# STATISTICAL ANALYSIS OF GROUND SURFACE TEMPERATURE SIMULATIONS IN THE NORTHWEST TERRITORIES TUNDRA.

PermafrostNet
NSERC | CRSNG

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

Carleton University

## INRODUCTION / BACKGROUND

Permafrost modelling can contribute to informing adaptation in permafrost regions by characterizing the subsurface thermal regime at different points in time. However, as models vary in their representation of physical phenomena, they also differ in performance at each location. This can make it difficult to make a justifiable comparison of two simulation products, or to distinguish improvement in 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* — Ground surface temperature simulations are produced using the modelling software GEOtop[2] forced with JRA55, MERRA-2, and ERA5 reanalysis data.

*OBSERVATIONS* — Observational ground temperature data from the NWT is collected from Carleton permafrost database (COLDASS) and NSERC PermafrostNet ERDDAP.

*ACCOMATIC* — The python package used to partition simulation and observational datasets and produce a suite of summary statistics used to generate model rankings is called **accomatic**[3]. Each simulation will be tested against a range of accordance measures, then split by season and terrain type.

## STUDY SITES

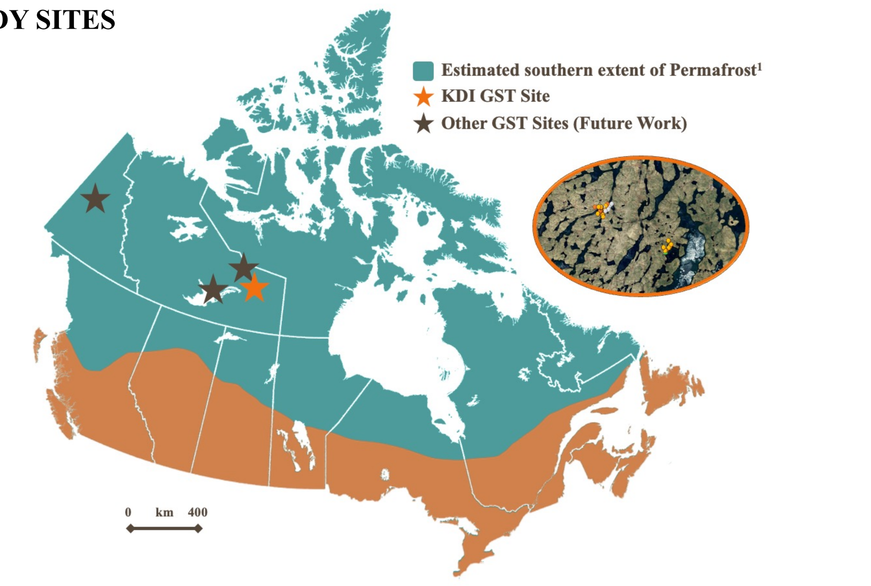


**Figure 1**: Ground Surface Temperature site clusters across Canada

## SIMULATED VARIABLE: GST

Ground surface temperature is measured roughly 10 cm below the ground surface. GST is inexpensive to measure relative to other permafrost variables while remaining highly representative of the underlying thermal regime. Characteristics to represent when modelling include:

(1) Topography
(2) Ground type *i.e.* subsurface materials
(3) Surface vegetation


**Mini loggers** that are used to measure GST.

## *t* - INTERVAL BOOTSTRAPPING

Observational datasets that can be used to test simulations are often spatially sparse and incomplete. To make use of incomplete datasets, we implement a *t*-interval bootstrap approach[4]. Fig 3 shows a visualization of how *n* windows of *t* days are randomly selected.

The bootstrap test provides a confidence interval around each mean accordance measure, showing that rankings between models can overlap.

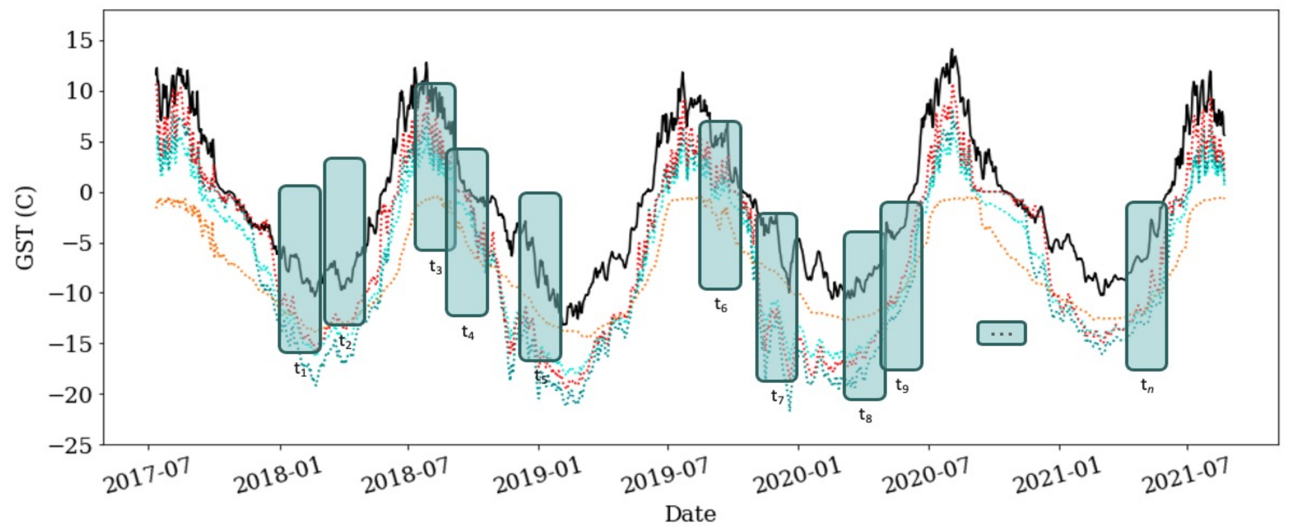


**Figure 3**: Daily mean GST data for the KDI cluster ($n$=23) with *t*-interval bootstrapping visualization superimposed in blue rectangles.

## TESTING CONDITIONS

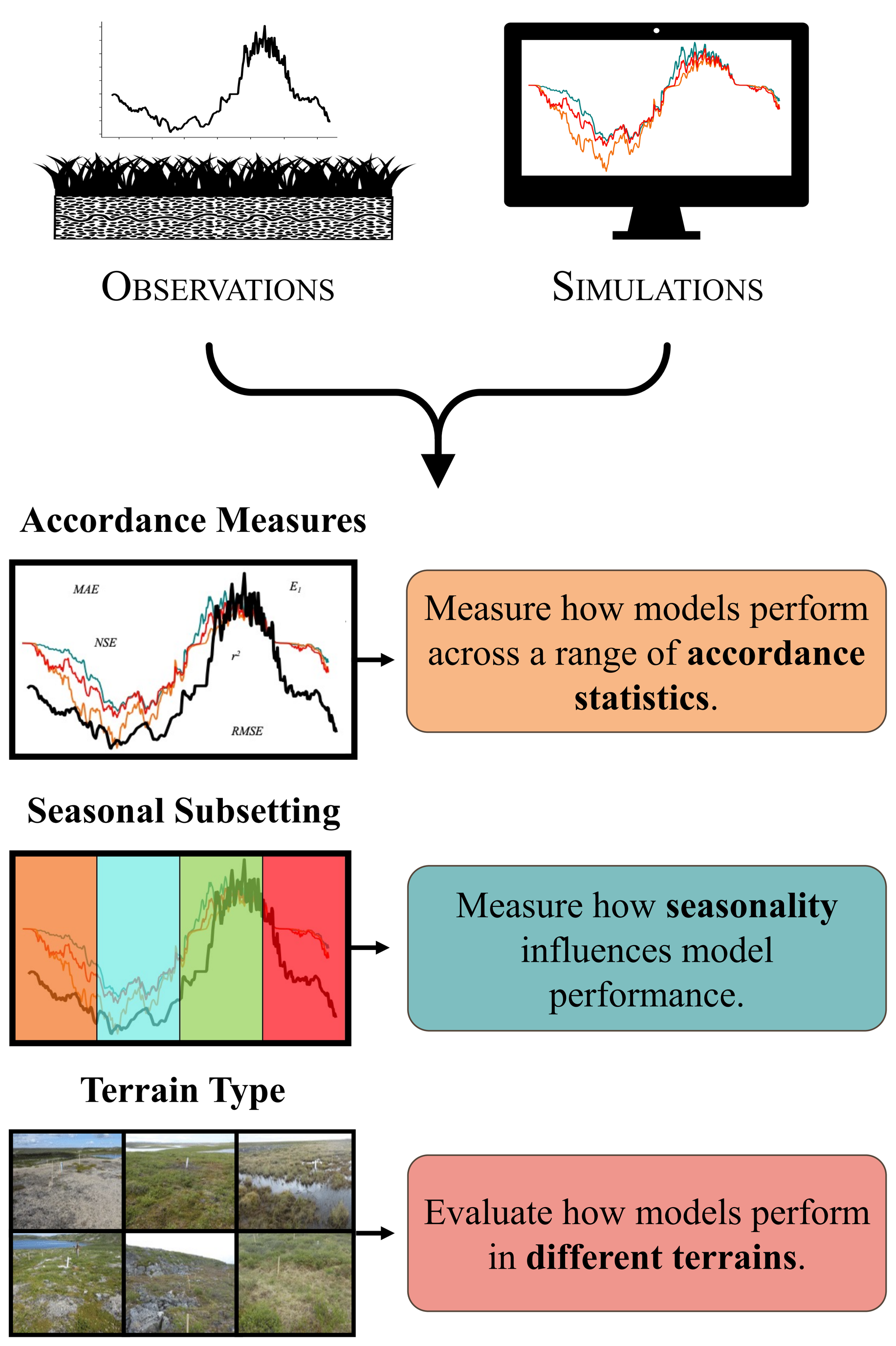


**Figure 2**: Visualization of testing conditions for GST simulation ranking, including a variety of accordance measures, seasons and terrain subsetting.

## RESULTS: SIMULATION PERFORMANCE ACROSS TESTING CONDITIONS

### Accordance Measures

Figure 4 below shows the results of three different accordance measures being used to evaluate simulation performance. While the *JRA-55* simulation performs best across each accordance measure shown here (*RMSE*, $R^2$, *BIAS*), it ranges in performance considerably, overlapping with the worst performing simulation (*ERA5*).

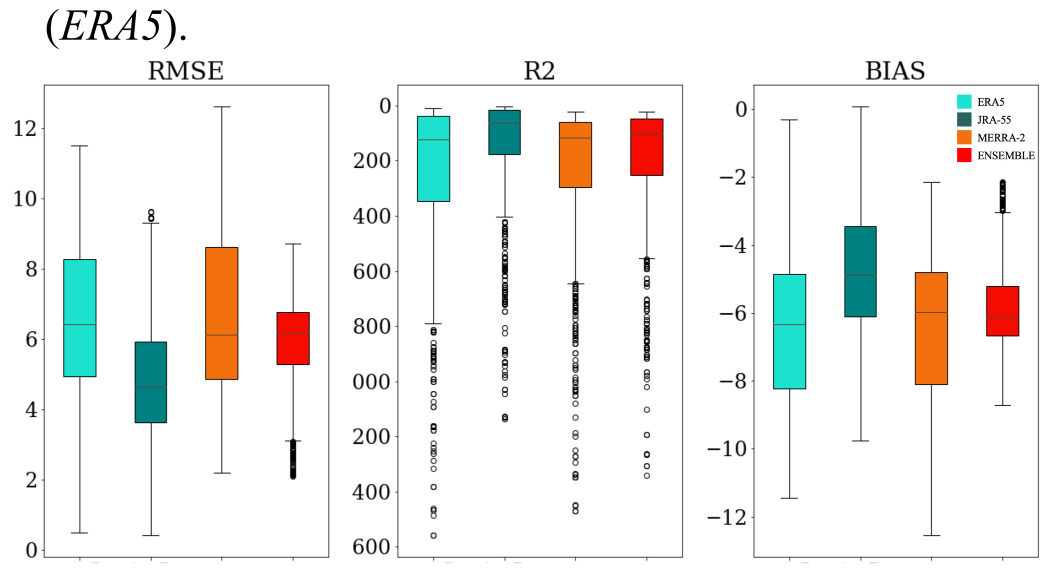


**Figure 4**: *t*-interval bootstrapping results for *RMSE*, $R^2$ and *BIAS* accordance measures.

While *RMSE* and *BIAS* show interpretable results, the t-interval bootstrap approach does not seem effective at capturing correlation (*R2* in Fig 4).

### Seasonal Subsetting

Figure 5 shows *RMSE* bootstrap results with a 0.95 confidence interval shown around each *RMSE* mean. While the *JRA-55* model performs best over all (Fig 4), it has a greater *RMSE* value than the *MERRA-2* simulation in Winter and Spring.

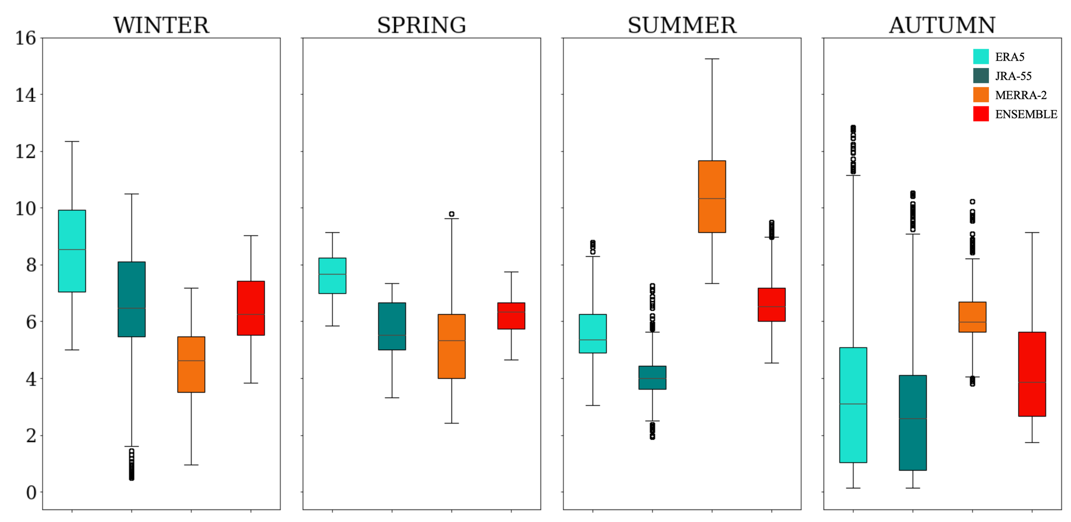


**Figure 5**: Seasonal *t*-interval bootstrapping results using the *RMSE* metric for four different simulations.

Additionally, though the *MERRA-2* model is ranked second in Fig 4, here in Fig 5 we see that it has a large *RMSE* of 10.1 in the Summer.

### Terrain Type

When simulations are evaluated by Terrain type, we can evaluate the strengths and weaknesses of model parameterization at these terrains of interest.

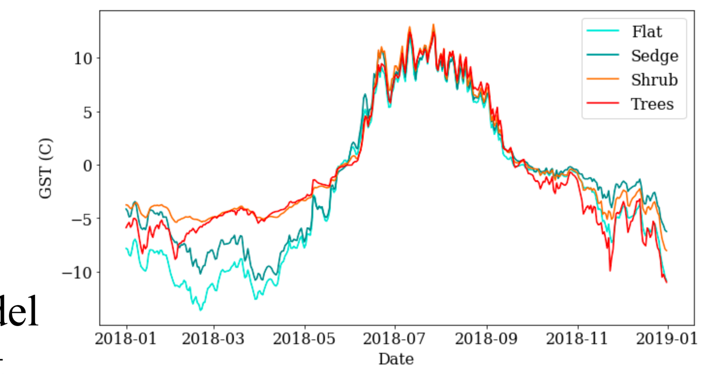


**Figure 6**: Daily mean GST for four proposed terrain types in the KDI region.

Fig 7 shows a how *RMSE* performance varies for each model across different categories of terrain: 'Flat' here describes a short-vegetation flat tundra terrain.

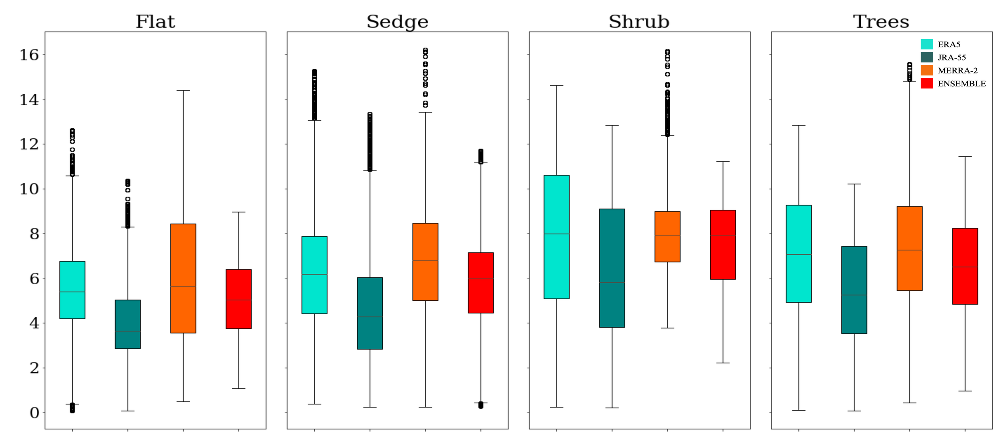


**Figure 7**: Terrain-type subsetting *t*-interval bootstrapping results using the *RMSE* metric for four different simulations.

## FUTURE WORK

This poster summarizes the findings of only the first iteration of using *accomatic* to evaluate model simulations and uses only a small subset of GST data. Future work includes:

1. Larger amount of GST data to allow for meaningful terrain type analysis (Fig 1)
2. More rigorous parameterization of individual sites in GEOtop.
3. Addition of CLASSIC model, driven by all three reanalysis datasets.
4. More in depth description of terrain type subsetting and classification metrics.
5. Additional analysis of seasonality (How do we define a season?)

## 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