Article

Multi-Tempo Forecasting of Soil Temperature Data; Application over Quebec, Canada

Mohammad Zeynoddin [1], Hossein Bonakdari [2,*], Silvio José Gumiere [1] and Alain N. Rousseau [3]

1 Department of Soils and Agri-Food Engineering, Université Laval, Quebec, QC G1V 0A6, Canada;
   mohammad.zeynoddin.1@ulaval.ca (M.Z.); silvio-jose.gumiere@fsaa.ulaval.ca (S.J.G.)
2 Department of Civil Engineering, University of Ottawa, Ottawa, ON K1N 6N5, Canada
3 Institut National de la Recherche Scientifique-Eau Terre Environnement INRS-ETE,
   Quebec, QC G1K 9A9, Canada; alain.rousseau@inrs.ca
* Correspondence: hossein.bonakdari@uottawa.ca; Tel.: +1-6135625800 (ext. 6016)

Abstract: The profound impact of soil temperature (T_S) on crucial environmental processes, including water infiltration, subsurface movement, plant growth, and its influence on land–atmosphere dynamics, cannot be undermined. While satellite and land surface model-based data are valuable in data-sparse areas, they necessitate innovative solutions to bridge gaps and overcome temporal delays arising from their dependence on atmospheric and hydro–meteorological factors. This research introduces a viable technique to address the lag in the Famine Early Warning Network Land Data Assimilation System (FLDAS). Notably, this approach exhibits versatility, proving highly effective in analyzing datasets characterized by significant seasonal trends, and its application holds immense value in watershed-scaled hydrological research. Leveraging the enhanced state-space (SS) method for forecasting in the FLDAS, this technique harnesses T_S datasets collected over time at various depths (0–10 cm, 10–40 cm, and 40–100 cm), employing a multiplicative SS model for modeling purposes. By employing the 1-step, 6-step, and 12-step-ahead models at different depths and 2 locations in Quebec, Canada, the outcomes showcased a performance with an average coefficient of determination (R^2) of 0.88 and root mean squared error (RMSE) of 2.073 °C for the dynamic model, R^2 of 0.834 and RMSE of 2.979 °C for the 6-step-ahead model, and R^2 of 0.921 and RMSE of 1.865 °C for the 12-step-ahead model. The results revealed that as the prediction horizon expands and the length of the input data increases, the accuracy of predictions progressively improves, indicating that this model becomes increasingly accurate over time.

Keywords: prediction; FLDAS; soil science; soil temperature; satellite; lag

1. Introduction

Soil temperature (T_S) is one of the essential environmental factors used in various disciplines such as engineering and hydraulic structures design, geotechnics, hydrological and meteorological studies, agriculture, forestry, and many other fields. The majority of the chemical and physical characteristics of soil are temperature-dependent. For instance, fluctuations in T_S directly affect water infiltration; thus, variations in T_S and infiltration rates follow similar patterns [1–3]. Soil temperature may affect other processes, such as groundwater discharge, land–atmosphere fluxes such as evapotranspiration, gas exchange, and surface runoff patterns [4–6]. Fluxes of greenhouse gases such as carbon dioxide (CO_2) and nitrous oxide (N_2O) directly depend on T_S [7,8], which relates to global warming on a large scale. Any alterations to atmospheric constituents mutually affect the soil temperature. Because the thermal energy balance between the soil and atmosphere is preserved, the soil can either act as a heat source or sink [9–12]. Similarly, the impact of T_S on the agriculture section cannot be ignored, e.g., the dependency of crops' growth on T_S, irrigation planning, and hydrological drought monitoring [13,14].

Estimating the T_S time series by using artificial intelligence (AI) models and exogenous variables is common. For instance, multilayer perceptron (MLP), an artificial neural network (ANN) model and a multivariate linear regression with air temperature (Ta), solar radiation (RS), relative humidity (RH), and precipitation inputs as inputs [15], MLP and an adaptive neuro–fuzzy inference system, genetic programming and an ANN with Ta, RS, RH, wind speed, and dew points inputs [16–18], a self-adaptive evolutionary extreme learning machine with Rs, Ta, and pressure [19], an emotional neural network, and a least square support vector machine [20] are some of the AI methods used in this field. Though these modeling methods are powerful in creating estimations, they are prone to model parameter tuning and input selection uncertainties [21,22]. To solve this problem, researchers developed remotely sensed data.

Further developments of remote sensing tools and land surface methods with extended historical temporal coverage have improved the quality and quantity of environmental data, particularly in regions with limited available data. However, land coverage, upwelling atmospheric emissions [23], cloud coverage [24], satellite machine failure, and consequent errors in the inputs of land surface models are most often responsible for missing values in remotely sensed data. Therefore, many methods have been proposed to fill in the dataset's gaps [25].

Various methods have been proposed to fill in gaps in remotely sensed data, including combining temporal steps to create a composite set. Weiss et al. [26] categorized data recovery methods into three classes: (a) spatial, (b) temporal sequences of images, and (c) spatio–temporal filling methods. Inverse distance weighting (IDW), kriging, cokriging, and their variants are widely recognized methods in spatial class [27,28]. IDW and its variants, such as harmony search, genetic algorithm (ga), and particle swarm–IDW, are commonly used and reliable approaches. A comparative study by Bărbulescu et al. [29] showed that the ga–IDW method outperformed other methods in 70% of the cases studied. The principal component analysis, stochastic models, and k nearest neighbor (kNN) are other popular statistical approaches for estimating missing values [30,31]. Weiss et al. used the spatio–temporal method to fill gaps in the land surface temperature and vegetation index, achieving an accuracy of $R^2 = 0.87$, however could not evaluate embedded uncertainties or account for temperature values impacted by cloudy days. This conclusion adds to the inherent uncertainty of their model.

Integrating remotely-sensed data with modeling backed by historical data decreases measurement limitations and variable-based method deficiencies, namely, input variable uncertainties and error accumulation [18,22], where both non-linear and linear models are applicable. However, using a method with a derivable explicit expression fast enough to build real-time estimations and reasonably accurate information is prioritized over other methods. The advanced state-space (SS) method in hydrology is a relatively less utilized linear method for modeling time series with intense periodic fluctuations. This model uses multiple in-line pre-processing methods to produce white noise data. It additionally can be used to remove time series anomalies [32] and forecast future steps using the out-of-sample forecasting ability of the model. Studies on the SS model as a forecasting method are limited. Reservoir inflow modeling, climatic data, catchment rainfalls, and land hotspots are the parameters modeled and forecasted by the SS method [33–36]. For example, Vijayakumar and Vennila [33] compared the autoregressive integrated moving average (ARIMA) model with the SS method. The model input included reservoir inflows, and the results indicated that the SS method could overcome the deficiencies of the ARIMA model. They stated that the ARIMA model did not successfully deal with the time series components observed in their data. For the forecast procedure, the trend component was removed before autocorrelation modeling. Additionally, they mentioned that the model could not forecast the "turning points". These deficiencies, alongside the linear nature of stochastic models and requisite preprocessing steps before modeling, represent other limitations.

In a recent study by Zeynoddin et al. [37], a comprehensive exploration of techniques for predicting time series data was undertaken, encompassing variable-based models, AI modeling approaches, and linear methods. They proposed a linear methodology that pre-processes T_S data using several methods and models the data dynamically via stochastic models. Comparative analysis with ML techniques [19] revealed that combining the SS method with the stochastic model yielded significantly more precise results. Building upon these findings, Zeynoddin and Bonakdari [38] introduced a structurally optimized deep learning method that integrated the SS method for long-term soil moisture forecasting. Remarkably, their study demonstrated that the SS method excelled at regenerating long-term time series features and substantially enhanced forecast accuracy compared to other methods. Moreover, their solution effectively mitigated the data size dependency problem often encountered in data-driven modeling approaches. While this proposed structure has limitations, such as its finite domain of model variables, exclusivity to seasonally structured data, and the complexity of the deep learning method, the authors hypothesized that the comprehensive and generalizable structure of the SS method for seasonal data modeling could be leveraged to fill the temporal gap in FLDAS datasets. These datasets, generated at a monthly resolution [39] with a two-month availability lag, as indicated on the GEE catalog website (Earth Engine Data Catalog: NASA_FLDAS_NOAH01_C_GL_M_V001#description), are influenced by the annual or monthly trends exhibited by the T_S time series [10], which temporally and spatially vary due to atmospheric and subsoil processes affecting thermal equilibrium.

The objectives of this paper are (a) to introduce an extraction method for the collection of monthly mean T_S data for two agricultural areas in Quebec, Canada and (b) to apply the powerful SS method for modeling. Following these are (c) a spatial validation analysis will be carried out to analyze the model's accuracy alternation and pattern evaluation in case a time series of different locations are used, and (d) a trend and long-term alternations analysis of the T_S variable will be carried out based on the model results and forecasts. Finally, (e) the results of the proposed FLDAS–state-space method will be comprehensively discussed in the following sections. Furthermore, this paper will outline the practical application of these results, including their integration into future T_S forecasts, thereby making a significant contribution to the field.

2. Time Series and Pre-Processing Review

Since this study uses a time series-based approach, a review of necessary methods for time series preparation is presented. Time series can have jumps, periods, and seasonal and non-seasonal trends deemed deterministic. They can additionally have a stochastic component, constituted of residuals of the deterministic characteristics subtracted from the time series' values. The deterministic component is the predictable part of the time series' pattern and can be detected and extracted. Detecting the mentioned patterns and analyzing T_S data entails employing several tests. These tests not only help to detect meaningful patterns, however they can additionally help to extract statistics and interpret the time series. Following pre-processing methods, models are applied to prepare data for forecasting [40].

The first attribute of a time series to be checked is stationarity. It means the absence of meaningful changes in mean, variance, and other statistics over time. Therefore, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is applied to T_S for assessing stationarity [41]. In the KPSS test, a regression equation is fitted to the data. The T_S series is stationary if the variance of the independent variables of that equation is null. Alternatively, it is stationary in trend if the deterministic term of the trend is not null and stationary in level if the deterministic term of the trend is equal to zero. Seasonal and non-seasonal trends in the T_S time series as one of the non-stationarity factors is checked by the Mann–Kendal and seasonal Mann–Kendal trend (MK and SMK) tests [42,43].

Jumps appear in time series as sudden increases or decreases and can be both induced by human and natural sources. To evaluate this factor, the non-parametric Mann–Whitney

(MW) test can be performed [42]. More details about pre-processing steps can be found in Zeynoddin et al. [22]. The autocorrelation function (ACF) is a tool to evaluate a dataset's randomness or assess whether a non-linear or linear model would be appropriate. This way, it can be found whether an observation is related to adjacent observations or not. Additionally, specific patterns in time series, such as sinusoidal fluctuations interpreted as periodicity, can be detected. Therefore, we can judge the time series' condition and interpret it. An ACF plot is an intuitive method to assess the data more conveniently. The ACF demonstrates the autocorrelation of data separated by a lag distance. Both positive and negative ACF values are considered as the correlation between lags. Still, for several lags, if the positive values of an ACF consecutively exceed specific values, in this case, the 95% significance levels, they are interpreted as significantly correlated lags. These significant correlations should be reduced or removed from the time series to obtain independent variables and reduce the impacts of the deterministic components in the modeling procedure. As mentioned earlier, the deterministic components are the predictable part of the series and are responsible for dependency in the time series. This part of the time series' structure can be separated before modeling and restored to modeled data. The stochastic component is the mere important part for modeling purposes. The significance intervals are defined based on the cumulative distribution function of the standard normal distribution and the 95% significance level.

3. Materials and Methods

3.1. Soil Temperature Collection and Analyses

3.1.1. Famine Early Warning Network Land Data Assimilation System

The Famine Early Warning Network Land Data Assimilation System (FLDAS) program improves the quality of measured values by combining hydroclimatic data and observations to develop rainfall, soil moisture, evapotranspiration, and soil temperature products at a monthly resolution. These products are hosted by Google Earth Engine and can be merged with variables related to food security assessment in data-sparse countries. The program is developed by the U.S. Geological Survey (USGS), NASA GSFC, Goddard Space Flight Center (GSFC), and the University of California Santa Barbara (UCSB). It can provide helpful information to water resources and agricultural stakeholders. The output of the FLDAS can be used to compare current and past hydrological conditions and can be merged with other land information systems to produce an ensemble of hydroclimatic variables.

The FLDAS utilizes the Noah version 3.6.1 land surface model (LSM), the Climate Hazards Group Infrared Precipitation with Station (CHIRPS) 6-hourly rainfall, and the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2). Moreover, the system uses the African Rainfall Estimation Algorithm Version 2 (RFE 2.0) and the CHIRPS for low latency. Using the Land Data Tool Kit (LD_{TK}), the FLDAS provides and preprocesses the land cover parameters needed for the Noah model. The LD_{TK} contains the products of the NASA Soil Moisture Active Passive (SMAP), the Moderate Resolution Imaging Spectroradiometer (MODIS), the Advanced Very High-Resolution Radiometer (AVHRR) from NOAA 19 satellites, the CHIRPS quasi-global rainfall data set obtained from high-resolution satellite imagery, MERRA2, and the Global Data Assimilation System (GDAS), which uses multiple sources of information such as balloon, aircraft, radar, and satellite data. This tool kit allows the FLDAS to use rich sources of high-resolution datasets, with near real-time observations. In return, the data are fed to the Noah LSM to produce high-accuracy outputs.

All datasets are available from 1982 to the present at a 0.1 arc degree and a monthly temporal resolution. A flexibility in using different products represents the advantage of the FLDAS, products such as FEWS NET operational rainfall. Using multiple surface models, normal assessment, and a fast and appropriate data distribution represent the other advantages of this system. The FLDAS datasets can be obtained from the Google Earth Engine (GEE), which contains a vast catalog of satellite and geospatial datasets. In this application programming interface (API), both Python and JavaScript languages can

be used to obtain and process the required data. Using the GEE API, various datasets for different fields of study can be accessed, such as global surface water changes, crop production, and drought monitoring [44–46]. Figure 1 shows the data collection and modeling procedural steps. Aside from the GEE API, many data transmission mechanisms are presented in this figure. In the GEE code editor, the following pseudocode and the corresponding commands in the Java language can be used to obtain the desired datasets (Appendix A Algorithm A1), or users can download the dataset from the developed application, SOILPARAM APP, by [47].
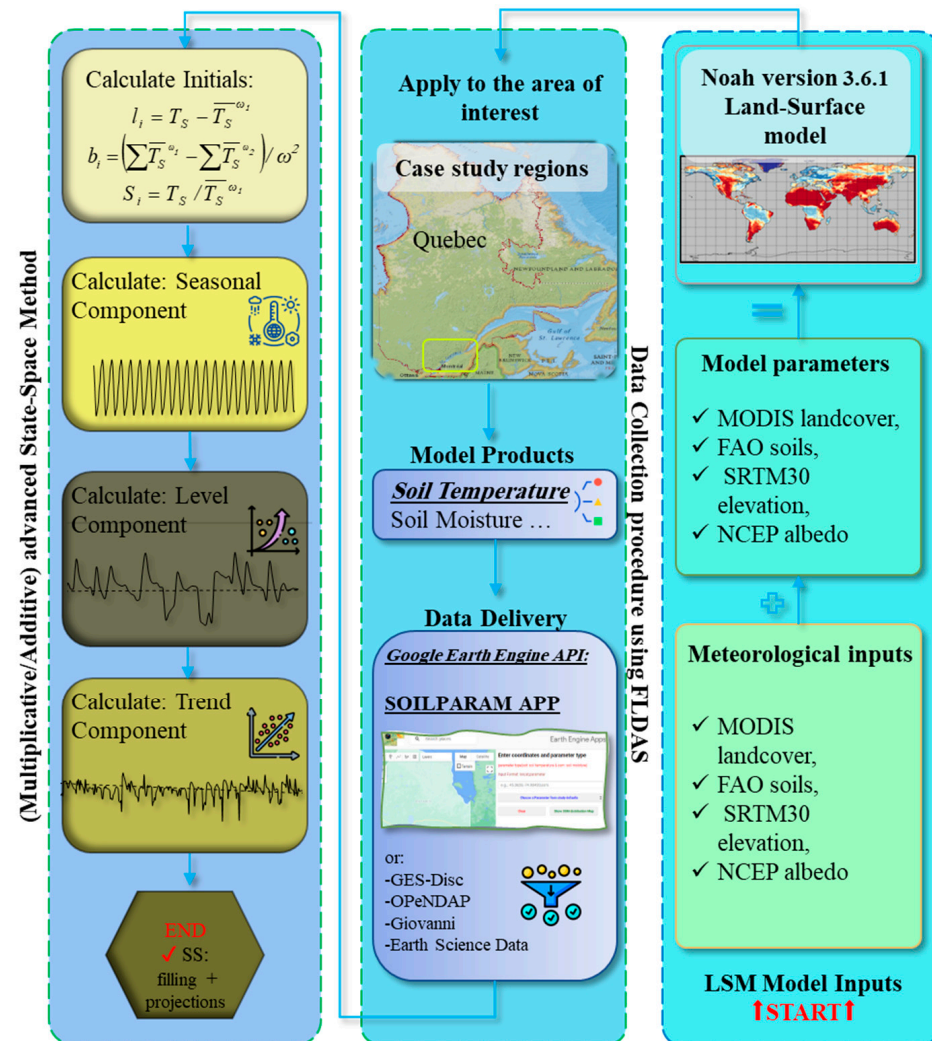

Figure 1. Data collection and modeling flowchart using FLDAS; ω_i is the *i*-th period, and S_i, l_i, and b_i are seasons, level, and trend initials, respectively, at *i*-th time. T_S is the values of observed data and mean data, respectively.

3.1.2. Case Study

Two locations in Quebec, Canada, were chosen for this study. The first location is situated in a suburb of Quebec City (St. Augustin), with a latitude of $46.73°$ and longitude of $-71.50°$. The second is in a suburb of Montreal (St. Mirabel), with a latitude of $45.67°$ and longitude of $-74.03°$. According to the Köppen classification, Quebec City's climate is "humid continental". This means that summers are hot, and winters are cold and snowy. Fall and spring temperatures are, additionally, moderate. The atmospheric temperature in this region varies on average between $-18\,°C$ and $25\,°C$. Montreal has a "warm-summer humid continental climate", based on the same classification. This classification means summers are hot and humid, and winters are icy and snowy. The other two seasons are

moderate. The air temperature ranges between −30 °C and 27 °C in some regions of the study areas [48]. Figure 2 shows the locations on a map.

To validate the results of the proposed methodology, the soil temperatures at two other locations were obtained, using the same method as the GEE API. These locations were selected based on the mapping of the land use of the St. Lawrence Lowlands report, published by Environment and Climate Change Canada in 2018. The Mauricie (MA) validation point is located between the first two study points (QC and MC), and the second validation point Lennoxville (LX) is located east of these points. These locations are two critical agricultural areas in Quebec, Canada. They are located between and southeast of the previous ones (Figure 2).
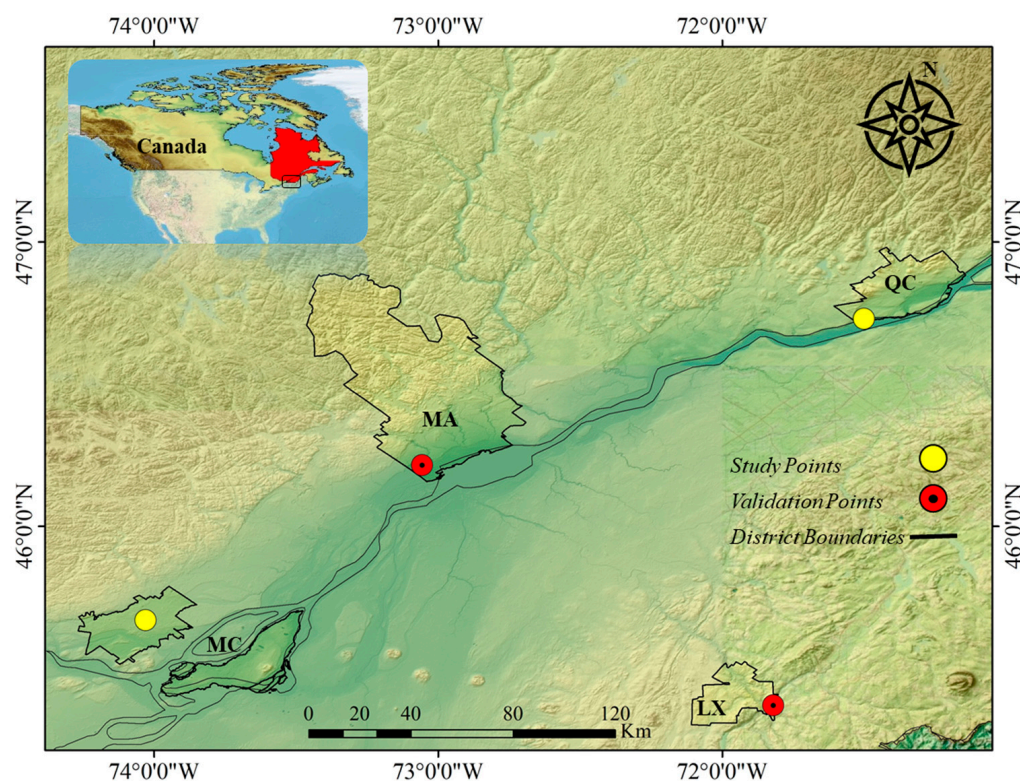

Figure 2. The chosen locations in Quebec, Canada; QC: Quebec City; MC: Montreal; MA: Mauricie; and LX: Lennoxville.

The time series plots for each location are displayed in Figures 3 and 4. The time series length was divided into training and testing intervals, with a ~75:25% ratio for modeling purposes. Hence, from a total of 445 months of data, 336 months are considered for model training and the rest, 109 months, for testing. In addition, two other sites were selected to obtain additional T_S data for further validation of the models. These locations were chosen based on the 2018 land cover mapping of the St. Lawrence Lowlands, published by "Environnement et Changement Climatique Canada". The first site is in Mauricie (MA), with a latitude of 46.216° and longitude of −73.056°, and the second one is in Lennoxville (LX), with a latitude of 45.37° and longitude of −71.82°. The differences between the study sites and validation sites are illustrated in Figure 5. The size of the datasets of the validation sites is the same as that of the study sites.
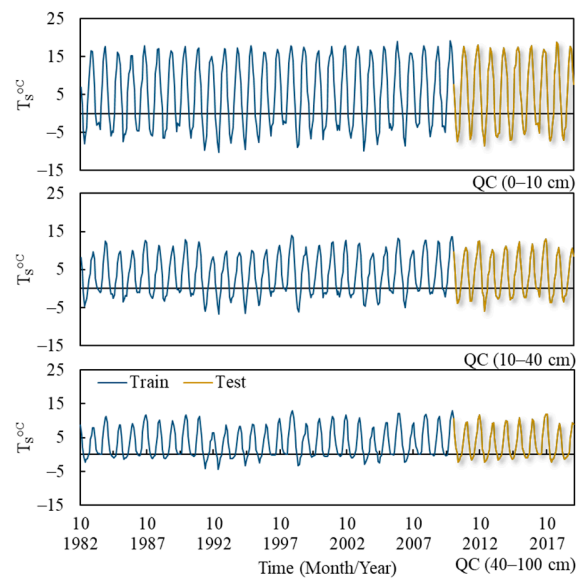
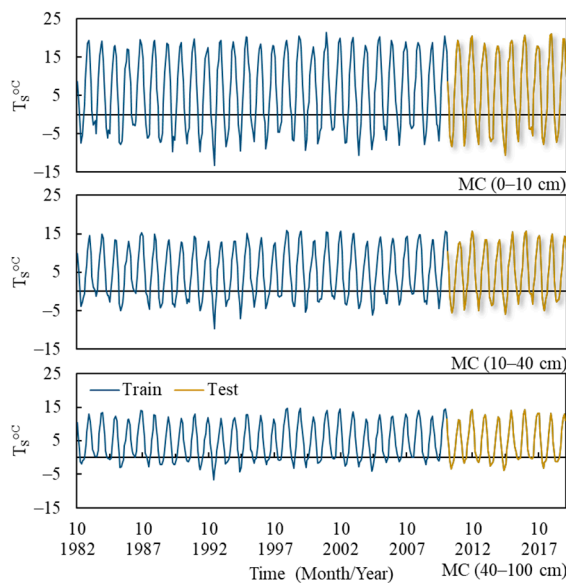**Figure 3.** Soil temperature (T_S) time series for all depths for the Quebec City (QC) study suburbs.

Figure 4. Soil temperature (T_S) time series for all depths for Montreal (MC) study suburbs.
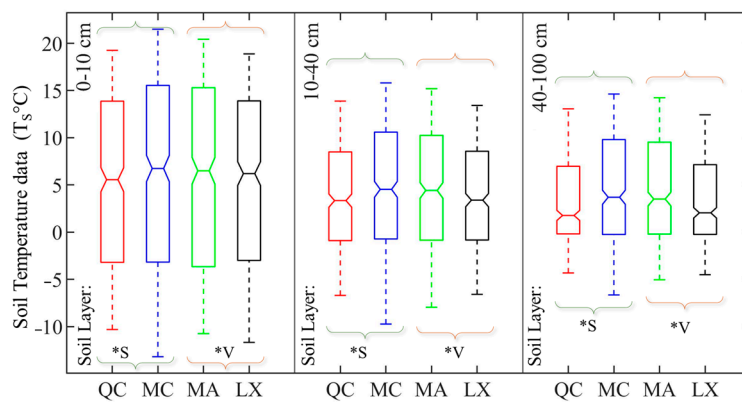

Figure 5. Statistical features of soil temperature datasets in studied locations; QC: Quebec City; MC: Montreal; MA: Mauricie; LX: Lennoxville; *V: validation; and *S: study points.

3.2. Modeling, Calibration, and Validation

3.2.1. State-Space Method

Smoothing methods are among the data processing approaches to stationarize time series and eliminate the deterministic components. Using specific algorithms, they can remove fluctuations and predictable patterns to enable predicting trends, seasonality, and future values. Among them are included moving average, simple exponential, linear exponential, seasonal exponential smoothing, and damped trend exponential smoothing [49–51].

The state-space (SS) method is a renowned, advanced, and robust triple exponential smoothing method used to remove trend, seasonality, and level in highly seasonal series and has additionally been used for forecasting data [52]. The SS method is divided into an additive form for cases where seasonality is roughly constant and a multiplicative form for series with seasonal trends. Each one has three smoothing parameters that need to be optimized using any optimization method, or trial and error. The multiplicative form of the method is as follows [53]:

$$l_t = \alpha(T_s/(S_{t-\omega})) + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{1}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{2}$$

$$S_t = \gamma(T_s/(l_{t-1} + b_{t-1})) + (1-\gamma)b_{t-1} \tag{3}$$

$$SS_{t+m} = (l_t + mb_t)S_{t-\omega+m} \tag{4}$$

In the above equations, l_t is level smoothing, b_t is trend, and S_t is seasonal components smoothing; A, β, and γ are additional smoothing parameters; Ω is seasonality, and m denotes forecasting time steps; SS_{t+m} is the forecasted data by the state-space method for m future lags; and $T - \omega + m$ assures the use of the latest historical data. Using real data to update forecasts, the model adapts itself to the new data, and the error is subsequently reduced. Therefore, the longer the training set is, the fewer errors will be generated. This feature can be considered a drawback too. In short time series, the model cannot adapt to data fluctuations, and therefore, accuracy may be reduced in multi-step forecasts. Compared to stochastic models, another advantage of this model is that it does not require normalization. The SS parameters can be obtained through either trial and error or optimization algorithms such as nonlinear generalized reduced gradient (GRG), which is applied to the three components of the SS method in this study.

3.2.2. Performance Metrics

The first 75% of the data is used in this study to calibrate the model parameters, and the last 25% is used for model validation by using the following indices: Coefficients of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE) are used to evaluate the model accuracy of the time series obtained from pre-processing of the original T_S data. To assess the modeling power, the Nash–Sutcliffe model efficiency (E_{N-S}) is additionally used [54]. The closer E_{N-S} is to a value of one, the more powerful the model is.

$$R^2 = \left(\frac{\left(\sum_{i=1}^{n} \left(T_{S,obsi} - \overline{T_{S,obsi}} \right) \left(T_{S,Pi} - \overline{T_{S,Pi}} \right) \right)}{\sqrt{\sum_{i=1}^{n} \left(T_{S,obsi} - \overline{T_{S,obsi}} \right)^2 \sum_{i=1}^{n} \left(T_{S,Pi} - \overline{T_{S,Pi}} \right)^2}} \right)^2 \tag{5}$$

$$RMSE = \sqrt{ \left(\sum_{i=1}^{n} \left(T_{S_{obs,i}} - T_{S_{P,i}} \right)^2 \right) / n } \tag{6}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| T_{S_{obs,i}} - T_{S_{P,i}} \right| \tag{7}$$

$$E_{N-S} = 1 - \left[\left(\sum_{i=1}^{N} \left(T_{S_{obs,i}} - T_{S_{P,i}} \right)^2 \right) / \left(\sum_{i=1}^{N} \left(T_{S_{obs,i}} - \overline{T_{S}}_{obs,i} \right)^2 \right) \right] \tag{8}$$

where $T_{S_{obs,i}}$, $T_{S_{P,i}}$ and $\overline{T_S}$ are the *i*th value of observed data, predicted soil temperature, and mean of data, respectively; *n* is the number of months.

4. Results Analysis

4.1. T_S Time Series

In order to retrieve the soil temperature data from the FLDAS, a set of commands was coded using the GEE API. The extracted time series were tested for stationarity, and the results are presented in Table 1 for Quebec City and Montreal, respectively. Since the focus was on the study points, and the validation points followed the same pattern, the results of these points are not presented. The SMK test results show significant seasonal trends in both T_S time series. For each soil layer (there are 3 layers with the following upper–lower depths of 0–10, 10–40, and 40–100 cm), the probabilities corresponding to the MK and MW test results are higher than the significance level of 5%. Thus, the presence of non-seasonal trends and jumps in the datasets are not significant. Figure 6 shows the ACF plots of all the series. The seasonal correlations can be seen as sinusoidal patterns, with peaks at lags 12, 24, 36, 48, 60, 72, and 84, equivalent to 7 seasonal lags with a repetition of 12 months. In these lags, the ACF values are higher than the 95% confidence interval, and therefore, these correlations are considered significant. They are found at all soil depths. Considering the significant periodicity that could be observed and the presence of a seasonal trend, the KPSS test failed to detect the impact of these non-stationary factors properly. All probability values for this test are higher than a significance level of 5%. It additionally indicates the importance of utilizing several tests for stationarity detection. In other words, by using the MK and SMK tests, the significance of the gradual changes over time was assessed. According to these results, these incremental changes were only significant seasonally. As the MW test demonstrated, no sudden changes were detected in the datasets.

On the other hand, the ACF plots showed intensive periodicity. Consequently, as the seasonal trend and periodicity are both prerequisites of the state-space multiplicative method, this method is deemed suitable for forecasting. On the other hand, if the series were trendless, then the additive method would have been appropriate.

Table 1. Stationarity survey for soil temperature time series of three soil layers (upper and lower depths of 0–10, 10–40, and 40–100 cm).

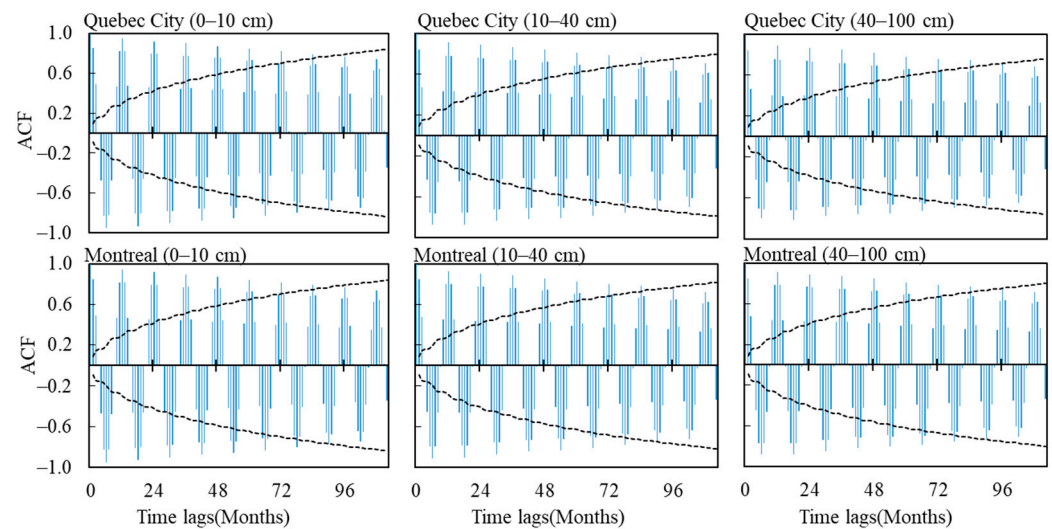| Tests	Jump	Quebec City (QC)			Jump	Montreal (MC)		
		Trend		Stationarity		Trend		Stationarity
depth cm	P_{MW}	P_{MK}	P_{SMK}	P_{KPSS}	P_{MW}	P_{MK}	P_{SMK}	P_{KPSS}
0–10	29.69	19.12	0.01	90.05	48.37	28.13	0.06	95.14
10–40	16.74	12.71	0.01	74.74	36.53	29.18	0.32	92.19
40–100	9.97	13.05	0.01	70.65	32.37	32.24	0.17	90.99

P = *p*-value > 5% = acceptable.

Figure 6. Autocorrelation function (ACF) plots of T_S data.

4.2. State-Space Modeling Results

In this study, as mentioned in the methodology section, the state-space (SS) method is employed to forecast the T_S of Montreal and Quebec City. This method requires three smoothing parameters to eliminate deterministic terms in the time series. Moreover, initial values are needed to obtain the SS formula's level, trend, and seasonal components. Therefore, the first period of the T_S time series was used to initiate the modeling, which in this case, was the first 12 months or the first seasonal lag. In addition, optional initial values were chosen for the smoothing parameters α, β, and γ, and they were optimized by the GRG solving method to achieve the most accurate results.

Three types of modeling were conducted to examine the SS method's modeling capability. The first was dynamic modeling, where the model was updated after each forecast. Next, a 6-step-ahead forecast (*fs6*) was performed, followed by a 12-step-ahead forecast (*fs12*) that covered the entire period. The results of the 1-step-ahead forecast (*fs1*) models are provided in Table 2. As seen, the SS model could forecast the test period with a coefficient of determination higher than 0.90 and with low and relatively close errors, except for the 40–100 cm depth in Quebec City. Additionally, a decreasing trend in the accuracy by increasing the depths was observed.

Table 2. State-space dynamic modeling (1-step-ahead) for both QC and MC locations.

Station	Depth cm	(α, β, γ)	R^2	RMSE °C	MAE °C	E_{NS}
Quebec City	0–10	(0.164, 0.006, 0.028)	0.970	1.633	1.340	0.966
	10–40	(0.348, 0.014, 0.008)	0.921	1.979	1.537	0.859
	40–100	(0.947, 0.010, 0.021)	0.632	2.826	2.074	0.568
Montreal	0–10	(0.168, 0.028, 0.101)	0.952	2.131	1.808	0.952
	10–40	(0.331, 0.014, 0.008)	0.902	2.119	1.682	0.894
	40–100	(0.000, 0.013, 0.907)	0.906	1.749	1.319	0.896

Tables 3 and 4 present the modeling results for the *fs6* and *fs12* modeling methods. As the soil depth increased, a general decline in accuracy could be seen, except for the first soil layer (10–40 cm) for both *fs6* and *fs12*. At this depth, the accuracy was lower than in the other two depths. According to the results, there are some variations in the magnitude of the errors with a maximum value of ~3 °C. For all depths, the average difference between forecasts and observations was ~2 °C except for Montreal (10–40 cm). For the Quebec City modeled time series, all sudden drops in accuracy were observed deeper in the soil profile (40–100 cm). A decrease in model accuracy was additionally observed in the Montreal model results. However, this reduction in accuracy of the first soil layer (10–40 cm) is

more significant than that of the second (40–100 cm), in contrast to the Quebec City model results, which had a uniform trend in accuracy reduction. The Nash–Sutcliffe coefficient additionally demonstrates the power and efficiency of the generated model in forecasting shallow depths. This sudden decline is due to extreme changes in the level component of the SS model, and the model could not adequately adapt over that period. Furthermore, the parameters are not optimized properly. The T_S datasets are highly periodic, and the SS seasonal parameter was expected to have a greater weight in the equation. However, on the contrary, for the middle soil layer in Montreal (10–40 cm), the involvement of the seasonal parameter in the model equation is less than the other 2 parameters.

Alongside assessing the models using indices such as R^2, RMSE, etc., the statistical features of the forecasts should be evaluated. Appropriate models should additionally reproduce the statistical characteristics of the original data [55]. To investigate the reason for drops in accuracy and evaluate the statistical aspects, box plots of forecasts vs. observed data were drawn (Figure 7). Some outliers were observed in QCT_Ss (40–100 cm) and T_S Ts (10–40 cm). These outliers were the reason for poor results at these depths and were created because of some extreme values in the level component during the training period. These extreme values were multiplied by the SS relationship's seasonal factor and seasonal coefficients. The interquartile areas and means of the T_S data in the Montreal time series were almost perfectly reproduced. However, these features for the Quebec City time series fluctuated by a few degrees centigrade.

Table 3. State-space modeling (6-steps-ahead) for both QC and MC locations.

Station	Depth cm	(α, β, γ)	R^2	RMSE $°$C	MAE $°$C	E_{NS}
	0–10	(0.164, 0.006, 0.028)	0.974	1.506	1.215	0.971
Quebec City	10–40	(0.223, 0.001, 0.015)	0.889	2.338	1.870	0.803
	40–100	(0.219, 0.003, 0.013)	0.638	2.956	1.948	0.527
	0–10	(0.168, 0.028, 0.101)	0.927	2.735	2.175	0.920
Montreal	10–40	(0.331, 0.014, 0.008)	0.670	6.591	4.255	0.025
	40–100	(0.000, 0.013, 0.907)	0.906	1.749	1.319	0.896

Table 4. State-space modeling (12-steps-ahead) for both QC and MC locations.

Station	Depth cm	(α, β, γ)	R^2	RMSE $°$C	MAE $°$C	E_{NS}
	0–10	(0.164, 0.006, 0.028)	0.976	1.385	1.116	0.975
Quebec City	10–40	(0.348, 0.014, 0.008)	0.929	2.097	1.745	0.842
	40–100	(0.947, 0.010, 0.021)	0.840	2.128	1.813	0.755
	0–10	(0.168, 0.028, 0.101)	0.969	1.710	1.378	0.969
Montreal	10–40	(0.331, 0.014, 0.008)	0.902	2.123	1.562	0.894
	40–100	(0.000, 0.013, 0.907)	0.906	1.749	1.319	0.896

Figures 8 and 9 display both locations' T_S forecasts and scatter plots. The 0–10 cm and 40–100 cm soil layers are presented as samples of the forecasting capability of the SS method. These figures show a good correlation between the forecasts and observed data, and most of the estimates are located within the 10% intervals of the scatter plots. For each time step, the dispersion in the forecasts of QC datasets increases with soil depth; however, the forecasts were more correlated for the *fs12* predictions than in the other time steps. The forecasts were more dispersed for the *fs6* forecasts of the QC 10–40 cm and 40–100 cm soil layers, as the *fs6* forecast of the QC 10–40 cm soil layer had the most dispersion. The same pattern was observed for the MC forecast datasets, albeit the *fs6* forecasts of the MC 10–40 cm soil layer had the most dispersion. According to the time series plots of the QC datasets, the SS method generally overestimated the highs and lows for all time steps of the 10–40 cm and 40–100 cm soil layers and underestimated the peaks of the 0–10 cm soil layer data. On the other hand, the peaks of the MC datasets were generally overestimated

by the SS method, whereas the lows were underestimated. The middle soil layer (10–40 cm) of both locations was poorly forecasted, compared to the other 2 layers.
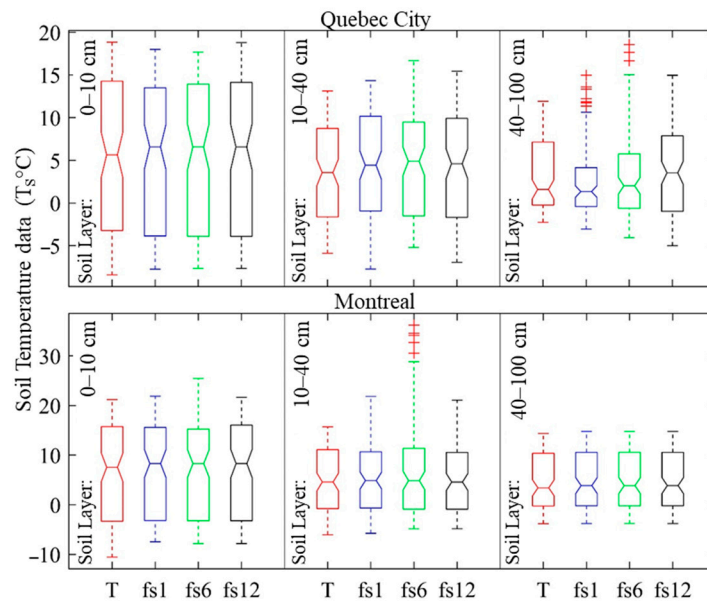

Figure 7. The box plot of the T_S data (T) vs. forecasts of different soil layer depths with different forecast steps (*fs*).
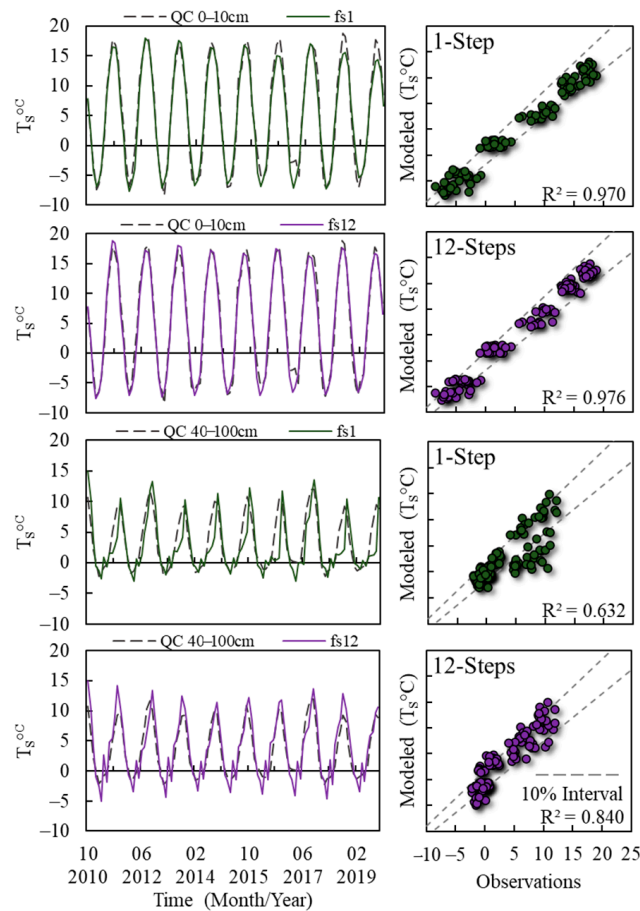

Figure 8. Forecasts vs. original T_S data plots for Quebec City (QC), based on the trained model from 1 October 2010 to 1 October 2019.
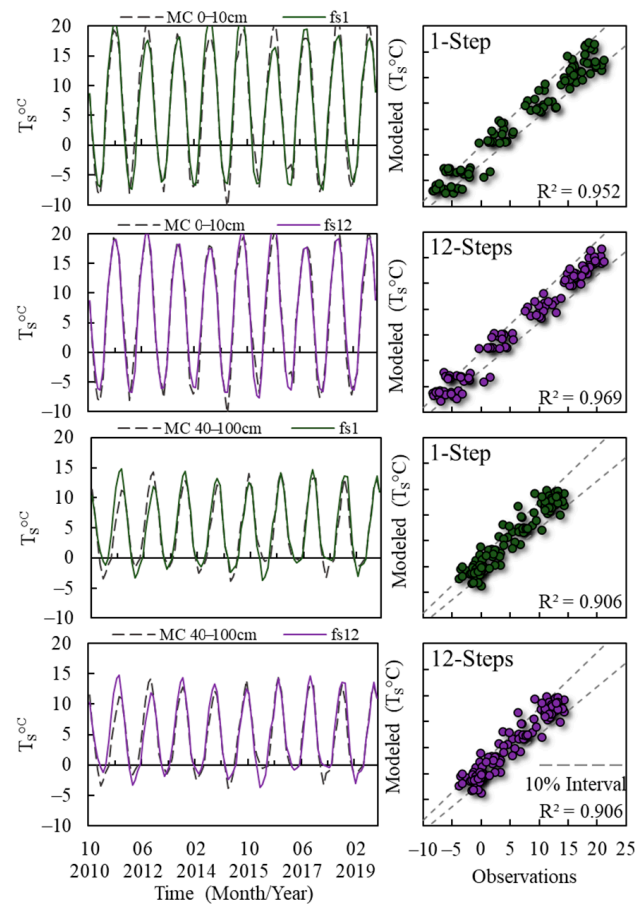
Figure 9. Forecasts vs. original T_S data plots for Montreal (MC), based on the trained model from 1 October 2010 to 1 October 2019.

The magnitude of the outliers seen in the forecasts' box plots are less than a few degrees, and as seen in the time series plots (Figures 8 and 9), they can be neglected. Moreover, these outliers were seen in the QC second soil layer (40–100 cm) *fs1* and *fs6* and MC first soil layer (10–40 cm) *fs6* while they were removed for the *fs12* forecasts. As mentioned earlier, the reasons for the formation of outliers were changes in level components and smoothing parameters. However, while the model was updated with the new real data in *fs12* forecasts, and the weight of the seasonal parameters increased, the level component was able to correct itself and dampen the errors; therefore, no outliers were produced. The R^2 additionally increased from 0.63 to 0.84. The self-correction feature of the SS model through time represents a positive feature; however, as it requires long time series, it can additionally be viewed as its weakness.

The SS method could not produce equally precise results for the last soil layer (40–100 cm). This performance is additionally valid for *fs6* periods, when compared to the different soil layer depths and forecasts' steps. The overall performance is acceptable for temporal lag compensation and future state forecasts. This method can produce reliable results, compared to black box AI methods and even other linear methods such as stochastic models. Since it has a simple structure, it is one of the most generalizable methods. The SS method was used to model data from October 1982 to October 2019. Given the fact that the FLDAS data were updated to October 2020, the estimated parameters were used to model data from November 2019 to October 2020 and perform an out-of-sample forecast to estimate T_S values from November 2020 to October 2021. The results are introduced in Figure 10. The shaded areas are the out-of-sample forecasts. As seen in this figure, the ~2–3 °C error remained in the forecasts of the 2020 data. Considering this error, an uncertainty bound equal to 3 °C can be expected for the 2021 values.
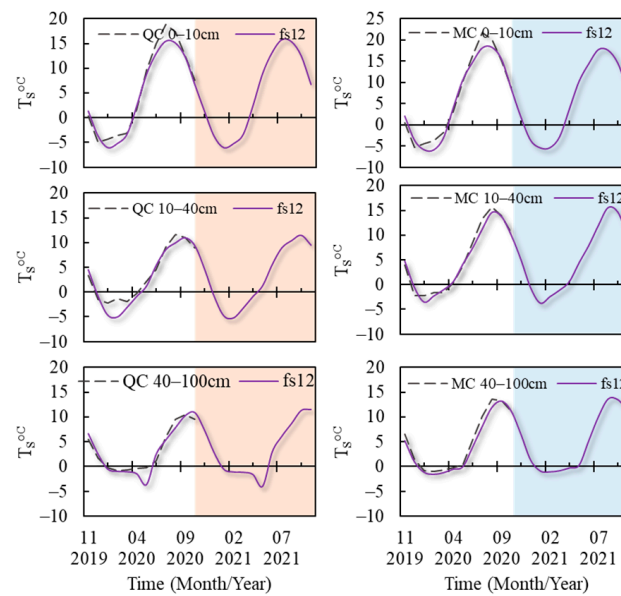
Figure 10. Future forecasts are based on the trained model parameters; shaded areas are the out-of-sample forecasts (12-step-ahead (*fs12*)) from 1 November 2020 to 1 October 2021.

5. Discussion

It was mentioned that a significant seasonal trend exists in these time series. The monthly values of different years were separated from the series to exhibit this seasonal trend and assess its significance. The MK test was then applied again. The test results indicated a trend in each month of the T_S time series, some of which are significant. In Quebec City, the months of August, September, October, and November and in Montreal, September, October, and November had a significant positive trend for different soil layer depths. Each location had a positive trend for seven months of the year. This means that the soil is generally getting warmer. Figure 11 presents the time series of the first soil layer (0–10 cm) for August, September, October, and November in Quebec City, along their linear trend lines, as examples of the progressive warming in the T_S time series.
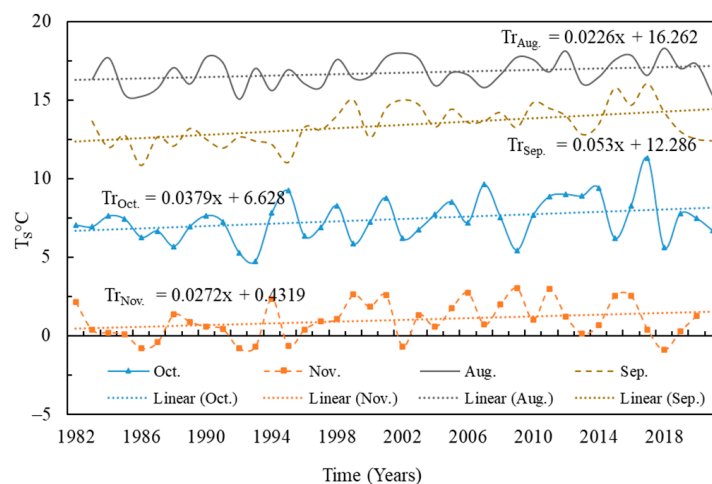

Figure 11. Quebec City (0–10 cm) monthly T_S time series, from October 1982 to October 2021; Tr.: trend line (linear).

Qian et al. [56] performed extensive research on the T_S trend and its association with the trend in climatical variable changes in Canada. They investigated the T_S trend in 30 stations across Canada, from 1958 to 2008. Their research showed a positive trend in T_S in 60% of stations, significantly in spring and summer means. They additionally reported a positive

correlation between T_S trend and changes trend in snow cover thickness, air temperature, rain and snowfall, and total precipitation. The correlation between these parameters was additionally surveyed in other studies [10,57]. Later, Assani et al. [58] performed a similar investigation on the maximum and minimum air temperature and rainfall, exclusively to Quebec province. They reported significant trends in these variables during summer, indicating that summer warming in southern Quebec is more pervasive than winter warming. Considering these findings, the progressive warming in soil temperature can be justified by reduced rainfall during the warm season and decreased soil moisture, especially in the mentioned months. Consequently, water exploitation typically increases during this season. The increase in air temperature and, at a large scale, the impact of global warming have caused changes in the overall trend of soil temperature. Some researchers even reported a stark trend in soil temperature compared to air temperature [59,60], which means the trend in the former is more noticeable than that in the latter. Monitoring and controlling soil moisture can be a short-term solution to controlling the T_S upward trend. However, unmanaged water exploitation can significantly reduce groundwater and decrease soil moisture. This factor is essential as it directly impacts the soil's thermal conductivity and can lower it to one-eighth [61]. Moreover, the heat-retaining feature of soil is negatively affected since the specific heat of pure water (4.2 J/g$°$C) is required compared to that of clay minerals (0.76 J/g$°$C), a dominant component of the soil [10]. To understand the magnitude of the impact, it should be mentioned that an increase of 1 $°$C of 1 gram of dry clay soil needs 1.1 J whereas 1 gram of wet clay soil needs 2.21 J (average) of energy under laboratory conditions with 25% of water content and 1300 kg/m^3 density. However, this value varies for different conditions and depends on soil texture, water content, and soil density [62].

Soil moisture and temperature are the two essential parameters that affect soil respiration. The relationship between these parameters and respiration can be estimated using several models, e.g., quadratic and exponential models. As mentioned earlier, these parameters have a strong correlation. Several studies have been carried out to distinguish them from one another. In this case, the estimation of soil moisture given the available soil temperature at multiple or limited depths has been debated by many scholars, and several methods have been proposed in this regard [63–66]. These parameters are essential in the long-term prediction of drought and food monitoring [39,67,68]. The other applications include irrigation planning, water resources management, flow forecasting, and hydrological land–atmospheric studies.

Higher temperatures are linked to climate change [69]. Crop water demand and evapotranspiration increase as the temperature rises. Consequently, soil moisture decreases, and soil droughts result from this phenomenon [70]. As the T_S level increases over time at both Quebec City and Montreal, soil moisture reduction disrupts irrigation scheduling in terms of amount and time intervals. More water pumping is needed to compensate for the reduced soil moisture, and the additional supply of water leads to more energy consumption [71]. However, the interactions between crop water requirement, soil temperature and moisture, evapotranspiration, and other hydro–climatic factors are complex. Access to the parameter records can additionally be limited. Thus, additional studies about these factors, their interactions, and the use of new data collection methods are required. Methods such as the FLDAS, development of land surface models, integration of deep learning methods in the processing and simulating of different hydro–climatic factors, as well as access to satellite data (i.e., data-driven Earth system science) are needed [72–74].

6. Validation of the SS Method

Access to high-quality field data has always been problematic, due to financial and operational resource constraints. Therefore, remotely sensed data and model-based simulations have been used for the past couple of decades. On the other hand, evaluating and validating these methods have been equally crucial for scholars and users. Hence, different validation studies are always carried out. In our case, the Noah LSM, the central part of the FLDAS, has been the center of attention for scholarly work with field-data alternatives. For

instance, Xia et al. [75] validated Noah soil temperature products by comparing them with field datasets from 137 stations in the United States over different time scales (e.g., daily, monthly, and annually) and different soil layer depths, from 0 to 200 cm. These authors declared that model products generally emulated observational field data. Although they found some bias for greater soil layer depths and a decreased accuracy of model results, they stated that Noah modeling "Kills" is acceptable. Other articles additionally validated the Noah products for different soil conditions and locations [76,77]. The FLDAS LD$_{TK}$ inputs are additionally monitored and evaluated by The National Aeronautics and Space Administration (NASA) and many other scholars. Therefore, in this paper, we concentrated on validating the results of our model at different locations. Following the assessments of the obtained time series characteristics for these two locations, the proposed methodology modeled the datasets. The results are provided in Tables 5 and 6 for the dynamic and 12-step-ahead forecasts.

Table 5. State-space modeling (1-step-ahead) for both validation locations (MA and LX).

Station	Depth cm	(α, β, γ)	R^2	RMSE °C	MAE °C	E_{NS}
Mauricie	0–10	(0.171, 0.028, 0.101)	0.962	1.883	1.560	0.961
	10–40	(0.160, 0.000, 0.000)	0.896	2.471	1.953	0.847
	40–100	(0.000, 0.013, 0.936)	0.853	2.095	1.536	0.839
Lennoxville	0–10	(0.184, 0.030, 0.040)	0.962	1.750	1.453	0.959
	10–40	(0.160, 0.000, 0.000)	0.867	2.067	1.468	0.845
	40–100	(0.001, 0.010, 0.927)	0.931	1.154	0.879	0.931

Table 6. State-space modeling (12-steps-ahead) for both validation locations (MA and LX).

Station	Depth cm	(α, β, γ)	R^2	RMSE °C	MAE °C	E_{NS}
Mauricie	0–10	(0.171, 0.028, 0.101)	0.972	1.629	1.291	0.971
	10–40	(0.160, 0.000, 0.000)	0.925	3.784	2.913	0.641
	40–100	(0.000, 0.013, 0.936)	0.853	2.095	1.536	0.839
Lennoxville	0–10	(0.184, 0.030, 0.040)	0.975	1.372	1.093	0.975
	10–40	(0.160, 0.000, 0.000)	0.929	1.556	1.195	0.912
	40–100	(0.001, 0.010, 0.927)	0.932	1.153	0.879	0.931

The same prediction pattern as QC and MC is seen in these tables. The reduction in accuracy with an increase in soil layer depth is obvious. An intercomparison of the models reveals that the smoothing parameters related to trend (β) and seasonal factor (γ) are decreased with an increase in the soil layer, as they are almost equal to 0 or very small for the 10–40 cm and 40–100 cm layers. For instance, in Tables 3 and 4, the seasonal parameter for soil layer 10–40 cm is equal to 0.008, and the trend factor is almost equal to 0.014 for the MC datasets. Nevertheless, after raising the seasonal parameter to 0.907 and increasing its involvement in the modeling equation, the correlation of the models was enhanced. In the QC results, this pattern can be seen for both β and γ parameters; however, it is not as obvious. After modeling the datasets of the validation points, the same problem is seen in the models. This means that these important factors are not adequately involved in the modeling procedure, which can be due to the nature of the data or the disability of the optimization method in finding the appropriate parameters. The reason for this problem can be investigated by using other datasets of a different nature (e.g., soil moisture) and/or changing the optimization method and using metaheuristic optimization methods to find the best-suited parameters (e.g., utilizing a genetic algorithm, particle swarm, or teacher–learner optimization methods) [38]. In the end, as with previous datasets, the future forecasts based on the trained model parameters were conducted, and the results are displayed in Figure 12. Shaded areas in this figure are the out-of-sample forecasts (12-step-ahead (*fs12*)) from 1 November 2020 to 1 October 2021.
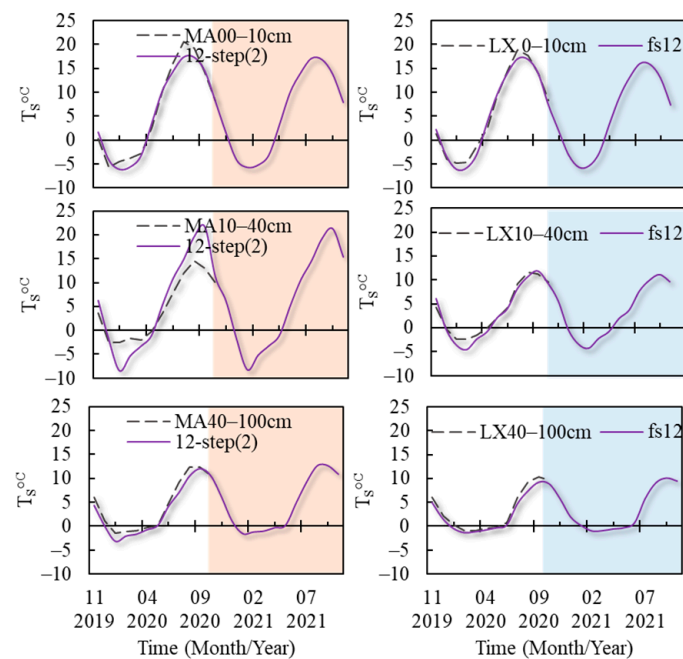
Figure 12. The future forecasts are based on the trained model parameters; shaded areas are the out-of-sample forecasts (12-step-ahead (*fs12*)) from 1 November 2020 to 1 October 2021.

7. Conclusions

Soil temperature (T_S) is crucial in environmental processes. However, obtaining high-quality field data is challenging due to resource constraints. Remotely sensed data and model-based simulations have been used as alternatives; however, they have limitations such as missing data and temporal lags. The FLDAS project addresses this issue by providing reliable data collection for regions without soil temperature measurements, although it still has a temporal lag. This study focused on T_S at two locations (QC and MC) across different soil layers. An SS method was developed to model the time series, and the proposed techniques were validated through 1-, 6-, and 12-step-ahead forecasts of T_S data. The results showed that the proposed SS model could fill in the gap with an average accuracy of R^2 of 0.88 and RMSE of 2.073 °C for the dynamic model, R^2 of 0.834 and RMSE of 2.979 °C for 6-steps-ahead, and R^2 of 0.921 and RMSE of 1.865 °C for different depths at the QC and MC locations, respectively. The SS method can adjust its parameters as the model progresses, resulting in increased accuracy with more forecasting steps. However, the model's level component is sensitive to extremes, and the SS seasonal parameter is not adequately incorporated, leading to over- and underestimations. Future studies should explore alternative optimization methods and different types of time series to address this issue. Additionally, a trend assessment revealed a significant warming trend in the datasets of both locations. It emphasizes the need for a close monitoring of T_S in these areas and consideration of influencing factors such as groundwater level, surface soil moisture, air temperature, and evapotranspiration. The proposed methodology does not require data preprocessing, as the SS model performs multiple preprocessing steps simultaneously, and it can be applied to any study location using the FLDAS. However, a major drawback is its reliance on historical data, and the FLDAS only provides monthly datasets starting from 1982. It is recommended to additionally model and compare the FLDAS data with other linear and machine learning models.

Appendix A

Algorithm A1. GEE T_S extraction pseudocode.

1	Define the point (**ee.Geometry.Point(lng, lat)**)
2	Set start time (**ee.Date('1982-01-01')**)
3	Set end time (**ee.Date('any date)**)
4	Calculate the number of months to process (**ee.Number(endDate-startDate)**)
5	Define the path by using Earth engine snippet (**ee.ImageCollection ('NASA/FLDAS/NOAH01/C/GL/M/V001')**)
6	Use each one of the following band names to access the desired data: **'SoilTemp00_10cm_tavg'** **'SoilTemp10_40cm_tavg'** **'SoilTemp40_100cm_tavg'**
7	Generate a sequence of numbers from start to end by (**ee.List.sequence(start, end, step,** *count***)**)
8	Map an algorithm over a collection by (**map(algorithm)**) The algorithm:
9	Calculate the offset from start date (**.advance(n,'month')**)
10	Advance just one month (**.advance(1,'month')**)
11	Filter the collection by a date range (**.filterDate(start,end)**)
12	Print the results

References

1. Iwata, Y.; Nemoto, M.; Hasegawa, S.; Yanai, Y.; Kuwao, K.; Hirota, T. Influence of rain, air temperature, and snow cover on subsequent spring-snowmelt infiltration into thin frozen soil layer in northern Japan. *J. Hydrol.* **2011**, *401*, 165–176. [CrossRef]
2. Panahi, M.; Khosravi, K.; Ahmad, S.; Panahi, S.; Heddam, S.; Melesse, A.M.; Omidvar, E.; Lee, C.-W. Cumulative infiltration and infiltration rate prediction using optimized deep learning algorithms: A study in Western Iran. *J. Hydrol. Reg. Stud.* **2021**, *35*, 100825. [CrossRef]
3. Jaynes, D.B. Temperature Variations Effect on Field-Measured Infiltration. *Soil Sci. Soc. Am. J.* **1990**, *54*, 305–312. [CrossRef]
4. Genxu, W.; Tianxu, M.; Juan, C.; Chunlin, S.; Kewei, H. Processes of runoff generation operating during the spring and autumn seasons in a permafrost catchment on semi-arid plateaus. *J. Hydrol.* **2017**, *550*, 307–317. [CrossRef]

5. Mittelbach, H.; Lehner, I.; Seneviratne, S.I. Comparison of four soil moisture sensor types under field conditions in Switzerland. *J. Hydrol.* **2012**, *430–431*, 39–49. [CrossRef]

6. Nanda, A.; Sen, S.; Sharma, A.N.; Sudheer, K.P. Soil Temperature Dynamics at Hillslope Scale—Field Observation and Machine Learning-Based Approach. *Water* **2020**, *12*, 713. [CrossRef]

7. Schindlbacher, A. Effects of soil moisture and temperature on NO, NO_2, and N_2O emissions from European forest soils. *J. Geophys. Res.* **2004**, *109*, 1137. [CrossRef]

8. Wei, S.; Zhang, X.; McLaughlin, N.B.; Liang, A.; Jia, S.; Chen, X.; Chen, X. Effect of soil temperature and soil moisture on CO2 flux from eroded landscape positions on black soil in Northeast China. *Soil Tillage Res.* **2014**, *144*, 119–125. [CrossRef]

9. Sauer, T.J.; Horton, R. Soil Heat Flux. In *Micrometeorology in Agricultural Systems*; Hatfield, J.L., Baker, J.M., Eds.; American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America: Madison, WI, USA, 2005; pp. 131–154. ISBN 9780891182689.

10. Shirvani, A.; Moradi-Choghamarani, F.; Zand-Parsa, S.; Moosavi, A.A. Analysis of long-term trends in air and soil temperature in a semi-arid region in Iran. *Environ. Earth Sci.* **2018**, *77*, 173. [CrossRef]

11. Hu, Q.; Feng, S. A Daily Soil Temperature Dataset and Soil Temperature Climatology of the Contiguous United States. *J. Appl. Meteorol.* **2003**, *42*, 1139–1156. [CrossRef]

12. Bilgili, M. Prediction of soil temperature using regression and artificial neural network models. *Meteorol. Atmos. Phys.* **2010**, *110*, 59–70. [CrossRef]

13. Huang, P.M.; Li, Y.; Sumner, M.E. *Handbook of Soil Sciences: Properties and Processes*; CRC Press: Boca Raton, FL, USA, 2011; ISBN 1439803064.

14. Dong, X.; Xu, W.; Zhang, Y.; Leskovar, D. Effect of Irrigation Timing on Root Zone Soil Temperature, Root Growth and Grain Yield and Chemical Composition in Corn. *Agronomy* **2016**, *6*, 34. [CrossRef]

15. Tabari, H.; Sabziparvar, A.-A.; Ahmadi, M. Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. *Meteorol. Atmos. Phys.* **2011**, *110*, 135–142. [CrossRef]

16. Kim, S.; Singh, V.P. Modeling daily soil temperature using data-driven models and spatial distribution. *Theor. Appl. Clim.* **2014**, *118*, 465–479. [CrossRef]

17. Kisi, O.; Sanikhani, H.; Cobaner, M. Soil temperature modeling at different depths using neuro-fuzzy, neural network, and genetic programming techniques. *Theor. Appl. Clim.* **2017**, *129*, 833–848. [CrossRef]

18. Mehdizadeh, S.; Behmanesh, J.; Khalili, K. Evaluating the performance of artificial intelligence methods for estimation of monthly mean soil temperature without using meteorological data. *Environ. Earth Sci.* **2017**, *76*, 59. [CrossRef]

19. Nahvi, B.; Habibi, J.; Mohammadi, K.; Shamshirband, S.; Al Razgan, O.S. Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature. *Comput. Electron. Agric.* **2016**, *124*, 150–160. [CrossRef]

20. Ebtehaj, I.; Bonakdari, H.; Samui, P.; Gharabaghi, B. Multi-depth daily soil temperature modeling: Meteorological variables or time series? *Theor. Appl. Clim.* **2023**, *151*, 989–1012. [CrossRef]

21. Post, J.; Hattermann, F.F.; Krysanova, V.; Suckow, F. Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. *Environ. Model. Softw.* **2008**, *23*, 125–138. [CrossRef]

22. Zeynoddin, M.; Bonakdari, H.; Ebtehaj, I.; Esmaeilbeiki, F.; Gharabaghi, B.; Haghi, D.Z. A reliable linear stochastic daily soil temperature forecast model. *Soil Tillage Res.* **2019**, *189*, 73–87. [CrossRef]

23. O'Neill, P.; Bindlish, R.; Chan, S.; Njoku, E.; Jackson, T. Algorithm Theoretical Basis Document. Level 2 & 3 Soil Moisture (Passive) Data Products 2018. Available online: https://smap.jpl.nasa.gov/system/internal_resources/details/original/316_L2_SM_P_ATBD_v7_Sep2015.pdf (accessed on 1 January 2023).

24. He, M.; Hu, Y.; Chen, N.; Wang, D.; Huang, J.; Stamnes, K. High cloud coverage over melted areas dominates the impact of clouds on the albedo feedback in the Arctic. *Sci. Rep.* **2019**, *9*, 9529. [CrossRef] [PubMed]

25. Sarafanov, M.; Kazakov, E.; Nikitin, N.O.; Kalyuzhnaya, A.V. A Machine Learning Approach for Remote Sensing Data Gap-Filling with Open-Source Implementation: An Example Regarding Land Surface Temperature, Surface Albedo and NDVI. *Remote Sens.* **2020**, *12*, 3865. [CrossRef]

26. Weiss, D.J.; Atkinson, P.M.; Bhatt, S.; Mappin, B.; Hay, S.I.; Gething, P.W. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS J. Photogramm. Remote Sens.* **2014**, *98*, 106–118. [CrossRef] [PubMed]

27. Eskelson, B.N.I.; Temesgen, H.; Lemay, V.; Barrett, T.M.; Crookston, N.L.; Hudak, A.T. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. *Scand. J. For. Res.* **2009**, *24*, 235–246. [CrossRef]

28. Bhattacharjee, S.; Mitra, P.; Ghosh, S.K. Spatial Interpolation to Predict Missing Attributes in GIS Using Semantic Kriging. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4771–4780. [CrossRef]

29. Bărbulescu, A.; Șerban, C.; Indrecan, M.-L. Computing the Beta Parameter in IDW Interpolation by Using a Genetic Algorithm. *Water* **2021**, *13*, 863. [CrossRef]

30. Lopes Martins, L.; Martins, W.A.; Rodrigues, I.C.D.A.; Freitas Xavier, A.C.; Moraes, J.F.L.D.; Blain, G.C. Gap-filling of daily precipitation and streamflow time series: A method comparison at random and sequential gaps. *Hydrol. Sci. J.* **2023**, *68*, 148–160. [CrossRef]

31. Chen, Y.-Y.; Chu, C.-R.; Li, M.-H. A gap-filling model for eddy covariance latent heat flux: Estimating evapotranspiration of a subtropical seasonal evergreen broad-leaved forest as an example. *J. Hydrol.* **2012**, *468–469*, 101–110. [CrossRef]

32. Cagueró-Escofet, M.À.; García, D.; Quevedo, J.; Puig, V.; Espin, S.; Roquet, J. A methodology and a software tool for sensor data validation/reconstruction: Application to the Catalonia regional water network. *Control Eng. Pract.* **2016**, *49*, 159–172. [CrossRef]

33. Vijayakumar, N.; Vennila, S. A comparative analysis of forecasting reservoir inflow using ARMA model and Holt winters exponential smoothening technique. *Int. J. Innov. Sci. Math.* **2016**, *4*, 85–90.

34. Heydari, M.; Benisi Ghadim, H.; Rashidi, M.; Noori, M. Application of Holt-Winters Time Series Models for Predicting Climatic Parameters (Case Study: Robat Garah-Bil Station, Iran). *Pol. J. Environ. Stud.* **2020**, *29*, 617–627. [CrossRef]

35. Puah, Y.J.; Huang, Y.F.; Chua, K.C.; Lee, T.S. River catchment rainfall series analysis using additive Holt–Winters method. *J. Earth Syst. Sci.* **2016**, *125*, 269–283. [CrossRef]

36. Wahyuningsih, S.; Goejantoro, R.; Rizki, N.A. Forecasting hotspots in East Kutai, Kutai Kartanegara, and West Kutai as early warning information. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *144*, 12022. [CrossRef]

37. Zeynoddin, M.; Ebtehaj, I.; Bonakdari, H. Development of a linear based stochastic model for daily soil temperature prediction: One step forward to sustainable agriculture. *Comput. Electron. Agric.* **2020**, *176*, 105636. [CrossRef]

38. Zeynoddin, M.; Bonakdari, H. Structural-optimized sequential deep learning methods for surface soil moisture forecasting, case study Quebec, Canada. *Neural Comput. Appl.* **2022**, *34*, 19895–19921. [CrossRef]

39. McNally, A.; Arsenault, K.; Kumar, S.; Shukla, S.; Peterson, P.; Wang, S.; Funk, C.; Peters-Lidard, C.D.; Verdin, J.P. A land data assimilation system for sub-Saharan Africa food and water security applications. *Sci. Data* **2017**, *4*, 170012. [CrossRef]

40. Hamilton, J. *Time Series Econometrics*; Princeton University Press: Princeton, NJ, USA, 1994.

41. Kwiatkowski, D.; Phillips, P.C.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econom.* **1992**, *54*, 159–178. [CrossRef]

42. Yue, S.; Pilon, P.; Phinney, B.; Cavadias, G. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Process.* **2002**, *16*, 1807–1829. [CrossRef]

43. Tekleab, S.G.; Kassew, A. Hydrologic responses to land use/Land cover change in the Kesem Watershed, Awash basin, Ethiopia. *J. Spat. Hydrol.* **2019**, *15*, 1–31.

44. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422. [CrossRef]

45. Klisch, A.; Atzberger, C. Operational Drought Monitoring in Kenya Using MODIS NDVI Time Series. *Remote Sens.* **2016**, *8*, 267. [CrossRef]

46. Lobell, D.B.; Thau, D.; Seifert, C.; Engle, E.; Little, B. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* **2015**, *164*, 324–333. [CrossRef]

47. Zeynoddin, M.; Bonakdari, H.; Gumiere, S.J.; Caron, J.; Rousseau, A.N. SOILPARAM 1.0: A Global-Scaled Enhanced Remote Sensing Application for Soil Characteristics Data Retrieval–Google Engine Environment, An Open-Source Treasure. In Proceedings of the 39th IAHR World Congress from Snow to Sea, Granada, Spain, 18–23 June 2022; pp. 5309–5319, ISBN 978-90-832612-1-8.

48. Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* **2018**, *5*, 180214. [CrossRef] [PubMed]

49. Taylor, J.W. Exponential smoothing with a damped multiplicative trend. *Int. J. Forecast.* **2003**, *19*, 715–725. [CrossRef]

50. Kalekar, P.S. Time series forecasting using holt-winters exponential smoothing. *Kanwal Rekhi Sch. Inf. Technol.* **2004**, *4329008*, 1–13.

51. Arroyo, J.; San Roque, A.M.; Maté, C.; Sarabia, A. Exponential smoothing methods for interval time series. In Proceedings of the 1st European Symposium on Time Series Prediction, Helsinki, Finland, 7–9 February 2007; pp. 231–240.

52. Goodwin, P. The holt-winters approach to exponential smoothing: 50 years old and going strong. *Foresight* **2010**, *19*, 30–33.

53. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; Otexts: Melbourne, Australia, 2013.

54. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]

55. Singh, V.P.; Frevert, D.K. *Mathematical Models of Small Watershed Hydrology and Applications*; Water Resources Publication: Littleton, CO, USA, 2002; ISBN 1887201351.

56. Qian, B.; Gregorich, E.G.; Gameda, S.; Hopkins, D.W.; Wang, X.L. Observed soil temperature trends associated with climate change in Canada. *J. Geophys. Res.* **2011**, *116*, 1–16. [CrossRef]

57. Chudinova, S.M.; Frauenfeld, O.W.; Barry, R.G.; Zhang, T.; Sorokovikov, V.A. Relationship between air and soil temperature trends and periodicities in the permafrost regions of Russia. *J. Geophys. Res.* **2006**, *111*, 1–15. [CrossRef]

58. Assani, A.A.; Maloney-Dumont, V.; Pothier-Champagne, A.; Kinnard, C.; Quéssy, J.-F. Comparison of the temporal variability of summer temperature and rainfall as it relates to climate indices in southern Quebec (Canada). *Theor. Appl. Clim.* **2019**, *137*, 2425–2435. [CrossRef]

59. Du, J.; Li, C.; Liao, J.; Lhak, P.; Lu, H. Soil temperature change at shallow layer in Lhasa from 1961 to 2005. *Arid Land Geogr.* **2007**, *6*, 826–831.

60. Zhang, T.; Barry, R.G.; Gilichinsky, D.; Bykhovets, S.S.; Sorokovikov, V.A.; Ye, J. An amplified signal of climatic change in soil temperatures during the last century at Irkutsk, Russia. *Clim. Chang.* **2001**, *49*, 41–76. [CrossRef]

61. Al Nakshabandi, G.; Kohnke, H. Thermal conductivity and diffusivity of soils as related to moisture tension and other physical properties. *Agric. Meteorol.* **1965**, *2*, 271–279. [CrossRef]

62. Abu-Hamdeh, N.H. Thermal Properties of Soils as affected by Density and Water Content. *Biosyst. Eng.* **2003**, *86*, 97–102. [CrossRef]

63. Lu, S.; Ju, Z.; Ren, T.; Horton, R. A general approach to estimate soil water content from thermal inertia. *Agric. For. Meteorol.* **2009**, *149*, 1693–1698. [CrossRef]

64. Steele-Dunne, S.C.; Rutten, M.M.; Krzeminska, D.M.; Hausner, M.; Tyler, S.W.; Selker, J.; Bogaard, T.A.; van de Giesen, N.C. Feasibility of soil moisture estimation using passive distributed temperature sensing. *Water Resour. Res.* **2010**, *46*, 234. [CrossRef]

65. Dong, J.; Steele-Dunne, S.C.; Ochsner, T.E.; van de Giesen, N. Determining soil moisture by assimilating soil temperature measurements using the Ensemble Kalman Filter. *Adv. Water Resour.* **2015**, *86*, 340–353. [CrossRef]

66. Zhang, S.-W.; Qiu, C.-J.; Xu, Q. Estimating Soil Water Contents from Soil Temperature Measurements by Using an Adaptive Kalman Filter. *J. Appl. Meteorol.* **2004**, *43*, 379–389. [CrossRef]

67. Svoboda, M.; LeComte, D.; Hayes, M.; Heim, R.; Gleason, K.; Angel, J.; Rippey, B.; Tinker, R.; Palecki, M.; Stooksbury, D.; et al. The Drought Monitor. *Bull. Am. Meteorol. Soc.* **2002**, *83*, 1181–1190. [CrossRef]

68. Fan, Y.; van den Dool, H. Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present. *J. Geophys. Res.* **2004**, *109*, 549. [CrossRef]

69. Turral, H.; Burke, J.; Faurès, J.-M. *Climate Change, Water and Food Security*; Food and Agriculture Organization of the United Nations (FAO): Rome, Italy, 2011; ISBN 9251067953.

70. Zhang, Z.; Pan, Z.; Pan, F.; Zhang, J.; Han, G.; Huang, N.; Wang, J.; Pan, Y.; Wang, Z.; Peng, R. The Change Characteristics and Interactions of Soil Moisture and Temperature in the Farmland in Wuchuan County, Inner Mongolia, China. *Atmosphere* **2020**, *11*, 503. [CrossRef]

71. Pasha, M.F.K.; Srinivasamurthy, N.; Singh, D.; Yeasmin, D.; Valenzuela, G. An Artificial Intelligence Model to Predict Crop Water Requirement Using Weather, Soil Moisture, and Plant Health Monitoring Data. In *World Environmental and Water Resources Congress 2020: Water Resources Planning and Management and Irrigation and Drainage*; American Society of Civil Engineers: Reston, VA, USA, 2020; pp. 9–14.

72. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef]

73. Su, Y.-S.; Ni, C.-F.; Li, W.-C.; Lee, I.-H.; Lin, C.-P. Applying deep learning algorithms to enhance simulations of large-scale groundwater flow in IoTs. *Appl. Soft Comput.* **2020**, *92*, 106298. [CrossRef]

74. Hu, C.; Wu, Q.; Li, H.; Jian, S.; Li, N.; Lou, Z. Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. *Water* **2018**, *10*, 1543. [CrossRef]

75. Xia, Y.; Ek, M.; Sheffield, J.; Livneh, B.; Huang, M.; Wei, H.; Feng, S.; Luo, L.; Meng, J.; Wood, E. Validation of Noah-Simulated Soil Temperature in the North American Land Data Assimilation System Phase 2. *J. Appl. Meteorol. Climatol.* **2013**, *52*, 455–471. [CrossRef]

76. Zheng, D.; van der Velde, R.; Su, Z.; Wen, J.; Wang, X.; Yang, K. Evaluation of Noah Frozen Soil Parameterization for Application to a Tibetan Meadow Ecosystem. *J. Hydrometeor.* **2017**, *18*, 1749–1763. [CrossRef]

77. Wang, L.; Li, X.; Chen, Y.; Yang, K.; Chen, D.; Zhou, J.; Liu, W.; Qi, J.; Huang, J. Validation of the global land data assimilation system based on measurements of soil temperature profiles. *Agric. For. Meteorol.* **2016**, *218–219*, 288–297. [CrossRef]