Forecasting daily soil temperature at different depths based on air temperature using Support Vector Machines, Extreme Gradient Boosting and Artificial Neural Networks,

Does Wavelet Transform have a significant effect on improving evaluation metrics?

By:

Siavash Kazembakhshi

Kazemba1@myumanitoba.ca

NOV 2020

Contents

Table of Figures

Table of Figures

Table of Equations

1- Abstract:

Soil temperature is a very important variable in climate change studies, agricultural meteorology and strongly influences agricultural activities and planning (e.g. the date and depth of sowing crops, frost protection). There are many physically based studies in the literature which model soil temperature, but few are easily applicable for use in the field. Simple and precise short-term forecasting of soil temperature with minimum data requirements is the main goal of this study. The daily data were collected from the St. Adolphe Station (49°40'50.1"N 97°06'48.6"W) from 2011 to 2019. The St. Adolphe station is in Manitoba province in Canada. The minimum, average, and maximum daily soil temperature at 5, 20, 50 and, 100 cm depth of today and two days ahead were forecasted based only on surface air temperatures of six days before and today using support vector machines (SVM), extreme gradient boosting (XGB), and artificial neural network (ANN) methods. In the next step, Wavelet transform (WT) was employed at the data preprocessing stage to see if it can improve the models' performance. The results of this study showed that using Wavelet Transform at data preprocessing stage can significantly improve the all the models performance at all depths. Moreover, the results showed that using wavelet transform at data preprocessing stage, can significantly improve ANN models performance at the depth of 50cm and 100cm.

2- Introduction:

Soil temperature deeply affects many of the physical, chemical and biological processes in soil and is one of the most important variables in agricultural meteorology, soil and geotechnical engineering and the climate change research (*Introduction to Environmental Soil Physics - 1st Edition*, n.d.).

From agricultural point of view, soil temperature not only plays a crucial role in plant growth, especially during germination and seeding emergence, but also large crop yield losses can happen if the soil temperature does not remain in suitable range (Araghi et al., 2017). It should be noticed that various seeds germination needs different soil temperature ranges and most soil organisms thrive at temperature between 25-30C. Moreover, cold soil conditions may prevent liquid water from trees (Mellander et al., 2004). (Domisch et al., 2001) also discovered that, whenever the soil temperature increases, the number of underground microbes will increase significantly which can cause a slight increase in biomass and carbohydrate percentages. They also found that, the new roots' sugar content also will increase when the soil temperature increases.

From climatic change point of view, it has been reported that soil warming has a greater impact on climatic changes than global atmospheric warming (*AR4 Climate Change 2007: Synthesis Report — IPCC*, n.d.). It has been revealed that many active soil layer properties have changed due to the changes in soil temperature which occur in seasons outside of winter. In general soil temperature is required as a variable which can affect capturing sensible and latent heat fluxes, the heat energy from the geothermal system, assessing sea ice and permafrost, determining CO2 and NH4 emissions patterns, microbial decomposition, and rates of organic matter decomposition, mineralization, and plant growth (*AR4 Climate Change 2007: Synthesis Report — IPCC*, n.d.) (WANG et al., 2006).

(Mitchell, n.d.) studied the effects of soil temperature on engineering properties of soil. They investigated the volume and pore water pressure variations in saturated soil due to changes in soil temperature. They showed that, increase in soil temperature can cause a volume of water to drain from saturated soil in fully

drained conditions and constant confining pressure. (Badache et al., 2016) showed that soil temperature is also important parameter in pavement design, pipelines and the installation of high voltage power cable facilities.

Despite a real need for having soil temperature in various fields of science, research in this area is surprisingly limited. Instead of direct measurements, modeling can be used to predict soil temperature in short-term and long-term (Xing et al., 2018). In recent years three types of models have been developed to estimate soil temperature: analytical model, numerical model, and data-driven model.

2-1- Analytical model:

Fourier developed a model that calculate soil temperature as a function of time of a year and depths in the early 18th century (Xing et al., 2018). In this model, Fourier considered the soil surface temperature as the boundary condition and used one-dimensional heat conduction equation while Lord kelvin constructed the same model with higher order harmonics (Narasimhan, 2010). Some research have been carried out to enhance these models accuracy in recent years (Elias et al., 2004)(Droulia et al., 2009). (Thoma et al., 2013)used measured soil surface heat flux as a boundary condition and proposed a similar soil temperature prediction model. He considered only convection and solar radiation at the earth surface in his model. In addition to the convection and solar radiation, (Cleall et al., 2015) (Ouzzane et al., 2015) (Badache et al., 2016) added sky radiation and transpiration of water vapor in their model. These models require some empirical parameters and variables such as annual average soil temperature, soil surface temperature amplitude and phase angel to predict soil temperature more accurately. In laboratory environment, these variables and parameters can be easily obtained using measured soil temperature or measured air temperature. By using this method, soil temperature can be easily obtained at a small case, but at a bigger scale, such as continental or global, another methods must be used due to limited available measured soil temperature or over simplified assumptions for air-soil temperature relations.

2-2- Numerical model:

Soil freezing/melting and snow cover at the ground surface can cause complex heat and mass transport in soil. Numerical methods are capable of elucidate such phenomena while analytical models only can consider heat transfer. Changes in temperature which are followed by changes in soil moisture, induce heat migration. A model was proposed by (Philip & De Vries, 1957) (De Vries, 1958a) (De Vries, 1958b) based on conservation theory of mass and energy which can depict the gas and liquid migration in soil. The moisture evaporation and heat and moisture distributions of the dry soil were simulated by (De Vries, 1958b). Belgheit et al. developed a model for heat transfer in unsaturated soils assuming soil surface temperature as constant. Moreover, by assuming the soil temperature as a sine function, he established a model for mass transfer in unsaturated soils. Due to effect of precipitation on soil moisture Richards also derived equations to describe a theory to unsaturated seepage (Belghit & Benyaich, 2014).

Using the finite element method and with the help of computer science developments, Richards equation was solved. A 2D finite element method was developed by Lam and Fredlund to elucidate the seepage effect, based on the theory of consolidation and water movement in unsaturated soil (Belghit & Benyaich, 2014). Rainfall's effect on thermal and moisture conditions of soil was investigated by Gao et al. He also developed a 1D soil heat and moisture migration model using Richards' equation (Belghit & Benyaich, 2014). Belghit concluded that soil freezing is also one important factor affecting soil heat and moisture transfer (Belghit & Benyaich, 2014). Herb et al investigated how different land covers affect soil surface temperature by simulation. He concluded that, 10 C temperature difference between paved surface and vegetation covered surface (Belghit & Benyaich, 2014). There are also studies that investigate how snow accumulation and melting can have effects on the moisture and heat transfer by considering low thermal conductivity and high emissivity of snow-covered surfaces.

2-3- Data-driven model:

Researchers and scientists have revealed that the data mining technology and mathematical statistics can help us to find a strong correlation between soil temperature and climatic data. Talaee utilized the adaptive neuro-fuzzy inference system (ANFIS) to find a correlation between soil temperature and mean/maximum/minimum air temperature, wind speed and precipitation (Hosseinzadeh Talaee, 2014). Ahmed and Rasul described the relationship between the seasonal daily air temperature and the seasonal daily soil temperature in the Faisalabad region by developing a regression model (Fahim Ahmad & Rasul, 2008). An Artificial Neural Networks (ANN) model was established to simulate the daily soil temperature by Mihalakakou et al (Mihalakakou, 2002). The effect of single-layer and multi-layer neural network in the ANN soil temperature estimation model was investigated by George et al (George, 2001). Bilgili et al. developed models for monthly average soil temperature predictions based on linear regression (LR), nonlinear regression (NLR), ANN methods and analyzed the correlation between the weather parameters and soil temperature. They considered effect of day of year on the soil temperature results of heating and cooling season respectively by grouping the weather parameters as inputs (Bilgili, 2010). Sungwon et al investigated how different input parameter groups can have effects on the accuracy of soil temperature prediction results. He also developed a multilayer perceptron (MLP) and an adaptive neuro fuzzy inference system (ANFIS) model to predict daily average soil temperature (Bilgili, 2010). Various data mining algorithms such as generalized regression neural network, radial basis neural network and multilayer perceptron neural network were compared by Kisi et al. it was concluded that these algorithms based on numerical relationship analysis can be used to estimate monthly average soil temperature (Bilgili, 2010) .

Three various algorithms including adaptive neuro-fuzzy inference systems (ANFISs), ANFIS with grid partition (ANFIS-GP), ANFIS with subtractive clustering (ANFIS-SC), and ANFIS with fuzzy c means (ANFIS-FCM) were used in long-term monthly air temperatures predictions by (Kisi, Demir, et al., 2017), in order to find the most accurate model. The month of the year, and geographical variables

(latitude, longitude, and altitude)) of 72 stations in Turkey were the model inputs. The three ANFIS methods performed better than Artificial Neural Networks (ANNs) and Multiple Linear Regression (MLR) in terms of accuracy and among the ANFIS methods the ANFIS-GP provide superior accuracy to ANFIS-SC and ANFIS-FCM models. They concluded that, long-term monthly air temperatures can be effectively predicted using ANFIS-GP algorithm and geographical inputs.

Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy inference system (ANFIS), and genetic programming (GP) were utilized to develop models to predict soil temperature at different depths by (Kisi, Sanikhani, et al., 2017). In the first part of the study, they compared ANN, ANFIS, and GP models of two stations at the depths of 10, 50, and 100 cm in terms of accuracy. They found that GP outperformed both ANN and ANFIS in estimating monthly soil temperature. In the second part, they investigated the effect of periodicity (month of the year) in models' accuracy. They concluded that, if we consider periodicity as an input in our models, we can increase the accuracy of our models.

Using support vector machine (SVM) algorithm, a new data-driven model was proposed by (Xing et al., 2018). They considered the ground temperature as superposition of annual average ground temperature predictions (long-term climates impact) and daily ground temperature amplitude predictions (short-term climates impact). Annual average soil temperature was calculated by air temperature, solar radiation, wind speed and relative humidity as inputs and daily soil temperature amplitudes was determined by air temperature, solar radiant and day of year as model inputs. They used their model to predict daily soil temperature at 16 sites located in arid or dry summer climates, warm climates and snow climates in united states. The new model's mean absolute error is 1.26C and root mean square error is 1.66C. Meanwhile, traditional SVM model's mean absolute error is 2.20C and root mean square error is 2.91C.

(Citakoglu, 2017) developed various models including the Artificial Neural Networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR), using maximum air temperature, minimum air temperature and mean precipitation as inputs and soil temperature as an output. They measured monthly soil temperature at 5, 10, 20, 50 at100 cm depths below the soil surface at 261

stations in turkey having records of at least 20 years. Mean absolute error (MAE), root mean squared error (RMSE) and determination coefficient (R^2) were used to validate the models. ANFIS (RMSE 1.99; MAE 1.09; R^2 0.98) is found to have a better performance compared to ANN (RMSE 5.80; MAE 1.89; R^2 0.93) and MLR (RMSE 8.89; MAE 2.36; R^2 0.93) in predicting soil temperature.

(Alizamir et al., 2020) established different models including Extreme Learning Machine (ELM), Artificial Neural Networks (ANN), Classification and Regression Trees (CART) and group method of data handling (GMDH) in order to predict soil temperature at 5, 10, ,50 and 100 cm below the ground surface. Different combinations of climatic variables and parameters were employed as model inputs including air temperature, relative humidity, solar radiation, and wind speed to estimate soil temperature. They utilized root mean square error and coefficient of determination to validate their models. Their result show that ELM performs better than the other four models in soil temperature prediction. They also concluded that by increasing the soil depth, the model's accuracy decrease.

("Journal of Geophysical Research," 1955) believed that most of the research that had been carried out to estimate soil temperature, focused on small spatial resolutions and modeling frameworks intended for higher spatial resolutions (much finer than 1 km^2) were lacking across areas of complex topography. Therefore, they proposed a simple modeling framework for estimating the soil temperature at high (i.e., 5*5m) spatial resolutions and high temporal (i.e., 1 h steps) resolutions in mountainous terrain using a few discrete air temperatures. They concluded that the spatial distribution of soil temperature can be simulated efficiently by using this approach with a root-mean-square error ranging from 2.1 to 2.9°C. Moreover, they declared that their approach could predict the daily and monthly variability of soil temperature well.

The effects of canopy, topography and ground litter were incorporated to develop an hybrid soil temperature model to predict daily spatial patterns of soil temperature in a forested landscape by (Kang et al., 2000). The model was originated from both empirical relationship between air and soil temperature, and heat transfer physics. Its inputs variables are extracted from a digital elevation model (DEM),

standard weather records and satellite imagery. Model-predicted soil temperatures fitted well with data measured at 10 cm soil depth at three sites: two hardwood forests and a bare soil area. By doing sensitivity analysis, it was revealed that the model was highly sensitive to leaf area index (LAI) and air temperature. Their results showed that spatial variability of soil temperature across landscape crucially depends on site-specific surface structures such as LAI and ground litter stores.

(Zheng et al., 1993) employed a linear regression model to estimate the daily mean soil temperature at depth of 10 cm using an 11-day running average of daily mean air temperature. After performing frequency analysis for 17 of 19 data sets, they concluded that between 77% and 96% of data were within 3.5C range centered on the measured soil temperature. Their results also showed that changes of soil temperature under snow cover were smaller than those without snow cover.

In this study, (Delbari et al., 2019) developed support vector regression (SVR)-based model in predicting daily soil temperature at 10, 30 and 100 cm depth at different climate conditions over Iran. They consider minimum air temperature, maximum air temperature, solar radiation, relative humidity, dew point, and the atmospheric pressure from five different stations as inputs. After performing correlation sensitivity for the input combinations and the effect of periodicity, they compared the obtained results to the result of a multiple linear regression (MLR) model. Their study showed that, both ANN and MLR models performed quite well at 10-cm depth, but at deeper layers especially 100cm depth, the SVR performed better than MLR.

In analytical and numerical models, the effects of various weather conditions on soil heat and mass transfer are modeled to calculate the ground temperature. These models are obtained using some complicated physical equations and analysis. As a result, not only the model computational cost is high but also the model development process is time consuming. On the other hand, data-driven models are originated from simple mathematical relations of weather inputs and soil temperature outputs.

Therefore, they are easy to build and requires less inputs and much less computational time. In overall, the data-driven models are more suitable for a wide variety engineering application and therefore, the main focus of this project is to build a data-driven model for soil temperature prediction.

3- Methods:

3-1 Support Vector Machines Regression (SVM-SVR):

(Cortes & Vapnik, 1995) developed SVM with the capabilities of wide variety of applications in machine learning domain which has two main branches that are called support vector classification (SVC) and support vector regression (SVR). SVR is employed in this study.

The SVR model uses the following approximation functions to predict a new output, where y denotes the new output, ω denotes the calculated weights, x is the high dimensional feature space, and ϕ is a Kernel function which will be discussed, and finally b which is intercept:

$$y = \omega . \phi (x) + b$$

Equation 3-1

The ω is obtained from minimizing the following equation in which C is the regularization parameter, ε and ε* are the positive and negative relaxation constant.

$$1/2||\omega|| + C \sum_{i=1}^{M} (\varepsilon_i + \varepsilon_i^*)$$

Equation 3-2

The above equation should be minimized subject to two conditions:

$$y_i - \omega . \phi (x) - b < \varepsilon + \varepsilon_i$$

Equation 3-3

$$\omega . \phi (x) + b - y_i < \varepsilon + \varepsilon_i^*$$

Equation 3-4

as can be seen, we have encountered a constrained optimization problem and therefore, the Lagrangian method must be employed to solve it.

As was mentioned, ϕ is a Kernel function which is basically a mathematical method of increasing the dimension of the input data features, with the intension to place the problem in a higher dimensional space, in which the problem may be solved by a linear equation.

3-2 Extreme gradient boosted trees (XGBoost):

The XGBoost is an ensemble tree algorithm that was firstly developed by (Chen & Guestrin, 2016), after that it was improved using gradient boosting (GB) decision method by (Friedman, 2002). It can deal with both classification and regression problems. The XGBoost is described as below:

Let $D = \{(x_i, y_i)\}$ is a dataset including of m samples as well as n features. The suggested tree ensemble model uses z additive functions for approximation the system response as:

$$\hat{y} = \phi(x_i) = \sum_{z=1}^{Z} f_z(x_i),\ f_z \in F$$

Equation 3-5

In which F is the space of regression trees.

The objective function of the XGBoost is supposed to get minimum defined as follows:

$$L^t = \sum_{i=1}^{M} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right)$$

Equation 3-6

In the equation above, l denotes the convex function (i.e. loss function) which is applied to determine the difference between exact and calculated values, y_i is considered as a measured value, $\hat{y}_i$ indicates the output value. For minimizing the errors, the number of iteration (t) is used, whereas Ω is the penalty factor for the complication of the regression tree approach.

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda||\omega||$$

Equation 3-7

3-3 Artificial Neural Networks (ANNs):

Artificial Neural Networks (ANNs) are one of the most commonly used method of artificial intelligence which has affected many areas of studies in recent years. They are also highly applicable to hydrology and meteorology for modeling and prediction (Hsu et al., 1995). The main structure of the ANNs is very similar to the human brain which can recognize patterns and learn from examples (Silverman & Dracup, 2000). In general, an ANN is constructed of a set of joined elements which are called neurons, with each element expressed by an activation function of sum of weighted inputs:

$$Y_i = f_i\left(\sum_i W_{ij}X_j + \theta_i\right)$$

Equation 3-8

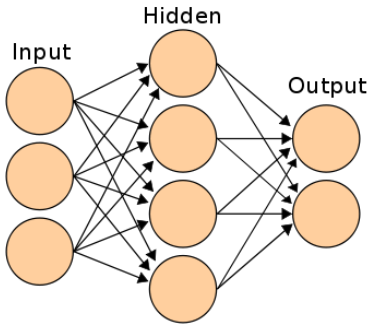
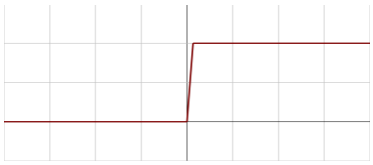


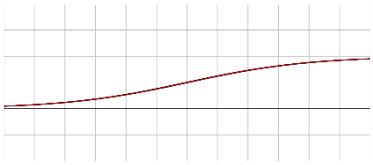
Figure 3-1 ANN Structure

Where Y_i is the output value, f_i is the activation function of the i^{th} neuron, W_{ij} is the connection weight between the i^{th} and j^{th} neuron and θ_i is the bias of i^{th} neuron. There are many various activation functions which the most used ones are defined as follows:

- Binary step function: $f(x) = \begin{cases} 0 & for\ x < 0 \\ 1 & for\ x \geq 0 \end{cases}$

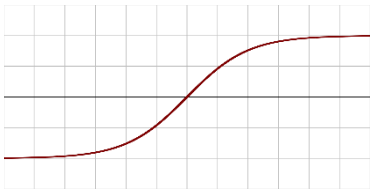


- Sigmoid function (Logistic function): $f(x) = \frac{1}{1+e^{-x}}$



The sigmoid function is the ideal activation function for binary classification.

- Hyperbolic tangent function: $f(x) = \tanh(x) = \frac{(e^{x}-e^{-x})}{(e^{x}+e^{-x})}$



- Rectified linear unit (RELU): $f(x) = \begin{cases} 0 & for\ x \leq 0 \\ x & for\ x > 0 \end{cases}$

ANNs are built of input, hidden and output layers, and each layer is connected the next layer (Figure 3). The values obtained from the output layer are compared to the target values and, based on output and target values, error functions or indices such as the root mean square error (RMSE) are calculated to find the optimal values for weights and biases of the ANN:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - Y_i)^2}$$

Equation 3-9

Where Y is the output of the model, T is the target value and N is the data point number. The main goal of the ANN model is to minimize the error function which is done by optimization algorithms. There are many optimization algorithms that can minimize the error function.

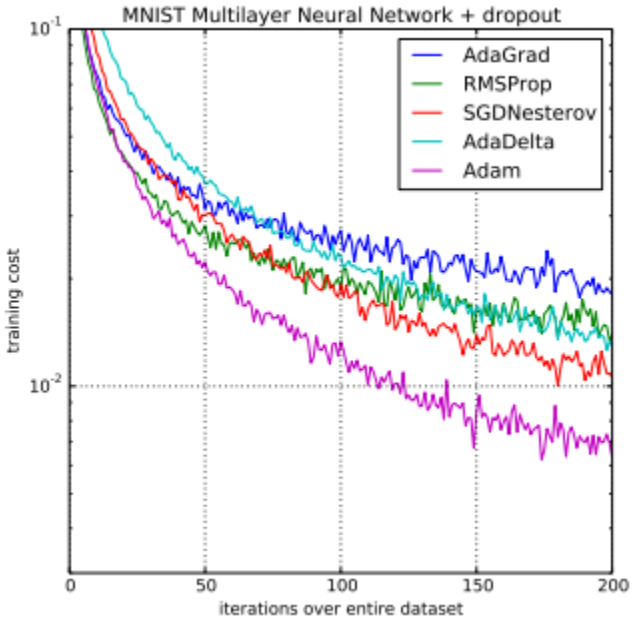


Figure 3-2: Comparison of Optimization Algorithms Training ANNs

In figure 4, we can wee that Adam optimization algorithm can achieve good results fast.

3-4 Wavelet transform (WT):

Studying meteorological variables in a frequency domain can have many advantages, since uniform or non-uniform oscillations are one of the natural attributes for many atmospheric phenomena and variables (Rakhecha et al., 2009). The Fourier transform was one of the first methods for studying fluctuations in signals, but it has limitations especially when the signal or time series has a non-uniform structure over time. The WT is an advanced modification of the short time Fourier transform in which the window

function is completely flexible and can be changed over time based on the shape and compactness of the signal or time series(*A First Course on Wavelets - Eugenio Hernandez, Guido Weiss - Google Books*, n.d.). The continuous WT is defined as follows (Lau & Hengyi Weng, 1995):

$$W(s,\alpha) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} x(t)\ \psi^*(\frac{t-\ \alpha}{s})dt$$

Equation 3-10

$$W_{(s,\alpha)}(t) = \frac{1}{\sqrt{s}}\psi\frac{t-\alpha}{s}$$

Equation 3-11

Where $W(s,\alpha)$ is the WT with scale s and time shift α; ψ represents the wavelet function and ψ^* denotes the complex conjugate. The flexibility of WT is founded on variations in s (Mallat, 2009). The other type of WT is a discrete WT (DWT), in which the scale is dyadic and, accordingly, the calculation process can be simplified although the results will be sufficiently accurate. The DWT is defined as follows (Partal & Küçük, 2006):

$$W(s,\alpha) = \frac{1}{2^{s/2}}\sum_{0}^{N-1} x(t)\,\psi(\frac{t}{2^s}-\alpha)$$

Equation 3-12

On applying the DWT to a time series, it decomposes into two new ancillary time series, called the approximation (A) and detail (D) components. Components A and D show the low and high frequencies, respectively, of the original time series. This decomposition process can be iterated at several levels, and component A is broken down into new A and D components at each decomposition level. In addition, there are very different wavelet functions (Mallat, 2009), but the Daubechies (db) is one of the most commonly used. There are different types of db, called dbN, such as db1 to db10. Increasing the number means more complexity in the shape and details of that db wavelet function type.

The DWT was employed in this study as a preprocessing step for the SVM, XGBoost and ANN models and therefore these models were named WSVM, WXGBoost and WANN.

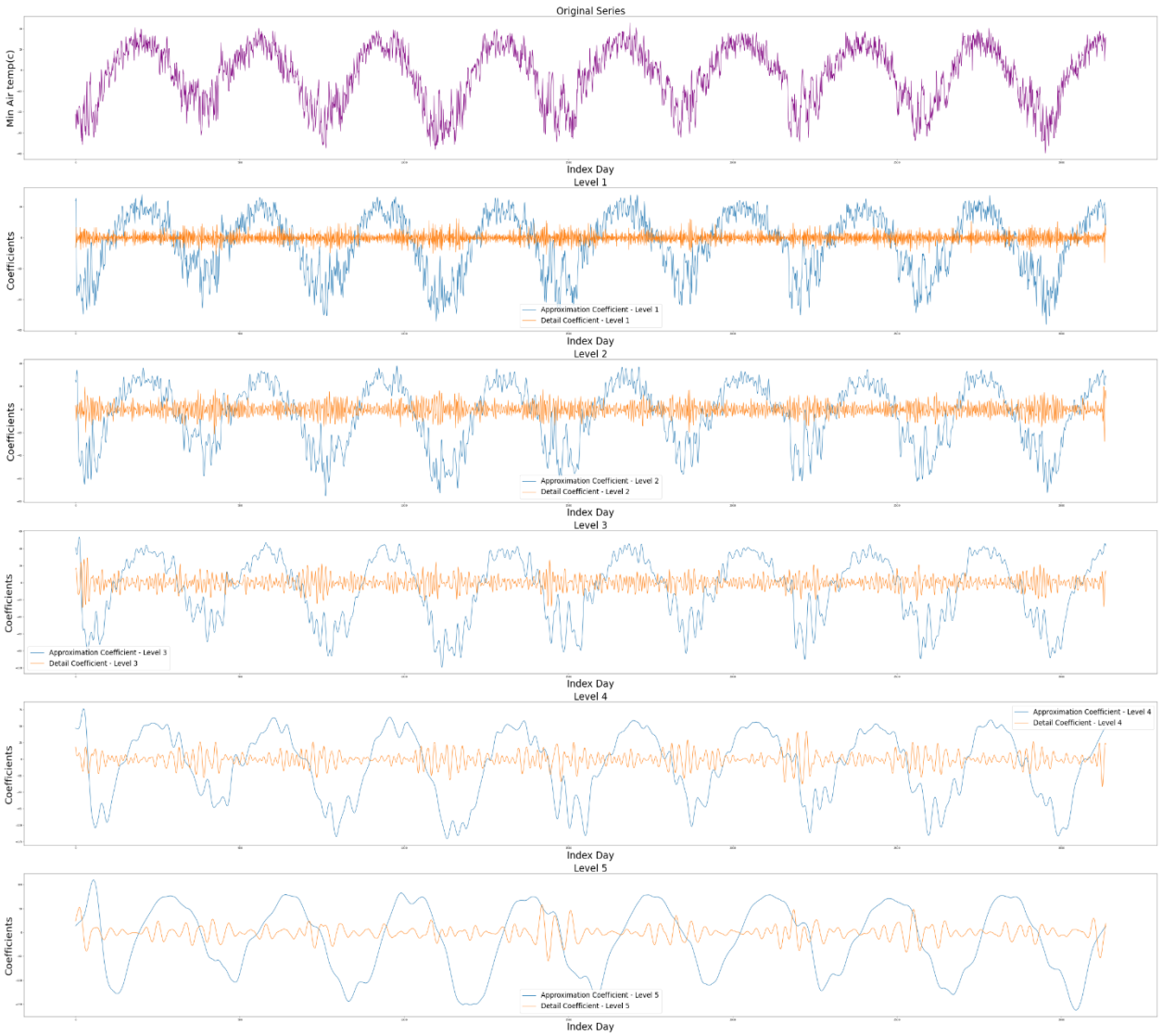


Figure 3-3: Wavelet transform – db5-5 levels

4- Materials:

Minimum, average and maximum daily soil and air temperature at depths of 5, 20, 50 and 100 cm below the surface were measured at the St. Adolphe Station (49°40'50.1"N 97°06'48.6"W) from 2011 to 2019. The St. Adolphe station is in Manitoba province in Canada. In Manitoba, the climate is moderately dry with sharp seasonal temperature changes. Winter temperatures of about −40 °F (−40 °C) may occasionally occur in any part of the province, and summer days of 100 °F (38 °C) are not unusual in the southern regions. The soil and air temperature were measured every minute in the station and the minimum, average and maximum amount were calculated. In this study, minimum, average and maximum air temperature today and six days before were selected as the inputs and minimum, average and maximum soil temperature at the depths of 5, 20, 50 and 100 cm of today and two days ahead were chosen to be forecasting targets. Here is the description of variable names:

AvgAir_T_TempC_Minus_6	average air temperature of 24 hours of 1 min avg's of six days ago deg C (°C)
MaxAir_T_TempC_Minus_6	maximum air temperature of 24 hours of 1 min avg's of six days ago deg C (°C)
MinAir_T_TempC_Minus_6	minimum air temperature of 24 hours of 1 min avg's of six days ago deg C (°C)
AvgAir_T_TempC_Minus_5	average air temperature of 24 hours of 1 min avg's of five days ago deg C (°C)
MaxAir_T_TempC_Minus_5	maximum air temperature of 24 hours of 1 min avg's of five days ago deg C (°C)
MinAir_T_TempC_Minus_5	minimum air temperature of 24 hours of 1 min avg's of five days ago deg C (°C)
AvgAir_T_TempC_Minus_4	average air temperature of 24 hours of 1 min avg's of four days ago deg C (°C)
MaxAir_T_TempC_Minus_4	maximum air temperature of 24 hours of 1 min avg's of four days ago deg C (°C)
MinAir_T_TempC_Minus_4	minimum air temperature of 24 hours of 1 min avg's of four days ago deg C (°C)
AvgAir_T_TempC_Minus_3	average air temperature of 24 hours of 1 min avg's of three days ago deg C (°C)
MaxAir_T_TempC_Minus_3	maximum air temperature of 24 hours of 1 min avg's of three days ago deg C (°C)
MinAir_T_TempC_Minus_3	minimum air temperature of 24 hours of 1 min avg's of three days ago deg C (°C)
AvgAir_T_TempC_Minus_2	average air temperature of 24 hours of 1 min avg's of two days ago deg C (°C)
MaxAir_T_TempC_Minus_2	maximum air temperature of 24 hours of 1 min avg's of two days ago deg C (°C)
MinAir_T_TempC_Minus_2	minimum air temperature of 24 hours of 1 min avg's of two days ago deg C (°C)
AvgAir_T_TempC_Minus_1	average air temperature of 24 hours of 1 min avg's of one days ago deg C (°C)
MaxAir_T_TempC_Minus_1	maximum air temperature of 24 hours of 1 min avg's of one days ago deg C (°C)
MinAir_T_TempC_Minus_1	minimum air temperature of 24 hours of 1 min avg's of one days ago deg C (°C)
AvgAir_T_TempC	average air temperature of 24 hours of 1 min avg's of today deg C (°C)
MaxAir_T_TempC	maximum air temperature of 24 hours of 1 min avg's of today deg C (°C)
MinAir_T_TempC	minimum air temperature of 24 hours of 1 min avg's of today deg C (°C)

Figure 4-1: Air temperatures – Inputs variable

Avg_Soil_TP05_TempC	average of 24 hours of 1 min avg's of soil temperature of today at 5 cm deg C (°C)
Max_Soil_TP05_TempC	maximum of 24 hours of 1 min avg's of soil temperature of today at 5 cm deg C (°C)
Min_Soil_TP05_TempC	minimum of 24 hours of 1 min avg's of soil temperature of today at 5 cm deg C (°C)
Avg_Soil_TP05_TempC_Plus_1	average of 24 hours of 1 min avg's of soil temperature of one day ahead at 5 cm deg C (°C)
Max_Soil_TP05_TempC_Plus_1	maximum of 24 hours of 1 min avg's of soil temperature of one day ahead at 5 cm deg C (°C)
Min_Soil_TP05_TempC_Plus_1	minimum of 24 hours of 1 min avg's of soil temperature of one day ahead at 5 cm deg C (°C)
Avg_Soil_TP05_TempC_Plus_2	average of 24 hours of 1 min avg's of soil temperature of two day ahead at 5 cm deg C (°C)
Max_Soil_TP05_TempC_Plus_2	maximum of 24 hours of 1 min avg's of soil temperature of two day ahead at 5 cm deg C (°C)
Min_Soil_TP05_TempC_Plus_2	minimum of 24 hours of 1 min avg's of soil temperature of two day ahead at 5 cm deg C (°C)

Figure 4-2: Soil temperature at depth 5 Cm – output variable

Avg_Soil_TP20_TempC	average of 24 hours of 1 min avg's of soil temperature of today at 20 cm deg C (°C)
Max_Soil_TP20_TempC	maximum of 24 hours of 1 min avg's of soil temperature of today at 20 cm deg C (°C)
Min_Soil_TP20_TempC	minimum of 24 hours of 1 min avg's of soil temperature of today at 20 cm deg C (°C)
Avg_Soil_TP20_TempC_Plus_1	average of 24 hours of 1 min avg's of soil temperature of one day ahead at 20 cm deg C (°C)
Max_Soil_TP20_TempC_Plus_1	maximum of 24 hours of 1 min avg's of soil temperature of one day ahead at 20 cm deg C (°C)
Min_Soil_TP20_TempC_Plus_1	minimum of 24 hours of 1 min avg's of soil temperature of one day ahead at 20 cm deg C (°C)
Avg_Soil_TP20_TempC_Plus_2	average of 24 hours of 1 min avg's of soil temperature of two day ahead at 20 cm deg C (°C)
Max_Soil_TP20_TempC_Plus_2	maximum of 24 hours of 1 min avg's of soil temperature of two day ahead at 20 cm deg C (°C)
Min_Soil_TP20_TempC_Plus_2	minimum of 24 hours of 1 min avg's of soil temperature of two day ahead at 20 cm deg C (°C)

Figure 4-3: Soil temperature at depth 20 Cm - output variable

Avg_Soil_TP50_TempC	average of 24 hours of 1 min avg's of soil temperature of today at 50 cm deg C (°C)
Max_Soil_TP50_TempC	maximum of 24 hours of 1 min avg's of soil temperature of today at 50 cm deg C (°C)
Min_Soil_TP50_TempC	minimum of 24 hours of 1 min avg's of soil temperature of today at 50 cm deg C (°C)
Avg_Soil_TP50_TempC_Plus_1	average of 24 hours of 1 min avg's of soil temperature of one day ahead at 50 cm deg C (°C)
Max_Soil_TP50_TempC_Plus_1	maximum of 24 hours of 1 min avg's of soil temperature of one day ahead at 50 cm deg C (°C)
Min_Soil_TP50_TempC_Plus_1	minimum of 24 hours of 1 min avg's of soil temperature of one day ahead at 50 cm deg C (°C)
Avg_Soil_TP50_TempC_Plus_2	average of 24 hours of 1 min avg's of soil temperature of two day ahead at 50 cm deg C (°C)
Max_Soil_TP50_TempC_Plus_2	maximum of 24 hours of 1 min avg's of soil temperature of two day ahead at 50 cm deg C (°C)
Min_Soil_TP50_TempC_Plus_2	minimum of 24 hours of 1 min avg's of soil temperature of two day ahead at 50 cm deg C (°C)

Figure 4-4: Soil temperature at depth 50 Cm - output variable

Avg_Soil_TP100_TempC	average of 24 hours of 1 min avg's of soil temperature of today at 100 cm deg C (°C)
Max_Soil_TP100_TempC	maximum of 24 hours of 1 min avg's of soil temperature of today at 100 cm deg C (°C)
Min_Soil_TP100_TempC	minimum of 24 hours of 1 min avg's of soil temperature of today at 100 cm deg C (°C)
Avg_Soil_TP100_TempC_Plus_1	average of 24 hours of 1 min avg's of soil temperature of one day ahead at 100 cm deg C (°C)
Max_Soil_TP100_TempC_Plus_1	maximum of 24 hours of 1 min avg's of soil temperature of one day ahead at 100 cm deg C (°C)
Min_Soil_TP100_TempC_Plus_1	minimum of 24 hours of 1 min avg's of soil temperature of one day ahead at 100 cm deg C (°C)
Avg_Soil_TP100_TempC_Plus_2	average of 24 hours of 1 min avg's of soil temperature of two day ahead at 100 cm deg C (°C)
Max_Soil_TP100_TempC_Plus_2	maximum of 24 hours of 1 min avg's of soil temperature of two day ahead at 100 cm deg C (°C)
Min_Soil_TP100_TempC_Plus_2	minimum of 24 hours of 1 min avg's of soil temperature of two day ahead at 100 cm deg C (°C)

Figure 4-5: Soil temperature at depth 100 cm - output variable

Data had some missing values and outliers that need to be dealt with. Some of them were handled using interpolation method and the rest were filled with mean of four previous and four next values.

Handling missing values and outliers using mean: Let x_n be an outlier or a missing value

$$x_n = \frac{x_{n-4} + x_{n-3} + x_{n-2} + x_{n-1} + x_{n+1} + x_{n+2} + x_{n+3} + x_{n+4}}{8}$$

Equation 4-1

Handling missing values and outliers using interpolation: Let x_n be an outlier or a missing value:

$$x_n = x_{n-1} + (t_n - t_{n-1}) \frac{(x_{n+1} - x_{n-1})}{(t_{n+1} - t_{n-1})}$$

Equation 4-2

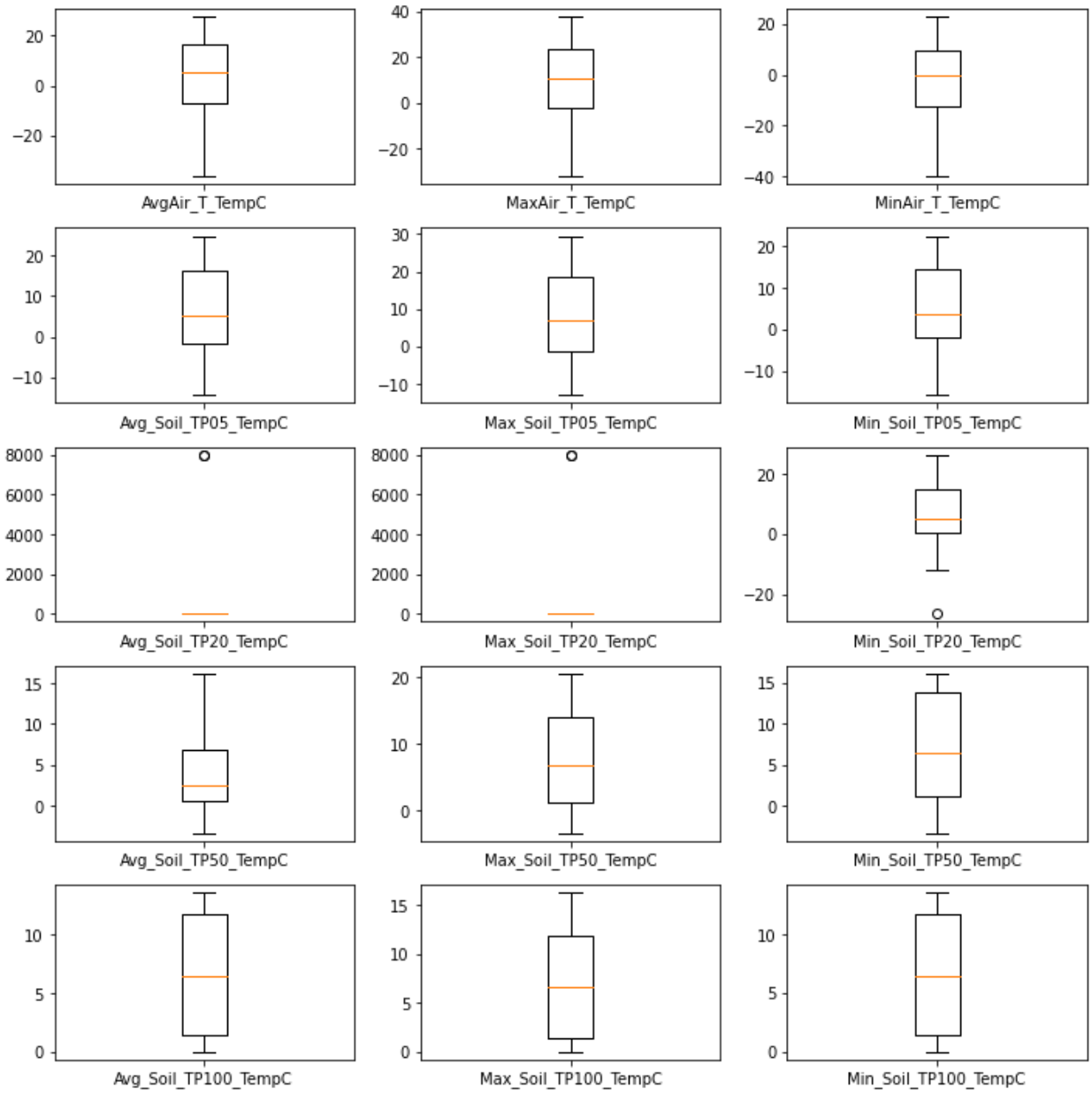


Figure 4-6: Output variables before handling missing and outlier values

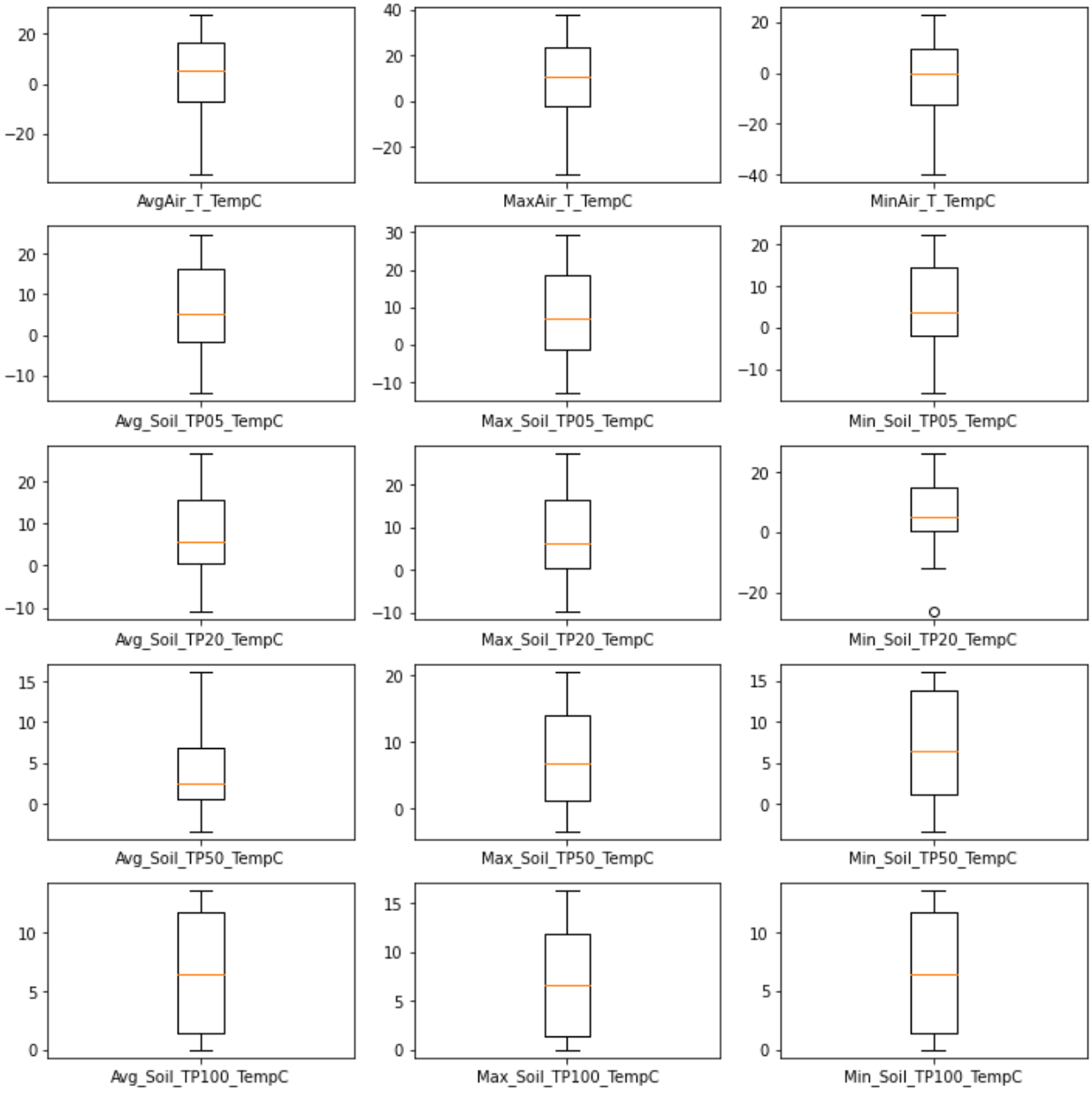


Figure 4-7: output variables after handling missing and outlier values

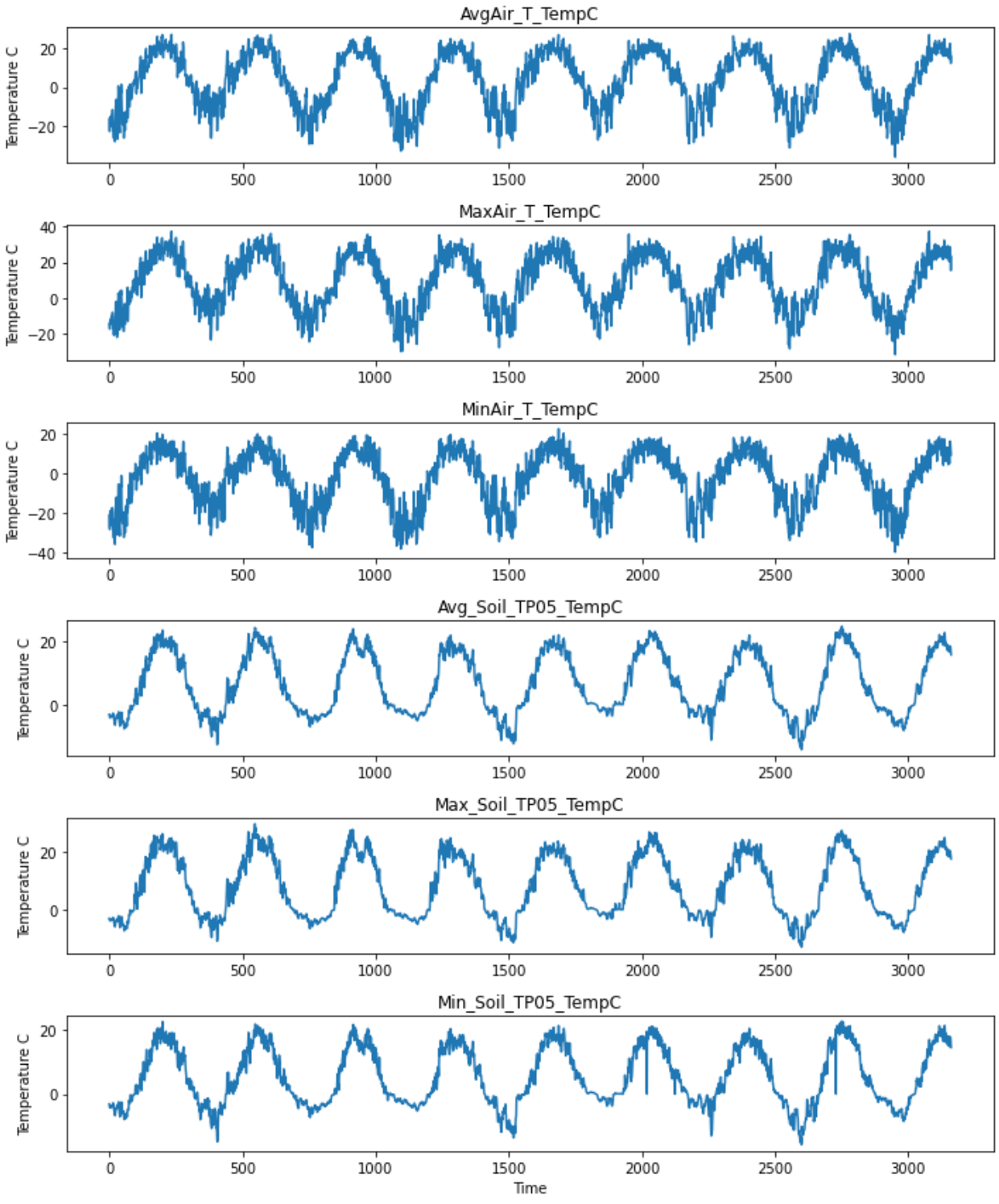


Figure 4-8: Air temperature and Soil temperature at the depth of 5cm

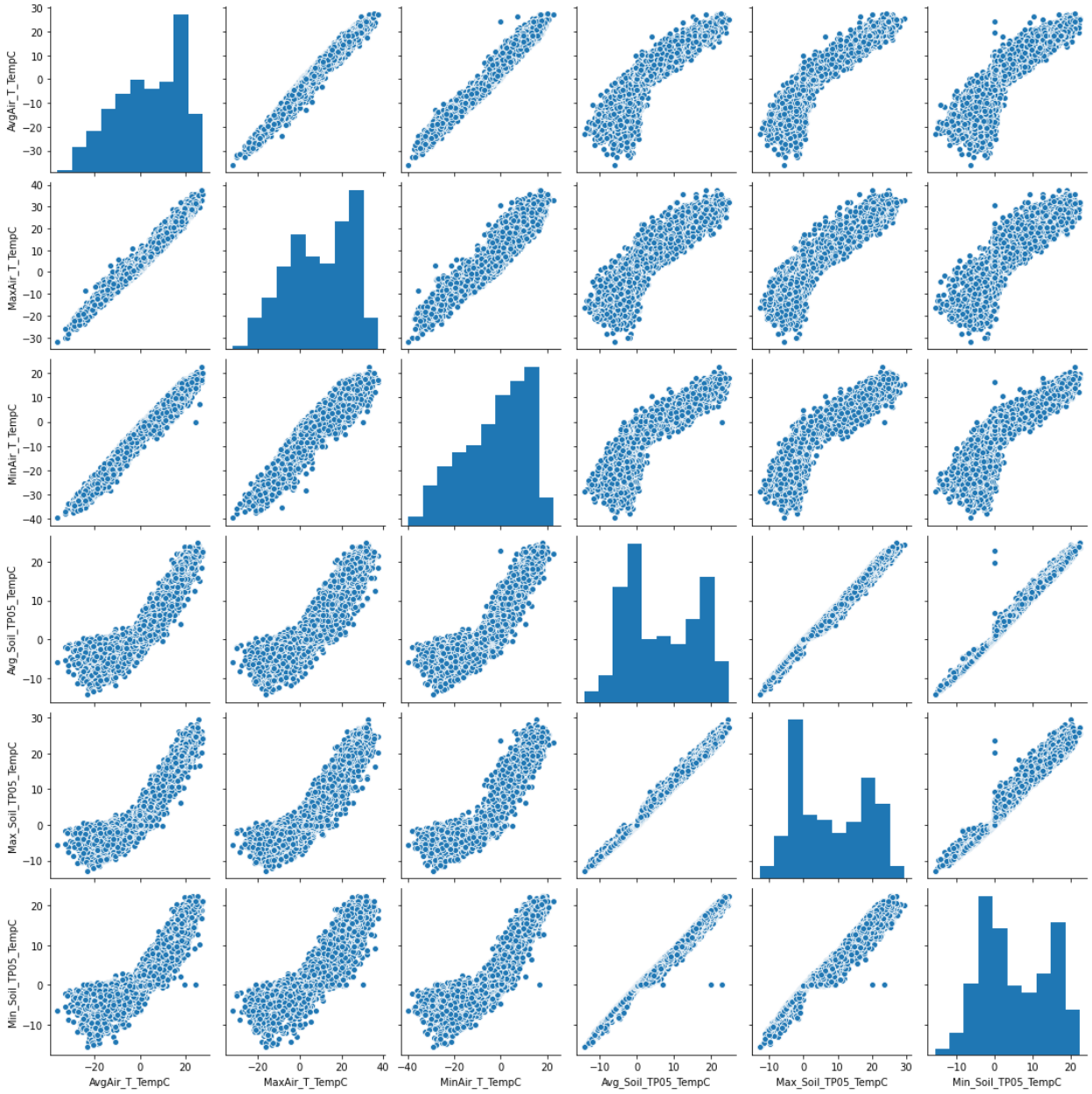


Figure 4-9: Air temperature and Soil temperature at the depth of 5cm - Correlation

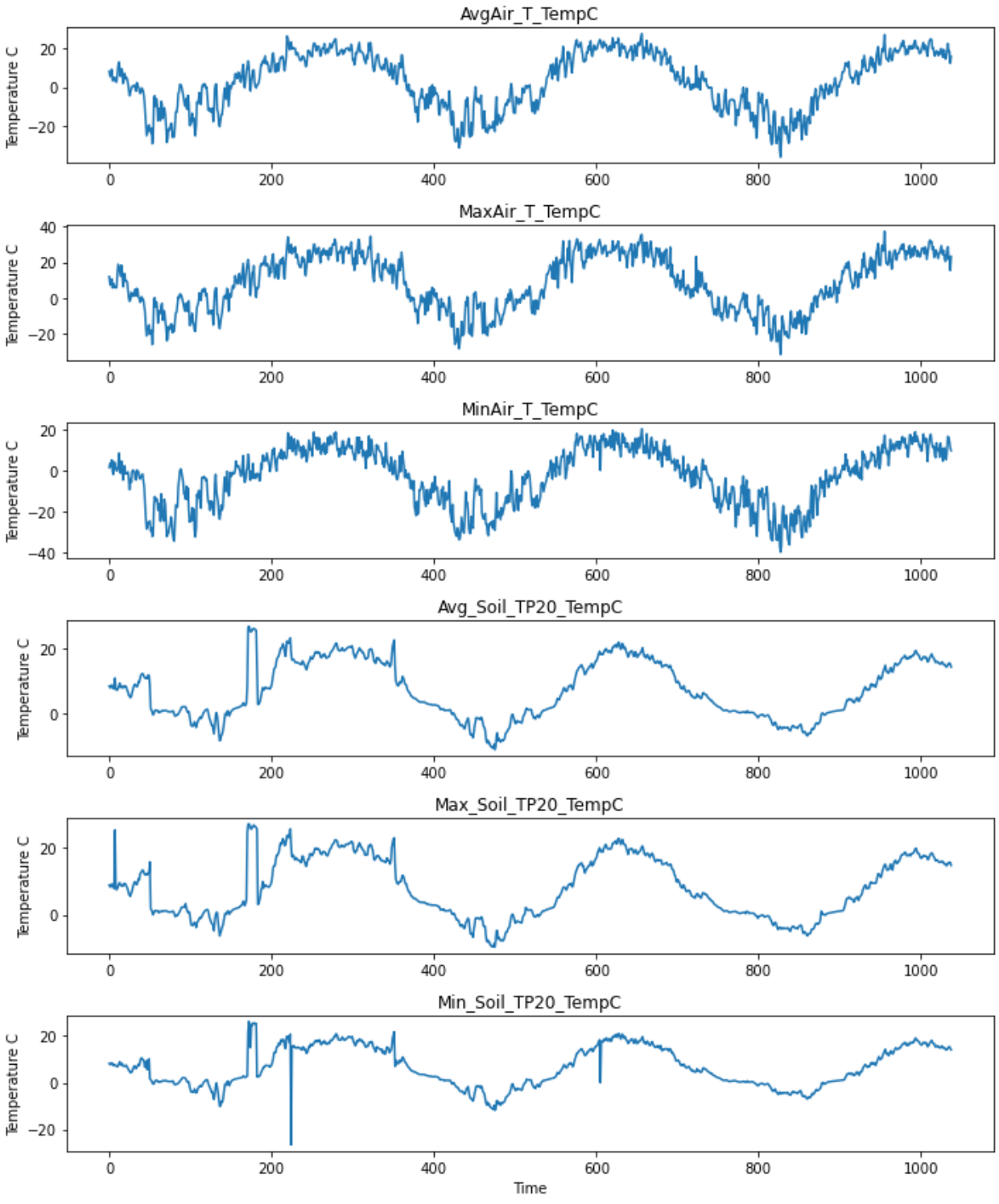


Figure 4-10:Air temperature and Soil temperature at the depth of 20cm - before handling outliers

As can be seen from the figure 13, there are some abrupt changes in the soil temperature which indicate outlier that need to be handled.

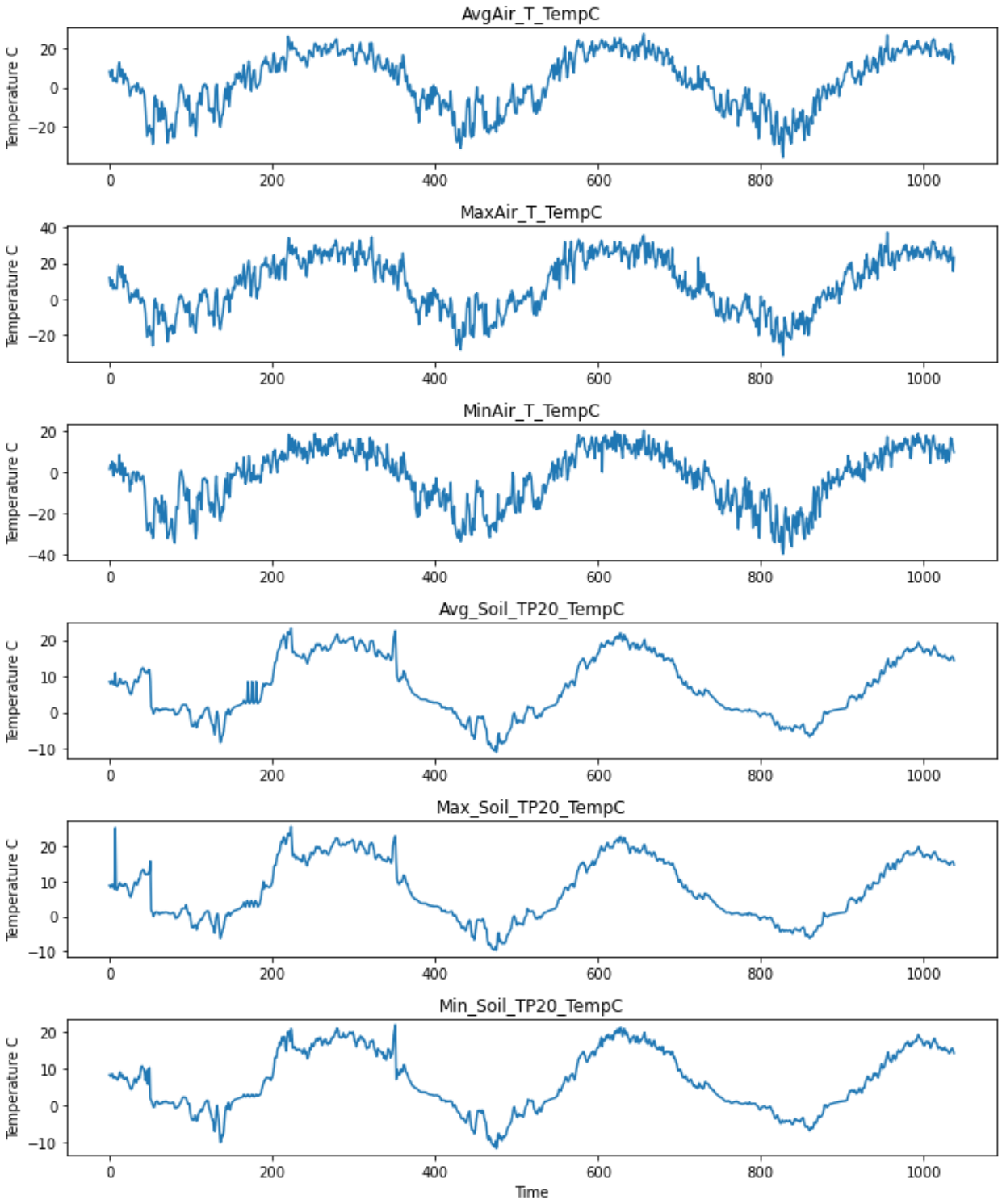


Figure 4-11: Air temperature and Soil temperature at the depth of 20cm- after handling outliers

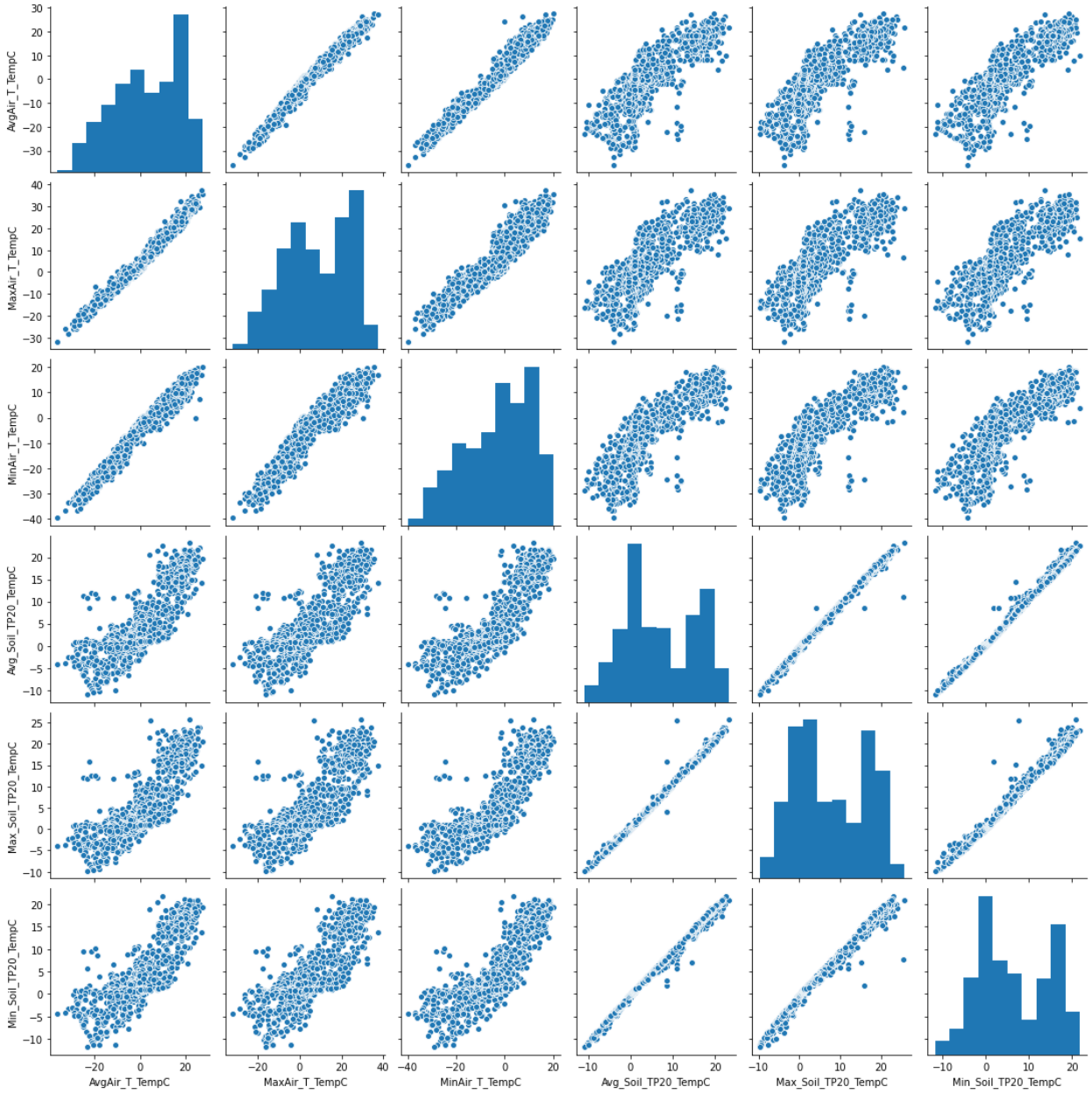


Figure 4-12:Air temperature and Soil temperature at the depth of 20cm - correlation

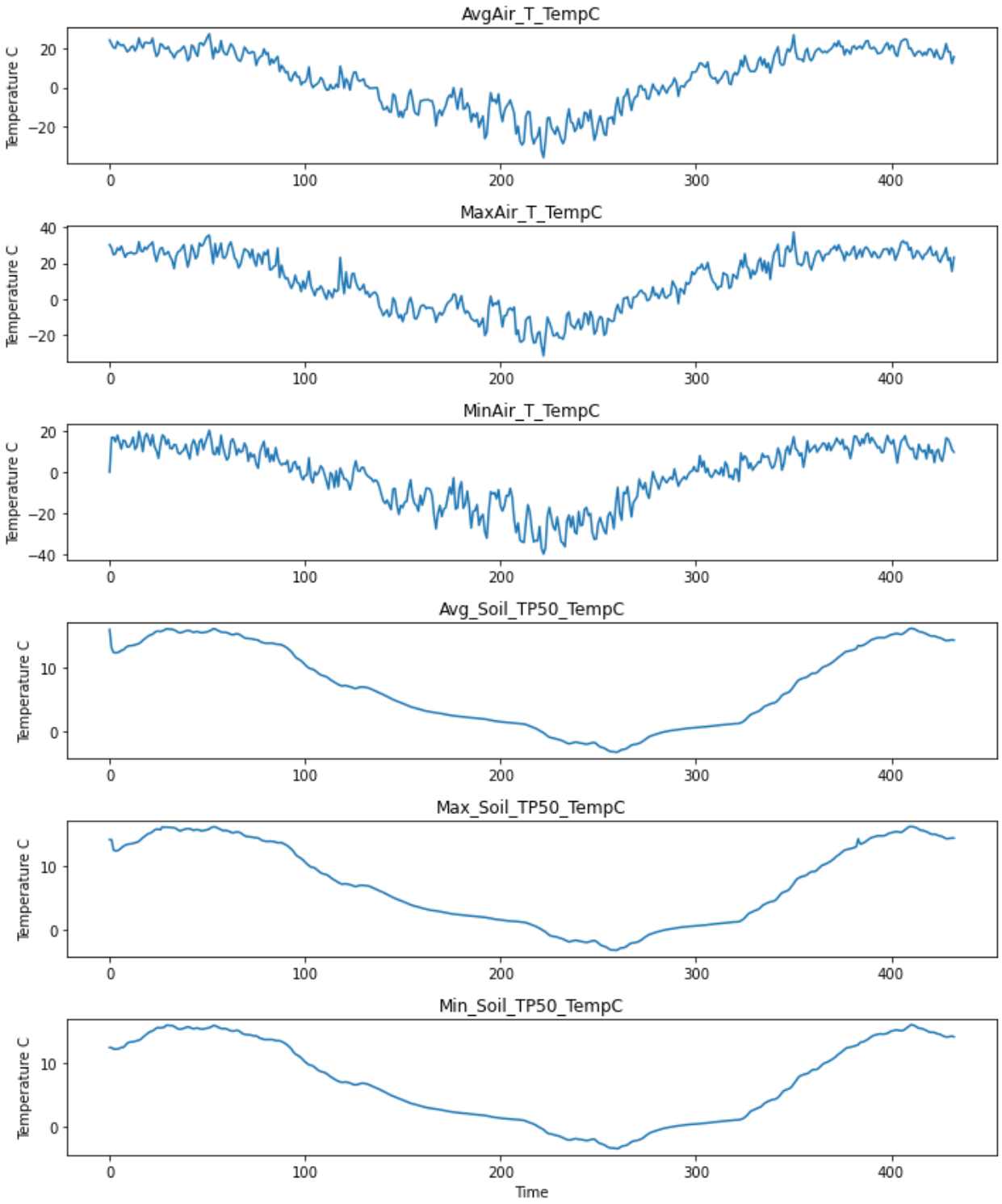


Figure 4-13: Air temperature and Soil temperature at the depth of 50cm

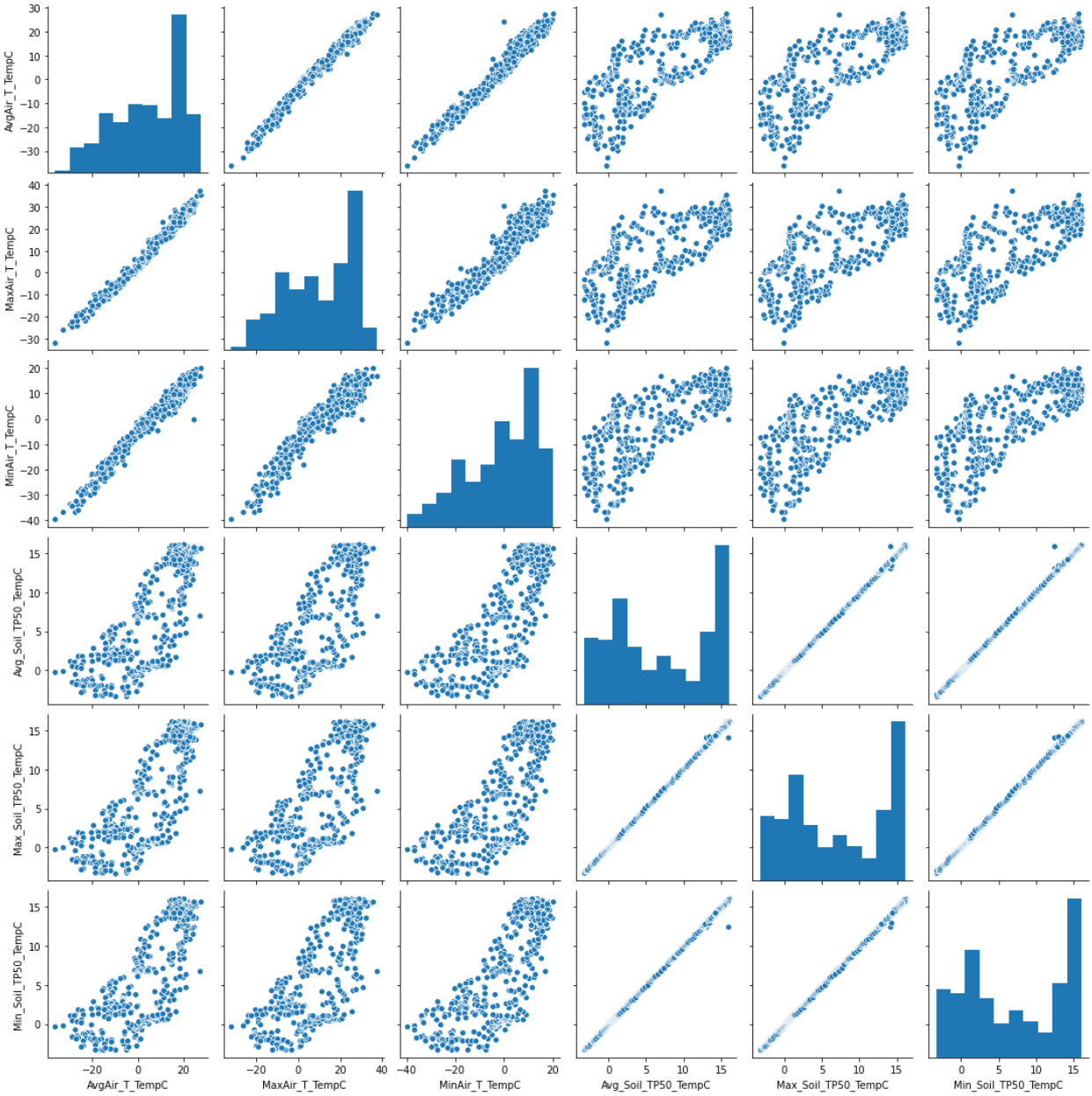


Figure 4-14: Air temperature and Soil temperature at the depth of 50cm - correlation

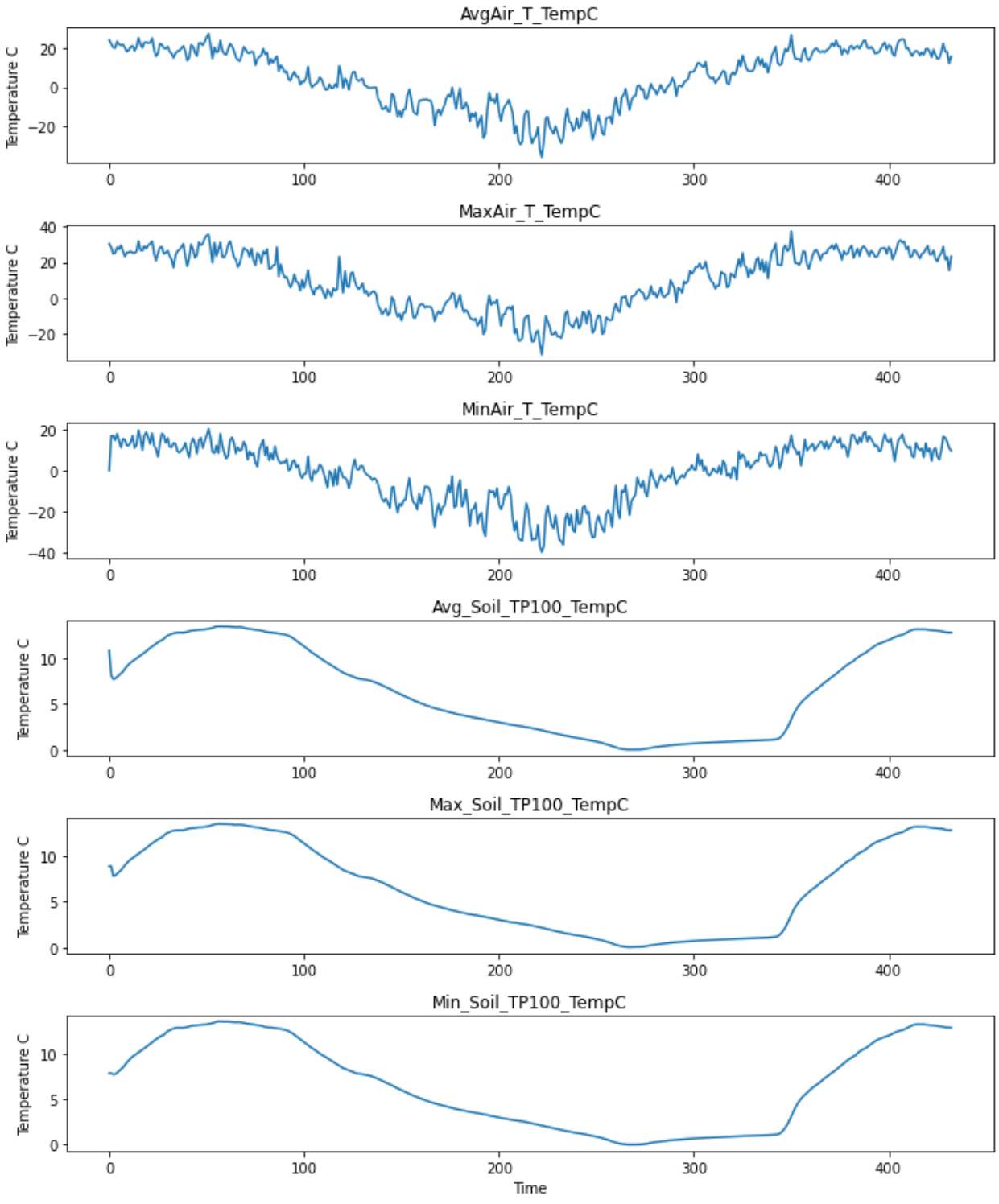


Figure 4-15: Air temperature and Soil temperature at the depth of 100cm

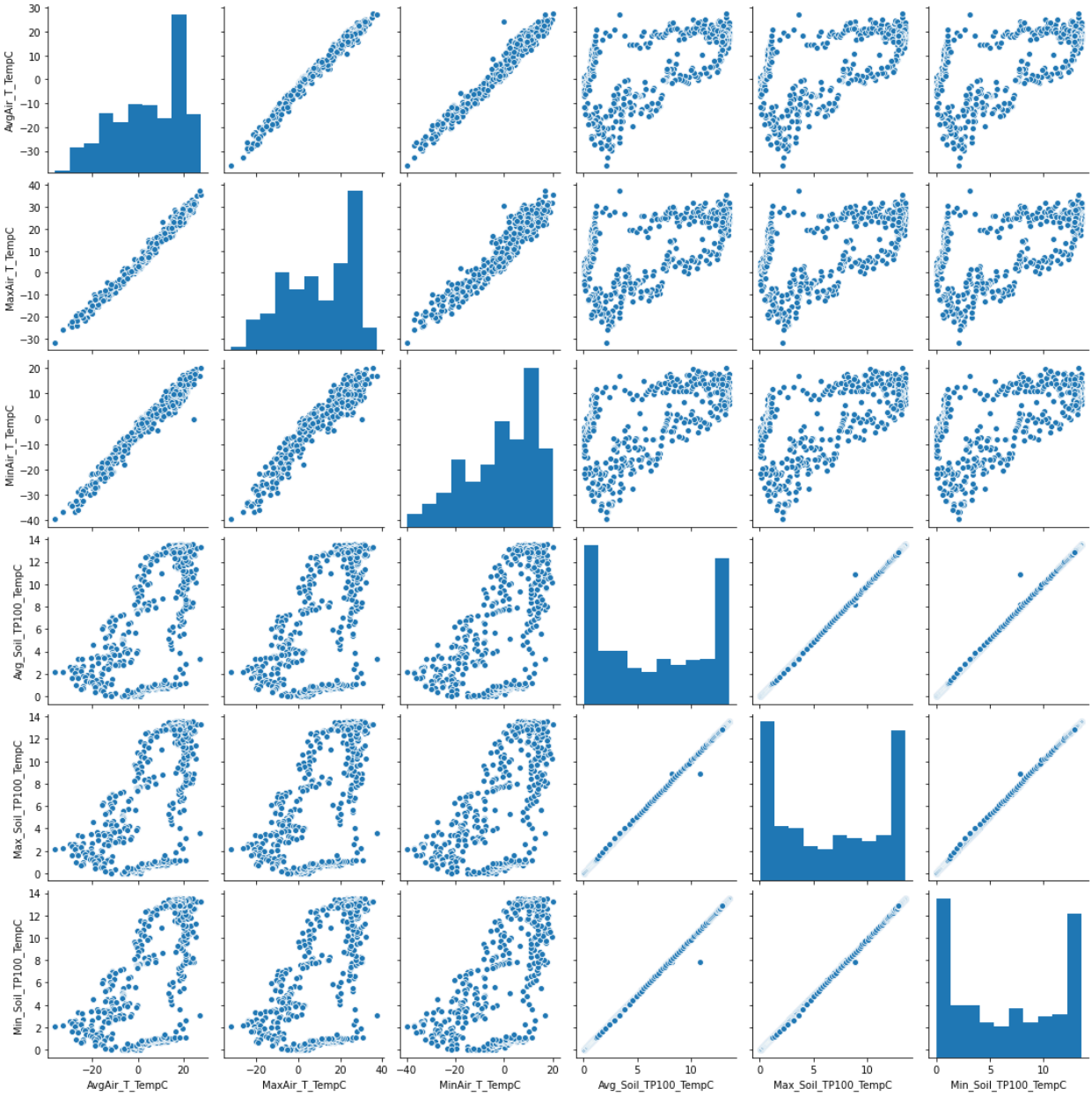


Figure 4-16: Air temperature and Soil temperature at the depth of 100cm – correlation

5- Feature engineering, Model development and evaluation metrics:

5-1 Feature engineering:

In this study, For the sake of computational cost, and avoid overfitting all the possible input variables will not be utilized in the model training and testing section. A threshold for correlation coefficient has been determined for each target variable and model. The variables that have the correlation coefficient higher than the determined correlation coefficient will be used as model inputs. Here are the details:

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2}\sqrt{\sum(y_i - \overline{y})^2}}$$

Equation 5-1

$$r = correlation\ coefficient$$

$$x_i = values\ of\ the\ possible\ input\ variable$$

$$\overline{x} = mean\ of\ the\ values\ of\ that\ possible\ variable$$

$$y_i = values\ of\ the\ target\ variable$$

$$\overline{y} = mean\ of\ the\ values\ of\ that\ target\ variable$$

SVR Model			WSVR Model(Wavelet Transform has been used)	
Target Variable	**Corrleation Coefficient Threshold**		**Target Variable**	**Corrleation Coefficient Threshold**
Avg_Soil_TP05_TempC	0.85		Avg_Soil_TP05_TempC	0.85
Max_Soil_TP05_TempC	0.85		Max_Soil_TP05_TempC	0.85
Min_Soil_TP05_TempC	0.85		Min_Soil_TP05_TempC	0.85
Avg_Soil_TP05_TempC_Plus_1	0.85		Avg_Soil_TP05_TempC_Plus_1	0.85
Max_Soil_TP05_TempC_Plus_1	0.85		Max_Soil_TP05_TempC_Plus_1	0.85
Min_Soil_TP05_TempC_Plus_1	0.85		Min_Soil_TP05_TempC_Plus_1	0.85
Avg_Soil_TP05_TempC_Plus_2	0.85		Avg_Soil_TP05_TempC_Plus_2	0.85
Max_Soil_TP05_TempC_Plus_2	0.85		Max_Soil_TP05_TempC_Plus_2	0.85
Min_Soil_TP05_TempC_Plus_2	0.85		Min_Soil_TP05_TempC_Plus_2	0.85
Avg_Soil_TP20_TempC	0.85		Avg_Soil_TP20_TempC	0.85
Max_Soil_TP20_TempC	0.85		Max_Soil_TP20_TempC	0.85
Min_Soil_TP20_TempC	0.85		Min_Soil_TP20_TempC	0.85
Avg_Soil_TP20_TempC_Plus_1	0.85		Avg_Soil_TP20_TempC_Plus_1	0.85
Max_Soil_TP20_TempC_Plus_1	0.85		Max_Soil_TP20_TempC_Plus_1	0.85
Min_Soil_TP20_TempC_Plus_1	0.85		Min_Soil_TP20_TempC_Plus_1	0.85
Avg_Soil_TP20_TempC_Plus_2	0.85		Avg_Soil_TP20_TempC_Plus_2	0.85
Max_Soil_TP20_TempC_Plus_2	0.85		Max_Soil_TP20_TempC_Plus_2	0.85
Min_Soil_TP20_TempC_Plus_2	0.85		Min_Soil_TP20_TempC_Plus_2	0.85
Avg_Soil_TP50_TempC	0.5		Avg_Soil_TP50_TempC	0.5
Max_Soil_TP50_TempC	0.5		Max_Soil_TP50_TempC	0.5
Min_Soil_TP50_TempC	0.5		Min_Soil_TP50_TempC	0.5
Avg_Soil_TP50_TempC_Plus_1	0.5		Avg_Soil_TP50_TempC_Plus_1	0.5
Max_Soil_TP50_TempC_Plus_1	0.5		Max_Soil_TP50_TempC_Plus_1	0.5
Min_Soil_TP50_TempC_Plus_1	0.5		Min_Soil_TP50_TempC_Plus_1	0.5
Avg_Soil_TP50_TempC_Plus_2	0.5		Avg_Soil_TP50_TempC_Plus_2	0.5
Max_Soil_TP50_TempC_Plus_2	0.5		Max_Soil_TP50_TempC_Plus_2	0.5
Min_Soil_TP50_TempC_Plus_2	0.5		Min_Soil_TP50_TempC_Plus_2	0.5
Avg_Soil_TP100_TempC	0.5		Avg_Soil_TP100_TempC	0.5
Max_Soil_TP100_TempC	0.5		Max_Soil_TP100_TempC	0.5
Min_Soil_TP100_TempC	0.5		Min_Soil_TP100_TempC	0.5
Avg_Soil_TP100_TempC_Plus_1	0.5		Avg_Soil_TP100_TempC_Plus_1	0.5
Max_Soil_TP100_TempC_Plus_1	0.5		Max_Soil_TP100_TempC_Plus_1	0.5
Min_Soil_TP100_TempC_Plus_1	0.5		Min_Soil_TP100_TempC_Plus_1	0.5
Avg_Soil_TP100_TempC_Plus_2	0.5		Avg_Soil_TP100_TempC_Plus_2	0.5
Max_Soil_TP100_TempC_Plus_2	0.5		Max_Soil_TP100_TempC_Plus_2	0.5
Min_Soil_TP100_TempC_Plus_2	0.5		Min_Soil_TP100_TempC_Plus_2	0.5

Figure 5-1: Correlation coefficient threshold for each target variable for both SVR and WSVR models

XGBoost Model			WXGBoost Model(Wavelet Transform has been used)	
Target Variable	**Corrleation Coefficient Threshold**		**Target Variable**	**Corrleation Coefficient Threshold**
Avg_Soil_TP05_TempC	0.85		Avg_Soil_TP05_TempC	0.85
Max_Soil_TP05_TempC	0.85		Max_Soil_TP05_TempC	0.85
Min_Soil_TP05_TempC	0.85		Min_Soil_TP05_TempC	0.85
Avg_Soil_TP05_TempC_Plus_1	0.85		Avg_Soil_TP05_TempC_Plus_1	0.85
Max_Soil_TP05_TempC_Plus_1	0.85		Max_Soil_TP05_TempC_Plus_1	0.85
Min_Soil_TP05_TempC_Plus_1	0.85		Min_Soil_TP05_TempC_Plus_1	0.85
Avg_Soil_TP05_TempC_Plus_2	0.85		Avg_Soil_TP05_TempC_Plus_2	0.85
Max_Soil_TP05_TempC_Plus_2	0.85		Max_Soil_TP05_TempC_Plus_2	0.85
Min_Soil_TP05_TempC_Plus_2	0.85		Min_Soil_TP05_TempC_Plus_2	0.85
Avg_Soil_TP20_TempC	0.85		Avg_Soil_TP20_TempC	0.85
Max_Soil_TP20_TempC	0.85		Max_Soil_TP20_TempC	0.85
Min_Soil_TP20_TempC	0.85		Min_Soil_TP20_TempC	0.85
Avg_Soil_TP20_TempC_Plus_1	0.85		Avg_Soil_TP20_TempC_Plus_1	0.85
Max_Soil_TP20_TempC_Plus_1	0.85		Max_Soil_TP20_TempC_Plus_1	0.85
Min_Soil_TP20_TempC_Plus_1	0.85		Min_Soil_TP20_TempC_Plus_1	0.85
Avg_Soil_TP20_TempC_Plus_2	0.85		Avg_Soil_TP20_TempC_Plus_2	0.85
Max_Soil_TP20_TempC_Plus_2	0.85		Max_Soil_TP20_TempC_Plus_2	0.85
Min_Soil_TP20_TempC_Plus_2	0.85		Min_Soil_TP20_TempC_Plus_2	0.85
Avg_Soil_TP50_TempC	0.5		Avg_Soil_TP50_TempC	0.5
Max_Soil_TP50_TempC	0.5		Max_Soil_TP50_TempC	0.5
Min_Soil_TP50_TempC	0.5		Min_Soil_TP50_TempC	0.5
Avg_Soil_TP50_TempC_Plus_1	0.5		Avg_Soil_TP50_TempC_Plus_1	0.5
Max_Soil_TP50_TempC_Plus_1	0.5		Max_Soil_TP50_TempC_Plus_1	0.5
Min_Soil_TP50_TempC_Plus_1	0.5		Min_Soil_TP50_TempC_Plus_1	0.5
Avg_Soil_TP50_TempC_Plus_2	0.5		Avg_Soil_TP50_TempC_Plus_2	0.5
Max_Soil_TP50_TempC_Plus_2	0.5		Max_Soil_TP50_TempC_Plus_2	0.5
Min_Soil_TP50_TempC_Plus_2	0.5		Min_Soil_TP50_TempC_Plus_2	0.5
Avg_Soil_TP100_TempC	0.5		Avg_Soil_TP100_TempC	0.5
Max_Soil_TP100_TempC	0.5		Max_Soil_TP100_TempC	0.5
Min_Soil_TP100_TempC	0.5		Min_Soil_TP100_TempC	0.5
Avg_Soil_TP100_TempC_Plus_1	0.5		Avg_Soil_TP100_TempC_Plus_1	0.5
Max_Soil_TP100_TempC_Plus_1	0.5		Max_Soil_TP100_TempC_Plus_1	0.5
Min_Soil_TP100_TempC_Plus_1	0.5		Min_Soil_TP100_TempC_Plus_1	0.5
Avg_Soil_TP100_TempC_Plus_2	0.5		Avg_Soil_TP100_TempC_Plus_2	0.5
Max_Soil_TP100_TempC_Plus_2	0.5		Max_Soil_TP100_TempC_Plus_2	0.5
Min_Soil_TP100_TempC_Plus_2	0.5		Min_Soil_TP100_TempC_Plus_2	0.5

Figure 5-2: Correlation coefficient threshold for each target variable for both XGBoost and WXGBoost models

ANN Model			WANN Model(Wavelet Transform has been used)	
Target Variable	**Corrleation Coefficient Threshold**		**Target Variable**	**Corrleation Coefficient Threshold**
Avg_Soil_TP05_TempC	0.85		Avg_Soil_TP05_TempC	0.85
Max_Soil_TP05_TempC	0.85		Max_Soil_TP05_TempC	0.85
Min_Soil_TP05_TempC	0.85		Min_Soil_TP05_TempC	0.85
Avg_Soil_TP05_TempC_Plus_1	0.85		Avg_Soil_TP05_TempC_Plus_1	0.85
Max_Soil_TP05_TempC_Plus_1	0.85		Max_Soil_TP05_TempC_Plus_1	0.85
Min_Soil_TP05_TempC_Plus_1	0.85		Min_Soil_TP05_TempC_Plus_1	0.85
Avg_Soil_TP05_TempC_Plus_2	0.85		Avg_Soil_TP05_TempC_Plus_2	0.85
Max_Soil_TP05_TempC_Plus_2	0.85		Max_Soil_TP05_TempC_Plus_2	0.85
Min_Soil_TP05_TempC_Plus_2	0.85		Min_Soil_TP05_TempC_Plus_2	0.85
Avg_Soil_TP20_TempC	0.85		Avg_Soil_TP20_TempC	0.85
Max_Soil_TP20_TempC	0.85		Max_Soil_TP20_TempC	0.85
Min_Soil_TP20_TempC	0.85		Min_Soil_TP20_TempC	0.85
Avg_Soil_TP20_TempC_Plus_1	0.85		Avg_Soil_TP20_TempC_Plus_1	0.85
Max_Soil_TP20_TempC_Plus_1	0.85		Max_Soil_TP20_TempC_Plus_1	0.85
Min_Soil_TP20_TempC_Plus_1	0.85		Min_Soil_TP20_TempC_Plus_1	0.85
Avg_Soil_TP20_TempC_Plus_2	0.85		Avg_Soil_TP20_TempC_Plus_2	0.85
Max_Soil_TP20_TempC_Plus_2	0.85		Max_Soil_TP20_TempC_Plus_2	0.85
Min_Soil_TP20_TempC_Plus_2	0.85		Min_Soil_TP20_TempC_Plus_2	0.85
Avg_Soil_TP50_TempC	0.5		Avg_Soil_TP50_TempC	0.5
Max_Soil_TP50_TempC	0.5		Max_Soil_TP50_TempC	0.5
Min_Soil_TP50_TempC	0.5		Min_Soil_TP50_TempC	0.5
Avg_Soil_TP50_TempC_Plus_1	0.5		Avg_Soil_TP50_TempC_Plus_1	0.5
Max_Soil_TP50_TempC_Plus_1	0.5		Max_Soil_TP50_TempC_Plus_1	0.5
Min_Soil_TP50_TempC_Plus_1	0.5		Min_Soil_TP50_TempC_Plus_1	0.5
Avg_Soil_TP50_TempC_Plus_2	0.5		Avg_Soil_TP50_TempC_Plus_2	0.5
Max_Soil_TP50_TempC_Plus_2	0.5		Max_Soil_TP50_TempC_Plus_2	0.5
Min_Soil_TP50_TempC_Plus_2	0.5		Min_Soil_TP50_TempC_Plus_2	0.5
Avg_Soil_TP100_TempC	0.5		Avg_Soil_TP100_TempC	0.5
Max_Soil_TP100_TempC	0.5		Max_Soil_TP100_TempC	0.5
Min_Soil_TP100_TempC	0.5		Min_Soil_TP100_TempC	0.5
Avg_Soil_TP100_TempC_Plus_1	0.5		Avg_Soil_TP100_TempC_Plus_1	0.5
Max_Soil_TP100_TempC_Plus_1	0.5		Max_Soil_TP100_TempC_Plus_1	0.5
Min_Soil_TP100_TempC_Plus_1	0.5		Min_Soil_TP100_TempC_Plus_1	0.5
Avg_Soil_TP100_TempC_Plus_2	0.5		Avg_Soil_TP100_TempC_Plus_2	0.5
Max_Soil_TP100_TempC_Plus_2	0.5		Max_Soil_TP100_TempC_Plus_2	0.5
Min_Soil_TP100_TempC_Plus_2	0.5		Min_Soil_TP100_TempC_Plus_2	0.5

Figure 5-3:Correlation coefficient threshold for each target variable for both ANN and WANN models

As can be seen from the tables, at the depth of 5cm and 20cm, for all the models, the correlation coefficients were set to the 0.85. on the other hand, since the data was scarce at the depth of 50cm and 100cm, and in order to train our models effectively, the correlation coefficients for all the models were set to 0.5 to allow more input variables.

In this study wavelet function type was set to db5 for every model in five levels which has neither a very simple nor a very complex shape.

5-2 Model development and hyperparameter tuning:

In this study, the total number of observations for all depths were divided into two groups: training set and testing set. 30% of the data was assigned to the model testing set and the rest used for model training. Using Grid Search function in python and 5-fold cross validation method, many hyperparameters were tested to find the best hyperparameters for each target variable. Here are all the hyperparameters options which were tested in the training stage:

SVR and WSVR+P1:Q21	
Kernel	["rbf", "linear","poly"]
C	[0.1,1,10,20,30,40]

Figure 5-4: Possible hyperparameters for SVR and WSVR

XGBoost and WXGBoost	
max_depth	[4, 5, 6, 7]
learning_rate	[0.5,0.1, 0.01, 0.05]
gamma	[0, 0.25, 1.0]
reg_lambda	[0, 1.0, 10.0]
n_estimators	[15,20,50,100,150]

Figure 5-5: Possible hyperparameters for XGBoost and WXGBoost

ANN and WANN	
batch_size	[128]
epochs	[200, 400]
neurons	[150, 250,350]
hidden_layers	[4,6,8]
optimizer	['Adam']
activation	['relu',"sigmoid"]

Figure 5-6: Possible hyperparameters for ANN and WANN

After hours of testing, the selected hyperparameters are as follow:

Target Variable	SVR Selected hyperparameters	WSVR Selected hyperparameters
Avg_Soil_TP05_TempC	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP05_TempC	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP05_TempC	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP05_TempC_Plus_1	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP05_TempC_Plus_1	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP05_TempC_Plus_1	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP05_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP05_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP05_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP20_TempC	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP20_TempC	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP20_TempC	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP20_TempC_Plus_1	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP20_TempC_Plus_1	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP20_TempC_Plus_1	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP20_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP20_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP20_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP50_TempC	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP50_TempC	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP50_TempC	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP50_TempC_Plus_1	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP50_TempC_Plus_1	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP50_TempC_Plus_1	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP50_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP50_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP50_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP100_TempC	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP100_TempC	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP100_TempC	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP100_TempC_Plus_1	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP100_TempC_Plus_1	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP100_TempC_Plus_1	C': 20, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Avg_Soil_TP100_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Max_Soil_TP100_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'
Min_Soil_TP100_TempC_Plus_2	C': 10, 'kernel': 'rbf'	C': 40, 'kernel': 'rbf'

Figure 5-7: Selected Hyperparameters for SVR and WSVR

Target Variable	XGBoost Selected hyperparameters	WXGBoost Selected hyperparameters
Avg_Soil_TP05_TempC	{'gamma': 0.25,'learning_rate': 0.05, 'max_depth': 4,'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0,'learning_rate': 0.05, 'max_depth': 7 , 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP05_TempC	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP05_TempC	{'gamma': 1.0,'learning_rate': 0.05, 'max_depth': 4,'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP05_TempC_Plus_1	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 10.0},	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP05_TempC_Plus_1	{'gamma': 0,'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP05_TempC_Plus_1	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 10.0},	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP05_TempC_Plus_2	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP05_TempC_Plus_2	{'gamma': 0, 'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP05_TempC_Plus_2	{'gamma': 1.0,'learning_rate': 0.1, 'max_depth': 4,'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 6, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP20_TempC	{'gamma': 0.25,'learning_rate': 0.05, 'max_depth': 4,'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0,'learning_rate': 0.05, 'max_depth': 7 , 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP20_TempC	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP20_TempC	{'gamma': 1.0,'learning_rate': 0.05, 'max_depth': 4,'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP20_TempC_Plus_1	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 10.0},	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP20_TempC_Plus_1	{'gamma': 0,'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP20_TempC_Plus_1	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 10.0},	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP20_TempC_Plus_2	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP20_TempC_Plus_2	{'gamma': 0, 'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP20_TempC_Plus_2	{'gamma': 1.0,'learning_rate': 0.1, 'max_depth': 4,'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 6, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP50_TempC	{'gamma': 0.25,'learning_rate': 0.05, 'max_depth': 4,'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0,'learning_rate': 0.05, 'max_depth': 7 , 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP50_TempC	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP50_TempC	{'gamma': 1.0,'learning_rate': 0.05, 'max_depth': 4,'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP50_TempC_Plus_1	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 10.0},	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP50_TempC_Plus_1	{'gamma': 0,'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP50_TempC_Plus_1	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 10.0},	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP50_TempC_Plus_2	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP50_TempC_Plus_2	{'gamma': 0, 'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP50_TempC_Plus_2	{'gamma': 1.0,'learning_rate': 0.1, 'max_depth': 4,'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 6, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP100_TempC	{'gamma': 0.25,'learning_rate': 0.05, 'max_depth': 4,'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0,'learning_rate': 0.05, 'max_depth': 7 , 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP100_TempC	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP100_TempC	{'gamma': 1.0,'learning_rate': 0.05, 'max_depth': 4,'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP100_TempC_Plus_1	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 10.0},	{'gamma': 0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP100_TempC_Plus_1	{'gamma': 0,'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP100_TempC_Plus_1	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 10.0},	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Avg_Soil_TP100_TempC_Plus_2	{'gamma': 0.25, 'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Max_Soil_TP100_TempC_Plus_2	{'gamma': 0, 'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 150, 'reg_lambda': 1.0},
Min_Soil_TP100_TempC_Plus_2	{'gamma': 1.0,'learning_rate': 0.1, 'max_depth': 4,'n_estimators': 50, 'reg_lambda': 1.0},	{'gamma': 1.0, 'learning_rate': 0.05, 'max_depth': 6, 'n_estimators': 150, 'reg_lambda': 1.0},

Figure 5-8:Selected Hyperparameters for XGBoost and WXGBoost

Target Variable	ANN Selected hyperparameters	WANN Selected hyperparameters
Avg_Soil_TP05_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150,'optimizer': 'Adam'},	{'activation': 'relu', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'opW101:W116timizer': 'Adam'}
Max_Soil_TP05_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Min_Soil_TP05_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400,'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP05_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 250, 'optimizer': 'Adam'},	{'activation': 'relu', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 350, 'optimizer': 'Adam'},
Max_Soil_TP05_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Min_Soil_TP05_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 200, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP05_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Max_Soil_TP05_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Min_Soil_TP05_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP20_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150,'optimizer': 'Adam'},	{'activation': 'relu', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'opW101:W116timizer': 'Adam'}
Max_Soil_TP20_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Min_Soil_TP20_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400,'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP20_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 250, 'optimizer': 'Adam'},	{'activation': 'relu', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 350, 'optimizer': 'Adam'},
Max_Soil_TP20_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Min_Soil_TP20_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 200, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP20_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Max_Soil_TP20_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Min_Soil_TP20_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP50_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150,'optimizer': 'Adam'},	{'activation': 'relu', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'opW101:W116timizer': 'Adam'}
Max_Soil_TP50_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Min_Soil_TP50_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400,'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP50_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 250, 'optimizer': 'Adam'},	{'activation': 'relu', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 350, 'optimizer': 'Adam'},
Max_Soil_TP50_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Min_Soil_TP50_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 200, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP50_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Max_Soil_TP50_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Min_Soil_TP50_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP100_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150,'optimizer': 'Adam'},	{'activation': 'relu', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'opW101:W116timizer': 'Adam'}
Max_Soil_TP100_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Min_Soil_TP100_TempC	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400,'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP100_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 250, 'optimizer': 'Adam'},	{'activation': 'relu', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 350, 'optimizer': 'Adam'},
Max_Soil_TP100_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Min_Soil_TP100_TempC_Plus_1	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 200, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},
Avg_Soil_TP100_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Max_Soil_TP100_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 350, 'optimizer': 'Adam'},
Min_Soil_TP100_TempC_Plus_2	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 6, 'neurons': 150, 'optimizer': 'Adam'},	{'activation': 'sigmoid', 'batch_size': 128, 'epochs': 400, 'hidden_layers': 4, 'neurons': 150, 'optimizer': 'Adam'},

Figure 5-9: Selected Hyperparameters for ANN and WANN

5-3 Evaluation metrics:

Once the models were developed, three statistical criteria, such as R^2 (coefficient of determination), MAE (mean absolute error), and MSE (mean squared error) were suggested to evaluate the models' performances, as follow:

$$R^2 = 1 - \frac{SSR}{SST}$$

Equation 5-2

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Equation 5-3

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Equation 5-4

Where y_i is the real value, $\hat{y}_i$ is the estimated value, n is the number of data points, SSR and SST are as follow:

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Equation 5-5

$$SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Equation 5-6

In which $\bar{y}$ is the mean of the real values.

6- Results:

In the presented work, the exactness off our data-driven models, SVR, WSVR, XGBoost, WXGBoost, ANN and WANN, is examined in mapping daily soil temperature of today and two days ahead at various depths. Models hyperparameters were examined using Grid Search function and 5-fold cross validation method and results are shown in the figures 26, 27 and 28. It should be noticed that testing set had not any role in model training (Hyperparameter tuning stage). Coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE) were utilized to assess the employed models. Here is the result of the models:

	SVR						WSVR					
	Train			Test			Train			Test		
Target Variable	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*
Avg_Soil_TP05_TempC	0.97	1.36	3.30	0.96	1.35	3.27	0.97	0.94	2.01	0.97	0.94	2.00
Max_Soil_TP05_TempC	0.98	1.45	3.51	0.97	1.44	3.48	0.97	1.14	2.51	0.97	1.13	2.50
Min_Soil_TP05_TempC	0.95	1.46	4.22	0.94	1.45	4.18	0.96	0.99	2.66	0.96	0.99	2.65
Avg_Soil_TP05_TempC_Plus_1	0.97	1.45	3.68	0.96	1.44	3.64	0.97	1.07	2.30	0.97	1.06	2.29
Max_Soil_TP05_TempC_Plus_1	0.97	1.65	4.50	0.96	1.63	4.46	0.96	1.32	3.25	0.96	1.31	3.23
Min_Soil_TP05_TempC_Plus_1	0.95	1.53	4.63	0.94	1.51	4.58	0.96	1.12	2.69	0.96	1.11	2.68
Avg_Soil_TP05_TempC_Plus_2	0.95	1.68	4.87	0.94	1.66	4.82	0.96	1.27	3.00	0.96	1.26	2.99
Max_Soil_TP05_TempC_Plus_2	0.95	1.84	5.90	0.94	1.82	5.84	0.96	1.51	4.15	0.96	1.50	4.13
Min_Soil_TP05_TempC_Plus_2	0.94	1.75	5.63	0.93	1.73	5.57	0.95	1.32	3.35	0.95	1.31	3.33
Avg_Soil_TP20_TempC	0.89	2.13	7.59	0.88	2.11	7.51	0.93	1.37	4.60	0.93	1.36	4.58
Max_Soil_TP20_TempC	0.89	2.12	7.78	0.88	2.10	7.70	0.93	1.38	4.92	0.93	1.37	4.90
Min_Soil_TP20_TempC	0.90	2.00	6.46	0.89	1.98	6.40	0.94	1.24	3.77	0.94	1.23	3.75
Avg_Soil_TP20_TempC_Plus_1	0.89	2.14	7.84	0.88	2.12	7.76	0.93	1.37	4.67	0.93	1.36	4.65
Max_Soil_TP20_TempC_Plus_1	0.88	2.19	9.08	0.87	2.17	8.99	0.92	1.45	5.86	0.92	1.44	5.83
Min_Soil_TP20_TempC_Plus_1	0.90	2.00	6.83	0.89	1.98	6.76	0.94	1.24	3.90	0.94	1.23	3.88
Avg_Soil_TP20_TempC_Plus_2	0.88	2.19	8.57	0.87	2.17	8.49	0.92	1.48	5.33	0.92	1.47	5.30
Max_Soil_TP20_TempC_Plus_2	0.87	2.20	9.39	0.86	2.18	9.30	0.92	1.45	5.86	0.92	1.44	5.83
Min_Soil_TP20_TempC_Plus_2	0.89	2.05	7.26	0.88	2.03	7.19	0.94	1.24	3.90	0.94	1.23	3.88
Avg_Soil_TP50_TempC	0.75	2.50	10.50	0.74	2.48	10.40	0.99	0.14	0.07	0.99	0.14	0.07
Max_Soil_TP50_TempC	0.77	2.43	9.37	0.76	2.41	9.28	0.99	0.13	0.04	0.99	0.13	0.04
Min_Soil_TP50_TempC	0.74	2.50	10.52	0.73	2.48	10.42	0.99	0.13	0.04	0.99	0.13	0.04
Avg_Soil_TP50_TempC_Plus_1	0.79	2.35	8.83	0.78	2.33	8.74	0.99	0.12	0.02	0.99	0.12	0.02
Max_Soil_TP50_TempC_Plus_1	0.79	2.35	8.87	0.78	2.33	8.78	0.99	0.12	0.03	0.99	0.12	0.03
Min_Soil_TP50_TempC_Plus_1	0.76	2.43	9.86	0.75	2.41	9.76	0.99	0.12	0.03	0.99	0.12	0.03
Avg_Soil_TP50_TempC_Plus_2	0.80	2.28	8.31	0.79	2.26	8.23	0.99	0.11	0.02	0.99	0.11	0.02
Max_Soil_TP50_TempC_Plus_2	0.80	2.27	8.34	0.79	2.25	8.26	0.99	0.12	0.02	0.99	0.12	0.02
Min_Soil_TP50_TempC_Plus_2	0.80	2.28	8.31	0.79	2.26	8.23	0.99	0.12	0.02	0.99	0.12	0.02
Avg_Soil_TP100_TempC	0.40	2.75	13.23	0.40	2.72	13.10	0.99	0.19	0.17	0.99	0.19	0.17
Max_Soil_TP100_TempC	0.47	2.66	11.83	0.47	2.63	11.71	0.99	0.19	0.17	0.99	0.19	0.17
Min_Soil_TP100_TempC	0.40	2.76	13.34	0.40	2.73	13.21	0.99	0.19	0.18	0.99	0.19	0.18
Avg_Soil_TP100_TempC_Plus_1	0.49	2.62	11.38	0.49	2.59	11.27	0.99	0.18	0.16	0.99	0.18	0.16
Max_Soil_TP100_TempC_Plus_1	0.49	2.61	11.29	0.49	2.58	11.18	0.99	0.18	0.15	0.99	0.18	0.15
Min_Soil_TP100_TempC_Plus_1	0.48	2.62	11.47	0.48	2.59	11.36	0.99	0.19	0.16	0.99	0.19	0.16
Avg_Soil_TP100_TempC_Plus_2	0.52	2.56	10.86	0.51	2.53	10.75	0.99	0.17	0.14	0.99	0.17	0.14
Max_Soil_TP100_TempC_Plus_2	0.53	2.55	10.79	0.52	2.52	10.68	0.99	0.17	0.13	0.99	0.17	0.13
Min_Soil_TP100_TempC_Plus_2	0.52	2.57	10.95	0.51	2.54	10.84	0.99	0.17	0.14	0.99	0.17	0.14

Figure 6-1: Models scores for both SVR and WSVR (best models are colored orange)

	XGBoost						WXGBoost					
	Train			Test			Train			Test		
Target Variable	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*
Avg_Soil_TP05_TempC	0.96	1.38	3.34	0.96	1.37	3.32	0.97	0.98	1.85	0.97	0.98	1.84
Max_Soil_TP05_TempC	0.96	1.47	3.58	0.96	1.46	3.56	0.97	1.15	2.35	0.97	1.14	2.34
Min_Soil_TP05_TempC	0.94	1.50	4.20	0.94	1.49	4.18	0.96	1.05	2.44	0.96	1.04	2.43
Avg_Soil_TP05_TempC_Plus_1	0.95	1.50	3.83	0.95	1.49	3.81	0.97	1.11	2.27	0.97	1.10	2.26
Max_Soil_TP05_TempC_Plus_1	0.95	1.69	4.70	0.95	1.68	4.68	0.96	1.38	3.40	0.96	1.37	3.38
Min_Soil_TP05_TempC_Plus_1	0.94	1.58	4.64	0.94	1.57	4.62	0.96	1.16	2.52	0.96	1.15	2.51
Avg_Soil_TP05_TempC_Plus_2	0.94	1.73	5.08	0.94	1.72	5.05	0.96	1.35	3.25	0.96	1.34	3.23
Max_Soil_TP05_TempC_Plus_2	0.94	1.89	6.15	0.94	1.88	6.12	0.95	1.61	4.59	0.95	1.60	4.57
Min_Soil_TP05_TempC_Plus_2	0.92	1.80	5.73	0.92	1.79	5.70	0.95	1.40	3.49	0.95	1.39	3.47
Avg_Soil_TP20_TempC	0.88	2.13	7.69	0.88	2.12	7.65	0.95	1.22	3.03	0.95	1.21	3.01
Max_Soil_TP20_TempC	0.89	2.13	7.62	0.89	2.12	7.58	0.95	1.23	3.04	0.95	1.22	3.02
Min_Soil_TP20_TempC	0.89	2.03	6.84	0.89	2.02	6.81	0.96	1.14	2.63	0.96	1.13	2.62
Avg_Soil_TP20_TempC_Plus_1	0.88	2.14	7.88	0.88	2.13	7.84	0.95	1.30	3.23	0.95	1.29	3.21
Max_Soil_TP20_TempC_Plus_1	0.87	2.17	8.65	0.87	2.16	8.61	0.94	1.39	4.22	0.94	1.38	4.20
Min_Soil_TP20_TempC_Plus_1	0.88	2.05	7.27	0.88	2.04	7.23	0.95	1.18	2.84	0.95	1.17	2.83
Avg_Soil_TP20_TempC_Plus_2	0.87	2.15	8.25	0.87	2.14	8.21	0.94	1.31	3.68	0.94	1.30	3.66
Max_Soil_TP20_TempC_Plus_2	0.87	2.18	8.87	0.87	2.17	8.83	0.93	1.43	4.69	0.93	1.42	4.67
Min_Soil_TP20_TempC_Plus_2	0.88	2.04	7.16	0.88	2.03	7.12	0.94	1.26	3.40	0.94	1.25	3.38
Avg_Soil_TP50_TempC	0.77	2.40	9.10	0.77	2.39	9.05	0.99	0.22	0.11	0.99	0.22	0.11
Max_Soil_TP50_TempC	0.77	2.38	9.00	0.77	2.37	8.96	0.99	0.20	0.09	0.99	0.20	0.09
Min_Soil_TP50_TempC	0.75	2.46	9.63	0.75	2.45	9.58	0.99	0.27	0.13	0.99	0.27	0.13
Avg_Soil_TP50_TempC_Plus_1	0.78	2.33	8.54	0.78	2.32	8.50	0.99	0.20	0.09	0.99	0.20	0.09
Max_Soil_TP50_TempC_Plus_1	0.79	2.31	8.25	0.79	2.30	8.21	0.99	0.29	0.15	0.99	0.29	0.15
Min_Soil_TP50_TempC_Plus_1	0.78	2.37	8.80	0.78	2.36	8.76	0.99	0.30	0.18	0.99	0.30	0.18
Avg_Soil_TP50_TempC_Plus_2	0.81	2.20	7.53	0.81	2.19	7.49	0.99	0.29	0.14	0.99	0.29	0.14
Max_Soil_TP50_TempC_Plus_2	0.80	2.27	8.04	0.80	2.26	8.00	0.99	0.30	0.16	0.99	0.30	0.16
Min_Soil_TP50_TempC_Plus_2	0.80	2.24	7.83	0.80	2.23	7.79	0.99	0.29	0.15	0.99	0.29	0.15
Avg_Soil_TP100_TempC	0.40	2.73	13.15	0.40	2.72	13.08	0.99	0.15	0.06	0.99	0.15	0.06
Max_Soil_TP100_TempC	0.41	2.71	13.11	0.41	2.70	13.04	0.99	0.14	0.06	0.99	0.14	0.06
Min_Soil_TP100_TempC	0.40	2.74	13.13	0.40	2.73	13.06	0.99	0.18	0.08	0.99	0.18	0.08
Avg_Soil_TP100_TempC_Plus_1	0.49	2.55	11.35	0.49	2.54	11.29	0.99	0.16	0.07	0.99	0.16	0.07
Max_Soil_TP100_TempC_Plus_1	0.45	2.66	12.17	0.45	2.65	12.11	0.99	0.20	0.07	0.99	0.20	0.07
Min_Soil_TP100_TempC_Plus_1	0.47	2.58	11.58	0.47	2.57	11.52	0.99	0.26	0.14	0.99	0.26	0.14
Avg_Soil_TP100_TempC_Plus_2	0.47	2.60	11.70	0.47	2.59	11.64	0.99	0.21	0.09	0.99	0.21	0.09
Max_Soil_TP100_TempC_Plus_2	0.48	2.56	11.59	0.48	2.55	11.53	0.99	0.21	0.08	0.99	0.21	0.08
Min_Soil_TP100_TempC_Plus_2	0.48	2.55	11.43	0.48	2.54	11.37	0.99	0.22	0.10	0.99	0.22	0.10

Figure 6-2: models scores for both XGBoost and WXGBoost (best models are colored orange)

	ANN						WANN					
	Train			Test			Train			Test		
Target Variable	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*	*R2*	*MAE*	*MSE*
Avg_Soil_TP05_TempC	0.97	1.47	3.51	0.96	1.46	3.48	0.97	1.00	1.98	0.97	0.99	1.96
Max_Soil_TP05_TempC	0.97	1.49	3.56	0.96	1.48	3.52	0.97	1.31	2.90	0.97	1.30	2.87
Min_Soil_TP05_TempC	0.95	1.52	4.19	0.94	1.50	4.15	0.96	1.12	2.87	0.96	1.11	2.84
Avg_Soil_TP05_TempC_Plus_1	0.96	1.48	3.75	0.95	1.47	3.71	0.97	1.11	2.20	0.97	1.10	2.18
Max_Soil_TP05_TempC_Plus_1	0.96	1.67	4.52	0.95	1.65	4.48	0.96	1.42	3.44	0.96	1.41	3.41
Min_Soil_TP05_TempC_Plus_1	0.95	1.56	4.50	0.94	1.54	4.46	0.96	1.37	3.37	0.96	1.36	3.34
Avg_Soil_TP05_TempC_Plus_2	0.95	1.73	5.03	0.94	1.71	4.98	0.96	1.32	3.05	0.96	1.31	3.02
Max_Soil_TP05_TempC_Plus_2	0.95	1.88	6.03	0.94	1.86	5.97	0.95	1.69	5.04	0.95	1.67	4.99
Min_Soil_TP05_TempC_Plus_2	0.93	1.78	5.66	0.92	1.76	5.60	0.95	1.42	3.48	0.95	1.41	3.45
Avg_Soil_TP20_TempC	0.90	2.10	7.09	0.89	2.08	7.02	0.94	1.25	3.62	0.94	1.24	3.58
Max_Soil_TP20_TempC	0.90	2.10	7.35	0.89	2.08	7.28	0.94	1.57	4.37	0.94	1.55	4.33
Min_Soil_TP20_TempC	0.90	2.00	6.49	0.89	1.98	6.43	0.94	1.43	3.56	0.94	1.42	3.52
Avg_Soil_TP20_TempC_Plus_1	0.90	2.08	7.11	0.89	2.06	7.04	0.95	1.16	2.97	0.95	1.15	2.94
Max_Soil_TP20_TempC_Plus_1	0.89	2.11	8.14	0.88	2.09	8.06	0.93	1.60	4.92	0.93	1.58	4.87
Min_Soil_TP20_TempC_Plus_1	0.90	2.00	6.66	0.89	1.98	6.59	0.94	1.35	3.47	0.94	1.34	3.44
Avg_Soil_TP20_TempC_Plus_2	0.89	2.09	7.60	0.88	2.07	7.52	0.93	1.60	4.69	0.93	1.58	4.64
Max_Soil_TP20_TempC_Plus_2	0.89	2.10	7.99	0.88	2.08	7.91	0.93	1.59	4.81	0.93	1.57	4.76
Min_Soil_TP20_TempC_Plus_2	0.90	2.03	6.87	0.89	2.01	6.80	0.94	1.49	4.28	0.94	1.48	4.24
Avg_Soil_TP50_TempC	0.78	2.40	9.09	0.77	2.38	9.00	0.99	0.11	0.02	0.99	0.11	0.02
Max_Soil_TP50_TempC	0.82	2.24	7.53	0.81	2.22	7.46	0.99	0.12	0.03	0.99	0.12	0.03
Min_Soil_TP50_TempC	0.78	2.41	8.99	0.77	2.39	8.90	0.99	0.14	0.04	0.99	0.14	0.04
Avg_Soil_TP50_TempC_Plus_1	0.83	2.20	7.21	0.82	2.18	7.14	0.99	0.20	0.07	0.99	0.20	0.07
Max_Soil_TP50_TempC_Plus_1	0.83	2.24	7.42	0.82	2.22	7.35	0.99	0.14	0.03	0.99	0.14	0.03
Min_Soil_TP50_TempC_Plus_1	0.81	2.35	8.07	0.80	2.33	7.99	0.99	0.13	0.03	0.99	0.13	0.03
Avg_Soil_TP50_TempC_Plus_2	0.82	2.22	7.73	0.81	2.20	7.65	0.99	0.13	0.02	0.99	0.13	0.02
Max_Soil_TP50_TempC_Plus_2	0.81	2.25	7.94	0.80	2.23	7.86	0.99	0.11	0.02	0.99	0.11	0.02
Min_Soil_TP50_TempC_Plus_2	0.81	2.28	8.01	0.80	2.26	7.93	0.99	0.13	0.03	0.99	0.13	0.03
Avg_Soil_TP100_TempC	0.53	2.60	10.64	0.52	2.57	10.53	0.99	0.09	0.01	0.99	0.09	0.01
Max_Soil_TP100_TempC	0.51	2.62	11.17	0.50	2.59	11.06	0.99	0.12	0.03	0.99	0.12	0.03
Min_Soil_TP100_TempC	0.48	2.58	11.85	0.48	2.55	11.73	0.99	0.14	0.04	0.99	0.14	0.04
Avg_Soil_TP100_TempC_Plus_1	0.48	2.63	11.55	0.48	2.60	11.44	0.99	0.07	0.01	0.99	0.07	0.01
Max_Soil_TP100_TempC_Plus_1	0.62	2.32	8.68	0.61	2.30	8.59	0.99	0.19	0.06	0.99	0.19	0.06
Min_Soil_TP100_TempC_Plus_1	0.61	2.47	8.92	0.60	2.45	8.83	0.99	0.14	0.04	0.99	0.14	0.04
Avg_Soil_TP100_TempC_Plus_2	0.48	2.58	11.54	0.48	2.55	11.43	0.99	0.18	0.06	0.99	0.18	0.06
Max_Soil_TP100_TempC_Plus_2	0.49	2.61	11.48	0.49	2.58	11.37	0.99	0.16	0.05	0.99	0.16	0.05
Min_Soil_TP100_TempC_Plus_2	0.47	2.63	11.73	0.47	2.60	11.61	0.99	0.11	0.03	0.99	0.11	0.03

Figure 6-3:models scores for both ANN and WANN (best models are colored orange)

6-1 Wavelet transform effect:

As can be seen from the results, considering not using wavelet transform, the model's evaluation metrics have worsened dramatically by increasing the depth. The MSE value climbs from about 3.5 at the depth of 5cm to about 11 at the depth of 100 cm for all the models. The reason is that, by increasing the depth, the correlation between air temperature and soil temperature declines.

On the other hand, using the wavelet transform at data preprocessing stage, not only causes growing coefficient of determination by at least 0.02, but also affects significantly our models performance at higher depths. By using wavelet transform at data preprocessing stage, we have reached the following scores for each of target variables:

		Testing		
Target Variable	**Outperformed Model**	*R3*	*MAE*	*MSE*
Avg_Soil_TP05_TempC	WXGBoost	0.97	0.98	1.84
Max_Soil_TP05_TempC	WXGBoost	0.97	1.14	2.34
Min_Soil_TP05_TempC	WXGBoost	0.96	1.04	2.43
Avg_Soil_TP05_TempC_Plus_1	WANN	0.97	1.10	2.18
Max_Soil_TP05_TempC_Plus_1	WSVR	0.96	1.31	3.23
Min_Soil_TP05_TempC_Plus_1	WXGBoost	0.96	1.15	2.51
Avg_Soil_TP05_TempC_Plus_2	WSVR	0.96	1.26	2.99
Max_Soil_TP05_TempC_Plus_2	WSVR	0.96	1.50	4.13
Min_Soil_TP05_TempC_Plus_2	WSVR	0.95	1.31	3.33
Avg_Soil_TP20_TempC	WXGBoost	0.95	1.21	3.01
Max_Soil_TP20_TempC	WXGBoost	0.95	1.22	3.02
Min_Soil_TP20_TempC	WXGBoost	0.96	1.13	2.62
Avg_Soil_TP20_TempC_Plus_1	WANN	0.95	1.15	2.94
Max_Soil_TP20_TempC_Plus_1	WXGBoost	0.94	1.38	4.20
Min_Soil_TP20_TempC_Plus_1	WXGBoost	0.95	1.17	2.83
Avg_Soil_TP20_TempC_Plus_2	WXGBoost	0.94	1.30	3.66
Max_Soil_TP20_TempC_Plus_2	WXGBoost	0.93	1.42	4.67
Min_Soil_TP20_TempC_Plus_2	WXGBoost	0.94	1.25	3.38
Avg_Soil_TP50_TempC	WANN	0.99	0.11	0.02
Max_Soil_TP50_TempC	WANN	0.99	0.12	0.03
Min_Soil_TP50_TempC	WSVR	0.99	0.13	0.04
Avg_Soil_TP50_TempC_Plus_1	WSVR	0.99	0.12	0.02
Max_Soil_TP50_TempC_Plus_1	WSVR	0.99	0.12	0.03
Min_Soil_TP50_TempC_Plus_1	WSVR	0.99	0.12	0.03
Avg_Soil_TP50_TempC_Plus_2	WSVR	0.99	0.11	0.02
Max_Soil_TP50_TempC_Plus_2	WANN	0.99	0.11	0.02
Min_Soil_TP50_TempC_Plus_2	WSVR	0.99	0.12	0.02
Avg_Soil_TP100_TempC	WANN	0.99	0.09	0.01
Max_Soil_TP100_TempC	WANN	0.99	0.12	0.03
Min_Soil_TP100_TempC	WANN	0.99	0.14	0.04
Avg_Soil_TP100_TempC_Plus_1	WANN	0.99	0.07	0.01
Max_Soil_TP100_TempC_Plus_1	WANN	0.99	0.19	0.06
Min_Soil_TP100_TempC_Plus_1	WANN	0.99	0.14	0.04
Avg_Soil_TP100_TempC_Plus_2	WANN	0.99	0.18	0.06
Max_Soil_TP100_TempC_Plus_2	WANN	0.99	0.16	0.05
Min_Soil_TP100_TempC_Plus_2	WANN	0.99	0.11	0.03

Figure 6-4: Selected models and their performances

As can be seen, the MSE criterion has decreased to about 0.03 at the depth of both 50cm and 100 cm. We can conclude that, when our target variable is more stable and has less fluctuation, or in other words the magnitude of high frequency in the target variable is small, using wavelet transform on input variables can significantly increase our models accuracy.

6-2 a comparison of WSVR, WXGBoost and, WANN:

In the higher depths, where we have more observations, WSVR and WXGBoost have outperformed significantly. We can conclude that, WSVR and WXGBoost require a huge amount of data for their training process. On the other hand, at higher depths, when data is scarce WANN outperforms the other models.

7- Conclusion:

The abilities of three machine learning methods, SVR, XGBoost, and ANN in estimating average, minimum and maximum daily soil temperature at different depths were compared utilizing average, minimum and maximum air temperature of six days before as input variables. Furthermore, the effect of wavelet transform at data preprocessing stage on model's performance was investigated. The following conclusions can be reached from application results:

- Using wavelet transform, at data preprocessing stage, has significantly improved the models performance at all depth, especially at higher depths where our target variable is more stable.
- At higher depths where data was scarce, WANN outperformed both WSVR and WXGBoost. However, at lower depths were data was abundant, WSVR and XGBoost performed better.

8-References:

A First Course on Wavelets - Eugenio Hernandez, Guido Weiss - Google Books. (n.d.). Retrieved September 13, 2020, from https://books.google.ca/books?hl=en&lr=&id=Dq95ngIhy0oC&oi=fnd&pg=PA1&dq=Hernández+and+Weiss,+1996&ots=mwFg2ul_gf&sig=bo__sTz32cA6rMZ5PfSEhwGFIXo#v=onepage&q=Hernández and Weiss%2C 1996&f=false

Alizamir, M., Kisi, O., Ahmed, A. N., Mert, C., Fai, C. M., Kim, S., Kim, N. W., & El-Shafie, A. (2020). Advanced machine learning model for better prediction accuracy of soil temperature at different depths. *PLoS ONE*, *15*(4), 1–25. https://doi.org/10.1371/journal.pone.0231055

AR4 Climate Change 2007: Synthesis Report — IPCC. (n.d.). Retrieved September 7, 2020, from https://www.ipcc.ch/report/ar4/syr/

Araghi, A., Mousavi-Baygi, M., Adamowski, J., Martinez, C., & van der Ploeg, M. (2017). Forecasting soil temperature based on surface air temperature using a wavelet artificial neural network. *Meteorological Applications*, *24*(4), 603–611. https://doi.org/10.1002/met.1661

Badache, M., Eslami-Nejad, P., Ouzzane, M., Aidoun, Z., & Lamarche, L. (2016). A new modeling approach for improved ground temperature profile determination. *Renewable Energy*, *85*, 436–444. https://doi.org/10.1016/j.renene.2015.06.020

Belghit, A., & Benyaich, M. (2014). Numerical Study of Heat Transfer and Contaminant Transport in an Unsaturated Porous Soil. *Journal of Water Resource and Protection*, *06*(13), 1238–1247. https://doi.org/10.4236/jwarp.2014.613113

Bilgili, M. (2010). Prediction of soil temperature using regression and artificial neural network models. *Meteorology and Atmospheric Physics*, *110*(1), 59–70. https://doi.org/10.1007/s00703-010-0104-x

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM*

References

SIGKDD International Conference on Knowledge Discovery and Data Mining, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Citakoglu, H. (2017). Comparison of artificial intelligence techniques for prediction of soil temperatures in Turkey. *Theoretical and Applied Climatology*, *130*(1–2), 545–556. https://doi.org/10.1007/s00704-016-1914-7

Cleall, P. J., Muñoz-Criollo, J. J., & Rees, S. W. (2015). Analytical Solutions for Ground Temperature Profiles and Stored Energy Using Meteorological Data. *Transport in Porous Media*, *106*(1), 181–199. https://doi.org/10.1007/s11242-014-0395-3

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/bf00994018

De Vries, D. A. (1958a). Simultaneous transfer of heat and moisture in porous media. *Eos, Transactions American Geophysical Union*, *39*(5), 909–916. https://doi.org/10.1029/TR039i005p00909

De Vries, D. A. (1958b). Simultaneous transfer of heat and moisture in porous media. *Eos, Transactions American Geophysical Union*, *39*(5), 909–916. https://doi.org/10.1029/TR039i005p00909

Delbari, M., Sharifazari, S., & Mohammadi, E. (2019). Modeling daily soil temperature over diverse climate conditions in Iran—a comparison of multiple linear regression and support vector regression techniques. *Theoretical and Applied Climatology*, *135*(3–4), 991–1001. https://doi.org/10.1007/s00704-018-2370-3

Domisch, T., Finér, L., & Lehto, T. (2001). Effects of soil temperature on biomass and carbohydrate allocation in Scots pine (Pinus sylvestris) seedlings at the beginning of the growing season. *Tree Physiology*, *21*(7), 465–472. https://doi.org/10.1093/treephys/21.7.465

Droulia, F., Lykoudis, S., Tsiros, I., Alvertos, N., Akylas, E., & Garofalakis, I. (2009). Ground temperature estimations using simplified analytical and semi-empirical approaches. *Solar Energy*,

83(2), 211–219. https://doi.org/10.1016/j.solener.2008.07.013

Elias, E. A., Cichota, R., Torriani, H. H., & de Jong van Lier, Q. (2004). ANALYTICAL SOIL-TEMPERATURE MODEL. *Soil Science Society of America Journal*, *68*(3), 784–788. https://doi.org/10.2136/sssaj2004.7840

Fahim Ahmad, M., & Rasul, G. (2008). PREDICTION OF SOIL TEMPERATURE BY AIR TEMPERATURE; A CASE STUDY FOR FAISALABAD. In *Pakistan Journal of Meteorology* (Vol. 5). http://www.globe.gov/tctg/sectionpdf.jsp?sectionId=92#page=10

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

George, R. K. (2001). Prediction of soil temperature by using artificial neural networks algorithms. *Nonlinear Analysis, Theory, Methods and Applications*, *47*(3), 1737–1748. https://doi.org/10.1016/S0362-546X(01)00306-6

Hosseinzadeh Talaee, P. (2014). Daily soil temperature modeling using neuro-fuzzy approach. *Theoretical and Applied Climatology*, *118*(3), 481–489. https://doi.org/10.1007/s00704-013-1084-9

Hsu, K. -l, Gupta, H. V., & Sorooshian, S. (1995). Artificial Neural Network Modeling of the Rainfall-Runoff Process. *Water Resources Research*, *31*(10), 2517–2530. https://doi.org/10.1029/95WR01955

Introduction to Environmental Soil Physics - 1st Edition. (n.d.). Retrieved September 6, 2020, from https://www.elsevier.com/books/introduction-to-environmental-soil-physics/hillel/978-0-12-348655-4

Journal of geophysical research. (1955). *Nature*, *175*(4449), 238. https://doi.org/10.1038/175238c0

Kang, S., Kim, S., Oh, S., & Lee, D. (2000). Predicting spatial and temporal patterns of soil temperature based on topography, surface cover and air temperature. *Forest Ecology and Management*, *136*(1–

3), 173–184. https://doi.org/10.1016/S0378-1127(99)00290-X

Kisi, O., Demir, V., & Kim, S. (2017). Estimation of long-term monthly temperatures by three different adaptive neuro-fuzzy approaches using geographical inputs. *Journal of Irrigation and Drainage Engineering*, *143*(12), 1–18. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001242

Kisi, O., Sanikhani, H., & Cobaner, M. (2017). Soil temperature modeling at different depths using neuro-fuzzy, neural network, and genetic programming techniques. *Theoretical and Applied Climatology*, *129*(3–4), 833–848. https://doi.org/10.1007/s00704-016-1810-1

Lau, K. M., & Hengyi Weng. (1995). Climate signal detection using wavelet transform: how to make a time series sing. *Bulletin - American Meteorological Society*, *76*(12), 2391–2402. https://doi.org/10.1175/1520-0477(1995)076<2391:CSDUWT>2.0.CO;2

Mallat, S. (2009). A Wavelet Tour of Signal Processing. In *A Wavelet Tour of Signal Processing*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-374370-1.X0001-8

Mellander, P. E., Bishop, K., & Lundmark, T. (2004). The influence of soil temperature on transpiration: A plot scale manipulation in a young Scots pine stand. *Forest Ecology and Management*, *195*(1–2), 15–28. https://doi.org/10.1016/j.foreco.2004.02.051

Mihalakakou, G. (2002). On estimating soil surface temperature profiles. *Energy and Buildings*, *34*(3), 251–259. https://doi.org/10.1016/S0378-7788(01)00089-5

Mitchell, J. K. (n.d.). *Temperature Effects on the Engineering Properties and Behavior of Soils*.

Narasimhan, T. N. (2010). Thermal conductivity through the 19th century. *Physics Today*, *63*(8), 36–41. https://doi.org/10.1063/1.3480074

Ouzzane, M., Eslami-Nejad, P., Badache, M., & Aidoun, Z. (2015). New correlations for the prediction of the undisturbed ground temperature. *Geothermics*, *53*, 379–384. https://doi.org/10.1016/j.geothermics.2014.08.001

References

Partal, T., & Küçük, M. (2006). Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). *Physics and Chemistry of the Earth*, *31*(18), 1189–1200. https://doi.org/10.1016/j.pce.2006.04.043

Philip, J. R., & De Vries, D. A. (1957). Moisture movement in porous materials under temperature gradients. *Eos, Transactions American Geophysical Union*, *38*(2), 222–232. https://doi.org/10.1029/TR038i002p00222

Rakhecha, P. R., Singh, V. P., Rakhecha, P. R., & Singh, V. P. (2009). Design Storm Estimation. In *Applied Hydrometeorology* (pp. 219–243). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9844-4_10

Silverman, D., & Dracup, J. A. (2000). Artificial neural networks and long-range precipitation prediction in California. *Journal of Applied Meteorology*, *39*(1), 57–66. https://doi.org/10.1175/1520-0450(2000)039<0057:ANNALR>2.0.CO;2

Thoma, E., Tsiros, I. X., Lykoudis, S., & Psiloglou, B. E. (2013). *Applications of Semi-Analytical Models for Estimating Soil Temperature* (pp. 757–763). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29172-2_107

WANG, C., WAN, S., XING, X., ZHANG, L., & HAN, X. (2006). Temperature and soil moisture interactively affected soil net N mineralization in temperate grassland in Northern China. *Soil Biology and Biochemistry*, *38*(5), 1101–1110. https://doi.org/10.1016/j.soilbio.2005.09.009

Xing, L., Li, L., Gong, J., Ren, C., Liu, J., & Chen, H. (2018). Daily soil temperatures predictions for various climates in United States using data-driven model. *Energy*, *160*, 430–440. https://doi.org/10.1016/j.energy.2018.07.004

Zheng, D., Hunt, E. R., & Running, S. W. (1993). A daily soil temperature model based on air temperature and precipitation for continental applications. *Climate Research*, *2*(3), 183–191.

References

https://doi.org/10.3354/cr002183

9-Appendix

Here is the GitHub repo for the codes of this project:

https://github.com/siavashkzb/Soil-temperature.git