AI-powered ground surface temperature forecasting for cold regions geotechnical applications

GeoCalgary 2022 October 2-5
Reflection on Resources

Ali Fatolahzadeh Gheysari
Department of Civil Engineering – University of Manitoba, Winnipeg, Manitoba, Canada
Pooneh Maghoul
Civil, Geological and Mining Eng. Department – Polytechnique Montreal, Montreal, Quebec, Canada
Ahmed Ashraf
Department of Electrical and Computer Engineering – University of Manitoba, Winnipeg, Manitoba, Canada
Ahmed Shalaby
Department of Civil Engineering – University of Manitoba, Winnipeg, Manitoba, Canada
Kshama Roy
DNV Canada Ltd., Calgary, Alberta, Canada

ABSTRACT
Ground surface temperature is an essential variable in cold region geotechnical engineering. Physics-based long-term simulations of surface energy budgets are associated with complexity, high variance, and computational intensity. This study proposes an alternative data-informed framework based on long short-term memory (LSTM) networks to predict ground surface temperatures from meteorological variables. The LSTM model was evaluated using monitoring data from a permafrost site in the Canadian Arctic and a mid-latitude non-permafrost site. Various aspects of the machine learning problem were studied using a series of sensitivity analyses. Long-term projections of ground surface temperature were presented for the two sites under both moderate and extreme climate change scenarios. Data scarcity was found to be one of the major challenges for the proposed framework. However, the growing number of stations and more reliable instrumentation will be in favor of data-driven methods. Provided that suitable training data are available, the data-driven framework shows several advantages over the physics-based simulations in forecasting ground surface temperature, and potentially other related variables.

RÉSUMÉ
La température à la surface du sol est une variable essentielle dans la géotechnique des régions froides. Les simulations à long terme basées sur des bilans d'énergie de la surface sont associées à une complexité, une variance élevée et une intensité de calcul. Cette étude propose un cadre alternatif basé sur des réseaux de mémoire longue à court terme (LSTM) pour prédire les températures à la surface du sol à partir de variables météorologiques. Le modèle LSTM a été évalué à l'aide de données de monitoring provenant d'un site de pergélisol dans l'Arctique canadien et d'un site sans pergélisol de latitude moyenne. Divers aspects du problème d'apprentissage automatique ont été étudiés à l'aide d'une série d'analyses de sensibilité. Des projections à long terme de la température à la surface du sol ont été présentées pour les deux sites selon des scénarios de changements climatiques modérés et extrêmes. La rareté des données s'est avérée être l'un des principaux défis du cadre proposé. Cependant, le nombre croissant de stations et une instrumentation plus fiable favoriseront les méthodes basées sur les données. À condition que des données de training appropriées soient disponibles, le cadre axé sur les données présente plusieurs avantages par rapport aux simulations basées sur la physique dans la prévision de la température à la surface du sol et potentiellement d'autres variables connexes.

1 INTRODUCTION

Information on ground temperature profiles is essential in many geotechnical engineering projects, particularly in cold regions. The annual variation of the ground temperature profile plays an important role in choosing the minimum burial depth for utilities to avoid freezing. Besides, it affects the thermal performance of shallow geothermal systems and is an indicator of potential frost heaving and thaw weakening, the detrimental phenomena that can affect roads, highways, railways, pipelines, etc. Furthermore, the stability of permafrost, permanently frozen ground covering the arctic and subarctic mainland, is also governed by ground temperature.

Climate change is a major threat of the century and will affect communities and infrastructure (IPCC, 2021). The costs due to the adverse effects of climate change in Canada are estimated to be between $21 billion and $43 billion per year, and nearly 70% of the infrastructure in the arctic zone will be affected by permafrost degradation by 2050 (Hjort et al., 2022; NRT, 2011). Therefore, determining the ground temperature profile and predicting its future trends during the service life of construction projects are essential steps in a sustainable and climate-smart geotechnical design.

The energy budget at the ground surface, known as the surface energy budget (SEB), is an interconnected system of energy fluxes and components. It results in variations of temperature in the ground and at its surface, temporally and spatially. The ground temperature profile is essential in cold region geotechnics. The extreme variations of ground temperature happen at the ground surface since it is in contact with the atmosphere and fluctuating climatic forcings. Ground Surface Temperature (GST) is often used in the assessment of freeze-thaw-induced engineering challenges, e.g., frost heaving, snow removal, ice control, and the operation of winter roads. In some civil engineering simulations, SEB can be equivalently modeled by a GST boundary condition, which reduces the number of model inputs and computation load, and makes the simulations feasible over a long-term study period (Kong et al., 2019). Mean annual ground surface temperature (MAGST) is a widely-used indicator in climate and permafrost studies (Smith & Riseborough, 2002). An increase in MAGST implies a thermal imbalance, which results in changes in the permafrost regime. Therefore, reliable forecasts on GST and its derivatives are highly prioritized in many sectors and can be remarkable assets in attaining climate resiliency.

The GST is often calculated using physics-based methods, i.e., solving SEB through analytical or numerical techniques. An analysis of SEB requires the inclusion of its involved components, such as air convection heat fluxes at the ground surface, solar radiation, and the insulation effects due to snow cover (Liu et al., 2019). The inputs of the model, for instance, air temperature and snow depth, can be obtained from weather station records to calculate either present or past GST. Nevertheless, long-term physics-based forecasting of GST requires inputs that reflect future climate trends. Climate models provide projections of meteorological variables, which can be used as boundary conditions in SEB analyses to calculate GST under different climate pathways. Note that some global climate models, such as Canadian Earth System Model (CanESM), predict temperatures at the ground surface, yet at the time of this study, the spatial resolution of the model is 2.8° (more than 300 km) (Swart et al., 2019), which is deficient in many applications.

Although the physics-based analysis of SEB is the most realistic approach to calculating ground temperature, it is subjected to challenges that limit its use in practice. SEB involves a multitude of components, physics, as well as site conditions, which require various inputs and detailed modeling of the land-atmosphere energy system. Omitting one component in the model may significantly affect the results. Moreover, the multiphysics simulation of SEB is computationally intensive. While the approach can be used to calculate the ground temperature at one site, simulating a geospatial mesh grid at small temporal increments, such as daily, may become infeasible over multidecade study periods. Due to these limitations, the GST is sometimes estimated from meteorological variables without conducting a SEB analysis. For example, the n-factor method can estimate the mean seasonal GST from air temperature. The n-factor is calculated from the number of days with freezing (or above-zero) temperatures in the soil and the air. The method has been used in many permafrost studies (Klene et al., 2001; Riseborough, 2003). However, it has several shortcomings. For example, it estimates GST from a single parameter, air temperature, while other components of SEB can severely influence ground temperature. Besides, interannual fluctuations of air temperature and long-term climate changes alter the freezing and thawing indices. Therefore, the n-factor is subjected to change, and the method is not suitable for GST forecasting.

Data-driven forecasting methods benefit from recent advancements in the management of large datasets and improved machine learning techniques. With the increasing trend of collecting and storing data, it is now possible to analyze the data for correlations, patterns, and trends. The regression approach, using artificial neural networks (ANN), has previously been used to estimate the GST from air temperature and other meteorological measurements (Tabari et al., 2011). However, temperature below the ground surface follows air temperature with a lag due to ground thermal mass (Beltrami, 2001; Gilpin & Wong, 1976; Saaly et al., 2020). The GST is also affected by the past meteorological conditions as it is a resultant of above and below-surface heat fluxes, such as solar radiation and air convection at the ground surface. This has partially been addressed by including the past meteorological conditions into the input features of the ANN. In other words, each entry of the training and test input sets can be a past sequence of input parameters (Gheysari et al., 2021). While this approach significantly improved the estimation accuracy, ANNs regard the data as X-Y points and not as a time series. Besides, efforts have been made in the past to forecast the GST using linear stochastic methods, e.g., autoregressive integrated moving average (ARIMA) (Zeynoddin et al., 2019). However, this approach analyzes the past trends only in soil temperature time series and omits the governing SEB components. Therefore, it is not suitable for long-term GST forecasting.

Recurrent neural networks (RNN) are a family of neural networks that are exclusively developed to capture temporal dynamic behavior and process time-series data. An RNN has a chain-like structure of repeating cells along a temporal sequence, in which the output from each step can affect the output of the next steps. This enables RNNs to store information and process inputs in a sequential format. Nevertheless, backpropagation through time in RNNs may result in vanishing or exploding gradients when the input sequence is relatively long. Long short-term memory (LSTM), gated recurrent unit (GRU), and their variants are specialized types of RNNs that mitigate the vanishing and exploding gradient problem. Therefore, LSTM and GRU can process long sequences (Hochreiter & Schmidhuber, 1997). LSTMs have been used in climate and earth science studies in the past, e.g., forecasting rainfall, sea surface temperature, and reconstruction of missing groundwater level data (Kratzert et al., 2018; Vu et al., 2021; Yang et al., 2018). The application of LSTM to estimate the GST from meteorological forcings has recently been discussed in the literature (Li et al., 2020). These studies, however, used a two-year training dataset, which may not be able to capture the interannual variability of weather. The accuracy of the estimations throughout the year, e.g., in different seasons and at annual peaks, is also not known. Therefore, providing long-term forecasts on the GST remains a challenge and needs further investigation.

The present study proposes a data-informed LSTM-based framework for long-term forecasting of the ground surface temperature from meteorological variables. The performance of the framework is evaluated for two sites in the Canadian Arctic as well as mid-latitudes. A series of models with different properties are trained and evaluated to understand the sensitivity to each parameter and as an attempt to reduce prediction errors. Projections of ground surface temperature are then presented for the two sites under both moderate and extreme climate change scenarios.

2 METHODOLOGY

The proposed data-driven forecasting framework consists of three main stages: training, test, and projection (Figure 1). First, an LSTM network is trained using a set of labeled data. In this study, the dataset includes meteorological variables (e.g., air temperature) and ground surface temperature, which were measured at the site. The data is divided into two subsets for training and test. A data loader feeds the training data into the LSTM in the form of sequences. In other words, each entry of the training/test/projection data sets consists of a label (GST) associated with a sequence of meteorological data. The model is then trained to detect patterns and correlations between GST and meteorological variables. A forward pass is subsequently performed using the test data. Comparing the measured and estimated GST for the test set, the performance of the model is evaluated. Finally, time series of meteorological variables are extracted from the outputs of climate models under different climate pathways. The projections are then fed to the trained model to forecast the GST at each climate pathway.

While this study uses the classic LSTM architecture, other variants of LSTM, such as LSTM with peephole connections and GRU, can also be similarly integrated into the forecasting framework (Britz et al., 2017; Chung et al., 2014).

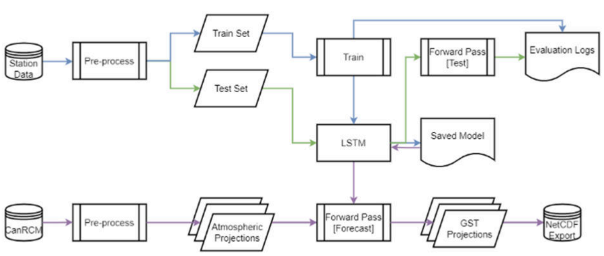

Figure 1. The proposed LSTM forecasting framework

2.1 LSTM theory

LSTMs, like other recurrent neural networks, are formed as chains of repeating modules. While a repeating block of RNNs can be as simple as a single layer, each repeating module of LSTM usually consists of additional interacting components: a cell and a forget gate, an input (update) gate, and an output gate, which control the flow of information into and out of the cell (Figure 2). The cell state is a means of storing information and accounts for the long-term memory of the network. The forget gate decides what information needs to be removed from the cell state. The input gate decides whether the cell state must be updated, and the output gate regulates the information passed to the next hidden state.

In a forward pass of a classic LSTM, the forget gate first inspects the input data, the output of the previous cell and decides whether the information should be kept or ignored:

$$\mathrm{f_t} = \sigma(\mathrm{W_f} \cdot [\mathrm{h_{t-1}}, \mathrm{x_t}] + \mathrm{b_f}) \quad [1]$$

where $\mathrm{f_t}$ is the forget gate activation vector, $\sigma()$ is the sigmoid function, x_t is the input vector, h_{t-1} is the hidden state vector of the previous cell, and $\mathrm{W_f}$ and $\mathrm{b_f}$ are the weight matrices and bias vector of the forget layer, respectively, which are learnt during training. The update gate decides which information needs to be added to the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad [2]$$

in which i_t is the input/update gate activation vector, and W_i and b_i are weight matrices and the bias vector of the update gate, respectively. A cell input candidate vector is then created as:

$$\widetilde{c_t} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad [3]$$

where $\widetilde{c_t}$ is the vector of new candidate values, $tanh()$ is the hyperbolic tangent function, and W_c and b_c are the weight matrices and bias vector, respectively. Having the forget and update activation vectors f_t and i_t, the updated candidates $\widetilde{c_t}$ and the previous cell's state c_{t-1}, the current cell state c_t is updated as:

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c_t} \quad [4]$$

Finally, the cell output h_t is calculated from the cell state c_t and the activation vector of the output gate o_t:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad [5]$$

$$h_t = o_t * \tanh(c_t) \quad [6]$$

where W_o and b_o are the weight matrices and bias vectors of the output layer, respectively, which are learned during the training.

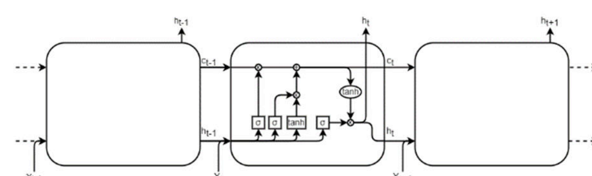

Figure 2. Repeating modules in a classic LSTM

2.2 Training and test data

The availability of suitable training data is one of the major challenges in machine learning practices. LSTM, and generally all types of RNN, require long uninterrupted training data as they process. In the case of meteorological data, missing data can be a common scenario, especially in remote stations or where maintenance is only feasible during a limited timeframe of a year. While short periods of missing values may be filled via interpolation, large gaps in a dataset may render it unusable for model training.

Having multiple training features, i.e., meteorological variables, improves model accuracy by capturing more components of SEB. However, choosing a training dataset is often a trade-off between temporal size and the number of features. Multiple features will increase the chance of missing data and narrows the criteria in the search for suitable datasets.

Some of the longest continuous measurements of ground temperature in northern Canada have been conducted by the Centre d'études Nordiques (CEN). Their data are published on the Nordicana D repository (CEN, 2021). Two sites, representing continuous arctic permafrost and mid-latitude (non-permafrost), are chosen for this study. The permafrost site is in the Northern Ellesmere Island in Nunavut, Canada (83.09N---74.13W). The dataset, named Nordicana D1, contains hourly measurements of air temperature, the ground temperature at different depths, and snow depth (CEN, 2020a). The simultaneous records of all channels are available from August 2005 to July 2019. There is a one-year gap in the GST channel. Nevertheless, the gap divides the data into two subsets of 3637 and 1096 daily averages in an approximately 75:25 proportion, which can be considered as the training and test sets, respectively.

The non-permafrost site is in the Grands-Jardins Park in Charlevoix, Quebec, Canada (47.68°N—70.85°W). The dataset, named Nordicana D6, includes hourly measurements of air temperature, the ground temperature at different depths, and snow depth from September 2008 to September 2016 (CEN, 2020b), 2928 days in total. However, only a portion of the dataset, from 2008 to 2012, contain snow depth measurements. Therefore, only one feature (air temperature) was used for model training. Note that other datasets with multiple meteorological variables were found. However, they either contained large gaps or were short at the time of the study.

The projections of meteorological parameters are taken from the outputs of Canadian regional climate model 4 (CanRCM4) as daily time series. The projection datasets were extracted under different emission scenarios so that the forward pass on the trained LSTM would forecast the GST in each case. This study uses the fifth coupled model intercomparison project's (CMIP5) representative concentration pathways (RCP), which are widely used in climate impact studies. In this regard, the medium stabilization and high-emission pathways, RCP 4.5 and RCP 8.5, were chosen, respectively. Other climate scenarios can be used for forecasts in a similar manner. For instance, CMIP6 climate scenarios, known as shared socioeconomic pathways (SSP) (O'Neill et al., 2014), can be used so that the model can forecast GST with respect to long-term socioeconomic trends and policies.

2.3 Model definition

The input and output dimensions of the LSTM are bound to the number of training features (e.g., air temperature) and the dependent variable (GST), respectively. However, additional properties of the model, such as the number of hidden units (H) and layers ($\mathrm{N_L}$), can affect the model's performance and therefore require a sensitivity analysis. In terms of input data, increasing the number of features ($\mathrm{N_f}$), e.g., including snow depth into the model, and the size of the training data ($\mathrm{N_t}$) can affect the model accuracy. Therefore, a sensitivity analysis is required to study the effect of these input parameters on the GST since temporally long uninterrupted measurements of GST and multivariate meteorological data are often limited.

The length of the input sequence (S) (Figure 3) determines how long the model will investigate the past in its memory. The memory of the LSTM is selective. In other words, even if a long sequence length is chosen, the model can selectively forget the excessive part that does not contribute to its performance. Nevertheless, a sensitivity analysis of sequence length can reveal how the past weather can affect ground temperature. It is expected that the introduction of a sequence lag (L) (Figure 3) will decrease the model performance since it omits the immediate past from being used in model training. However, a sensitivity analysis on sequence lag might give some insights into the temporal correlation between GST and meteorological variables.

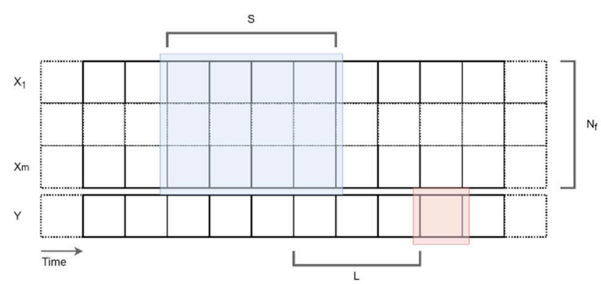

Figure 3. LSTM input data structure

3 RESULTS

The LSTM models of this study are developed using pyTorch, an open-source machine learning library in Python (Paszke et al., 2019). A manual random seed was used with the purpose of reproducibility of the results. To assess and compare the models quantitatively, the results are discussed in terms of root mean square error (RMSE).

3.1 Evaluation and cross-validation

Both predicted and actual (field measured) ground surface temperatures are plotted over the test set for both sites (Figures 4 and 5). The predictions generally conform with field measurements. For the Nordicana D1 site, where GST has an annual variation of about 50 °C, the model, with (M=2) and without (M=1) inclusion of snow cover, slightly underpredicted the peak summertime GST. On the other hand, the evaluation of the model at the D6 site shows closer conformity with the actual measurements.

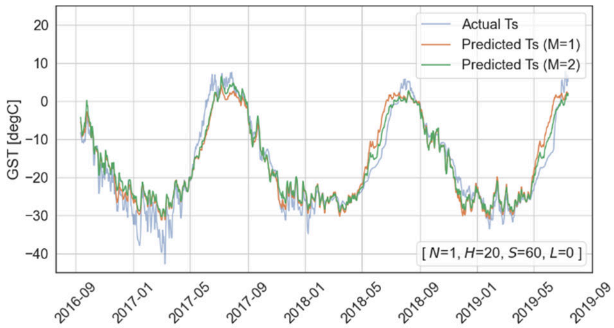

Figure 4. Model evaluation - Nordicana D1 dataset

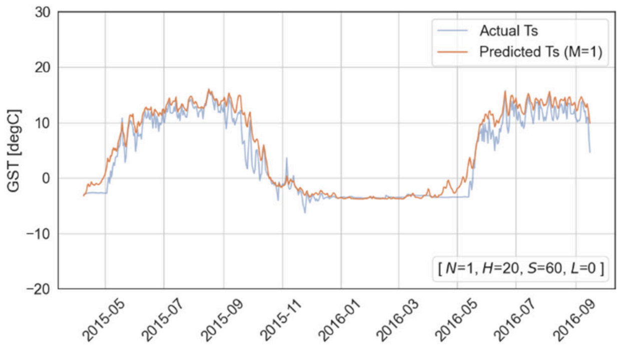

Figure 5. Model evaluation - Nordicana D6 dataset

3.2 Sensitivity analyses

A comparison of models trained by the D1 dataset with and without the inclusion of snow depth at various model settings (Figures 6 to 9) indicates a decrease in prediction error when both air temperature and snow depth are included as training features. This is because snow cover acts as an insulation layer and affects surface energy balance. However, as previously explained, while the inclusion of additional features improves the performance of the model, it requires uninterrupted time series of each feature, which is not feasible at many locations due to limited datasets. LSTM, like any other supervised learning approach, requires labeled data for training and evaluation. Therefore, the training and test datasets should contain the same meteorological variables (features) and ground surface temperature (label). Besides, to perform a forward pass to forecast long-term GST, the projection data should contain the same features as the training and test sets. Many downscaled results of climate prediction models contain only a limited number of meteorological variables. Therefore, while the inclusion of multiple features is suggested, it is often bound to available data.

A sensitivity analysis was performed by training models on input sequences (S) ranging from one day to 120 days (Figure 6). The models with a shorter length of input sequence resulted in high errors, while the models with longer input sequences had lower errors, which did not change past S=60 days. It implies that the very recent input data significantly affects how the model performs. In other words, the immediate past has more weight on the model memory. Although long input sequences can be used in a conservative manner, they significantly increase the training time, and therefore may not be suitable when computation time is of concern.

The inclusion of sequence lag (L) caused an increase in the prediction error of all models (Figure 7). The optimum prediction error is attained when L is zero. This is because the introduction of the lag disregards the input entries of the immediate past, which were shown to be of a significant influence. Note that the introduction of sequence lag may imply a short-term forecasting scheme. In other words, it may be inferred that when the model is trained with $\mathrm{L}=\mathrm{L_i}$ and $S=S_i$, the current $\mathrm{S_i}$-day time history of meteorological variables can be fed to the model, to predict the GST in the $\mathrm{L_i}$th day in the future. However, the future GST is governed by future meteorological conditions. The scheme, therefore, may not be suitable for short-term forecasting as it estimates the future GST from the past and present weather conditions and does not incorporate future meteorological conditions.

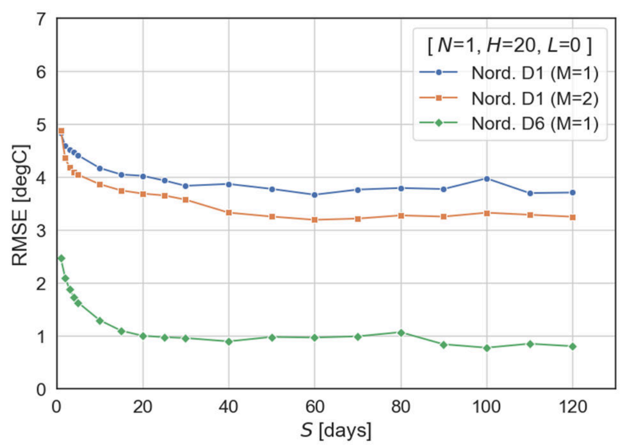

Figure 6. Prediction error vs. length of the input sequence

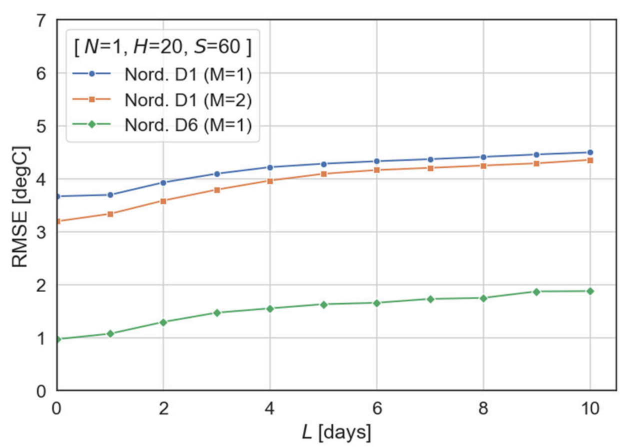

Figure 7. Prediction error vs. lag in the input sequence

The proposed approach was applied to smaller subsets of D1 and D6 datasets to assess the impact of the size of the training set on the prediction error. The results indicate smaller prediction errors when the models were trained with larger subsets of the data, yet the trends do not stabilize even when the entire training data were used

(Figure 8). Here the analysis was limited to the available monitoring data from the two sites. Larger training data are required to determine the optimum N_t and detect overfitting. For the model parameters, performing a sensitivity analysis on the number of hidden cells did not reveal any impact on forecast errors (Figure 9) or the training time.

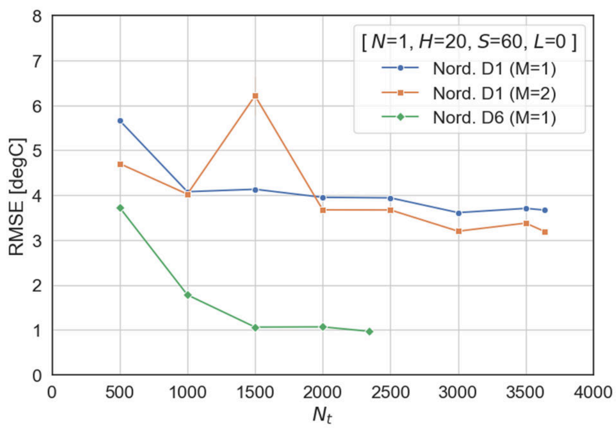

Figure 8. Prediction error vs. size of training set

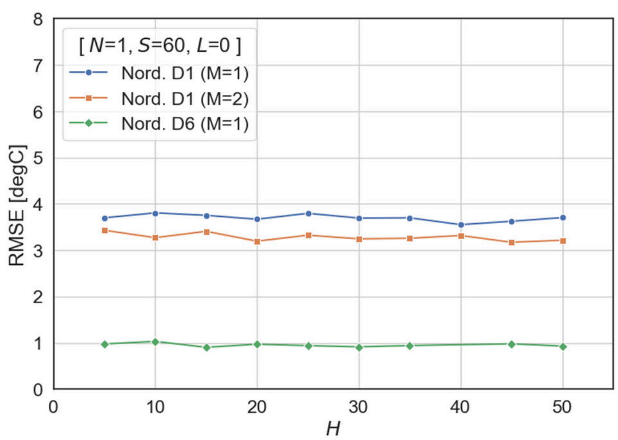

Figure 9. Prediction error vs. number of hidden cells

3.3 Long-term forecasts

Based on the sensitivity analyses, the models that resulted in the lowest errors over the test data were chosen for the long-term forecasts. Projections of daily air temperatures at the two sites were obtained from the outputs of CanRCM4, having a 0.22° horizontal grid resolution, equivalent to 25 km. Through bilinear interpolation, time series at each site were extracted for RCP4.5 and RCP8.5 climate scenarios. The projections of air temperature were then normalized and fed to the LSTM models to forecast daily GST under the two scenarios.

The predicted ground surface temperatures are presented as annual averages for Ellesmere Island and Grands-Jardins Park (Figures 10 and 11) until 2100. The results reflect the different climates at the two sites, with lower MAGST at the Ellesmere Island station. The forecasts under RCP4.5 and RCP8.5 scenarios diverge after 2050, with RCP8.5 causing a higher rise in MAGST by the end of the century. The results also show a higher increase in MAGST at the arctic site under both scenarios.

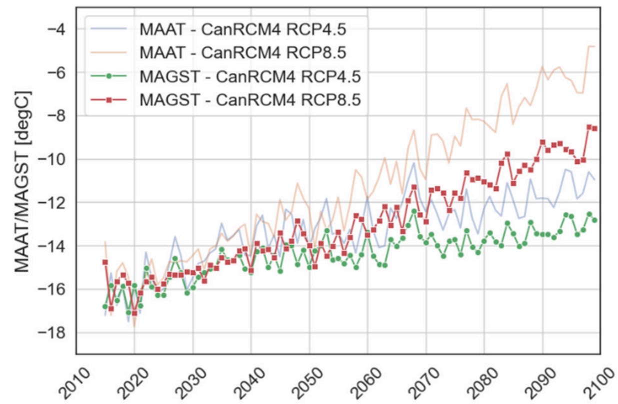

Figure 10. MAGST Projections - Ellesmere Island, NU

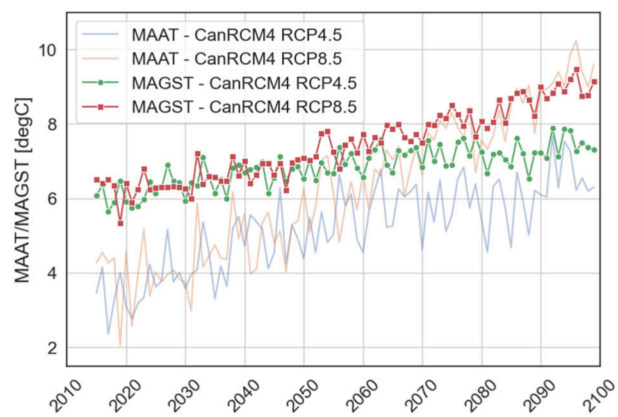

Figure 11. MAGST Projections - Grands-Jardins Park, QC

4 DISCUSSION

The data-driven approach, if provided with a robust training dataset, can have several advantages over physics-based simulations. Once trained, the machine learning model can predict over any projection set, i.e., various climate scenarios, through a forward pass, without any alteration to the trained model, while any change to the boundary conditions in a physics-based simulation requires the model to be run again. Compared to physics-based numerical simulations, the computational intensity of a forward pass over a trained model is almost nil. Therefore, the data-driven approach can easily generate predictions over very long periods in various climate scenarios or an ensemble. Here, only two CMIP5 climate scenarios were used for demonstration. However, the framework can simply use the CMIP6 climate scenarios, which include varying assumptions about human development, to forecast long-term ground temperatures under different socioeconomic pathways.

Despite the advantages, data-driven models, especially if trained with small datasets with a few features, may suffer from significant prediction bias. Numerical simulations, on the other hand, are more prone to variance. For example, thermal simulation of the ground surface may result in extremely high or low temperatures if one component of the surface energy budget is not omitted in the model. Data-driven models do not experience this issue as their predictions are usually within or close to the range of training data. Therefore, once their bias is within the acceptable tolerance, the low variance of data-driven methods may be considered an advantage.

5 CONCLUSION

A data-driven forecasting framework was proposed, which can forecast ground surface temperature by learning from past ground and air temperatures as well as other meteorological measurements. It can flexibly incorporate various components of the surface energy budget if they have been previously measured at the site. The proposed framework can predict the ground surface temperature in various climate change scenarios, corresponding to the projections of meteorological variables provided to the model, eliminating extra computation effort. The framework also inherently reflects the underlying drives of climate scenarios into the predicted ground surface temperature.

Similar to any data-driven method, the availability of data is a major challenge for the proposed framework. Long uninterrupted measurements of ground surface and other meteorological variables are still very limited. However, the increasing number of stations, development of more reliable instrumentation, and new remote sensing technologies will be in favor of data-driven methods. For example, in Canada, many agricultural weather stations in the prairies began recording soil temperature and meteorological variables since the 2010s and soon can provide datasets that are sufficient for data-driven forecasts. Besides, with the recent innovations in land data assimilation technology, reanalysis data products can be used to address the data scarcity and to train data-informed prediction models.

The proposed framework intrinsically inherits the uncertainties associated with climate prediction models, as it predicts ground surface temperature from the projections of atmospheric variables. Nevertheless, the alternative approach, i.e., physics-based simulations of the surface energy budget, also requires meteorological projections as boundary conditions and hence inherits the same uncertainties. If suitable data is provided, data-driven models have several advantages over numerical simulations in terms of speed, low variance, and scalability. Therefore, the proposed framework can be advantageous where many forecasts are required, such as statistical analyses, ensemble forecasting, and making predictions over large areas.

Although the proposed framework focuses on the long-term prediction of ground surface temperature, the same methodology can be applied to forecast other parameters if past observations are available for model training. Some potential examples are ground temperature profile and soil moisture. In short, forecasting ground surface temperature using data-driven methods, despite some limitations, can be considered an advantageous alternative to physics-based methods as it addresses the drawbacks that are associated with numerical simulations and can facilitate climate impact assessments on northern Canada's infrastructure.

ACKNOWLEDGEMENT

This research was supported by the New Frontiers in Research Fund – Exploration Grant [NFRF-2018-00966] and the Mitacs E-Accelerate program.

REFERENCES

Beltrami, H. (2001). On the relationship between ground temperature histories and meteorological records: a report on the Pomquet station. *Global and Planetary Change*, *29*(3), 327–348. https://doi.org/https://doi.org/10.1016/S0921-8181(01)00098-4

Britz, D., Goldie, A., Luong, M.-T., & Le, Q. (2017). Massive Exploration of Neural Machine Translation Architectures. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1442–1451. https://doi.org/10.18653/v1/D17-1151

CEN. (2020a). *Climate station data from Northern Ellesmere Island in Nunavut, Canada, v. 1.7 (2002-2019). Nordicana D1*. https://doi.org/10.5885/44985SL-8F203FD3ACCD4138

CEN. (2020b). *Climate station data from the Grands-Jardins park in Charlevoix, Quebec, Canada, v. 1.3 (2004-2019). Nordicana D6*. https://doi.org/10.5885/45156SL-0D3A7DFC670946AD

CEN. (2021). *The Nordicana D collection*. http://www.cen.ulaval.ca/nordicanad/en_index.aspx

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv Preprint ArXiv:1412.3555*.

Gheysari, A. F., Holländer, H. M., Maghoul, P., & Shalaby, A. (2021). Sustainability, climate resiliency, and mitigation capacity of geothermal heat pump systems in cold regions. *Geothermics*, *91*(September 2020). https://doi.org/10.1016/j.geothermics.2020.101979

Gilpin, R. R., & Wong, B. K. (1976). "Heat-Valve" Effects in the Ground Thermal Regime. *Journal of Heat Transfer*, *98*(4), 537–542. https://doi.org/10.1115/1.3450597

Hjort, J., Streletskiy, D., Doré, G., Wu, Q., Bjella, K., & Luoto, M. (2022). Impacts of permafrost degradation on infrastructure. *Nature Reviews Earth & Environment*, *3*(1), 24–38. https://doi.org/10.1038/s43017-021-00247-8

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*, *In*

Press, 3949.

Klene, A. E., Nelson, F. E., Shiklomanov, N. I., & Hinkel, K. M. (2001). The n-factor in natural landscapes: Variability of air and soil-surface temperatures, Kuparuk River Basin, Alaska, U.S.A. *Arctic, Antarctic, and Alpine Research*, *33*(2), 140–148. https://doi.org/10.2307/1552214

Kong, X., Doré, G., & Calmels, F. (2019). Thermal modeling of heat balance through embankments in permafrost regions. *Cold Regions Science and Technology*, *158*, 117–127. https://doi.org/https://doi.org/10.1016/j.coldregions.2018.11.013

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022.

Li, Q., Zhao, Y., & Yu, F. (2020). A Novel Multichannel Long Short-Term Memory Method With Time Series for Soil Temperature Modeling. *IEEE Access*, *8*, 182026–182043. https://doi.org/10.1109/access.2020.3028995

Liu, H., Maghoul, P., Bahari, A., & Kavgic, M. (2019). Feasibility study of snow melting system for bridge decks using geothermal energy piles integrated with heat pump in Canada. *Renewable Energy*, *136*, 1266–1280. https://doi.org/https://doi.org/10.1016/j.renene.2018.09.109

NRT. (2011). *Paying the price: the economic impacts of climate change for Canada* (Vol. 4). National Round Table on the Environment and the Economy.

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., & van Vuuren, D. P. (2014). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, *122*(3), 387–400. https://doi.org/10.1007/s10584-013-0905-2

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*.

Riseborough, D. W. (2003). Thawing and freezing indices in the active layer. *Proceedings, Eighth International Conference on Permafrost*, 953–958.

Saaly, M., Bobko, K., Maghoul, P., Kavgic, M., & Holländer, H. (2020). Energy performance of below-grade envelope of an institutional building in cold regions. *Journal of Building Engineering*, *27*(July 2019), 100911. https://doi.org/10.1016/j.jobe.2019.100911

Smith, M. W., & Riseborough, D. W. (2002). Climate and the limits of permafrost: a zonal analysis. *Permafrost and Periglacial Processes*, *13*(1), 1–15. https://doi.org/https://doi.org/10.1002/ppp.410

Swart, N. C., Cole, J. N. S., Kharin, V. V, Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., & Hanna, S. (2019). The Canadian earth system model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873.

Tabari, H., Sabziparvar, A.-A., & Ahmadi, M. (2011). Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. *Meteorology and Atmospheric Physics*, *110*(3), 135–142. https://doi.org/10.1007/s00703-010-0110-z

Vu, M. T., Jardani, A., Massei, N., & Fournier, M. (2021). Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network. *Journal of Hydrology*, *597*, 125776. https://doi.org/10.1016/j.jhydrol.2020.125776

Yang, Y., Dong, J., Sun, X., Lima, E., Mu, Q., & Wang, X. (2018). A CFCC-LSTM Model for Sea Surface Temperature Prediction. *IEEE Geoscience and Remote Sensing Letters*, *15*(2), 207–211. https://doi.org/10.1109/LGRS.2017.2780843

Zeynoddin, M., Bonakdari, H., Ebtehaj, I., Esmaeilbeiki, F., Gharabaghi, B., & Zare Haghi, D. (2019). A reliable linear stochastic daily soil temperature forecast model. *Soil and Tillage Research*, *189*, 73–87. https://doi.org/10.1016/j.still.2018.12.023