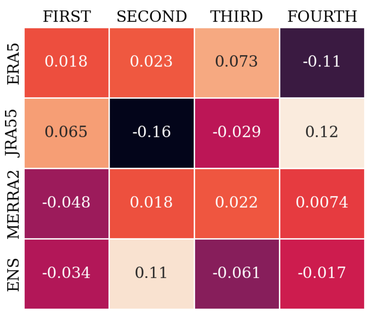


Figure 6: Difference between model performance at the surface and deeper depths.

RESULTS / DISCUSSION

- ACCOMATIC can be used to communicate model performance at various spatiotemporal scales, for instance:
  - *ENS + GEOtop* performed the **best**, ranking 1st 56% of the time, while *JRA55 + GEOtop* was a close 2nd
  - *ERA5 + GEOtop* performed the **worst**, ranking last 96% of the time
- These results can be subset by season or terrain, to better identify where a model is struggling.

FUTURE WORK

- Next, this method could be **tailored to other variables** of interest as ACCOMATIC is currently specific to ground surface temperature.
- Applying this method **using different modelling** software (e.g. *CLASSIC*, *FreeThaw1D*)
- Incorporating this method seamlessly into a comprehensive simulation workflow.

COLLABORATORS / ACKNOWLEDGEMENTS

Northwest Territories
Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

GTPEM Development / Technical Support: Brown, N., Cao, B.
Committee Members: Gruber, S., Kokelj, S., Lantz, T., Melton, J.

REFERENCES

1. Heginbottom, J. A. (1995). Canada-permafrost. *National Atlas of Canada*.
2. Rigon, R., Bertoldi, G., & Over, T. M. (2006). GEOtop: A distributed hydrological model with coupled water and energy budgets. *Journal of Hydrometeorology*, *7*(3), 371-388.
3. Padilla-Ramirez, L. (2020). *Evaluating the effect of accordance method choice on the selection of ground surface temperature models in a subarctic environment.* [Master's Dissertation, Carleton University]. Carleton University Research Virtual Environment.
4. Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The american statistician*, *69*(4), 371-386.

CONTACT INFO

Hannah Macdonell, M.Sc. D.S. Candidate. Geography Department, Carleton University
Email: Hannah.Macdonell@CARLETON.ca

Dr. Stephan Gruber, Professor & CRC. Geography Department, Carleton University
Email: StephanGruber@CUNET.CARLETON.CA