Using Machine Learning with Partial Dependence Analysis to Investigate Coupling Between Soil Moisture and Near-surface Temperature

Jared T. Trok[1], Frances V. Davenport[2], Elizabeth A. Barnes[2], and Noah S. Diffenbaugh[1]

[1]Stanford University
[2]Colorado State University

April 4, 2023

Abstract

Soil moisture influences near-surface air temperature by partitioning downwelling radiation into latent and sensible heat fluxes, through which dry soils generally lead to higher temperatures. The strength of this coupled soil moisture-temperature (SM-T) relationship is not spatially uniform, and numerous methods have been developed to assess SM-T coupling strength across the globe. These methods tend to involve either idealized climate-model experiments or linear statistical methods which cannot fully capture nonlinear SM-T coupling. In this study, we propose a nonlinear machine learning-based approach for analyzing SM-T coupling and apply this method to various mid-latitude regions using historical reanalysis datasets. We first train convolutional neural networks (CNNs) to predict daily maximum near-surface air temperature (TMAX) given daily SM and geopotential height fields. We then use partial dependence analysis to isolate the average sensitivity of each CNN's TMAX prediction to the SM input under daily atmospheric conditions. The resulting SM-T relationships broadly agree with previous assessments of SM-T coupling strength. Over many regions, we find nonlinear relationships between the CNN's TMAX prediction and the SM input map. These nonlinearities suggest that the coupled interactions governing SM-T relationships vary under different SM conditions, but these variations are regionally dependent. We also apply this method to test the influence of SM memory on SM-T coupling and find that our results are consistent with previous studies. Although our study focuses specifically on local SM-T coupling, our machine learning-based method can be extended to investigate other coupled interactions within the climate system using observed or model-derived datasets.

Using Machine Learning with Partial Dependence Analysis to Investigate Coupling Between Soil Moisture and Near-surface Temperature

Jared T. Trok[1], Frances V. Davenport[1,2,3], Elizabeth A. Barnes[2], and Noah S. Diffenbaugh[1,4]

[1]Department of Earth System Science, Stanford University, Stanford, CA, USA.

[2]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

[3]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, USA.

[4]Doerr School of Sustainability, Stanford University, Stanford, CA, USA.

Corresponding author: Jared Trok (trok@stanford.edu)

Key Points:

- We investigate land-atmosphere interactions by applying machine learning techniques to reanalysis datasets
- Partial dependence analysis reveals new insights into nonlinear summertime soil moisture-temperature coupling and soil moisture memory
- These relationships broadly agree with previous studies, supporting machine learning as a method for quantifying surface-atmosphere coupling

Abstract

Soil moisture influences near-surface air temperature by partitioning downwelling radiation into latent and sensible heat fluxes, through which dry soils generally lead to higher temperatures. The strength of this coupled soil moisture-temperature (SM-T) relationship is not spatially uniform, and numerous methods have been developed to assess SM-T coupling strength across the globe. These methods tend to involve either idealized climate-model experiments or linear statistical methods which cannot fully capture nonlinear SM-T coupling. In this study, we propose a nonlinear machine learning-based approach for analyzing SM-T coupling and apply this method to various mid-latitude regions using historical reanalysis datasets. We first train convolutional neural networks (CNNs) to predict daily maximum near-surface air temperature (TMAX) given daily SM and geopotential height fields. We then use partial dependence analysis to isolate the average sensitivity of each CNN's TMAX prediction to the SM input under daily atmospheric conditions. The resulting SM-T relationships broadly agree with previous assessments of SM-T coupling strength. Over many regions, we find nonlinear relationships between the CNN's TMAX prediction and the SM input map. These nonlinearities suggest that the coupled interactions governing SM-T relationships vary under different SM conditions, but these variations are regionally dependent. We also apply this method to test the influence of SM memory on SM-T coupling and find that our results are consistent with previous studies. Although our study focuses specifically on local SM-T coupling, our machine learning-based method can be extended to investigate other coupled interactions within the climate system using observed or model-derived datasets.

Plain Language Summary

Soil moisture content influences air temperature by controlling evaporation at the soil surface. Dry soils reduce evaporation which warms the surface and leads to higher air temperatures. Conversely, wet soils generally lead to cooler temperatures. This process results in a coupled relationship between soil moisture and temperature. Soil moisture-temperature (SM-T) coupling occurs everywhere but is especially strong in certain areas of the world. Over recent decades, numerous methods have been developed to measure regional differences in SM-T coupling strength. These studies agree on certain "hot spots" where this coupling relationship is particularly strong. However, these previous studies rely on idealized climate model experiments or linear statistics which cannot fully capture nonlinear SM-T coupling. To address this, we apply nonlinear machine learning techniques to investigate SM-T coupling. Our results show that this method captures the nonlinear characteristics of SM-T coupling and agrees well with previously documented coupling hot spots. Our method also provides a framework for using machine learning to investigate other coupled processes in the Earth system.

1 Introduction

Since the early 1980's, climate model experiments have confirmed that soil moisture content (SM) influences near-surface air temperature by modulating the surface energy budget (Shukla & Mintz, 1982). This coupled relationship between soil moisture and temperature (hereafter, "SM-T coupling") results from complex interactions between the land surface and the atmosphere. In regions with strong SM-T coupling, SM content controls the partitioning of downwelling radiation into latent and sensible heat fluxes, resulting in a positive feedback mechanism through which dry soils lead to higher temperatures and further soil drying, while wet soils generally lead to cooler temperatures (Seneviratne et al., 2010). Second-order positive feedback mechanisms have also been observed between soil moisture, boundary layer growth, 1000–500-hPa thickness, and near-surface temperature (Fischer et al., 2007; Miralles et al., 2014; Quesada et al., 2012; Seneviratne et al., 2010). These SM-T coupling mechanisms tend to be strongest in transitional regimes between wet and dry climates, which is consistent with the theoretical framework of Seneviratne et al., (2010). In wet and dry climate regimes, near-surface temperature is less sensitive to SM (i.e., decoupled) since evapotranspiration is limited by radiation and soil properties, respectively (Seneviratne et al., 2010). However, in transitional climate regimes, near-surface temperature is highly sensitive to SM content because small changes in SM influence evapotranspiration, which directly affects latent and sensible heat fluxes (Seneviratne et al., 2010). Together with SM content, differences in soil characteristics (e.g., albedo, porosity, texture) and land cover type also drive regional differences in SM-T coupling strength (Dennis and Berbery 2021; Hirsch et al. 2014).

SM-T coupling has both local (Durre et al., 2000; J. Liu & Pu, 2019) and non-local (i.e., downwind) effects (Schwingshackl et al., 2018; Seneviratne et al., 2013; Vautard et al., 2007) that occur on daily, monthly, and seasonal time scales (Durre et al., 2000; Fischer et al., 2007; Koster et al., 2006a; J. Liu & Pu, 2019; Vautard et al., 2007). Deep soil layers (10-200 cm) have longer SM memory (Wu & Dickinson, 2004), which makes these layers more important for monthly- and seasonal-scale SM-T coupling (Koster et al., 2006a). In contrast, the uppermost soil layer (< 10 cm) has the greatest influence on daily-scale SM-T coupling (J. Liu & Pu, 2019). Further, the potential for SM-T coupling is highest during daylight hours in the summer months (due in large part to the maximum of downwelling solar radiation; Durre et al., 2000; Koster et al., 2006a; J. Liu & Pu, 2019), which makes daily-scale SM-T coupling especially relevant for producing extreme daily maximum summer temperatures (Diffenbaugh et al., 2007; Miralles et al., 2014; Schwingshackl et al., 2017; Seneviratne et al., 2010; Vogel et al. 2017). As a result, we focus our analysis primarily on daily-scale coupling between top-layer SM and daily maximum 2-meter temperature in the summer months.

Over the past two decades, many studies have quantified regional differences in SM-T coupling strength using observational (Chen et al., 2019; Dirmeyer, 2011; Koster et al., 2009; Mei & Wang, 2012; Miralles et al., 2012; Spennemann et al., 2018; Teuling et al., 2009) and model-derived datasets (Fischer et al., 2007; Jaeger et al., 2009; Koster et al., 2006a, 2009; Mei & Wang, 2012; Ruscica et al., 2014; Schwingshackl et al., 2017; Seneviratne, Lüthi, et al., 2006). Global assessments of SM-T coupling strength typically involve comparing climate model simulations under different soil moisture scenarios (e.g., Fischer et al., 2007; Koster et al., 2006a; Seneviratne, Lüthi, et al., 2006) or analyzing linear statistics (e.g., correlation coefficients) between land-surface and/or atmospheric variables (e.g., Diffenbaugh & Ashfaq, 2010; Dirmeyer, 2011; Jaeger et al., 2009; Seneviratne, Lüthi, et al., 2006; Teuling et al., 2009). Regardless of the methodology, previous assessments broadly agree on certain transitional climate regimes as “hot spots” of SM-T coupling (e.g., the US Southern Great Plains, the Sahel region in Africa, areas of the Indian subcontinent). However, these studies consistently disagree on the relative magnitudes of SM-T coupling strength within certain regions. Inconsistencies between SM-T coupling studies can result from numerous sources, including climate model disagreement (Gevaert et al., 2018), model initializations (Fischer et al., 2007), experimental design (e.g., potential sea surface temperature effects; Koster et al., 2006a), and differences between climate model and reanalysis datasets (e.g., stronger SM-evaporative fraction coupling in reanalysis compared to climate models; Mei & Wang, 2012). In climate model-based assessments of SM-T coupling, additional inconsistencies can be caused by differences in model parameterization of soil hydraulic properties, plant hydraulic properties, vegetation type, and land use (Dennis and Berbery 2021; Hirsch et al. 2014).

Importantly, analyses of SM-T coupling strength (e.g., Dirmeyer, 2011; Fischer et al., 2007; Jaeger et al., 2009; Koster et al., 2006a; Menendez et al., 2019; Miralles et al., 2012; Ruscica et al., 2014; Seneviratne, Lüthi, et al., 2006; Teuling et al., 2009) have tended to use idealized climate model experiments and/or linear statistical methods to explain SM-T coupling. However, evidence suggests that the sensitivity of temperature to SM changes for different values of SM (Benson & Dirmeyer, 2021; Jaeger & Seneviratne, 2011; Seneviratne et al., 2010). This nonlinear relationship between temperature and SM is difficult to estimate using climate model experiments, requiring a large number of sensitivity experiments with slightly perturbed SM conditions repeated over numerous different atmospheric initializations (Fischer et al., 2007; Seneviratne et al., 2010). There is thus an opening for nonlinear statistical methods that can comprehensively assess SM-T coupling relationships without requiring extensive climate model simulations.

Deep neural networks have recently surged in popularity for their ability to learn complex nonlinear interactions between input and output variables (LeCun et al., 2015). Convolutional neural networks (CNNs) are one particular form of deep learning architecture that are designed to analyze gridded input data such as images and geospatial data (LeCun et al., 1989). To date, CNNs have been used extensively in the geosciences for image classification (Chilson et al., 2019; Davenport & Diffenbaugh, 2021; Jergensen et al., 2019; Lagerquist et al., 2019; Y. Liu et al., 2016; Wang et al., 2016; Wimmers et al., 2019), model parameterization (Bolton & Zanna, 2019; Han et al., 2020; Larraondo et al., 2019; Pan et al., 2019), and forecasting (Ham et al., 2019; Jacques-Dumas et al., 2021) applications. CNN models contain thousands (or millions) of trainable weights which are optimized during the training process to ensure that the CNN's output predictions closely resemble the target data. In addition, these CNN models utilize nonlinear mathematical functions to represent the complex nonlinear relationships between the geospatial input maps and output predictions. After the training process is complete, machine-learning (ML) model interpretation and visualization methods can be used to aid in interpreting the predictions of trained CNNs (e.g., layer-wise relevance propagation, S. Bach et al., 2015; backward optimization, Olah et al., 2017; etc.). These ML interpretation methods have been used in the geosciences to confirm that a model's predictions are based on the inputs in a physically meaningful way (Davenport & Diffenbaugh, 2021; Diffenbaugh & Barnes, 2023; Gagne et al., 2019; McGovern et al., 2019). More recently, studies have also begun to use ML interpretation methods to gain new insights into physical processes (Barnes, Mayer, et al., 2020; Barnes, Toms, et al., 2020; Toms et al., 2020; Zhang et al., 2021).

Although applications of ML interpretation techniques are increasingly commonplace in the geosciences, these techniques have the potential to give non-physical and/or misleading results (Mamalakis et al. 2022; Ebert-Uphoff and Hilburn 2020). Typically, the results of ML interpretation methods are deemed trustworthy by visually comparing results against prior knowledge. This works well in cases where the processes are well understood and a ground-truth comparison is available (Davenport & Diffenbaugh, 2021; Gagne et al., 2019; McGovern et al., 2019). However, it remains difficult to validate ML interpretation results when investigating new or poorly understood processes. Recently, the construction of synthetic benchmark datasets where the discoverable relationships are known a priori have been proposed as a way to assess the fidelity of ML interpretation results (Ebert-Uphoff & Hilburn, 2020; Mamalakis et al., 2022). Here, we show that by applying ML interpretation techniques to modified versions of our training dataset we can validate our results and gain additional insights into physical processes.

Partial dependence plots (PDPs; Friedman, 2001) are a common ML interpretation technique which can be used to visualize the nonlinear relationships that a model has learned between the input and output variables (Goldstein et al., 2015; Jergensen et al., 2019; McGovern et al., 2019). However, PDPs are rarely used to analyze deep-learning architectures (such as CNNs) for geoscience applications (Zhang et al., 2021). PDPs are infeasible for most deep-learning applications (especially those with a large number of inputs) because they require an assumption of independence between all input variables (McGovern et al., 2019). If variables are strongly correlated, certain combinations of input variables will not likely occur in nature, and the CNN will be forced to extrapolate beyond the training dataset in order to calculate the PDP (which can yield non-physical results). Additionally, in order to apply PDPs to CNNs we must have a physically meaningful way to sort geospatial input maps along a continuous axis (which can be difficult depending on the application). In spite of these limitations, PDPs show promise as a tool for analyzing CNNs to better understand complex nonlinear relationships within geospatial datasets, provided that the input variables are not too strongly correlated, and that the application is focused on quantifying the relationship between the output prediction and some quantity calculated from the input maps.

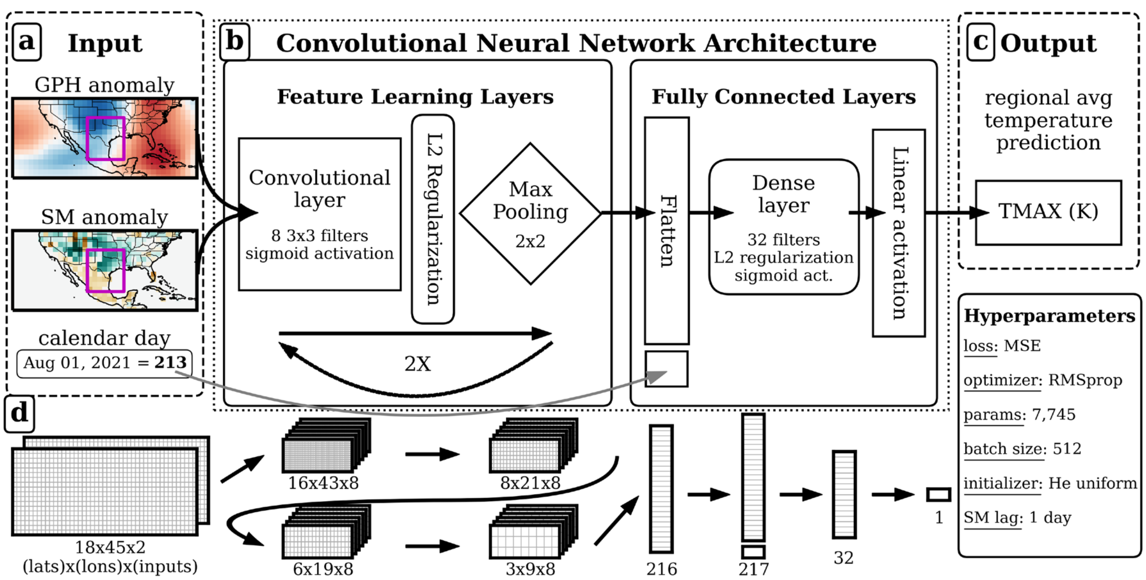

Figure 1. Schematic of the convolutional neural networks used in this analysis. (**a**) Model is given the following inputs: 500 millibar geopotential height (GPH) anomaly map, 0-7 cm volumetric soil moisture (SM) fraction anomaly map, and an integer input corresponding to the calendar day (normalized to fall between 0 and 1). Pink box shows the temperature prediction region. (**b**) The spatial input maps undergo feature learning as they are passed through a convolutional layer with 8 3x3 filters using sigmoid activation, followed by an L2 regularization layer (to reduce overfitting), and a 2x2 max pooling layer. These three feature learning layers repeat twice. The output from the feature learning layers is then flattened, and the normalized calendar day input is concatenated onto the end. The flattened vector is passed through a fully-connected dense layer with 32 neurons, L2 regularization, and sigmoid activation. Lastly, we use a linear activation function which outputs (**c**) the predicted TMAX. (**d**) The input and output size of each layer in the convolutional neural network. (**e**) Several hyperparameters used to construct and train each model.

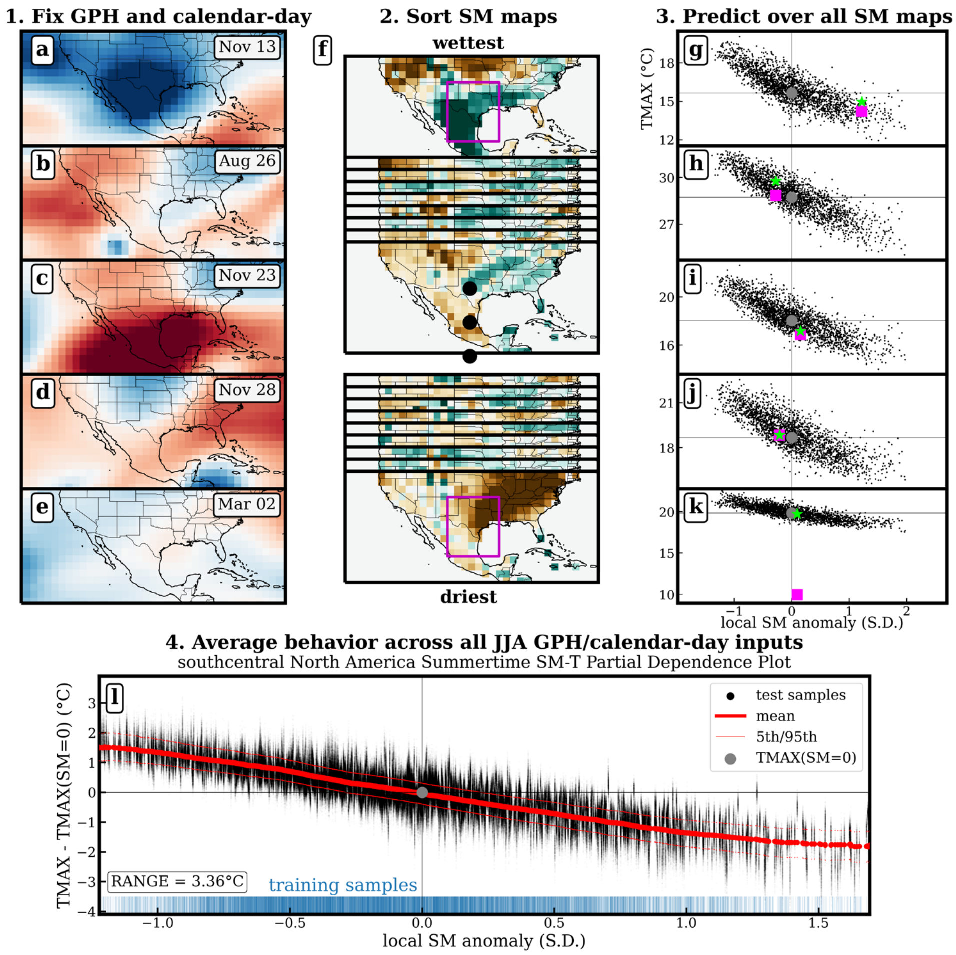

Figure 2. Schematic showing how partial dependence analysis is used to derive the nonlinear soil moisture-temperature coupling relationship that the convolutional neural network has learned through the training process. Shown is an example from a region in southcentral North America. (**1**) We take a single 500 millibar geopotential height map and the calendar day on which that map occurs. (**2**) We then pair this single GPH/calendar-day combination with every possible soil moisture anomaly input map (in the testing dataset) sorted from driest-wettest (**f**) according to local SM anomaly (area-weighted average of all non-ocean grid cells inside the pink box). (**3**) We then pass these new input combinations through a trained convolutional neural network to obtain daily maximum temperature (TMAX) predictions for a single GPH/calendar-day combination over the entire range of SM anomaly maps. (**4**) We repeat steps (**1**)-(**3**) and average the behavior across all summertime GPH/calendar-day combinations (in the 8-year testing dataset) to obtain the nonlinear soil moisture-temperature coupling relationship (**l**) that the convolutional neural network has learned through the training process. The 5 GPH/calendar-day examples (**a**-**e**) are chosen for lowest GPH anomaly, median GPH anomaly, highest GPH anomaly, model best-hit, and model worst-miss, respectively. The corresponding temperature predictions for these 5 examples are given in (**g**)-(**k**). The pink marker in (**g**)-(**k**) indicates the actual ERA5-Land temperature that occurred on that particular day. The green marker in (**g**)-(**k**) shows the model predicted temperature. The black marker in (**g**)-(**k**) shows the average model prediction for SM anomalies near zero, or TMAX(SM=0). Model predictions for each GPH/calendar-day combination in the testing dataset are shifted by TMAX(SM=0), then averaged to obtain the SM-T relationship in (**l**). We also include a rug plot showing the distribution of SM anomalies in the training dataset. Soil moisture anomalies are calculated as standard deviations (S.D.) from the calendar-day mean.

In this study, we apply partial dependence analysis to investigate daily-scale nonlinear SM-T coupling relationships over sixteen midlatitude regions in the Northern and Southern Hemispheres. Over each prediction region, we train a CNN to predict daily maximum temperature using several input variables, including atmospheric pressure patterns and soil moisture (Figure 1). Next, we use partial dependence plots (PDPs) to visualize how the CNN's temperature prediction changes as we vary the SM input (while holding all other inputs constant;

Figure 2). The resulting SM-T PDP shows the average sensitivity of the CNN's daily temperature prediction to the SM input. To ensure that these SM-T relationships are robust, we confirm that each CNN meets minimum performance criteria and compare our SM-T PDPs against those obtained from modified versions of our training datasets where we systematically reduce and/or eliminate the potential for SM-T coupling.

2 Data and Methods

2.1 Datasets

We construct two neural network training datasets which use daily mean 500-hPa geopotential height (GPH) anomalies and daily mean surface-layer volumetric soil moisture fraction (SM) anomalies as predictors of regional average daily maximum 2-meter air temperature (TMAX) over the 1979-2021 period. We focus on daily TMAX (as opposed to daily minimum or daily mean temperature) because the coupling between surface-layer SM and 2-meter temperature is most relevant during daylight hours (when SM controls the partitioning of downwelling solar radiation into sensible and latent heat fluxes).

Our primary dataset consists of GPH, SM, and TMAX from the ERA5/ERA5-Land historical reanalysis (ERA5, Hersbach et al., 2018; ERA5-Land, Muñoz-Sabater et al., 2021) provided by the European Centre for Medium-Range Weather Forecasts and downloaded from the Copernicus Climate Change Service Climate Data Store. We use ERA5 hourly 500-hPa geopotential provided globally at 0.25°×0.25° horizontal resolution. We then divide the geopotential by Earth's gravitational acceleration (9.80665 m s^{-2}) to obtain hourly 500-hPa GPH fields in meters above mean sea level. We use ERA5-Land hourly 0-7 cm SM fraction and hourly 2-meter air temperature provided globally at 0.1°×0.1° horizontal resolution. We then aggregate the ERA5/ERA5-Land hourly fields to obtain daily mean GPH, daily mean SM, and daily TMAX. Lastly, we convert the ERA5 GPH and SM fields to a T62 gaussian grid at 1.875°×1.875° horizontal resolution to match the resolution of our comparison dataset, and to reduce computational expense.

Our comparison dataset (used in supplemental analysis) consists of GPH, SM, and TMAX from the NCEP/DOE Reanalysis II (NCEP; Kanamitsu et al., 2002) historical reanalysis downloaded from the NOAA Physical Science Laboratory data archive at https://psl.noaa.gov. Daily mean 0-10 cm SM fraction and daily 2-meter TMAX are available globally on a T62 gaussian grid at 1.875°×1.875° horizontal resolution. Using bilinear interpolation, we convert the NCEP daily mean 500-hPa GPH fields from a 2.5°×2.5° rectangular grid to the T62 gaussian grid to match the SM and TMAX fields (regridding performed using NetCDF Operators; Zender, 2008).

Since this analysis focuses on land-atmosphere interactions at daily timescales, we first subtract the 1979-2021 area-weighted regional-mean linear trends from the GPH, SM, and TMAX fields in both datasets (Cattiaux et al., 2013). By subtracting spatially averaged trends, we avoid the impacts of uniform tropospheric thermal expansion and near-surface warming on our training datasets, while still preserving the non-uniform spatial trends in GPH and SM that are important drivers of regional TMAX (Horton et al., 2015; Swain et al., 2016). For both ERA5 and NCEP, we then use the daily mean GPH and SM maps to calculate daily standardized anomalies (i.e., z-scores) by subtracting grid-cell calendar-day means and dividing by grid-cell calendar-day standard deviations. All missing SM values (non-land grid cells) are assigned a zero anomaly to avoid numerical issues with missing values during neural network training.

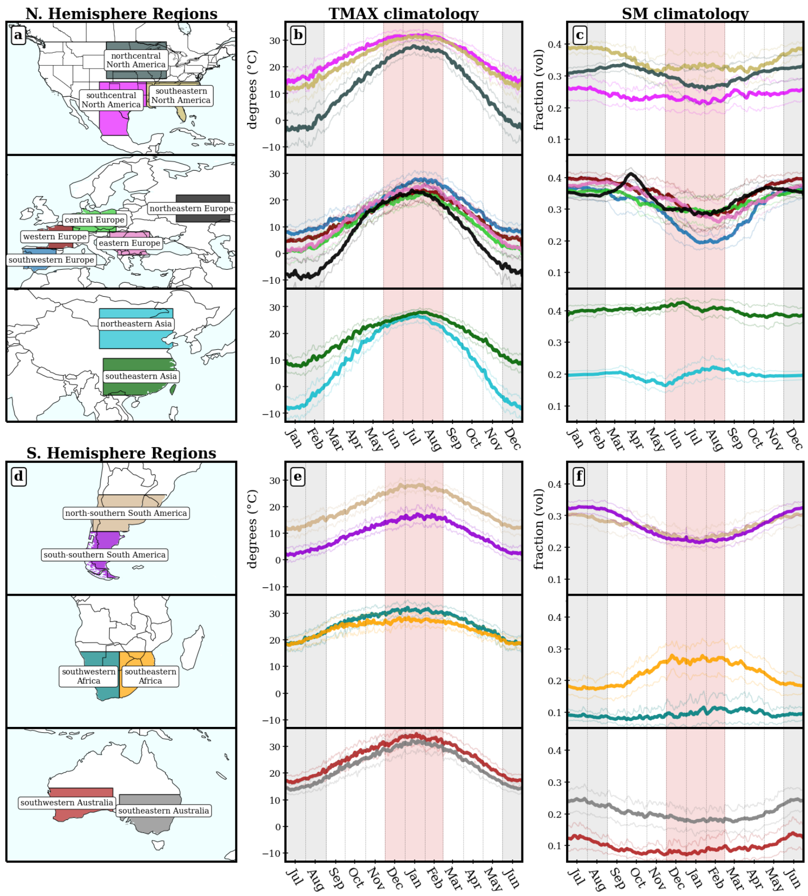

Figure 3. (**a**) Northern Hemisphere regions included in this analysis alongside 1979-2021 regional climatologies of (**b**) daily maximum 2-meter temperature (TMAX), and (**c**) volumetric soil moisture fraction (SM). (**d**, **e**, **f**) Same as (**a**, **b**, **c**) but for Southern Hemisphere regions. Red shading indicates summer months in each hemisphere over which this study analyzes soil moisture-temperature coupling. Gray shading indicates winter months removed from all subsequent analyses. Thin colored lines show +/- 1 standard deviation. TMAX and SM climatologies derived from ERA5-Land dataset.

2.2 Regions

We define sixteen prediction regions chosen to encompass a wide range of mid-latitude climate regimes, including known land-atmosphere coupling "hot spots" (as proposed by, e.g., Fischer et al., 2007; Koster et al., 2006b; Mei & Wang, 2012; Seneviratne, Lüthi, et al., 2006). The sixteen midlatitude regions (Figure 3) are: northcentral North America (38°N-49°N, 86°W-104°W), southcentral North America (21°N-37°N, 92°W-106°W), southeastern North America (25°N-37°N, 75°W-92°W), southwestern Europe (36°N-43°N, 10°W-1°E), western Europe (43°N-50°N, 5°W-6°E), central Europe (48°N-55°N, 6°E-19°E), eastern Europe (41°N-48°N, 17°E-29°E), northeastern Europe (51°N-59°N, 37°E-53°E), northeastern Asia (36°N-48°N, 99°E-121°E), southeastern Asia (22°N-33°N, 100°E-122°E), north-southern South America (30°S-41°S, 51°W-73°W), south-southern South America (41°S-55°S, 63°W-76°W), southwestern Africa (20°S-35°S, 12°E-25°E), southeastern Africa (20°S-35°S, 25°E-36°E), southwestern Australia (25°S-36°S, 112°E-133°E), and southeastern Australia (27°S-39°S, 135°E-154°E). The extent of the prediction regions (roughly 800-1100 km across) is determined based on the approximate size of mid-latitude weather patterns.

Over each of these prediction regions (Figure 3), we construct neural network training datasets (as detailed in Section 2.1). Each regional CNN uses standardized GPH and SM anomaly maps as predictors of regional average TMAX. We calculate regional average TMAX by taking an area-weighted mean over all non-ocean grid cells that fall within the region bounds. In order to provide sufficient spatial context for each regional TMAX prediction, we use broad GPH and SM anomaly input maps of 45 longitude points × 18 latitude points (at 1.875°×1.875° horizontal resolution), centered around the prediction region (see Figure 1 for an example of these input maps). Our choice of CNN input size (i.e., 45 longitude points × 18 latitude points) is based on the approach of Davenport and Diffenbaugh (2021), who showed that a CNN input map extending 35 degrees latitudinally and 85 degrees longitudinally provides sufficient spatial context for classifying GPH patterns associated with extreme precipitation over a similarly sized mid-latitude prediction region in the US Midwest.

2.3 Convolutional Neural Network (CNN) Architecture

We train a separate CNN regression model (Figure 1) to predict average daily TMAX over each prediction region using daily 500-hPa GPH anomalies, daily surface-layer SM anomalies, and calendar-day inputs. For each day in the training set, the neural network receives the calendar day (normalized to fall between 0 and 1) and a 3-dimensional spatial input matrix (18×45×2; lat×lon×inputs) containing the GPH map from the day of the prediction and the SM anomaly map from 1 day prior to prediction. We use SM inputs from 1 day prior to the prediction in order to avoid potential impacts of daily TMAX on daily SM. The spatial inputs then undergo feature learning as they are passed through two convolutional layers (8 3×3 filters with sigmoid activation) each followed by a 2×2 max pooling layer. After feature learning, the resulting feature maps are flattened into a 1-dimensional vector and the normalized calendar-day input is concatenated to the end. This vector is then passed through a fully-connected (dense) layer (32 neurons with sigmoid activations) followed by a final dense layer with linear activations which output a single TMAX prediction. The TMAX predictions are then compared to the target TMAX values from the training dataset, and CNN layer weights (initialized with He uniform; He et al., 2015) are adjusted using RMSprop (Hinton, Srivastava, & Swersky, 2012) in order to minimize the loss function (mean squared error; MSE). To reduce overfitting during the training process, we use L2 activity regularization on both convolutional layers and the dense layer. We also use early stopping with a patience threshold of 100 epochs which halts the training process and returns the optimal weights when validation loss stops improving. After the training process is complete, we save the model weights and use the trained model to predict TMAX over all days.

Prior to neural network training, we randomly split the 43-year datasets into training (27-year), validation (8-year), and testing (8-year) subsets while keeping calendar years intact. By keeping calendar years intact, we further reduce the potential for overfitting between chronologically adjacent days in different subsets which may look nearly identical due to slow day-to-day variations in SM, GPH, and TMAX. Each subset consists of randomly selected years (instead of a consecutive N-year period) to avoid potential impacts of interdecadal climate variability, land use change, anthropogenic climate forcing, and trends in land-atmosphere interactions which could otherwise prevent a fair evaluation of our model. We use different training/validation/testing subsets for each region in order to ensure that the target TMAX distributions are roughly equivalent between each subset. To avoid potential impacts of snow cover on land-atmosphere coupling (Dutra et al., 2011; Henderson et al., 2018), we remove the three canonical winter months in each hemisphere (December-January-February in the Northern Hemisphere and June-July-August in the Southern Hemisphere). This yields a total of 7425 training samples, 2200 validation samples, and 2200 testing samples for each Northern Hemisphere region; and 7378 training samples, 2186 validation samples, and 2186 testing samples for each Southern Hemisphere region. During training, model parameters are fit to the training data and hyperparameters are adjusted to minimize loss on the validation data. Once the training is complete, we predict TMAX on the unseen testing subset.

We optimize CNN architecture and hyperparameters using scikit-learn's GridSearchCV (Pedregosa et al., 2011), including: layer number/organization, filter number/size, loss function, optimizer, activation functions, weight initializers, and batch size. Additional hyperparameters such as initial learning rate, learning rate decay, and L2 activity regularization factor are optimized separately for each regional model in order to minimize loss on the validation subset. Due to the non-uniform nature of TMAX distributions, we use the DenseWeight/DenseLoss algorithm (Steininger et al., 2021) to perform imbalanced regression by weighting the loss function for each sample using weights inversely proportional to sample frequencies (calculated via kernel density estimation). The DenseWeight hyperparameter (which controls the degree of weighting) is optimized separately for each regional CNN and substantially improves model performance on extreme TMAX days. Although sinusoidal-based positional encoding is commonly used to encode temporal cycles as a CNN input variable, this method forces a seasonal symmetry in the input data. Given that a region's seasonal cycle of TMAX and SM are not symmetric (e.g., Figure 3), we instead use a normalized calendar-day integer input for this prediction task. We use Tensorflow with Keras 2.7.0 (Tensorflow Developers, 2021) to construct and train each model.

2.4 Evaluating CNN Performance

Prior to using the regional CNNs to quantify SM-T coupling strength, we must first evaluate whether the CNNs are sufficiently accurate to represent SM-T coupling at daily timescales over the respective regions. To that end, we first ensure that each CNN meets two criteria: (1) the CNN accurately predicts TMAX at daily timescales, and (2) the SM input contributes substantially to overall CNN performance at daily timescales.

To determine if a CNN meets these criteria for a given region, we compare the performance of our CNN against two model performance baselines:

a. Seasonal climatology: comparison between the calendar-day mean TMAX and the actual daily TMAX on individual calendar days;

b. CNN without SM input: performance of a CNN model trained with GPH and calendar-day inputs but no SM input maps.

We first compare the performance metrics (e.g., R^2, MAE, MSE) of our CNNs with those of the seasonal-climatology baseline. Any model which outperforms the seasonal-climatology baseline should, to some degree, be able to predict daily TMAX anomalies from the seasonal cycle. Then, to justify whether the SM input contributes to overall model skill at daily timescales, we compare the performance of our CNNs with all input variables against the CNN-without-SM baseline. The difference in skill between these models helps to quantify how much the SM input contributes to overall model skill at daily timescales. If the CNN with all input variables outperforms the CNN-without-SM baseline, and both of these CNN models outperform the seasonal-climatology baseline, then we can more confidently use the full CNN to assess SM-T coupling at daily timescales.

2.5 Evaluating Coupling Strength Using Partial Dependence

After training and evaluating our CNNs, we apply partial dependence analysis (Figure 2; Friedman, 2001) to visualize the nonlinear relationships between each CNN's summertime TMAX predictions and the average local SM anomaly calculated from the SM input maps. Although the training datasets include data from all nine non-winter months in each hemisphere, we only assess SM-T coupling over the three canonical summer months (when the potential for SM-T coupling is highest; Koster et al., 2006b). First, we select a single GPH anomaly map and the corresponding calendar-day input from a summer day in the testing dataset (Figure 2a-e). Holding this GPH and calendar-day input constant, we pair these fixed inputs with every daily SM map (in the testing dataset) sorted from driest to wettest according to the prediction region's average SM anomaly (area-weighted mean over all non-ocean grid cells; Figure 2f). Then, we use each trained CNN to predict TMAX from these newly constructed input combinations and visualize the results to assess how the CNN's TMAX prediction depends on the average local SM anomaly under daily GPH conditions (Figure 2g-k). We repeat this process for all summer days in the testing dataset (8 years) and compute the two-sided moving average (200 points on either side) to obtain the smoothed regional summertime SM-T partial dependence plot (PDP) that the CNN has learned through the training process (Figure 2l). Our two-sided moving average is calculated using smaller window sizes near the extreme SM anomalies to ensure an equal number of points on each side. We also remove the 10 driest and 10 wettest SM anomaly maps (in the testing dataset) from the PDP calculation in order to avoid biasing the results at extreme SM anomalies that are underrepresented in the training dataset. Areas of the PDP plot with non-zero SM-T PDP slope indicate where the CNN's TMAX prediction is sensitive to the local SM anomaly calculated from the SM input map (McGovern et al., 2019). We also use the vertical extent (range) of our SM-T PDPs as a relative indicator of SM-T coupling strength.

In order to compare the effects of SM anomalies across different days, we compute PDPs using centered TMAX predictions (Goldstein et al., 2015). For each day, we calculate the change in TMAX relative to the model's average prediction near climatological SM conditions (i.e., TMAX(SM=0)). We estimate TMAX(SM=0) each day by averaging the closest 200 daily predictions that fall on either side of the calendar-day mean SM anomaly (i.e., SM=0). Estimation of TMAX(SM=0) is largely insensitive to the choice of window size (i.e., 200 predictions on either side).

Because partial dependence analysis also relies on the assumption that all input variables are independent from one another (Friedman, 2001), we use SM and GPH calendar-day anomalies to remove seasonal variability. However, there still remains the potential for interaction effects between SM and GPH which may cause the CNN to learn different SM-T relationships for different GPH inputs. In this case, SM-T PDP curves can be misleading since they would average out these divergent SM-T relationships. We address this issue by including density plots of daily TMAX predictions alongside the PDPs. From these density plots, we can confirm that the PDPs are not averaging out divergent SM-T relationships caused by a violation of the independence assumption between SM and GPH inputs (Goldstein et al., 2015).

To assess the fidelity of our PDP-based approach, we apply the PDP method (Figure 2) to modified versions of our training datasets (i.e., baseline datasets) in which we have eliminated the potential for SM-T coupling. We construct a single baseline dataset by randomly shuffling the 1979-2021 daily SM input maps while leaving the GPH and calendar-day inputs untouched. Then, we train a new CNN using this baseline dataset, save the model weights, and apply the PDP method to obtain a baseline SM-T relationship. We repeat this process for numerous baseline datasets, each created with a different random seed. Randomizing the SM maps removes any statistical link between SM inputs and TMAX outputs within these baseline datasets. Therefore, we expect each baseline SM-T relationship to have zero slope. Using an approach similar to Buja et al., (2009) and Wickham et al. (2010), we then compare the true PDPs against 100 baseline PDPs (each obtained from a different baseline dataset) to determine whether the true PDP exhibits a relationship with SM beyond that of random noise.

3 Results

We show TMAX and SM climatologies calculated from the ERA5-Land dataset (1979-2021) for each of the sixteen mid-latitude regions (Figure 3). All sixteen regions experience their highest temperatures during the summer months and lowest temperatures during the winter months (Figures 3b and 3e). However, there are large regional differences in the magnitude of the TMAX seasonal cycle, ranging from ±10°C in southwestern Africa and southeastern Africa to ±35°C in northeastern Europe and northeastern Asia. Although SM seasonal cycles differ substantially between regions, nearly all regions experience their driest SM conditions in the summer months (with the exception of northeastern Asia, southeastern Asia, southwestern Africa, and southeastern Africa; Figures 3c and 3f). For most regions, we find that the TMAX and SM climatologies also show these patterns in the NCEP/DOE Reanalysis II dataset (Figure S5). (See Methods for additional information about region selection.)

3.1 CNN Model Evaluation

For each region, we compare the performance of our CNN regression models against two model performance baselines (detailed in Section 2.4; Figures 4-6). Across all regions, the CNN-without-SM baselines outperform the seasonal-climatology baseline (i.e., Figures 4-6(b) vs. Figures 4-6(c)), ranging from a minimum RMSE reduction of 10.4% in southeastern Asia to a maximum RMSE reduction of 50.7% in southwestern Europe. We also find that our CNN models with all input variables (including SM inputs) outperform the CNN-without-SM baselines (i.e., Figures 4-6(a) vs. Figures 4-6(b)), ranging from a minimum RMSE reduction of 8.1% in northcentral North America to a maximum RMSE reduction of 24.8% in southwestern Africa. These improvements in CNN model skill indicate that both GPH inputs and SM inputs each provide unique information that is useful for predicting TMAX at daily timescales. Therefore, we find that all regional CNNs satisfy the necessary criteria (detailed in Section 2.4) to confidently use partial dependence analysis to assess daily-scale SM-T coupling. (We further analyze each CNN's ability to predict daily TMAX anomalies ($0.38 \leq R^2 \leq 0.80$) as opposed to absolute values, and the TMAX seasonal cycle ($0.92 \leq R^2 \leq 0.99$, $0.58^oC \leq RMSE \leq 1.16^oC$); Figures S1 and S2.)

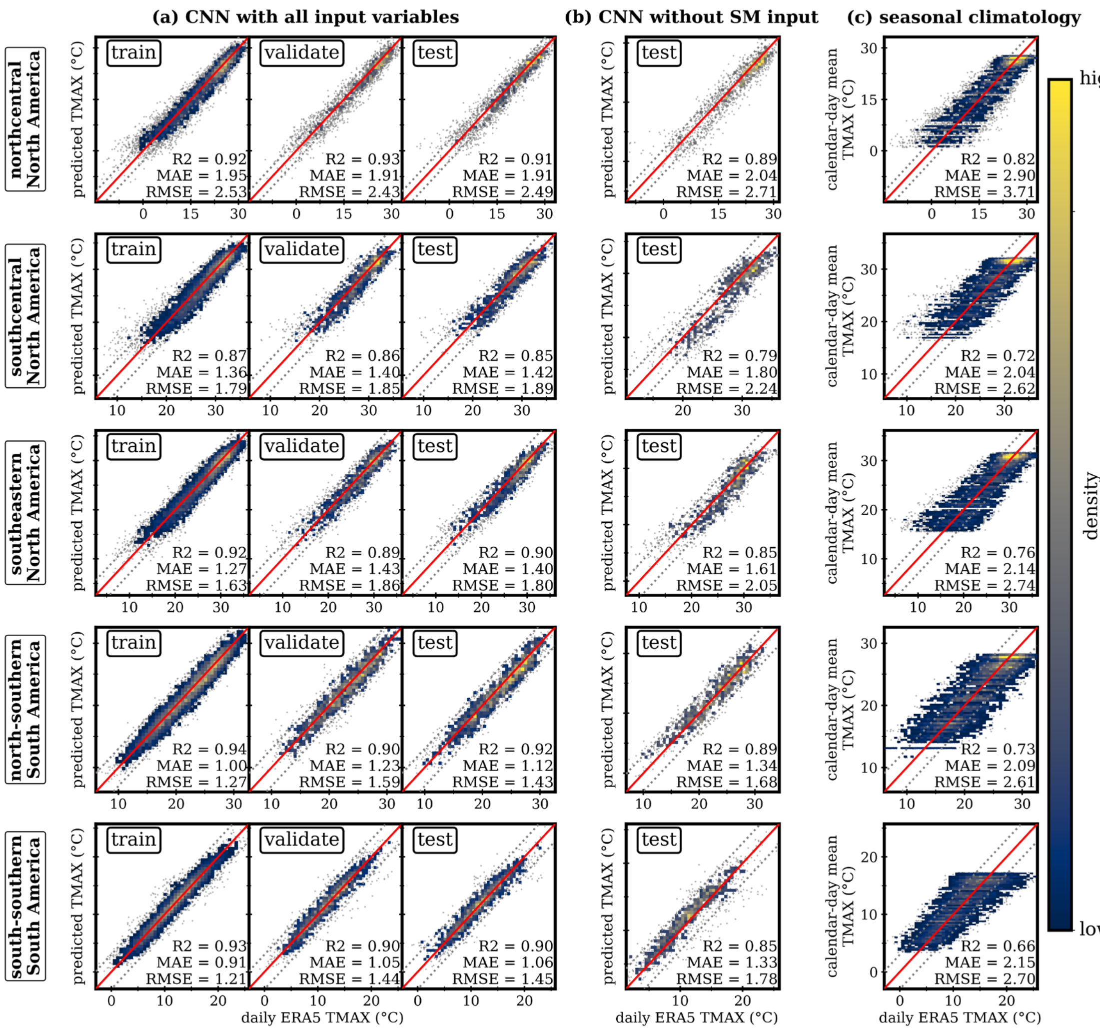

Figure 4. CNN model skill comparison for North and South American regions. (a) Comparison between ERA5-Land TMAX and predicted TMAX from convolutional neural networks (CNNs) trained with daily geopotential height anomaly maps, soil moisture anomaly maps (SM), and normalized calendar day inputs. Model performance is shown separately for the 27-year training subset (used to fit CNN weights), the 8-year validation subset (used to optimize hyperparameters), and the 8-year testing subset (unseen data left out of the training process). See Methods for more details on the training, validation, and testing subsets. (b) Same as (a) but for CNNs trained without the SM inputs. Model performance is shown for the 8-year testing subset. (c) The seasonal climatology of TMAX as shown by comparing the ERA5-Land daily TMAX and the calendar-day mean TMAX each day (averaged over 1979-2021). Each subplot shows the coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE). Correct predictions fall along the 1-1 line (red). Gray dotted lines show +/- 3 degrees C prediction errors.

We also find large differences in model performance between regions (Figures 4-6). These differences are most obvious between seasonal-climatology baselines, where RMSE ranges from $2.40^oC – 3.86^oC$ (southeastern Asia-northeastern Europe) and R^2 ranges from 0.34-0.88 (southeastern Africa-northeastern Asia; Figures 4-6(a)). These regional differences in model performance can be explained by the statistics of the underlying TMAX target data. In general,

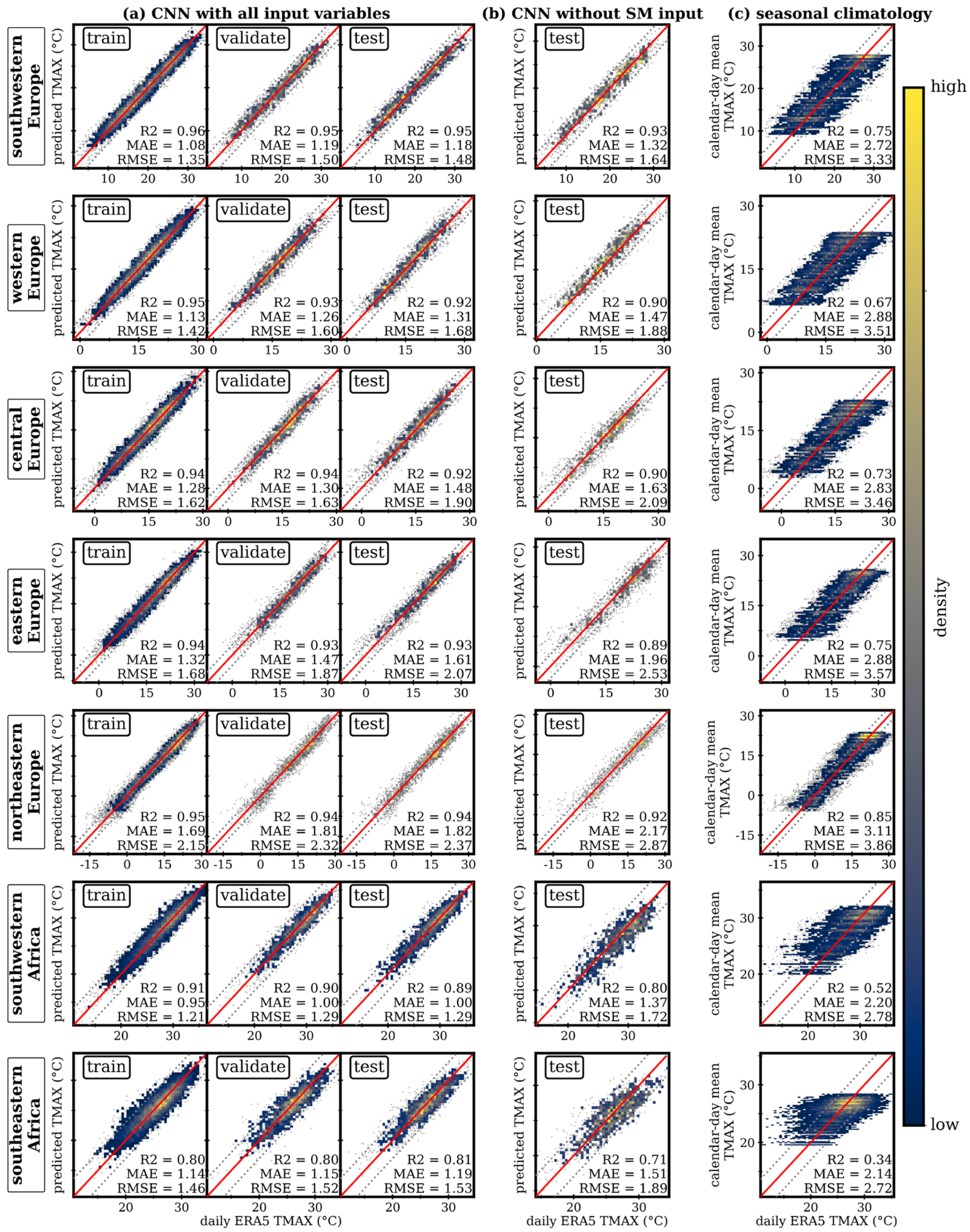

Figure 5. Same as Figure 4, but for regions in Europe and Africa.

the skill metrics (R^2, MAE, and RMSE) of the seasonal-climatology baseline are determined by the magnitude of the region's seasonal cycle and the standard deviation of the daily anomalies about the seasonal cycle. For example, regions with strong TMAX seasonal cycles (northcentral North America, northeastern Europe, and northeastern Asia; Figure 3) exhibit higher R^2 values relative to regions with weak TMAX seasonal cycles (southeastern Africa and southwestern Africa). Meanwhile, regions with low TMAX standard deviations about the seasonal cycle (southeastern Asia, north-southern South America, southeastern Africa, and southcentral North

America; Figure 3) tend to have lower RMSE than regions with high TMAX standard deviations about the seasonal cycle (northeastern Europe, northcentral North America, central Europe, and eastern Europe).

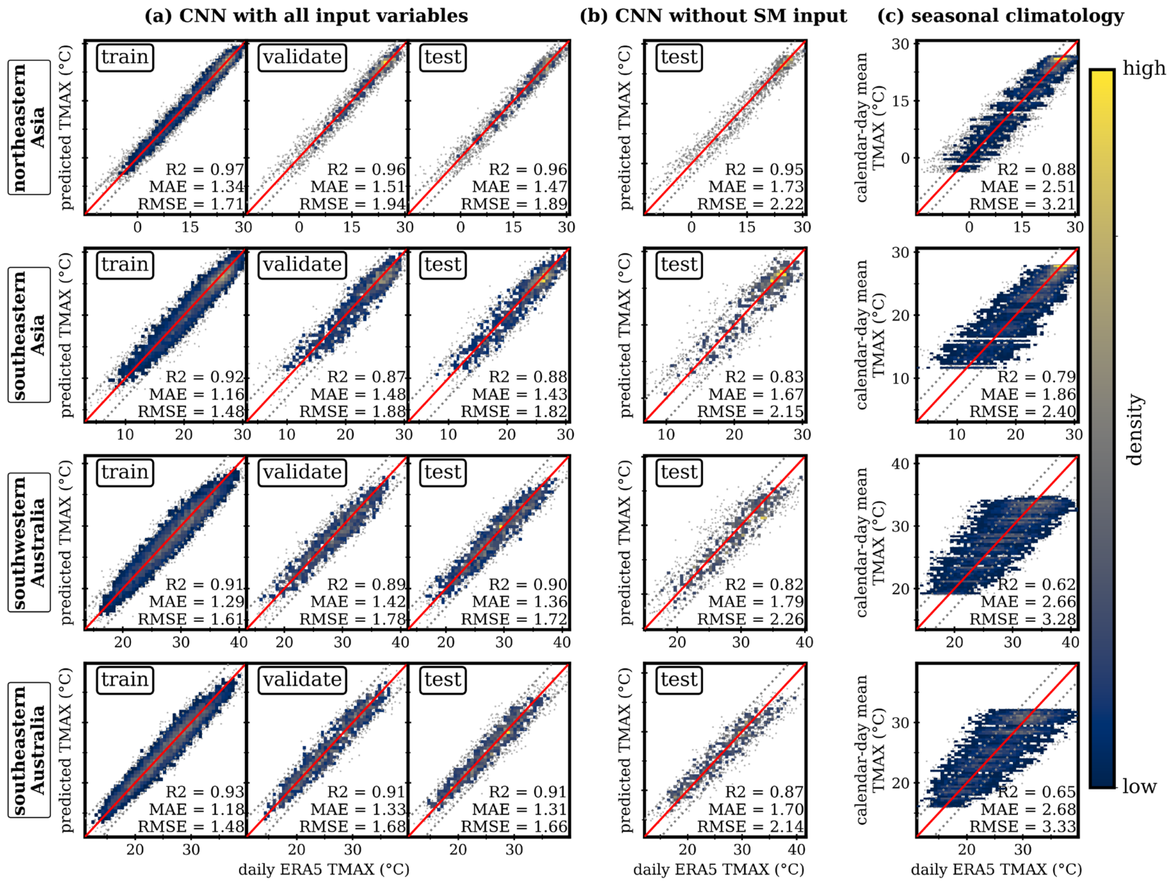

Figure 6. Same as Figure 4, but for regions in Eastern Asia and Australia.

3.2 Using Partial Dependence to Investigate SM-T Coupling

After evaluating the performance of each regional CNN (Figure 4-6), we apply the partial dependence analysis method (Figure 2) to obtain the ERA5 summertime SM-T relationships for each region (Figure 7). The resulting nonlinear SM-T partial dependence plots (PDPs) quantify how the CNN's average TMAX prediction depends on the average SM input, with areas of nonzero PDP slope indicating that the prediction is sensitive to the local SM anomaly calculated from the SM input map. Across all sixteen regions, we find that the SM-T PDPs are negatively sloped across the entire SM domain (aside from a positive slope in northcentral North America for wet SM anomalies; Figure 7). This pattern indicates that the CNNs tend to predict higher TMAX values when SM conditions are drier, and lower TMAX values when SM conditions are wetter.

Despite these overall similarities in the PDP shapes, there are also distinct regional differences in the ERA5 SM-T relationship (Figure 7). For some regions (e.g., eastern Europe, southeastern North America), we find that the slope of the SM-T relationship is relatively constant across the entire range of SM anomalies. Other regions exhibit nonlinear SM-T relationships indicating that the CNN has learned a different relationship between the SM input map and the TMAX output under different magnitudes of SM anomaly. In many regions, this nonlinear behavior is observed over a large portion of the SM range (e.g., southwestern Australia, northcentral North America), while other regions experience nonlinear SM-T behavior only during the most extreme SM conditions (e.g., the relatively flat PDP slope in southeastern Australia during extreme wet conditions). To assess the uncertainty associated with each regional SM-T relationship, we visualize the distribution of local SM anomalies in the training dataset to identify particular ranges of SM conditions where SM-T relationships may have higher uncertainty due to underrepresentation in the training dataset (Figure 7). Additionally, we find that the 5th-95th percentile ranges are narrowest near the origin (SM=0) and become wider near the tails of the SM distribution indicating that the SM-T relationships are more uncertain during extreme SM conditions where there are fewer testing samples available.

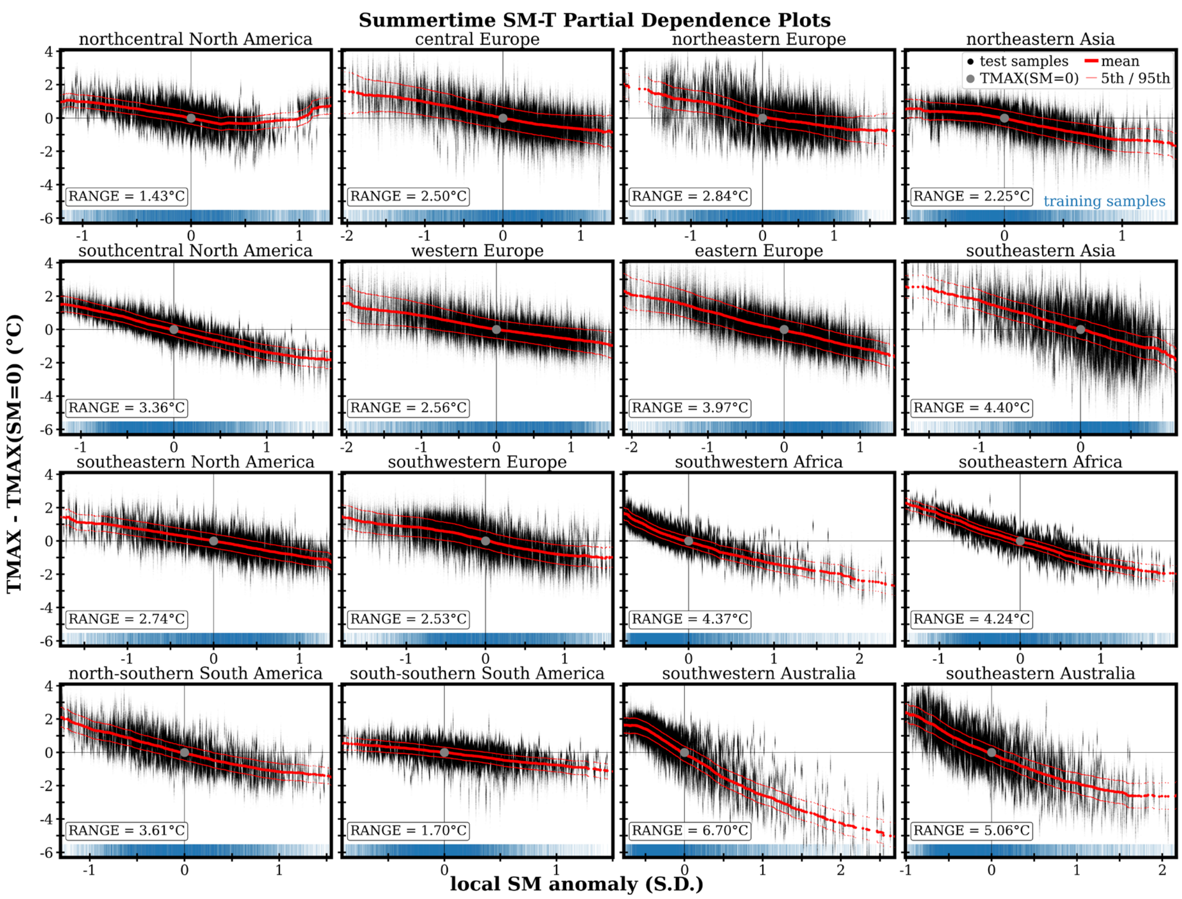

Figure 7. Soil moisture-temperature (SM-T) relationships obtained through partial dependence analysis of convolutional neural networks (method detailed in Figure 2). The smoothed moving average (thick red line) shows the average behavior of the neural network's prediction as the SM input varies from dry (negative) to wet (positive) anomalies. Also shown are the moving 5th and 95th percentiles of the temperature predictions (thin red lines). The SM-T relationships shown are calculated from the testing dataset. We also include a rug plot showing the distribution of SM anomalies in the training dataset. For each subplot, we calculate the range (vertical extent) of the mean SM-T relationship. Soil moisture anomalies are calculated as standard deviations (S.D.) from the calendar-day mean.

The vertical extent (range) of these SM-T relationships can be used as a relative measure of regional SM-T coupling strength by estimating the overall potential for SM to influence the

CNN's TMAX prediction on a typical summer day. In North America, we find that southcentral North America has an SM-T coupling strength of approximately 3.4°C, much higher than both northcentral North America (1.4°C) and southeastern North America (2.7°C). In Europe, we find the strongest coupling in eastern Europe (4.0°C) and northeastern Europe (2.8°C), and weaker coupling in central Europe (2.5°C) and western Europe (2.6°C), and southwestern Europe (2.5°C). Additionally, we find that southeastern Asia (4.4°C) has stronger coupling than northeastern Asia (2.3°C), and north-southern South America (3.6°C) has stronger coupling than south-southern South America (1.7°C), whereas southeastern Africa (4.2°C) and southwestern Africa (4.4°C) have approximately equal coupling. Finally, southeastern Australia (5.1°C) and southwestern Australia (6.7°C) have the strongest overall coupling. (We also show sub-regional variations in SM-T coupling for southcentral North America; Figure S3.)

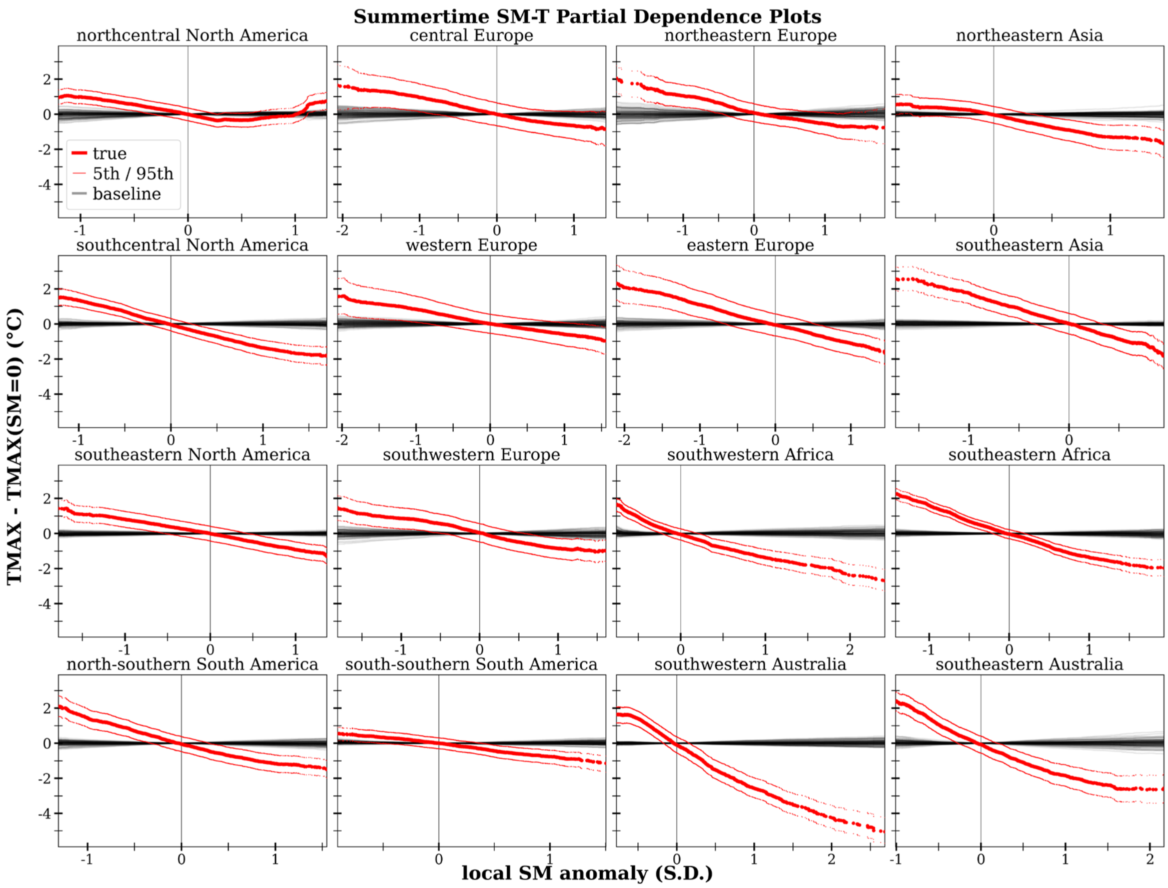

Figure 8. Regional soil moisture-temperature (SM-T) relationships obtained through partial dependence analysis (method detailed in Figure 2) of convolutional neural networks (CNNs) trained to predict regional daily maximum temperature (TMAX) given geopotential height, calendar-day, and soil moisture inputs. Each regional subplot shows 101 SM-T partial dependence plots (PDPs), consisting of the true SM-T PDP (red; Figure 7) and 100 baseline SM-T PDPs (black) derived from CNNs trained with shuffled soil moisture inputs (each shuffled using a different random seed). Also shown are the moving 5th and 95th percentiles of the true SM-T PDP (thin red lines). Soil moisture anomalies are calculated as standard deviations (S.D.) from the calendar-day mean.

To determine whether each PDP exhibits an SM-T relationship beyond that of random noise, we compare the true ERA5 SM-T PDPs (Figure 7) against 100 baseline PDPs calculated from 100 different CNNs trained with randomly shuffled SM input maps–each with a different random seed (Figure 8). (For illustration, we show a separate example of one of these baseline PDPs along with a density plot of TMAX predictions in Figure S4.) Across the regions, all 100 baseline SM-T PDPs have approximately zero slope over the entire SM domain, with no single baseline PDP exhibiting a coupling strength greater than 1.2°C (northeastern Europe). We also find that the vast majority of points along the true regional SM-T PDPs lie far outside the range of the baselines. Wet SM anomalies (0.5-1.0 standard deviations) in northcentral North America are the only notable exceptions for which a substantial portion of the regional PDP falls within the range of the baselines (Figure 8).

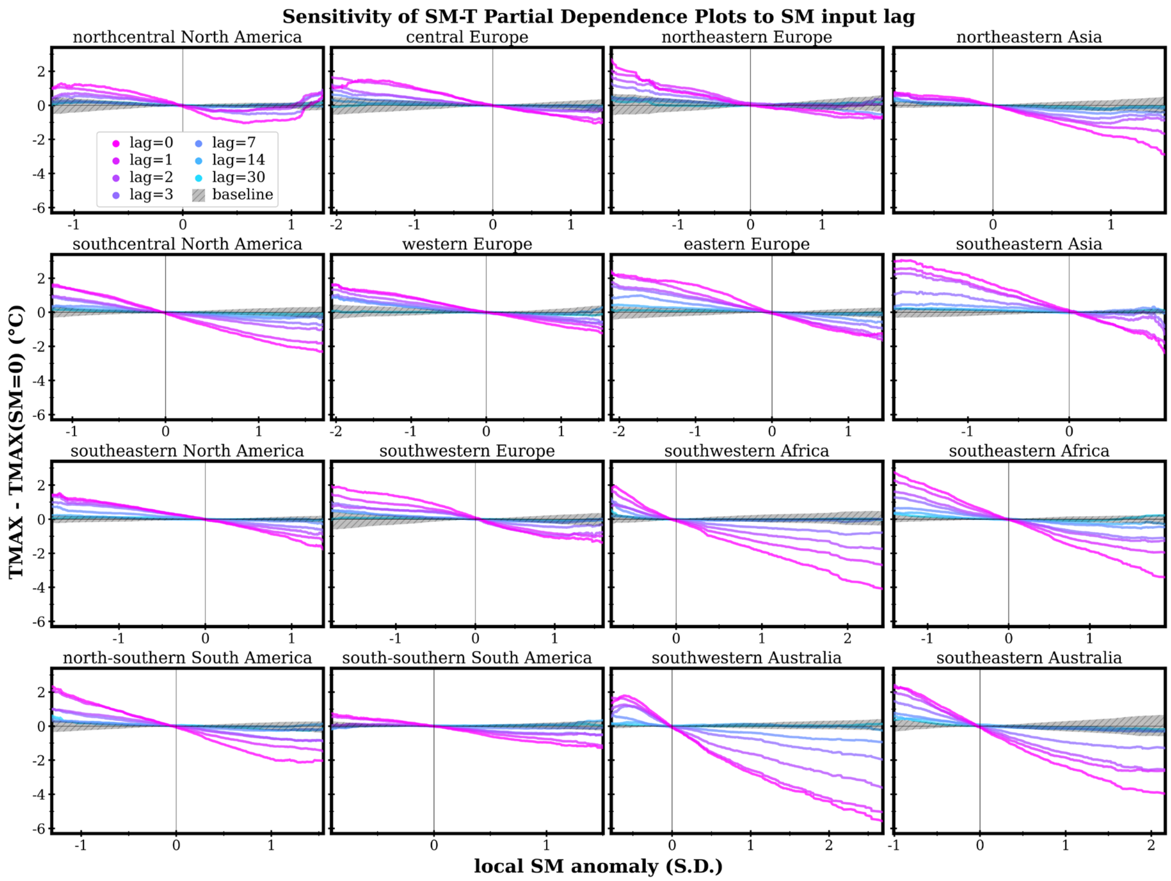

Figure 9. Regional soil moisture-temperature (SM-T) partial dependence relationships obtained using the method detailed in Figure 2 (but for CNNs trained with various levels of soil moisture input lag). Each regional subplot shows SM-T relationships derived from 7 different CNNs trained to predict daily TMAX given the following inputs: calendar day, daily GPH anomaly map, and a single day's SM anomaly map lagged by 0-30 days prior to the prediction day. After the training process, CNN weights are saved and used to calculate the SM-T PDPs as in Figure 2. Colors show SM-T relationships for CNNs trained with SM input lags of 0, 1, 2, 3, 7, 14, and 30 days. Hatching shows the range of 100 baseline PDPs trained with shuffled SM maps (Figure 8). Soil moisture anomalies are calculated as standard deviations (S.D.) from the calendar-day mean.

We also analyze the sensitivity of regional SM-T relationships to the choice of SM input lag (Figure 9). Specifically, we show SM-T PDPs derived from seven different CNNs each trained with different levels of SM input lag. (For example, lag = 3 implies that the CNN is trained to predict TMAX using the calendar day and GPH input from the prediction day, and the SM anomaly map from 3 days prior to the prediction day.) In general, although the PDP shape is similar across input lags, almost all regions experience a monotonic attenuation of SM-T coupling strength (amplitude) as SM input lag increases from 0 to 30 days. This attenuation is expected, based on the autocorrelation timescales of top-layer SM. However, the rate of attenuation varies between the regions. For example, over many regions (south-southern South America, northcentral North America, northeastern Europe), this attenuation is quite strong and

SM-T relationships fall within the range of baseline PDPs for SM lags greater than 3 days. For other regions, this attenuation is much weaker, and we find SM-T coupling relationships that fall outside the range of baseline PDPs at 7-day SM lags (central Europe, northeastern Asia, southwestern Africa), and 14-day SM lags (southcentral North America, southeastern North America, eastern Europe, western Europe, southwestern Europe, southeastern Asia, southeastern Africa, north-southern South America, southwestern Australia, southeastern Australia). Indeed, for extremely dry SM anomalies, some regions exhibit SM-T relationships beyond random noise for SM lags up to 30 days (southwestern Australia, southeastern Australia, southeastern Africa**,** north-southern South America).

We repeat our analysis for all sixteen regions using the NCEP/DOE Reanalysis II dataset over the same time period (1979-2021) at the same 1.875°×1.875° horizontal resolution (Figures S5-S11). Despite some notable differences in northcentral North America, the resulting NCEP SM-T relationships are consistent with the ERA5 analysis for regional PDP shape (Figure 7 vs. S9), SM-T coupling strengths, comparison with baseline PDPs (Figure 8 vs. S10), and the attenuation of coupling strength with input lag (Figure 9 vs. S11).

4 Discussion

We use CNNs (Figure 1) to predict daily average TMAX over 16 mid-latitude regions, and apply partial dependence analysis (Friedman, 2001; Figure 2) to investigate regional SM-T coupling relationships using the ERA5 and NCEP reanalysis datasets. Prior to conducting the partial dependence analysis, we first determine whether the CNN is sufficiently accurate to represent SM-T coupling at daily timescales. This is especially important since CNN model skill metrics vary widely between regions (Figures 4-6). As described in the Methods, our two criteria are that the CNN predicts daily TMAX anomalies from the seasonal cycle, and that the SM input contributes to overall CNN performance at daily timescales. After careful model evaluation, we find that all regional CNNs satisfy these criteria (Figures 4-6).

We also find that overall model performance is closely tied to the statistics of the underlying TMAX target data. For instance, a simple model which predicts the calendar-day mean TMAX each day has high R^2 and low MSE when asked to predict over a region characterized by a strong TMAX seasonal cycle with low variance about the seasonal cycle (e.g., southeastern Asia seasonal-climatology baseline model; Figure 6). Despite good performance metrics, this same model is not suitable for partial dependence analysis of daily-scale SM-T coupling because it fails to predict daily TMAX anomalies from the seasonal cycle. As a result, we stress the importance of thoroughly evaluating the CNN model skill (as suggested in Section 2.3) to assess performance at various timescales (Figures S1 and S2). Furthermore, we suggest the use of multiple CNNs with different input combinations to verify that each input variable contributes to overall model performance at the desired timescale (Figures 4-6). The results of these verification tests provide confidence in using the regional CNNs to quantify daily-scale SM-T coupling using partial dependence analysis (Figure 2).

Our SM-T PDPs show that the CNN TMAX predictions are sensitive to the local SM anomaly over the prediction region (Figure 7 and S9). Additionally, the SM-T PDPs are negatively sloped and roughly monotonic (aside from wet SM anomalies in northcentral North America), with each CNN predicting warmer temperatures associated with dry SM anomalies and cooler temperatures associated with wet SM anomalies. The general shapes of these SM-T

PDPs (Figure 7 and S9) are consistent with the well-understood land-atmosphere interactions through which SM conditions modulate the local surface energy budget and influence near-surface temperatures (Alfaro et al., 2006; Dirmeyer, 2011; J. Liu & Pu, 2019; Seneviratne et al., 2010). Previous studies rely on linear statistical methods (such as the correlation between evapotranspiration and temperature) to assess regional differences in land-atmosphere coupling strength (Dirmeyer, 2011; Jaeger et al., 2009; Koster et al., 2004, 2006a, 2009; Miralles et al., 2012; Seneviratne, Lüthi, et al., 2006; Teuling et al., 2009). While these linear methods are well-suited for quantifying regional differences in coupling strength, evidence from climate models and observations suggest that the actual influence of SM on temperature is nonlinear (Benson & Dirmeyer, 2021; Fischer et al., 2007; Jaeger & Seneviratne, 2011; Schwingshackl et al., 2017; Seneviratne et al., 2010).

To allow for the potential of these nonlinear SM-T relationships, our method uses CNN machine learning models to quantify the sensitivity of TMAX to SM across a range of different SM values. We find that the SM-T relationships derived from partial dependence analysis (Figure 2) are approximately linear for some regions (e.g., eastern Europe, southeastern North America) and nonlinear for other regions (e.g., southwestern Australia, northcentral North America, southeastern Australia) (Figure 7). These results suggest that the land-atmosphere interactions that couple daily SM conditions and near-surface TMAX vary under different ranges of SM anomaly, but these variations are regionally dependent. When evaluating these SM-T relationships, it is important to consider that the PDP behavior is most uncertain at the tails of the SM distribution where the 5th-95th percentile ranges are widest and where the underrepresentation of extreme SM anomalies in the training dataset limits the CNNs ability to learn the relationship between SM and TMAX.

In order to compare our results more directly with previous assessments of SM-T coupling, we use the vertical extent (range) of our SM-T PDPs (Figure 7) as a relative indicator of SM-T coupling strength. Using this metric, we find much stronger JJA SM-T coupling in southcentral North America compared to northcentral North America and southeastern North America. This agrees with previous assessments of land-atmosphere coupling strength using climate models (Koster et al., 2006a, 2009; Seneviratne, Lüthi, et al., 2006) and observational datasets (Dirmeyer, 2011; Miralles et al., 2012). These results are also consistent with Teuling et al., (2009) and Schwingshackl et al., (2017), who used observational and reanalysis datasets, respectively, to classify southcentral North America and northcentral North America as regions with a high potential for strong SM-T coupling and southeastern North America as a region with little potential for SM-T coupling.

In Europe, we find that our northeastern Europe and eastern Europe regions have the strongest PDP-based SM-T coupling strength, while our central Europe, western Europe, and southwestern Europe regions have the weakest (Figure 7). This hierarchy of coupling strength in Europe is consistent with Fischer et al., (2007), whose regional climate model experiments identified the strongest 2003 JJA SM-T coupling in eastern Europe (followed by central Europe and western Europe), and the weakest coupling in southwestern Europe (with northeastern Europe not considered in their domain). Seneviratne, Lüthi, et al., (2006) and Jaeger et al., (2009) also found strong JJA SM-T coupling in eastern Europe and northeastern Europe, with weaker coupling in western Europe and central Europe. Our results are also consistent with Teuling et al., (2009) who found the potential for strong SM-T coupling in eastern Europe and northeastern Europe. However, numerous previous studies (Dirmeyer, 2011; Jaeger et al., 2009;

Miralles et al., 2012; Seneviratne, Lüthi, et al., 2006; Teuling et al., 2009) all identified strong SM-T coupling over southwestern Europe, in contrast to our PDP-based results (although Seneviratne, Lüthi, et al., 2006, warn that certain coupling metrics, like the correlation of evapotranspiration and 2-meter temperature, may not be meaningful in regions with small evapotranspiration like southwestern Europe).

In the Southern Hemisphere, our PDP-based SM-T coupling strengths show weak coupling in south-southern South America, and strong coupling in north-southern South America, southwestern Africa, southeastern Africa, southwestern Australia, and southeastern Australia (Figure 7). These PDP-based coupling strengths are remarkably consistent with Dirmeyer, (2011), who analyzed coupling between latent heat flux and SM to identify regions with strong SM-T coupling potential. Our results are also consistent with Schwingshackl et al., (2017), who identified south-southern South America as a wet SM regime during DJF, and all other regions (north-southern South America, southwestern Africa, southeastern Africa, southwestern Australia, and southeastern Australia) as transitional SM regimes. In South America, Miralles et al., (2012) found approximately equal coupling across central and south-southern South America, although numerous other studies (e.g., Baker et al., 2021; Dirmeyer, 2011; Menendez et al., 2019; Ruscica et al., 2014; Spennemann et al., 2018) report that land-atmosphere coupling is much stronger in north-southern South America compared to south-southern South America. The results of Miralles et al., (2012) also support our conclusion that SM-T coupling is much stronger in Africa and Australia compared to South America.

The most notable differences between our results and previous assessments of regional SM-T coupling strengths occur in eastern Asia. Using both the ERA5 and NCEP datasets, we find substantially stronger PDP-based SM-T coupling in our southeastern Asia region compared to our northeastern Asia region (Figures 7 and S9, respectively). Previous studies report roughly equal (Koster et al., 2006a) or substantially stronger coupling in northeastern Asia (Dirmeyer, 2011; Koster et al., 2009; Miralles et al., 2012; Schwingshackl et al., 2017; Seneviratne, Lüthi, et al., 2006; Teuling et al., 2009), which conflicts with our ERA5 and NCEP results.

We extend our partial dependence analysis to modified versions of our training dataset, which yields additional insights into the timescale of SM memory within the SM-T relationship. We find a monotonic attenuation of PDP-based coupling strength with increasing SM input lag (Figure 9). The overall reduction in SM-T coupling strength is likely a consequence of limited SM memory as the SM input becomes less physically relevant to actual conditions on the prediction day. Our results also agree with previous studies which suggest that wet SM anomalies decay faster than dry SM anomalies (Orth & Seneviratne, 2012; Song et al., 2019), resulting in longer SM memory for extreme dry conditions (Orth & Seneviratne, 2012). Specifically, we find that in 12 of the 16 regions, the SM-T relationship remains outside the range of random noise at longer lags for dry anomalies than for wet anomalies (Figure 9). In addition, we find regional differences in the timescale of decay in PDP-based coupling strength as SM input lag increases (Figure 9). For example, southeastern Africa (among other regions) exhibits an SM-T relationship beyond random noise at lags up to 14 days. However, SM-T relationships in south-southern South America, northcentral North America, and northeastern Europe fall within the range of random noise beyond 3-day SM lags. These regional differences in PDP attenuation agree reasonably well with Seneviratne, Koster, et al., (2006), who found long SM memories across southern Africa, Australia, Europe, North America, and north-southern South America, but substantially shorter SM memory in northeastern Asia and south-

southern South America. Seneviratne, Koster, et al., (2006) also found long SM memory in southeastern Asia which conflicts with our ERA5 and NCEP results. Overall, these results suggest that incorporating additional temporal SM information from 7-, 14-, or even 30-days prior to the TMAX prediction could improve the CNN's ability to predict TMAX.

Our analysis focuses specifically on SM-T coupling over midlatitude regions; however, the physical processes that regulate SM-T interactions may be different in tropical and high-latitude regions. Therefore, though the flexibility of our machine learning-based framework makes it deployable to other regions, we do not claim that our results can be applied to other areas of the globe (such as in the tropics or high latitudes) without further investigation. We also acknowledge that there may exist different configurations of machine learning model (e.g., long short-term memory network), hyperparameters, and input variables that are able to achieve better performance than the CNNs used in this study. Regardless, our results show that these CNNs capture SM-T relationships that broadly agree with previous assessments of SM-T coupling. We also recognize that our regional assessment of SM-T coupling fails to capture fine-scale spatial differences in coupling found in previous studies (e.g., Koster et al., 2006b; Miralles et al., 2012). However, our framework could be readily extended to assess coupling at finer spatial resolutions by calculating SM-T relationships over smaller subregions (Figure S3) and/or using input data with finer spatial resolution. Though we focus specifically on the relationship between surface-layer SM and TMAX (which is most relevant for daily-scale SM-T coupling), our analysis could also be modified to assess coupling between numerous other land-surface and atmospheric variables (e.g., coupling between latent heat flux and daily mean temperature, coupling between evapotranspiration and precipitation).

Although our PDPs quantify the average impact of local SM conditions on the CNN's TMAX prediction, there may be other processes correlated with SM conditions whose effect on temperature is incorrectly attributed to SM. One way to address this would be to repeat this analysis using a different land-surface variable in place of SM (e.g., latent heat flux or evapotranspiration) and compare the corresponding coupling relationships with temperature. Another approach would be to include additional atmospheric and land-surface variables as CNN inputs and hold them constant during the PDP calculation to isolate the effect of SM alone on temperature. However, adding additional variables would run the risk of violating the independence assumption between input variables. Indeed, although we use standardized calendar-day anomalies for SM and GPH inputs to avoid seasonal dependencies with the calendar-day inputs, a side effect is that our PDPs are calculated in terms of standardized SM anomalies instead of the raw SM fraction values. Since each SM grid cell's calendar-day mean and standard deviation fluctuates throughout the summer, we cannot convert SM anomalies directly back to SM fraction values, which prevents us from being able to compare the magnitude of the PDP slope directly between regions.

Finally, like all SM-T coupling assessments, our results are also dataset-dependent. Although it represents an improvement over the land component of previous reanalyses, the ERA5-Land surface-layer soil moisture dataset used in this analysis has a known wet-bias and exhibits regional differences in agreement (i.e., correlation) when compared to 5-cm in situ observations of SM across Europe, North America, and Australia (Muñoz-Sabater et al. 2021). As a result, the SM-T relationships presented here may be more representative of the real world in regions where the ERA5-Land SM closely matches observations, and less representative in regions where the ERA5-Land SM has higher uncertainty. Regardless, while the results presented here are limited to the datasets that were analyzed, our framework could easily be extended to quantify SM-T relationships using a wide range of datasets from climate models, reanalyses, remote sensing, and/or gridded observations.

5 Conclusions

We present a new approach for quantifying soil moisture-temperature (SM-T) coupling which uses convolutional neural network (CNN) machine learning models and partial dependence plots (PDPs) to visualize nonlinear SM-T relationships over 16 mid-latitude regions in the Northern and Southern Hemispheres. From these regional SM-T relationships, we find that the CNNs predict warmer temperatures when the soils are dry and cooler temperatures when the soils are wet, which is consistent with well-understood land-atmosphere interactions in the mid-latitudes. We also find that our relative measure of SM-T coupling strength broadly agrees with previous assessments of regional SM-T coupling. Though our approach is designed to allow for the potential of nonlinear SM-T relationships, we find that the SM-T PDPs are approximately linear over several regions, such as eastern Europe and southeastern North America. That said, other regions exhibit pronounced nonlinear behavior across a large portion of the SM range (e.g., southwestern Australia, northcentral North America). This nonlinearity suggests that the coupled interactions governing the SM-T relationship vary under different SM conditions, but these variations are regionally dependent. Taken together, our results show that PDPs can be combined with CNNs to create a powerful tool for quantifying nonlinear SM-T coupling relationships.

In particular, we find that applying machine-learning interpretation and visualization techniques (i.e., PDPs) to modified versions of our training datasets can yield new insights into physical processes, such as the nonlinear characteristics of SM memory, which is a vital component of long-term SM-T coupling. For example, in accordance with previous studies, we find that SM memory fades monotonically over time, and that wet SM anomalies fade faster than dry anomalies. More research is required to understand the full potential for PDPs to reveal regional differences in the nonlinear properties of SM memory, with implications for seasonal forecasting of temperature and precipitation.

Partial dependence analysis has only recently been applied to CNNs for geoscience applications. However, we suggest that many complex climate processes have the potential to be studied by analyzing CNNs with PDPs as long as enough high-quality training data are available. For example, given sufficient training data, our analysis could be extended to investigate climate-driven changes in SM-temperature and SM-precipitation coupling at daily and seasonal timescales using climate model simulations under historic and future climate change scenarios. Likewise, PDPs with CNNs could be used to explore non-local coupling relationships between land, ocean, and atmospheric conditions which can improve our understanding of complex climate processes such as the El Nino Southern Oscillation. More generally, our results show that PDPs can be an effective tool for quantifying nonlinear coupling relationships between the CNN’s output prediction and quantities calculated from the input maps. We emphasize that, for each of these potential applications, even if the training data appears to be adequate, each CNN model must be thoroughly evaluated to ensure that the model is trustworthy and is representative of physical processes in the real world.

Coupled interactions in the Earth system are important drivers of climate variability and extreme weather events, but many of these coupled processes are still not fully understood.

Based on our results, partial dependence analysis is a promising pathway for using CNNs to investigate these nonlinear coupled interactions, with important implications for model development, model parameterization, and seasonal forecasting.

Open Research

The hourly ERA5 (Hersbach, H., et al., 2018) and ERA5-Land (Muñoz Sabater, J., 2021) data are available from the Copernicus Climate Change Service Climate Data Store and can be accessed from their website at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels and https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land, respectively. The daily mean NCEP/DOE Reanalysis II data (Kanamitsu et al. 2002) provided by the NOAA PSL, Boulder, Colorado, USA, is available from their website at https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html. Analysis code will be available on Zenodo via DOI (set at time of publication).

Acknowledgments

We thank three anonymous reviewers for insightful and constructive comments. Computational resources were provided by the Stanford Research Computing Center and Stanford's Center for Computational Earth and Environmental Sciences. This work was supported by Stanford University and NSF CAREER grant AGS-1749261.

References

Alfaro, E. J., Gershunov, A., & Cayan, D. (2006). Prediction of Summer Maximum and Minimum Temperature over the Central and Western United States: The Roles of Soil Moisture and Sea Surface Temperature. Journal of Climate, 19(8), 1407–1421. https://doi.org/10.1175/JCLI3665.1

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PloS One, 10(7), e0130140. https://doi.org/10.1371/journal.pone.0130140

Baker, Jessica C. A., Dayana Castilho de Souza, Paulo Y. Kubota, Wolfgang Buermann, Caio A. S. Coelho, Martin B. Andrews, Manuel Gloor, Luis Garcia-Carreras, Silvio N. Figueroa, and Dominick V. Spracklen. (2021). "An Assessment of Land–Atmosphere Interactions over South America Using Satellites, Reanalysis, and Two Global Climate Models." Journal of Hydrometeorology 22 (4): 905–22.

Barnes, E. A., Mayer, K. J., Rader, J., Toms, B. A., & Ebert-Uphoff, I. (2020). Leveraging Interpretable Neural Networks for Scientific Discovery. 2020, A069–03. https://ui.adsabs.harvard.edu/abs/2020AGUFMA069...03B

Barnes, E. A., Toms, B., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2020). Indicator Patterns of Forced Change Learned by an Artificial Neural Network. Journal of Advances in Modeling Earth Systems, 12(9), e2020MS002195. https://doi.org/10.1029/2020MS002195

Benson, D. O., & Dirmeyer, P. A. (2021). Characterizing the Relationship between Temperature and Soil Moisture Extremes and Their Role in the Exacerbation of Heat Waves over the Contiguous United States. Journal of Climate, 34(6), 2175–2187. https://doi.org/10.1175/JCLI-D-20-0440.1

Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. Journal of Advances in Modeling Earth Systems, 11(1), 376–399. https://doi.org/10.1029/2018ms001472

Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.-K., Swayne, D. F., & Wickham, H. (2009). Statistical inference for exploratory data analysis and model diagnostics. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences, 367(1906), 4361–4383. https://doi.org/10.1098/rsta.2009.0120

Cattiaux, J., Douville, H., & Peings, Y. (2013). European temperatures in CMIP5: origins of present-day biases and future uncertainties. Climate Dynamics, 41(11), 2889–2907. https://doi.org/10.1007/s00382-013-1731-y

Chen, Ajiao, Huade Guan, Okke Batelaan, Xinping Zhang, and Xinguang He. (2019). "Global Soil Moisture-air Temperature Coupling Based on GRACE-derived Terrestrial Water Storage." Journal of Geophysical Research 124 (14): 7786–96.

Chilson, C., Avery, K., McGovern, A., Bridge, E., Sheldon, D., & Kelly, J. (2019). Automated detection of bird roosts using NEXRAD radar data and Convolutional Neural Networks. Remote Sensing in Ecology and Conservation, 5(1), 20–32. https://doi.org/10.1002/rse2.92

Davenport, F. V., & Diffenbaugh, N. S. (2021). Using machine learning to analyze physical causes of climate change: A case study of U.s. midwest extreme precipitation. Geophysical Research Letters, 48(15). https://doi.org/10.1029/2021gl093787

Developers, T. (2021). TensorFlow [Software]. https://doi.org/10.5281/zenodo.5593257

Dennis, Eli J., and Ernesto Hugo Berbery. (2021). "The Role of Soil Texture in Local Land Surface–Atmosphere Coupling and Regional Climate." Journal of Hydrometeorology 22 (2): 313–30.

Diffenbaugh, N. S., & Ashfaq, M. (2010). Intensification of hot extremes in the United States. Geophysical Research Letters, 37(15). https://doi.org/10.1029/2010gl043888

Diffenbaugh, Noah S., and Elizabeth A. Barnes. (2023). "Data-Driven Predictions of the Time Remaining until Critical Global Warming Thresholds Are Reached." Proceedings of the National Academy of Sciences of the United States of America 120 (6): e2207183120.

Diffenbaugh, N. S., Pal, J. S., Giorgi, F., & Gao, X. (2007). Heat stress intensification in the Mediterranean climate change hotspot. Geophysical Research Letters, 34(11). https://doi.org/10.1029/2007gl030000

Dirmeyer, P. A. (2011). The terrestrial segment of soil moisture-climate coupling. Geophysical Research Letters, 38(16). https://doi.org/10.1029/2011gl048268

Durre, I., Wallace, J. M., & Lettenmaier, D. P. (2000). Dependence of Extreme Daily Maximum Temperatures on Antecedent Soil Moisture in the Contiguous United States during Summer. Journal of Climate, 13(14), 2641–2651. https://doi.org/10.1175/1520-0442(2000)013<2641:DOEDMT>2.0.CO;2

Dutra, E., Schär, C., Viterbo, P., & Miranda, P. M. A. (2011). Land-atmosphere coupling associated with snow cover. Geophysical Research Letters, 38(15). https://doi.org/10.1029/2011gl048435

Ebert-Uphoff, I., & Hilburn, K. (2020). Evaluation, Tuning, and Interpretation of Neural Networks for Working with Images in Meteorological Applications. Bulletin of the American Meteorological Society, 101(12), E2149–E2170. https://doi.org/10.1175/BAMS-D-20-0097.1

Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D., & Schär, C. (2007). Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave. Journal of Climate, 20(20), 5081–5099. https://doi.org/10.1175/JCLI4288.1

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189–1232. http://www.jstor.org/stable/2699986

Gagne, D. J., II, Haupt, S. E., Nychka, D. W., & Thompson, G. (2019). Interpretable Deep Learning for Spatial Analysis of Severe Hailstorms. Monthly Weather Review, 147(8), 2827–2845. https://doi.org/10.1175/MWR-D-18-0316.1

Gevaert, A. I., D. G. Miralles, R. A. M. de Jeu, J. Schellekens, and A. J. Dolman. (2018). "Soil Moisture-Temperature Coupling in a Set of Land Surface Models." Journal of Geophysical Research 123 (3): 1481–98.

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. Journal of Computational and Graphical Statistics: A Joint Publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America, 24(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. Nature, 573(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7

Han, Y., Zhang, G. J., Huang, X., & Wang, Y. (2020). A moist physics parameterization based on deep learning. Journal of Advances in Modeling Earth Systems, 12(9). https://doi.org/10.1029/2020ms002076

Henderson, G. R., Peings, Y., Furtado, J. C., & Kushner, P. J. (2018). Snow–atmosphere coupling in the Northern Hemisphere. Nature Climate Change, 8(11), 954–963. https://doi.org/10.1038/s41558-018-0295-6

Hersbach, Bell, & Berrisford. (2018). ERA5 hourly data on pressure levels from 1979 to present. (c3s) Climate Data …. [Dataset]

He, Zhang, Ren, & Sun. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE. http://openaccess.thecvf.com/content_iccv_2015/html/He_Delving_Deep_into_ICCV_2015_paper.html

Hinton, Srivastava, & Swersky. (2012). Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. Cited on. http://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf

Hirsch, A. L., A. J. Pitman, and J. Kala. (2014). "The Role of Land Cover Change in Modulating the Soil Moisture-Temperature Land-Atmosphere Coupling Strength over Australia." Geophysical Research Letters 41 (16): 5883–90.

Horton, D. E., Johnson, N. C., Singh, D., Swain, D. L., Rajaratnam, B., & Diffenbaugh, N. S. (2015). Contribution of changes in atmospheric circulation patterns to extreme temperature trends. Nature, 522(7557), 465–469. https://doi.org/10.1038/nature14550

Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P., & Bouchet, F. (2021). Deep Learning-based Extreme Heatwave Forecast. In arXiv [cs.LG]. arXiv. http://arxiv.org/abs/2103.09743

Jaeger, E. B., & Seneviratne, S. I. (2011). Impact of soil moisture–atmosphere coupling on European climate extremes and trends in a regional climate model. Climate Dynamics, 36(9), 1919–1939. https://doi.org/10.1007/s00382-010-0780-8

Jaeger, E. B., Stöckli, R., & Seneviratne, S. I. (2009). Analysis of planetary boundary layer fluxes and land-atmosphere coupling in the regional climate model CLM. Journal of Geophysical Research, 114(D17). https://doi.org/10.1029/2008jd011658

Jergensen, G. E., McGovern, A., Lagerquist, R., & Smith, T. (2019). Classifying convective storms using machine learning. Weather and Forecasting, 35(2), 537–559. https://doi.org/10.1175/waf-d-19-0170.1

Kanamitsu, M., Ebisuzaki, W., & Woollen, J. (2002). Ncep–doe amip-ii reanalysis (r-2). Bulletin of the American Meteorological Society. [Dataset]. https://journals.ametsoc.org/view/journals/bams/83/11/bams-83-11-1631.xml

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., … GLACE Team. (2004). Regions of strong coupling between soil moisture and precipitation. Science, 305(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Koster, R. D., Schubert, S. D., & Suarez, M. J. (2009). Analyzing the Concurrence of Meteorological Droughts and Warm Periods, with Implications for the Determination of Evaporative Regime. Journal of Climate, 22(12), 3331–3341. https://doi.org/10.1175/2008JCLI2718.1

Koster, R. D., Sud, Y. C., Guo, Z., Dirmeyer, P. A., Bonan, G., Oleson, K. W., Chan, E., Verseghy, D., Cox, P., Davies, H., Kowalczyk, E., Gordon, C. T., Kanae, S., Lawrence, D., Liu, P., Mocko, D., Lu, C.-H., Mitchell, K., Malyshev, S., … Xue, Y. (2006a). GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. Journal of Hydrometeorology, 7(4), 590–610. https://doi.org/10.1175/JHM510.1

Koster, R. D., Sud, Y. C., Guo, Z., Dirmeyer, P. A., Bonan, G., Oleson, K. W., Chan, E., Verseghy, D., Cox, P., Davies, H., Kowalczyk, E., Gordon, C. T., Kanae, S., Lawrence, D., Liu, P., Mocko, D., Lu, C.-H., Mitchell, K., Malyshev, S., … Xue, Y. (2006b). GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. Journal of Hydrometeorology, 7(4), 590–610. https://doi.org/10.1175/JHM510.1

Lagerquist, R., McGovern, A., & Gagne, D. J., II. (2019). Deep Learning for Spatially Explicit Prediction of Synoptic-Scale Fronts. Weather and Forecasting, 34(4), 1137–1160. https://doi.org/10.1175/WAF-D-18-0183.1

Larraondo, P. R., Renzullo, L. J., Inza, I., & Lozano, J. A. (2019). A data-driven approach to precipitation parameterizations using convolutional encoder-decoder neural networks. In arXiv [physics.ao-ph]. arXiv. http://arxiv.org/abs/1903.10274

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. Advances in Neural Information Processing Systems, 2.

https://proceedings.neurips.cc/paper/1989/hash/53c3bce66e43be4f209556518c2fcb54-Abstract.html

Liu, J., & Pu, Z. (2019). Does soil moisture have an influence on near-surface temperature? Journal of Geophysical Research, 124(12), 6444–6466. https://doi.org/10.1029/2018jd029750

Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., & Collins, W. (2016). Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. In arXiv [cs.CV]. arXiv. http://arxiv.org/abs/1605.01156

Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2022). Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. Environmental Data Science, 1, e8. https://doi.org/10.1017/eds.2022.7

McGovern, A., Lagerquist, R., Gagne, D. J., Eli Jergensen, G., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. Bulletin of the American Meteorological Society, 100(11), 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1

Mei, R., & Wang, G. (2012). Summer Land–Atmosphere Coupling Strength in the United States: Comparison among Observations, Reanalysis Data, and Numerical Models. Journal of Hydrometeorology, 13(3), 1010–1022. https://doi.org/10.1175/JHM-D-11-075.1

Menéndez, Claudio G., Julián Giles, Romina Ruscica, Pablo Zaninelli, Tanea Coronato, Magdalena Falco, Anna Sörensson, Lluís Fita, Andrea Carril, and Laurent Li. (2019). "Temperature Variability and Soil–atmosphere Interaction in South America Simulated by Two Regional Climate Models." Climate Dynamics 53 (5): 2919–30.

Miralles, D. G., Teuling, A. J., van Heerwaarden, C. C., & Vilà-Guerau de Arellano, J. (2014). Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nature Geoscience, 7(5), 345–349. https://doi.org/10.1038/ngeo2141

Miralles, D. G., van den Berg, M. J., Teuling, A. J., & de Jeu, R. A. M. (2012). Soil moisture-temperature coupling: A multiscale observational analysis. Geophysical Research Letters, 39(21). https://doi.org/10.1029/2012gl053703

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Jean-Noël Thépaut. (2021). ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. [Dataset]. Earth System Science Data, 13(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Olah, C., Mordvintsev, A., & Schubert, L. (2017). Feature Visualization. Distill, 2(11). https://doi.org/10.23915/distill.00007

Orth, R., & Seneviratne, S. I. (2012). Analysis of soil moisture memory from observations in Europe. Journal of Geophysical Research. https://doi.org/10.1029/2011JD017366

Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving precipitation estimation using convolutional neural network. Water Resources Research, 55(3), 2301–2321. https://doi.org/10.1029/2018wr024090

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & Others. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, 2825–2830.

https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://githubhelp.com

Quesada, B., Vautard, R., Yiou, P., Hirschi, M., & Seneviratne, S. I. (2012). Asymmetric European summer heat predictability from wet and dry southern winters and springs. Nature Climate Change, 2(10), 736–741. https://doi.org/10.1038/nclimate1536

Ruscica, R. C., Sörensson, A. A., & Menéndez, C. G. (2014). Hydrological links in Southeastern South America: soil moisture memory and coupling within a hot spot. International Journal of Climatology, 34(14), 3641–3653. https://doi.org/10.1002/joc.3930

Ryu, Y., Baldocchi, D. D., Ma, S., & Hehn, T. (2008). Interannual variability of evapotranspiration and energy exchange over an annual grassland in California. Journal of Geophysical Research, D: Atmospheres, 113(D9). https://doi.org/10.1029/2007JD009263

Schwingshackl, C., Hirschi, M., & Seneviratne, S. I. (2017). Quantifying Spatiotemporal Variations of Soil Moisture Control on Surface Energy Balance and Near-Surface Air Temperature. Journal of Climate, 30(18), 7105–7124. https://doi.org/10.1175/JCLI-D-16-0727.1

Schwingshackl, C., Hirschi, M., & Seneviratne, S. I. (2018). A theoretical approach to assess soil moisture–climate coupling across CMIP5 and GLACE-CMIP5 experiments. Earth System Dynamics Discussions, 1–26. https://doi.org/10.5194/esd-2018-34

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. Earth-Science Reviews, 99(3), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Seneviratne, S. I., Koster, R. D., Guo, Z., Dirmeyer, P. A., Kowalczyk, E., Lawrence, D., Liu, P., Mocko, D., Lu, C.-H., Oleson, K. W., & Verseghy, D. (2006). Soil Moisture Memory in AGCM Simulations: Analysis of Global Land–Atmosphere Coupling Experiment (GLACE) Data. Journal of Hydrometeorology, 7(5), 1090–1112. https://doi.org/10.1175/JHM533.1

Seneviratne, S. I., Lüthi, D., Litschi, M., & Schär, C. (2006). Land–atmosphere coupling and climate change in Europe. Nature, 443(7108), 205–209. https://doi.org/10.1038/nature05095

Seneviratne, S. I., Wilhelm, M., Stanelle, T., Hurk, B., Hagemann, S., Berg, A., Cheruy, F., Higgins, M. E., Meier, A., Brovkin, V., Claussen, M., Ducharne, A., Dufresne, J.-L., Findell, K. L., Ghattas, J., Lawrence, D. M., Malyshev, S., Rummukainen, M., & Smith, B. (2013). Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment. Geophysical Research Letters, 40(19), 5212–5217. https://doi.org/10.1002/grl.50956

Shukla, J., & Mintz, Y. (1982). Influence of Land-Surface Evapotranspiration on the Earth's Climate. Science, 215(4539), 1498–1501. https://doi.org/10.1126/science.215.4539.1498

Song, Y. M., Wang, Z. F., Qi, L. L., & Huang, A. N. (2019). Soil moisture memory and its effect on the surface water and heat fluxes on seasonal and interannual time scales. Journal of Geophysical Research, 124(20), 10730–10741. https://doi.org/10.1029/2019jd030893

Spennemann, P. C., M. Salvia, R. C. Ruscica, A. A. Sörensson, F. Grings, and H. Karszenbaum. (2018). "Land-Atmosphere Interaction Patterns in Southeastern South America Using Satellite Products and Climate Models." International Journal of Applied Earth Observation and Geoinformation 64 (February): 96–103.

Steininger, M., Kobs, K., Davidson, P., Krause, A., & Hotho, A. (2021). Density-based weighting for imbalanced regression. Machine Learning, 110(8), 2187–2211. https://doi.org/10.1007/s10994-021-06023-5

Swain, D. L., Horton, D. E., Singh, D., & Diffenbaugh, N. S. (2016). Trends in atmospheric patterns conducive to seasonal precipitation and temperature extremes in California. Science Advances, 2(4), e1501344. https://doi.org/10.1126/sciadv.1501344

Teuling, A. J., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., Buchmann, N., Ammann, C., Montagnani, L., Richardson, A. D., Wohlfahrt, G., & Seneviratne, S. I. (2009). A regional perspective on trends in continental evaporation. Geophysical Research Letters, 36(2). https://doi.org/10.1029/2008gl036584

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the geosciences: Applications to earth system variability. Journal of Advances in Modeling Earth Systems, 12(9). https://doi.org/10.1029/2019ms002002

Vautard, R., Yiou, P., D'Andrea, F., de Noblet, N., Viovy, N., Cassou, C., Polcher, J., Ciais, P., Kageyama, M., & Fan, Y. (2007). Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit. Geophysical Research Letters, 34(7). https://doi.org/10.1029/2006GL028001

Vogel, M. M., R. Orth, F. Cheruy, S. Hagemann, R. Lorenz, B. J. J. M. van den Hurk, and S. I. Seneviratne. (2017). "Regional Amplification of Projected Changes in Extreme Temperatures Strongly Controlled by Soil Moisture-Temperature Feedbacks." Geophysical Research Letters 44 (3): 1511–19.

Wang, L., Scott, K. A., Xu, L., & Clausi, D. A. (2016). Sea Ice Concentration Estimation During Melt From Dual-Pol SAR Scenes Using Deep Convolutional Neural Networks: A Case Study. IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society, 54(8), 4524–4533. https://doi.org/10.1109/TGRS.2016.2543660

Wickham, H., Cook, D., Hofmann, H., & Buja, A. (2010). Graphical inference for Infovis. IEEE Transactions on Visualization and Computer Graphics, 16(6), 973–979. https://doi.org/10.1109/TVCG.2010.161

Wimmers, A., Velden, C., & Cossuth, J. H. (2019). Using Deep Learning to Estimate Tropical Cyclone Intensity from Satellite Passive Microwave Imagery. Monthly Weather Review, 147(6), 2261–2282. https://doi.org/10.1175/MWR-D-18-0391.1

Wu, W., & Dickinson, R. E. (2004). Time Scales of Layered Soil Moisture Memory in the Context ofLand–Atmosphere Interaction. Journal of Climate, 17(14), 2752–2764. https://doi.org/10.1175/1520-0442(2004)017<2752:TSOLSM>2.0.CO;2

Zender, C. S. (2008). Analysis of self-describing gridded geoscience data with netCDF Operators (NCO). Environmental Modelling & Software, 23(10), 1338–1342. https://doi.org/10.1016/j.envsoft.2008.03.004

Zhang, G., Wang, M., & Liu, K. (2021). Deep neural networks for global wildfire susceptibility modelling. Ecological Indicators, 127, 107735. https://doi.org/10.1016/j.ecolind.2021.107735

Journal of Geophysical Research: Atmospheres

Supporting Information for

Using Machine Learning with Partial Dependence Analysis to Investigate Coupling Between Soil Moisture and Near-surface Temperature

Jared T. Trok[1], Frances V. Davenport[1,2,3], Elizabeth A. Barnes[2], and Noah S. Diffenbaugh[1,4]

[1]Department of Earth System Science, Stanford University, Stanford, CA, USA. [2]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA. [3]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, USA. [4]Doerr School of Sustainability, Stanford University, Stanford, CA, USA.

Contents of this file

Figures S1 to S11

Introduction

The following document consists of four figures (S1-S4) supporting our primary analysis (using the 1979-2021 ERA5 and ERA5-Land historical reanalysis datasets), and seven figures (S5-S11) showing the results of our secondary analysis (using the 1979-2021 NCEP/DOE R2 historical reanalysis dataset). Figure S1 compares the mean seasonal cycle of daily maximum 2-meter temperature (TMAX) between the ERA5 dataset and convolutional neural network (CNN) predictions. Figure S2 shows each CNN's ability to predict daily TMAX anomalies from the seasonal cycle. Figure S3 shows subregional variability in soil moisture-temperature relationships obtained through partial dependence analysis (using southcentral North America as an example). Figure S4 provides examples of regional soil moisture-temperature relationships obtained through partial dependence analysis of CNNs trained using datasets with randomly shuffled soil moisture input maps. Figures S5 through S11 present the results of our analysis when applied to the NCEP/DOE R2 historical reanalysis dataset (1979-2021).

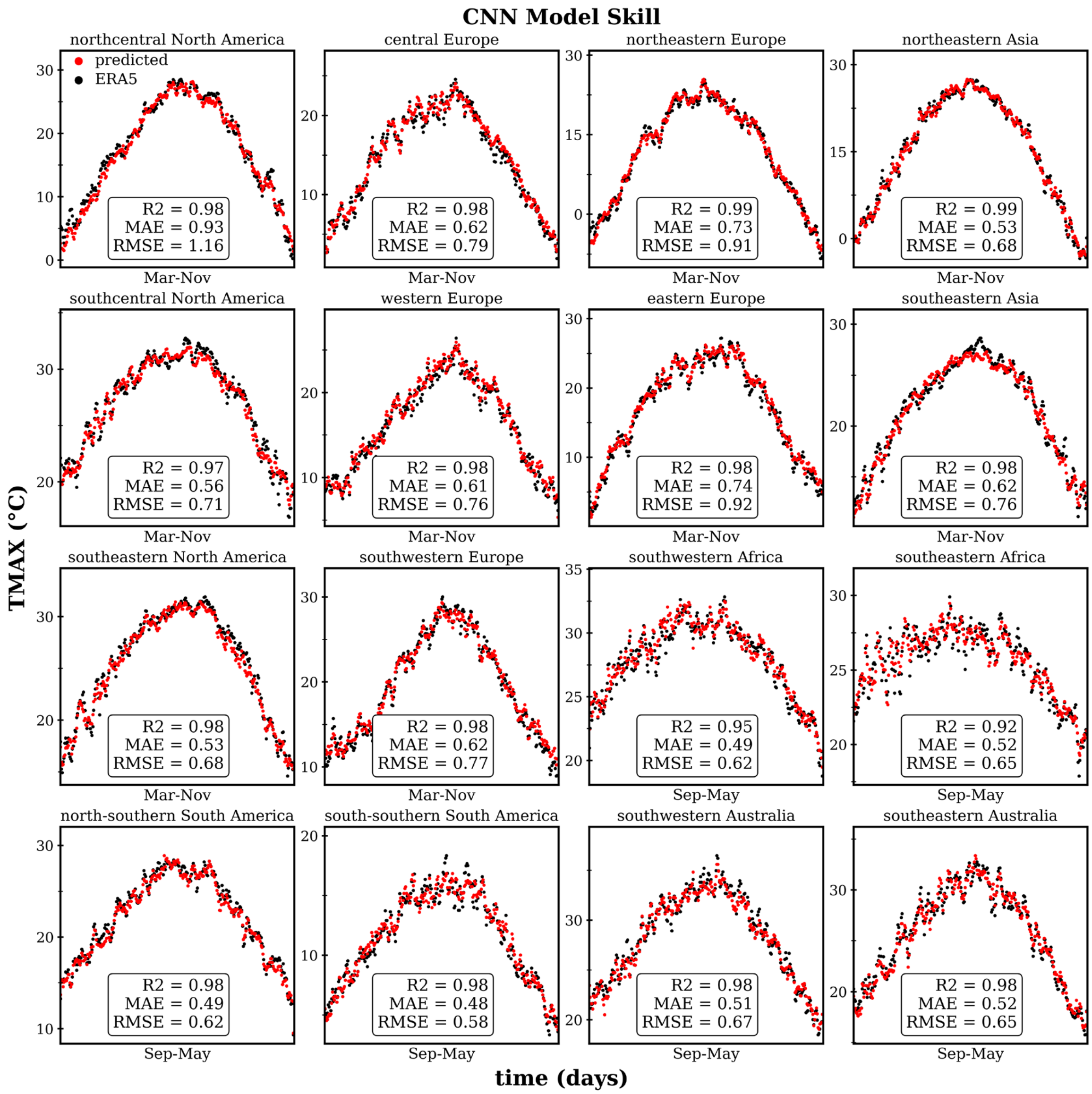

Figure S1. Comparison of the annual temperature cycle between (black) ERA5-Land daily maximum 2-meter temperatures (TMAX) and (red) convolutional neural network TMAX predictions. Each subplot shows the mean seasonal cycle of TMAX over snow-free months averaged across all 8 years in the testing subset. The coefficient of determination (R2), mean absolute error (MAE), and mean squared error (MSE) is shown for each region. Snow-free months are defined as March-November (Mar-Nov) in the Northern Hemisphere and September-May (Sep-May) in the Southern Hemisphere.

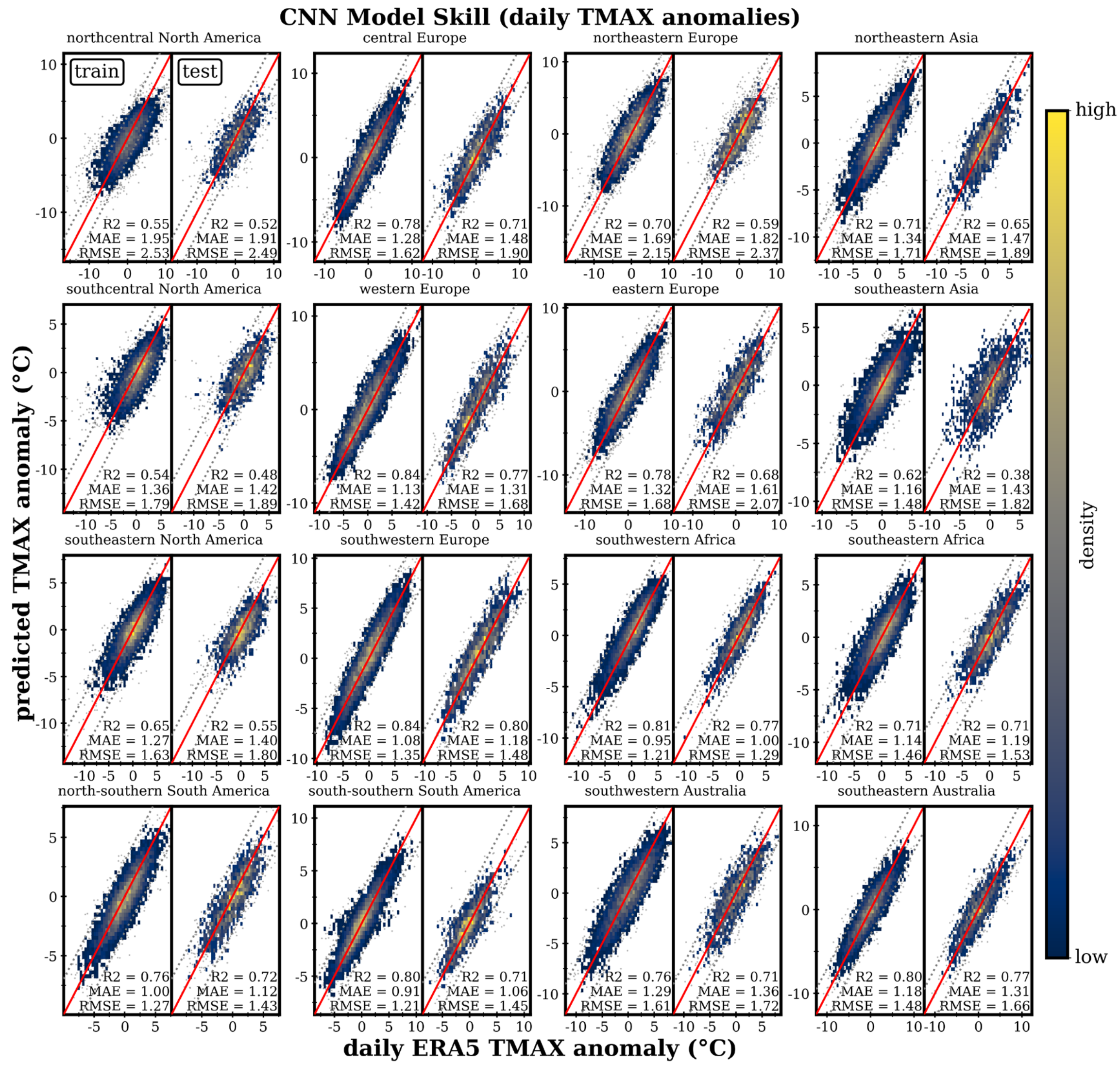

Figure S2. Comparison between model predicted temperature anomaly and ERA5-Land temperature anomaly for each convolutional neural network trained to predict daily maximum 2-meter temperature (TMAX) over a region. Each regional subplot shows the coefficient of determination (R2), mean absolute error (MAE), and mean squared error (MSE) for both the training dataset (left) and the testing dataset (right). Daily TMAX anomalies calculated as deviations from the ERA5 seasonal cycle (see Figure S1). Correct predictions fall along the 1-1 line (red). Gray dotted lines show +/- 3 degrees C prediction errors.

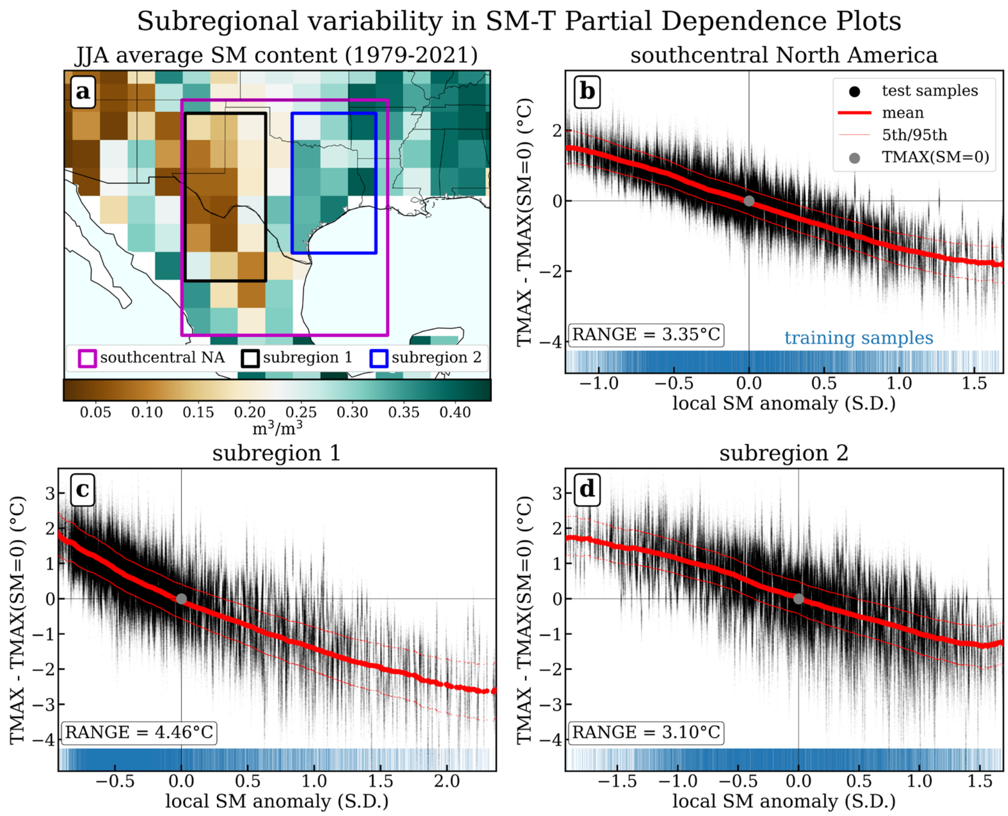

Figure S3. Subregional variability in soil moisture-temperature (SM-T) relationships obtained through partial dependence analysis (method detailed in Figure 2) of convolutional neural networks. (a) Average summertime (June-August) SM content over southcentral North America and two subregions which exhibit dry (subregion 1) and wet (subregion 2) summertime SM conditions, respectively. (b, c, d) SM-T relationships calculated from CNNs trained to predict TMAX over (b) southcentral North America (reproduced from Figure 7), (c) subregion 1, and (d) subregion 2. The smoothed moving average (thick red line) shows the average behavior of the neural network's TMAX prediction as the SM input varies from dry (negative) to wet (positive) local SM anomalies. Also shown are the moving 5th and 95th percentiles of the temperature predictions (thin red lines). The SM-T relationships shown are calculated from the testing dataset. We also include a rug plot showing the distribution of SM anomalies in the training dataset. For each region and subregion, we calculate the range (vertical extent) of the mean SM-T relationship. The local soil moisture anomalies (x-axis) are calculated as standard deviations (S.D.) from the calendar-day mean and averaged over all non-ocean grid cells within the region bounds.

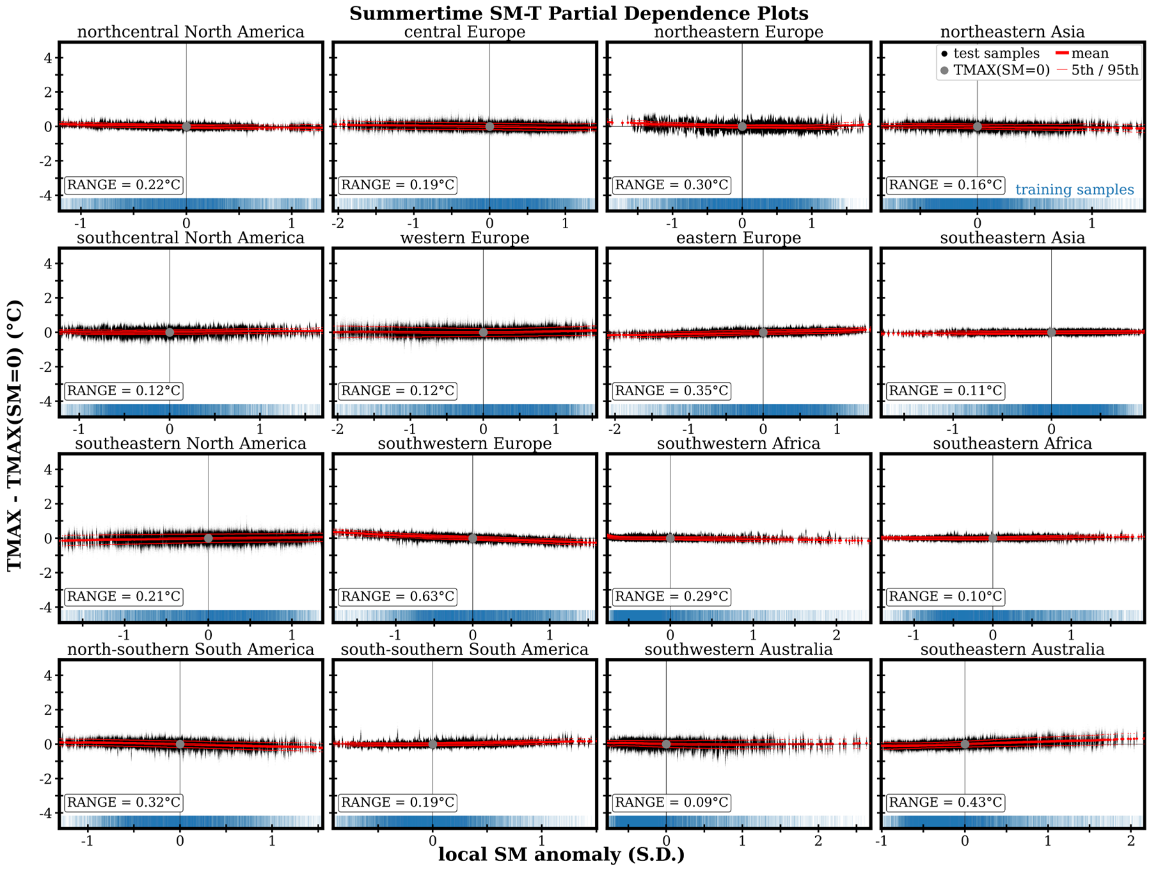

Figure S4. Soil moisture-temperature (SM-T) relationships obtained through partial dependence analysis (method detailed in Figure 2) of convolutional neural networks trained with randomly shuffled soil moisture input maps. The smoothed moving average (thick red line) shows the average behavior of the neural network's prediction as the SM input varies from dry (negative) to wet (positive) local SM anomalies. Also shown are the moving 5th and 95th percentiles of the temperature predictions (thin red lines). The SM-T relationships shown are calculated from the testing dataset. We also include a rug plot showing the distribution of SM anomalies in the training dataset. For each subplot, we calculate the range (vertical extent) of the mean SM-T relationship. Soil moisture anomalies are calculated as standard deviations (S.D.) from the calendar-day mean.

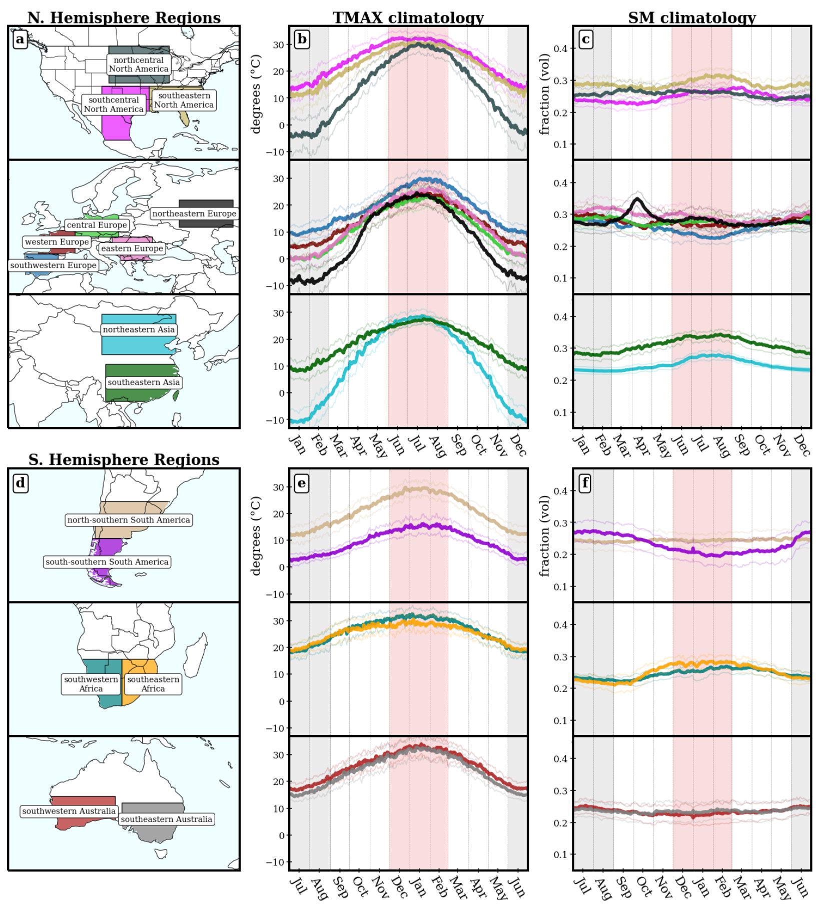

Figure S5. Same as Figure 3, except climatologies are calculated using the 1979-2021 NCEP/DOE R2 (NCEP) reanalysis dataset. (a) Northern Hemisphere regions included in this analysis alongside 1979-2021 regional climatologies of (b) daily maximum 2-meter temperature (TMAX), and (c) volumetric soil moisture fraction (SM). Thin lines show +/- 1 standard deviation from climatological mean. (d, e, f) Same as (a, b, c) but for Southern Hemisphere regions. Red shading indicates summer months in each hemisphere over which this study analyzes soil moisture-temperature coupling. Gray shading indicates winter months removed from all subsequent analyses.

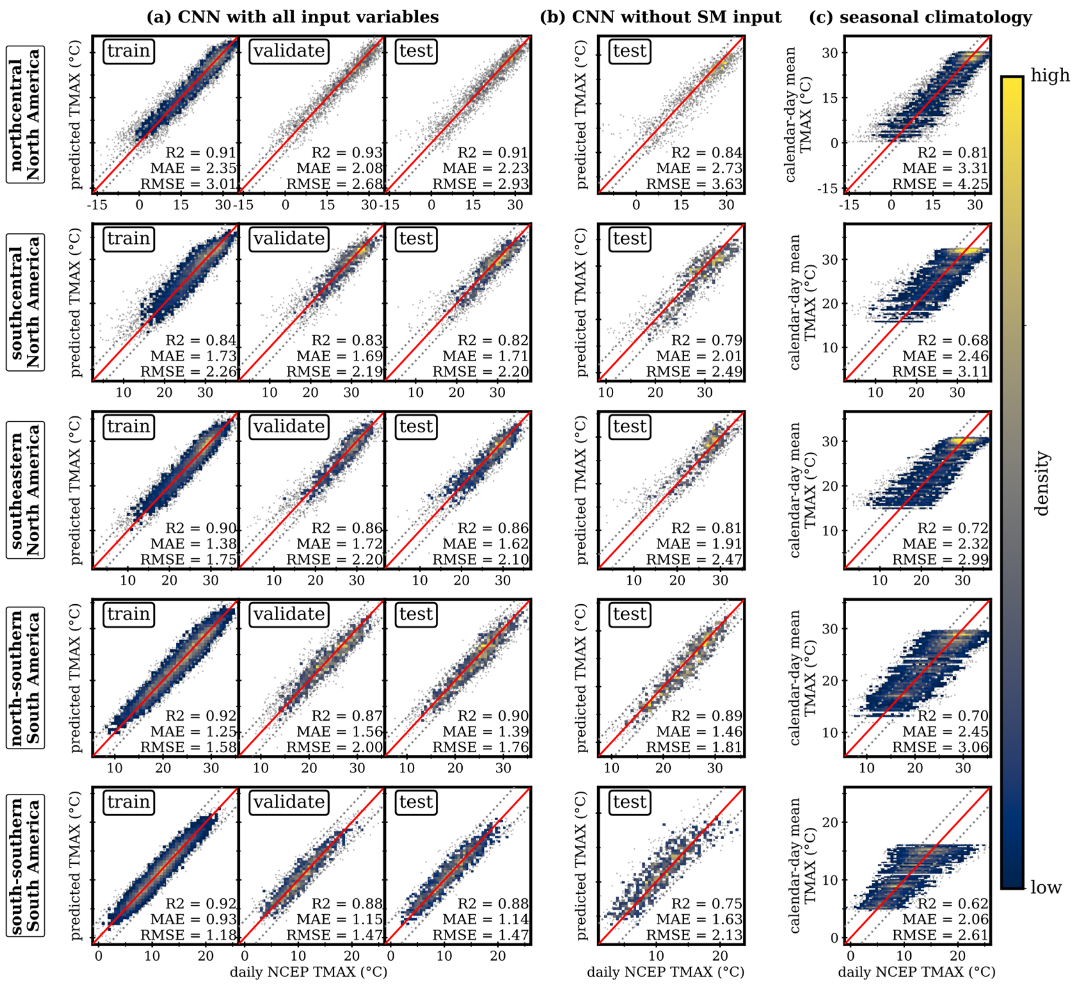

Figure S6. Same as Figure 4, but for convolutional neural networks trained using the 1979-2021 NCEP/DOE R2 (NCEP) reanalysis dataset. CNN model skill comparison for North and South American regions. (a) Comparison between ERA5-Land TMAX and predicted TMAX from convolutional neural networks (CNNs) trained with daily geopotential height anomaly maps, soil moisture anomaly maps (SM), and normalized calendar day inputs. Model performance is shown separately for the 27-year training subset (used to fit CNN weights), the 8-year validation subset (used to optimize hyperparameters), and the 8-year testing subset (unseen data left out of the training process). See Methods for more details on the training, validation, and testing subsets. (b) Same as (a) but for CNNs trained without the SM inputs. Model performance is shown for the 8-year testing subset. (c) The seasonal climatology of TMAX as shown by comparing the ERA5-Land daily TMAX and the calendar-day mean TMAX each day (averaged over 1979-2021). Each subplot shows the coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE). Correct predictions fall along the 1-1 line (red). Gray dotted lines show +/- 3 degrees C prediction errors.

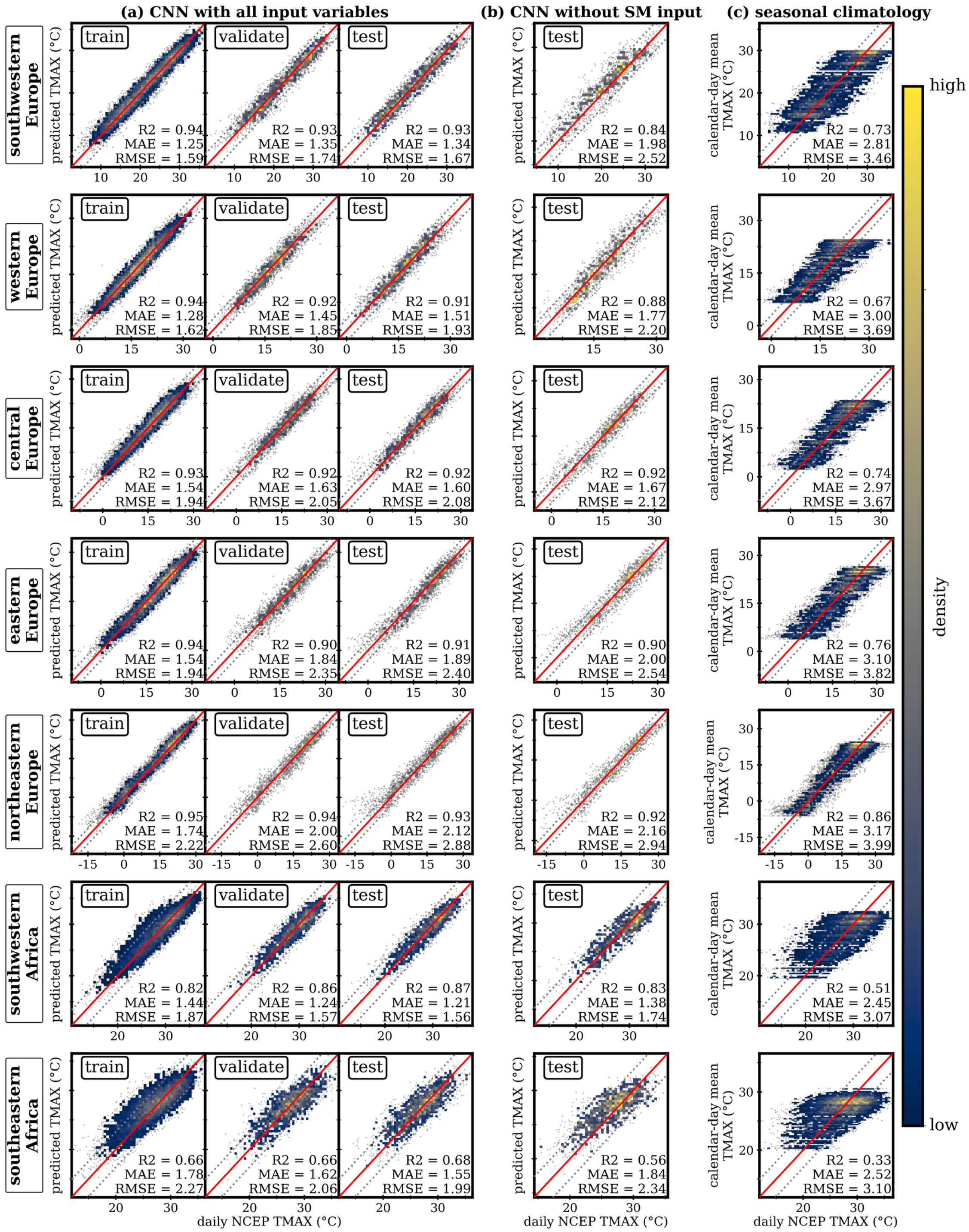

Figure S7. Same as Figure 5, but for convolutional neural networks trained using the 1979-2021 NCEP/DOE R2 (NCEP) reanalysis dataset. CNN model skill comparison for regions in Europe and Africa.

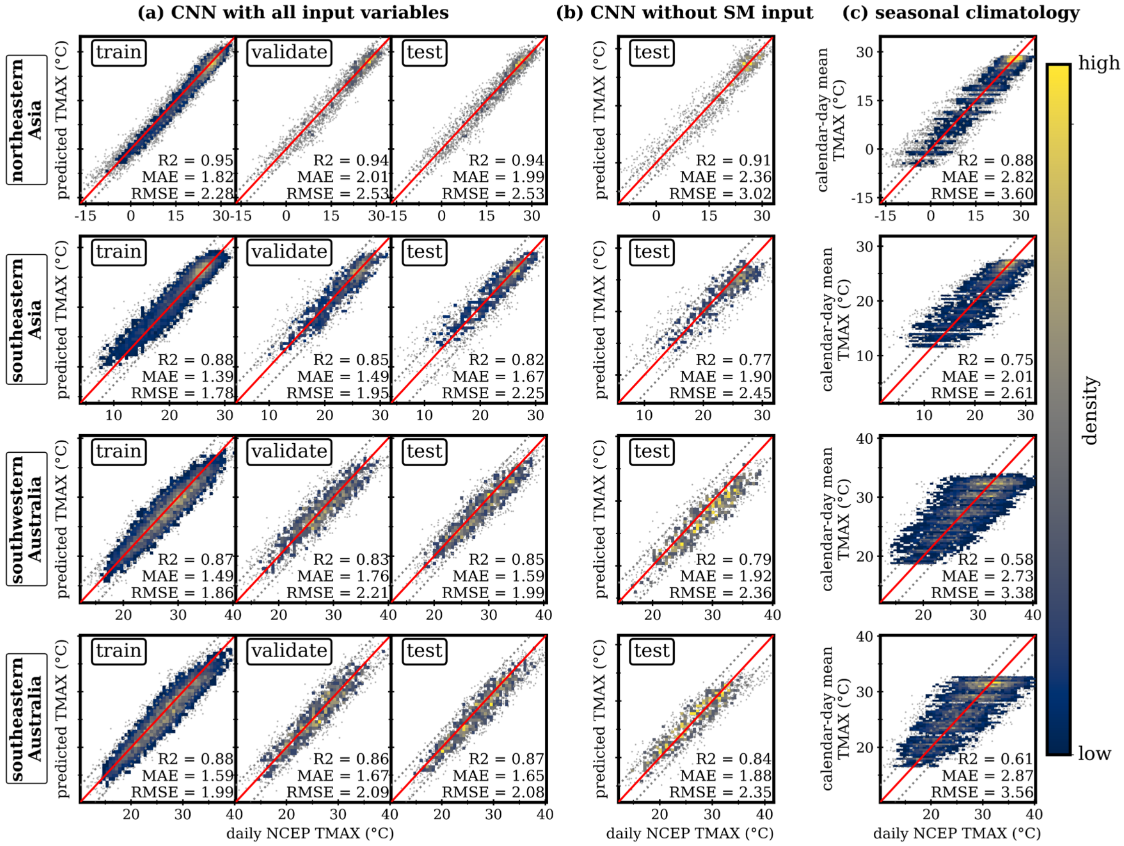

Figure S8. Same as Figure 6, but for convolutional neural networks trained using the 1979-2021 NCEP/DOE R2 (NCEP) reanalysis dataset. CNN model skill comparison for regions in Eastern Asia and Australia.

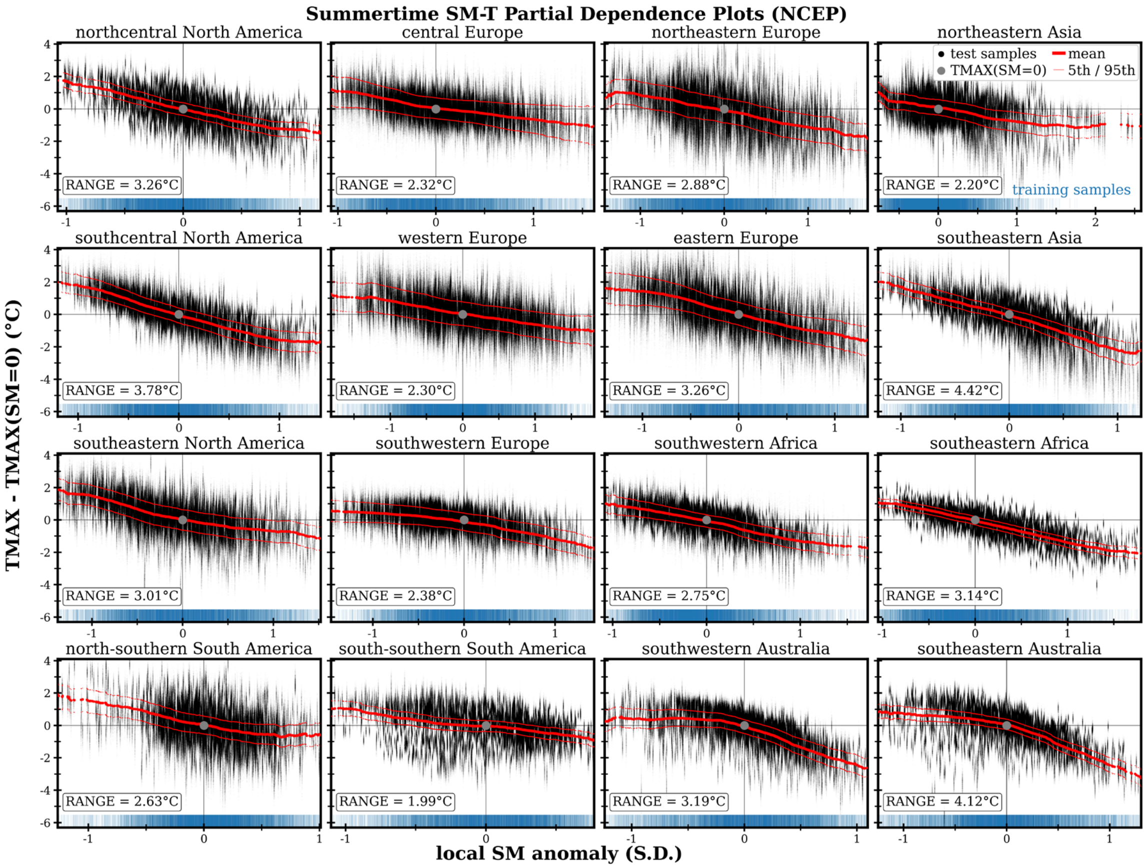

Figure S9. Same as Figure 7, but for convolutional neural networks trained using the 1979-2021 NCEP/DOE R2 (NCEP) reanalysis dataset. Soil moisture-temperature (SM-T) relationships obtained through partial dependence analysis of convolutional neural networks (method detailed in Figure 2). The smoothed moving average (thick red line) shows the average behavior of the neural network's prediction as the SM input varies from dry (negative) to wet (positive) anomalies. Also shown are the moving 5th and 95th percentiles of the temperature predictions (thin red lines). The SM-T relationships shown are calculated from the testing dataset. We also include a rug plot showing the distribution of SM anomalies in the training dataset. For each subplot, we calculate the range (vertical extent) of the mean SM-T relationship. Soil moisture anomalies are calculated as standard deviations (S.D.) from the calendar-day mean.

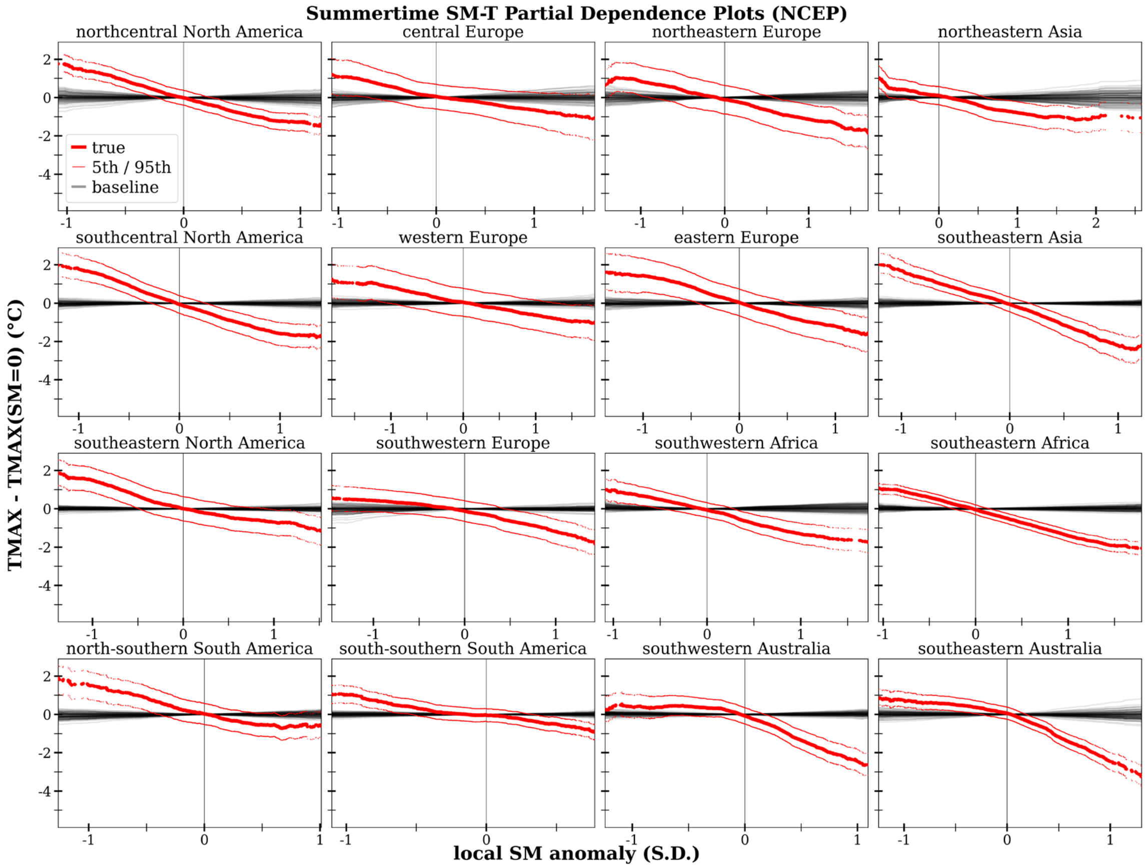

Figure S10. Same as Figure 8, but for convolutional neural networks trained using the 1979-2021 NCEP/DOE R2 (NCEP) reanalysis dataset. Regional soil moisture-temperature (SM-T) relationships obtained through partial dependence analysis (method detailed in Figure 2) of convolutional neural networks (CNNs) trained to predict regional daily maximum temperature (TMAX) given geopotential height, calendar-day, and soil moisture inputs. Each regional subplot shows 101 SM-T partial dependence plots (PDPs), consisting of the true SM-T PDP (red; Figure 7) and 100 baseline SM-T PDPs (black) derived from CNNs trained with shuffled soil moisture inputs (each shuffled using a different random seed). Also shown are the moving 5th and 95th percentiles of the true SM-T PDP (thin red lines). Soil moisture anomalies are calculated as standard deviations (S.D.) from the calendar-day mean.

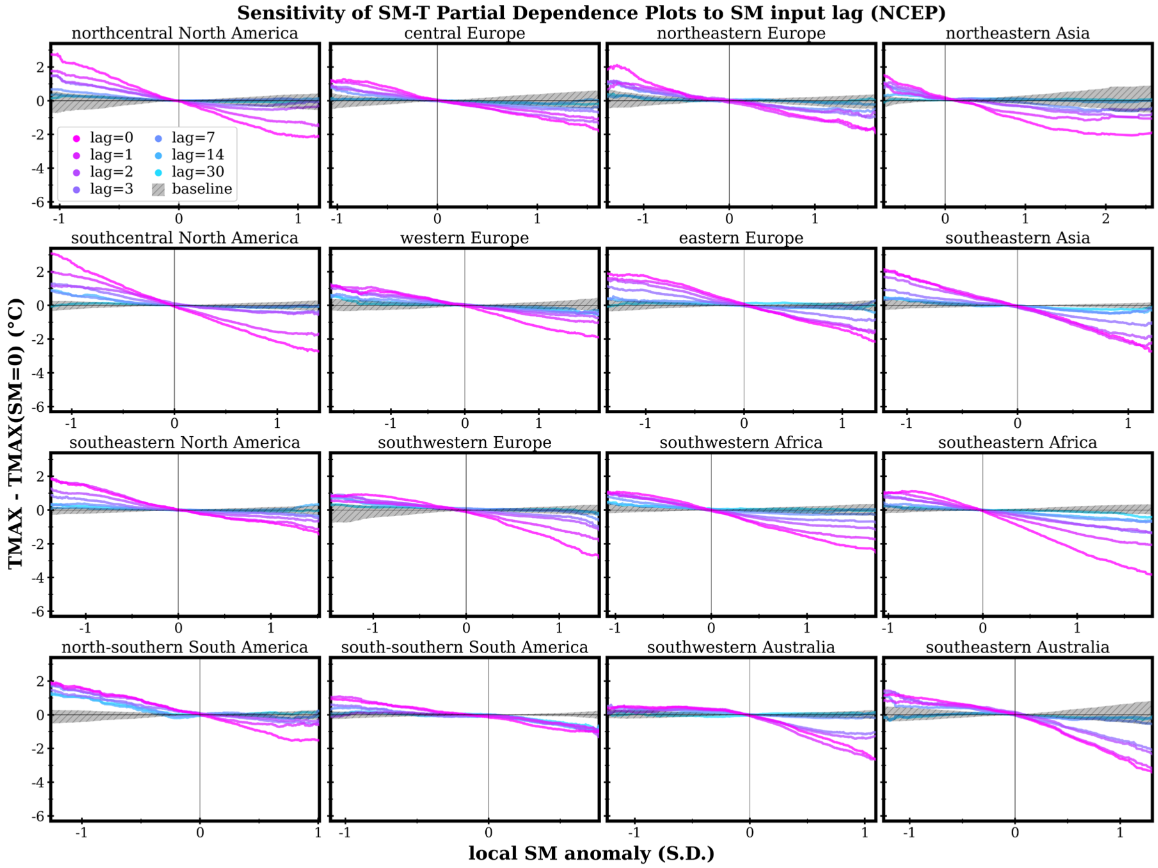

Figure S11. Same as Figure 9, but for convolutional neural networks trained using the 1979-2021 NCEP/DOE R2 (NCEP) reanalysis dataset. Regional soil moisture-temperature (SM-T) partial dependence relationships obtained using the method detailed in Figure 2 (but for CNNs trained with various levels of soil moisture input lag). Each regional subplot shows SM-T relationships derived from 7 different CNNs trained to predict daily TMAX given the following inputs: calendar day, daily GPH anomaly map, and a single day's SM anomaly map lagged by 0-30 days prior to the prediction day. After the training process, CNN weights are saved and used to calculate the SM-T PDPs as in Figure 2. Colors show SM-T relationships for CNNs trained with SM input lags of 0, 1, 2, 3, 7, 14, and 30 days. Hatching shows the range of the 100 baseline PDPs (Figure 8). Soil moisture anomalies are calculated as standard deviations (S.D.) from the calendar-day mean.