



Predicting near surface air temperature using a machine learning based hybrid model approach across Israel

Evyatar Erell (1,2), Bin Zhou (2,4), Ian Hough (1,3), and Itai Kloog (1)

(1) Department of Geography and Environmental Development, Ben-Gurion University, Israel (erell@bgu.ac.il), (2) Jacob Blaustein Institutes for Desert Research, Ben-Gurion University, Israel, (3) Institute for Advanced Biosciences, Université Grenoble Alpes, Grenoble, France, (4) Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany

Global temperature increases over the last decades have increased heat exposure among populations worldwide. An accurate estimate of the resulting impacts on human health demands temporally explicit and spatially resolved monitoring of near-surface air temperature. Neither ground-based nor satellite-based observations can achieve this individually, but the combination of the two provides synergistic opportunities. In this study, we propose a machine learning-based hybrid model to derive $1 \times 1 \text{ km}^2$ gridded near-surface air temperature (T_a) from land surface temperature (T_s) across the complex terrain of Israel. This approach further advances a multi-stage statistical modelling scheme based on linear mixed effects model detailed in (Kloog et al., 2014; Rosenfeld et al., 2017) which has been validated in the USA, France, and Israel. We first applied a random forest regression to impute missing T_s grid cells from the Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua and Terra satellites. Predictor variables for the random forest model include T_s from the geostationary Spinning Enhanced Visible and InfraRed Imager (SEVIRI) satellite and synoptic variables from European Centre for Medium-Range Weather Forecasts' (ECMWF) ERA-5 reanalysis datasets. We evaluated the imputation model's performance using spatial and non-spatial five-fold cross validations. Next, we used linear mixed effect models to calibrate T_a obtained from weather stations and imputed gap-free T_s , taking into account other explanatory variables including Normalized Difference Vegetation Index (NDVI), elevation, and population density. We used this calibration to predict T_a for all grid cells throughout the study area and quantified the accuracy of our predictions using ten-fold cross validation. The newly proposed model outperforms previous ones and provides excellent computationally efficient predictions of air temperature from land surface temperature. This helps further minimize exposure misclassification in epidemiological studies.

Kloog, I., Nordio, F., Coull, B. A., & Schwartz, J. (2014). Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. *Remote Sensing of Environment*, 150, 132–139. <https://doi.org/10.1016/j.rse.2014.04.024>

Rosenfeld, A., Dorman, M., Schwartz, J., Novack, V., Just, A. C., & Kloog, I. (2017). Estimating daily minimum, maximum, and mean near surface air temperature using hybrid satellite models across Israel. *Environmental Research*, 159(March), 297–312. <https://doi.org/10.1016/j.envres.2017.08.017>