SUPPLEMENTARY MATERIAL TO: Myeni et al. S Afr J Sci. 2025;121(5/6), Art. #18235

HOW TO CITE:

Myeni L, Nkhase T, Rapolaki R, Bello Z, Moeletsi ME. Estimation of soil temperature for agricultural applications in South Africa using machine-learning methods [supplementary material]. S Afr J Sci. 2025;121(5/6), Art. #18235. https://doi.org/10.17159/sajs.2025/18235/suppl

---

Artificial neural networks

Artificial neural networks (ANNs) are one of the most widely used machine-learning approaches due to their simplicity and effectiveness in solving sophisticated non-linear problems such as the estimation of agrometeorological variables, including soil temperature (ST).[1-3] In this study, a multilayer perceptron learning algorithm with four layers – the input layer, two hidden layers and an output layer – was used to train the ANN models. Each layer had a specific number of neurons which are independent of each other, and they are connected to the next layer. Initial weights are set randomly during the feedforward process and are then adjusted during backpropagation until the best results are obtained. The learning rate was set to 0.0009 and the number of epochs was set to 100. The activation function used for the input layer and hidden layers was "relu" and for the output layer was "linear' following a similar procedure undertaken in previous studies such as Feng et al.[1] and Imanian et al.[3] Detailed information about the application of ANNs in the estimation of ST can be found in existing literature.[4-6]

Multiple linear regression

Linear regression is based on the assumption that a linear relationship exists between the predictor (independent variable) and output (dependent variable).[7] If more than one predictor is used to predict a dependent variable, it is referred to as multiple linear regression. In the current study, multiple linear regressions with the dependent variable (i.e. ST at a specific depth) against meteorological data as the explanatory variables were developed following a similar procedure outlined in existing literature.[3,6,8]

Decision tree and random forest

A decision tree is based on the set of predictors used to build a binary tree. The random feature selection is used to split a tree node, and the final result is a prediction of a decision tree. To improve decision tree model efficiency, Breiman[9] proposed "bagging" to ensemble decision trees for both classification and regression problems where data are partitioned into subsets of predictors. For model development and training, a random subset of the training data set is randomly taken from the main data set. Generally, on each subset, splits are repeated indefinitely until the splitting does not improve predictions or the nodes contain similar samples.[10] In most cases, two-thirds of the data are used to develop a model and the remainder, commonly known as the out-of-bag (OOB) sample, are used to find the error rate on OOB samples.[11] The final output is obtained by averaging the predictions of all the decision trees (aggregation) that were constructed for the bootstrapped samples from the original data set. The training data set for each decision tree is taken as a subset of a data set ensemble and this reduces the effect of noisy data.[12] An additional benefit of random forest is that it can train multiple decision trees which aid in parallel computing, which is very useful in the meteorological field.[12] In this study, the number of estimators or decision trees was set to 45 for the training of random forest. This implies that the average of 45 decision trees was taken as a final output instead of one decision tree, to avoid overfitting. Detailed information about the application of decision tree and random forest in the estimation of ST can be found in existing literature.[1,3,13-15]

Supplementary table 1: Statistical results of the comparisons between the estimated and measured soil temperatures in specific climatic conditions by different models

Climatic zone	Soil depth (cm)	n	RF				ANN				MLR				DT			
			RMSE	MAE	CCC	*NSE*	RMSE	MAE	CCC	*NSE*	RMSE	MAE	CCC	*NSE*	RMSE	MAE	CCC	*NSE*
BSh	10	847	1.51	1.17	0.96	0.93	1.51	1.19	0.97	0.93	1.74	1.36	0.95	0.91	2.18	1.64	0.93	0.86
	20	847	1.44	1.10	0.96	0.93	1.51	1.18	0.96	0.92	1.76	1.36	0.94	0.89	1.99	1.50	0.93	0.86
	30	847	1.44	1.09	0.96	0.92	1.48	1.16	0.96	0.92	1.84	1.44	0.93	0.87	2.00	1.49	0.92	0.85
	40	847	1.53	1.15	0.95	0.91	1.55	1.23	0.95	0.90	1.98	1.54	0.91	0.84	2.14	1.57	0.91	0.82
	60	847	1.45	1.08	0.95	0.91	1.50	1.17	0.95	0.90	1.96	1.51	0.91	0.82	2.12	1.57	0.90	0.80
	80	847	1.42	1.05	0.95	0.90	1.53	1.18	0.94	0.87	2.02	1.56	0.89	0.80	1.98	1.40	0.91	0.81
BSk	10	588	1.91	1.43	0.95	0.91	2.11	1.62	0.94	0.89	2.34	1.83	0.93	0.87	1.91	1.43	0.95	0.81
	20	588	1.88	1.38	0.95	0.90	2.11	1.60	0.93	0.87	2.40	1.90	0.91	0.84	2.68	1.90	0.90	0.80
	30	588	1.75	1.31	0.95	0.90	2.03	1.58	0.93	0.87	2.31	1.83	0.91	0.83	2.36	1.65	0.91	0.82
	40	588	1.64	1.22	0.95	0.90	1.83	1.43	0.93	0.88	2.32	1.88	0.89	0.81	2.39	1.66	0.90	0.79
	60	588	1.62	1.19	0.95	0.90	1.84	1.42	0.93	0.87	2.25	1.78	0.89	0.81	2.22	1.56	0.91	0.81
	80	588	1.65	1.25	0.94	0.90	1.88	1.49	0.93	0.86	2.29	1.81	0.89	0.80	2.43	1.73	0.88	0.77
Cfa	10	210	0.89	0.67	0.97	0.95	1.18	0.92	0.96	0.91	1.65	1.25	0.91	0.83	1.11	0.80	0.96	0.92
	20	210	1.30	1.00	0.94	0.89	1.55	1.21	0.92	0.84	1.47	1.15	0.92	0.85	1.72	1.32	0.90	0.80
	30	210	1.23	0.94	0.94	0.89	1.55	1.21	0.91	0.82	1.39	1.07	0.92	0.85	1.62	1.23	0.90	0.80
	40	210	1.17	0.90	0.94	0.88	1.51	1.19	0.90	0.80	1.35	1.06	0.91	0.84	1.54	1.17	0.89	0.79
	60	210	1.07	0.81	0.94	0.88	1.41	1.05	0.88	0.76	1.41	1.08	0.89	0.79	1.39	1.03	0.90	0.80
	80	210	1.04	0.78	0.94	0.89	1.55	1.18	0.88	0.75	1.43	1.10	0.88	0.79	1.37	1.02	0.90	0.81
Cwb	10	288	1.29	0.97	0.97	0.95	1.71	1.33	0.95	0.90	1.59	1.25	0.96	0.92	1.89	1.34	0.94	0.88
	20	288	1.14	0.84	0.97	0.95	1.74	1.34	0.94	0.88	1.56	1.19	0.95	0.90	1.58	1.15	0.95	0.90
	30	288	1.10	0.83	0.97	0.94	1.76	1.35	0.93	0.85	1.54	1.19	0.94	0.89	1.45	1.05	0.95	0.90
	40	288	1.01	0.77	0.97	0.94	1.93	1.48	0.90	0.80	1.59	1.21	0.92	0.86	1.43	1.05	0.94	0.89
	60	288	0.98	0.75	0.97	0.94	1.80	1.40	0.90	0.81	1.50	1.19	0.93	0.87	1.39	1.01	0.94	0.88
	80	288	0.83	0.65	0.98	0.95	1.63	1.26	0.91	0.83	1.40	1.08	0.93	0.87	1.20	0.87	0.95	0.91

RF, random forest; ANN, artificial neural network; MLR, multiple linear regression; DT, decision tree; n*, number of observations; RMSE, root mean square error; MAE, mean bias error; CCC, concordance correlation coefficient; NSE, Nash–Sutcliffe efficiency*

Supplementary table 2: Statistical results of the comparisons between the estimated and measured soil temperatures at specific depths by different models at various stations

Station name	RF				ANN				MLR				DT			
	RMSE	MAE	CCC	NSE	RMSE	MAE	CCC	NSE	RMSE	MAE	CCC	NSE	RMSE	MAE	CCC	NSE
Bainsvlei	1.32	1.02	0.97	0.95	1.36	1.06	0.97	0.95	2.10	1.69	0.92	0.86	1.90	1.37	0.94	0.89
University of Limpopo	1.08	0.78	0.96	0.92	1.25	0.96	0.95	0.90	1.45	1.32	0.92	0.85	1.50	1.08	0.92	0.86
Bothaville	1.39	1.07	0.96	0.93	1.40	1.17	0.96	0.92	1.79	1.39	0.94	0.88	1.89	1.32	0.94	0.87
Gravelotte	1.41	1.07	0.93	0.87	1.44	1.17	0.92	0.85	1.71	1.35	0.89	0.81	2.00	1.46	0.89	0.74
Tosca	1.31	0.98	0.98	0.95	1.34	1.06	0.98	0.95	1.66	1.43	0.95	0.91	1.64	1.20	0.95	0.92
Bronkhorstspruit	1.48	1.07	0.94	0.90	1.53	1.19	0.96	0.89	1.76	1.36	0.93	0.85	1.04	0.79	0.97	0.95
Mandeni	1.57	1.19	0.90	0.80	1.41	1.09	0.91	0.83	1.51	1.14	0.90	0.80	1.19	0.91	0.94	0.89
Mean	1.37	1.03	0.95	0.90	1.39	1.10	0.95	0.90	1.71	1.38	0.92	0.85	1.59	1.16	0.94	0.87

RF, random forest; ANN, artificial neural network; MLR, multiple linear regression; DT, decision tree; n*, number of observations; RMSE, root mean square error; MAE, mean bias error; CCC, concordance correlation coefficient; NSE, Nash–Sutcliffe efficiency*

References

1. Feng Y, Cui N, Hao W, Gao L, Gong D. Estimation of soil temperature from meteorological data using different machine learning models. Geoderma. 2019;338:67–77. https://doi.org/10.1016/j.geoderma.2018.11.044
2. Bayatvarkeshi M, Bhagat SK, Mohammadi K, Kisi O, Farahani M, Hasani A, et al. Modeling soil temperature using air temperature features in diverse climatic conditions with complementary machine learning models. Comput Electron Agric. 2021;185, Art. #106158. https://doi.org/10.1016/j.compag.2021.106158
3. Imanian H, Cobo JH, Payeur P, Shirkhani H, Mohammadian A. A comprehensive study of artificial intelligence applications for soil temperature prediction in ordinary climate conditions and extremely hot events. Sustainability. 2022;14(13), Art. #8065. https://doi.org/10.3390/su14138065
4. Nahvi B, Habibi J, Mohammadi K, Shamshirband S, Al Razgan OS. Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature. Comput Electron Agric. 2016;124:150–160. https://doi.org/10.1016/j.compag.2016.03.025
5. Zare Abyaneh H, Bayat Varkeshi M, Golmohammadi G, Mohammadi K. Soil temperature estimation using an artificial neural network and co-active neuro-fuzzy inference system in two different climates. Arab J Geosci. 2016;9:1–10. https://doi.org/10.1007/s12517-016-2388-8
6. Citakoglu H. Comparison of artificial intelligence techniques for prediction of soil temperatures in Turkey. Theor Appl Climatol. 2017;130(1–2):545–556. https://doi.org/10.1007/s00704-016-1914-7
7. Modabber-Azizi S, Salarijazi M, Ghorbani K. Estimation of seasonal and annual river flow volume based on temperature and rainfall by multiple linear and Bayesian quantile regressions. Időjárás. 2022;126(4):567–582. https://doi.org/10.28974/idojaras.2022.4.6

8. Khosravi K, Golkarian A, Barzegar R, Aalami MT, HeddaM S, Omidvar E, et al. Multi-step ahead soil temperature forecasting at different depths based on meteorological data: Integrating resampling algorithms and machine learning models. Pedosphere. 2023;33(3):479–495. https://doi.org/10.1016/j.pedsph.2022.06.056
9. Breiman L. Random forests. Mach Learn. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324
10. Li B, Liang S, Liu X, Ma H, Chen Y, Liang T, et al. Estimation of all-sky 1 km land surface temperature over the conterminous United States. Remote Sens Environ. 2021;266, Art. #112707. https://doi.org/10.1016/j.rse.2021.112707
11. Souaissi Z, Ouarda TBMJ, St-Hilaire A. Non-parametric, semi-parametric, and machine learning models for river temperature frequency analysis at ungauged basins. Ecol Inform. 2023;75, Art. #102107. https://doi.org/10.1016/j.ecoinf.2023.102107
12. Lao P, Liu Q, Ding Y, Wang Y, Li Y, Li M. Rainrate estimation from FY-4a cloud top temperature for mesoscale convective systems by using machine learning algorithm. Remote Sens. 2021;13(16), Art. #3273. https://doi.org/10.3390/rs13163273
13. Dong J, Huang G, Wu L, Liu F, Li S, Cui Y, et al. Modelling soil temperature by tree-based machine learning methods in different climatic regions of China. Appl Sci. 2022;12(10), Art. #5088. https://doi.org/10.3390/app12105088
14. Sihag P, Esmaeilbeiki F, Singh B, Pandhiani SM. Model-based soil temperature estimation using climatic parameters: The case of Azerbaijan Province, Iran. Geol Ecol Landsc. 2020;4(3):203–215. https://doi.org/10.1080/24749508.2019.1610841
15. Sattari MT, Avram A, Apaydin H, Matei O. Soil temperature estimation with meteorological parameters by using tree-based hybrid data mining models. Mathematics. 2020;8(9), Art. #1407. https://doi.org/10.3390/math8091407