



Towards daily maximum heat index estimation across the conterminous United States using satellite-derived products

Timothy Pede & Giorgos Mountrakis

To cite this article: Timothy Pede & Giorgos Mountrakis (2022) Towards daily maximum heat index estimation across the conterminous United States using satellite-derived products, International Journal of Remote Sensing, 43:8, 2861-2884, DOI: [10.1080/01431161.2022.2072180](https://doi.org/10.1080/01431161.2022.2072180)

To link to this article: <https://doi.org/10.1080/01431161.2022.2072180>



View supplementary material [↗](#)



Published online: 31 May 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Towards daily maximum heat index estimation across the conterminous United States using satellite-derived products

Timothy Pede and Giorgos Mountrakis

Department of Environmental Resources Engineering, SUNY College of Environmental Science and Forestry, Syracuse, New York, USA

ABSTRACT

Satellite-derived land surface temperature (LST) is widely utilized to study urban heat islands in the context of human health and thermal exposure. However, there is growing evidence to suggest that LST may be a poor indicator of apparent temperature, or the human-perceived equivalent temperature that reflects both heat and humidity. Moreover, heat index (HI), the apparent temperature metric used by the US National Weather Service, has yet to be computed at an increased spatial resolution and coverage beyond weather station observations. The goals of this study were to: 1) assess the extent to which HI can be estimated by a combination of Moderate Resolution Imaging Spectroradiometer (MODIS) sensor and other satellite-derived products available at continental scales, and 2) determine which factors are most important in this estimation. Specifically, daily maximum 1-km HI from May through September of 2012 was modelled across the conterminous United States as a function of MODIS LST, precipitable water vapor (PWV), and near-infrared indices, in addition to static variables capturing land cover, topographic, and locational factors. The derived model was capable of estimating HI within a reasonable level of error ($R^2 = 0.83$, RMSE = 4.4°F). This is the first time that HI has been directly estimated using exclusively remotely sensed products and validated over a large spatial extent. Analysis of individual variables indicated that LST and PWV were, by far, the most important factors for estimation. However, the incorporation of additional parameters further improved model performance (R^2 : +0.14, RMSE: -1.6°F). We hope that our work will eventually result in a national HI product assisting researchers in a variety of fields, including epidemiology, building energy demand, and environmental justice. Further work to interpolate cloud-contaminated satellite observations and downscale estimates to a 60-m resolution would considerably increase the utility of this HI estimation methodology.

ARTICLE HISTORY

Received 10 January 2022

Accepted 27 April 2022

KEYWORDS

Heat index; MODIS; Land surface temperature (LST); Precipitable water vapor (PWV); Apparent temperature; Random forest

CONTACT Timothy Pede tjpede@syrr.edu Department of Environmental Resources Engineering, SUNY College of Environmental Science and Forestry, Syracuse, New York, USA

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/01431161.2022.2072180>

© 2022 Informa UK Limited, trading as Taylor & Francis Group

1. Introduction

On an average annual basis, extreme heat is the deadliest natural disaster in the US, killing 600 people per year (US Centre for Disease Control and Prevention (CDC) (2013)). Projections indicate that heat waves will become more frequent, more severe, and longer lasting in the US (US Meehl 2004). Heat Index (HI) is the standard apparent temperature metric used by the US National Weather Service (US NWS) and based on extensive biometeorology studies (Steadman 1979; Rothsfuz 1990). HI is computed as a 2nd order polynomial function of air temperature and humidity (US National Weather Service (US NWS) (2014)). The US NWS advises caution with HIs above 80°F, as prolonged exposure or physical activity can lead to heat exhaustion. Sunstroke is possible for HIs above 90°F and highly likely for HIs above 105°F (US National Weather Service (US NWS) (2019)).

A spatially explicit HI product derived from satellite imagery would allow for the study of variation in heating trends and thermal exposure across metropolitan and rural areas at a much higher spatial resolution than provided by in-situ weather stations. While weather stations are available, especially in developed regions, the vast majority do not record humidity. For instance, the Global Historic Climatology Network contains over 10,000 stations for the conterminous United States (CONUS) but does not contain humidity observations. This compares to the Local Climatology Dataset that does include humidity, but only encompasses 1,400 stations. Thus, to date, high resolution HI information has been limited. Several areas of research could benefit from this information and include:

- (1) Heat-related illness and identifying heat vulnerable populations (i.e., Klein Rosenthal, Kinney and Metzger 2014; Sabrin, Karimi and Nazari 2020).
- (2) Forecasting peak electrical demand and assessing grid vulnerability (i.e., Auffhammer and Mansur 2016; Ortiz, González and Lin 2018; Maia-Silva, Kumar and Nateghi 2020).
- (3) Environmental justice and quantifying thermal inequalities (i.e., Ahmed 2018; Mitchell and Chakraborty 2018).

Numerous researchers have employed land surface temperature (LST) from the Moderate Resolution Imaging Spectroradiometer (MODIS) to estimate daily maximum air temperature (Tmax) using a statistical framework (Table 1), meaning Tmax was empirically estimated as a function of one or more independent variables. Empirically estimating relative humidity (RH) from MODIS imagery is less common (Table 1). All but Li and Zha (2018) incorporated some form of MODIS moisture product, such as precipitable water vapour (PWV) or atmospheric moisture (AM). Recondo et al. (2013) was the only author to derive models for both Tmax and RH, which could have theoretically been used to derive apparent temperature. The study performed by Ho et al. (2016) is most relevant to this analysis, as they estimated HUMEDIX, a Canadian measurement of apparent temperature, using a combination of Landsat LST, MODIS PWV, and auxiliary data related to urbanization level.

The goal of this analysis was to propose a novel framework for estimating HI across the CONUS via remotely sensed products. The objectives were twofold: (1) determine the extent to which HI can be estimated using a combination of static and dynamic satellite

Table 1. Overview of studies that estimated daily maximum air temperature (Tmax) or relative humidity (RH) using a statistical framework.

Tmax Studies	Predictor Variables	Method	Location	R ²	RMSE/MAE
Zhang et al. (2011)	LST (Ad, An), Elev (by land cover)	Linear Regression	China	0.87	-
Benali et al. (2012)	LST (Td, Tn), Day length	Linear Regression (Bootstrapping)	Portugal	0.92	3.3°F
Emamifar, Rahimikhoob and Noroozi (2013)	LST (Td), Julian day, Solar radiation	Regression Tree	Khuzestan province	0.84–0.92	-
Kim and Han (2013)	LST (Td), NDWI	Stepwise Linear Regression	South Korea	0.82	5.2°F
Lin et al. (2012)	LST (Ad), PWV (Ad), NDVI, EVI, Solar angle, Elev, Lat, Lon	Stepwise Linear Regression	East Africa	0.79	3.4°F
Recondo et al. (2013)	LST (Td), NDVI, Elev, Julian day	Regression	Spain	0.89	4.7°F
Rhee and Im (2014)	LST (Ad, An) - Diurnal Curve (by land cover)	Robust Regression	South Korea	0.64–0.87	-
Xu, Knudby and Ho (2014)	LST (Ad), NDVI, NDWI, Lat, Lon, Dist, ocean, Elev, Albedo, and Solar radiation	Sin Curve Fit	British Columbia	0.74	3.6°F
Zeng et al. (2015a)	LST (Td, Tn), Elev (by land cover)	Random Forest	US Corn Belt	0.76–0.83	4.0–4.1°F
Noi, Kappas and Degener (2016)	LST (Ad, An, Td, Tn), NDVI, Elev, Lon, Lat, Day length, Julian day, View angle	Linear Regression	Northern Vietnam	0.93	2.6°F
Noi, Degener and Kappas (2017)	LST (Ad, An, Td, Tn), Elev, Julien day	Linear Regression	Northwest Vietnam	0.93	2.7°F
Rosenfeld et al. (2017)	LST (An), NDVI, Population density, Dist. water, %Road, %Open, Elev, Slope, Aspect	Cubist Regression	Israel	0.9	3.4°F
Shi et al. (2017)	LST (Td), NDVI	Linear Regression (Fixed Effects)	Yangtze River Delta	0.85–0.90	-
Didari and Zand-Parisa (2018)	LST (Ad, Td), NDVI, Elev, Lat	Linear Regression	Iran	0.93	4.7°F
Yoo et al. (2018)	LST (Ad, An, Td, Tn), Elev, Aspect, Solar radiation, NDVI, Lat, Lon, % Impervious	Stepwise Linear Regression	Los Angeles/ Seoul	0.77–0.85	2.2–3.1°F
Zhang et al. (2018)	LST (Ad, An, Td, Tn), Elev	Random Forest	Tibetan Plateau	-	9.5–13.5°F
Phan et al. (2019)	LST (Ad, An, Td, Tn)	Linear Regression	northwest Vietnam	0.74–0.94	2.0–3.4°F
Chung et al. (2020)	LST (Ad, An, Td, Tn)	Linear Regression	South Korea	0.92	8.6°F
dos Santos (2020)	LST (Td), SA, NDVI, Albedo, Lat, Lon, Dist coast, Julien day, Elev	Artificial Neural Network	London, England	0.54	4.4°F
Hough et al. (2020)	LST (Ad, An, Td, Tn, Landsat), NDVI, Elev, Population density, Climate Region	Artificial Neural Network	France	0.95	3.3°F

(Continued)

Table 1. (Continued).

Tmax Studies	Predictor Variables			Method	Location	R ²	RMSE/MAE
Serra et al. (2020)	LST (Td, Tn)	NDVI, NDBI, Elev, Aspect, Slope, Lat, Lon, Dist. ocean		Stepwise Linear Regression	Barcelona, Spain	0.92	2.0–3.4°F
Shen et al. (2020)	LST (Td), NDVI, Julien day, Lat, Lon, Elev, Albedo, View angle, Land cover, Population density, Road density, Wind speed, Soil moisture content, Evaporation from soil			Deep Belief Network	China	0.97	3.6°F
RH Studies	Predictor Variables			Method	Location	R²	RMSE/MAE
Peng et al. (2006)	AT (Ad), AM (Ad), NDVI			Quadratic Regression	Peninsular Malaysia	0.90–0.99	-
Sofan, Sugiharto and Hasnaeni (2010)	PWW (Ad)			Linear Regression	Java Island	0.71–0.85	-
Adab et al. (2013)	PWW (Ad), Elev			Polynomial Regression	Zarin Gol Valley	-	5.80%
Lin et al. (2013)	AT (Ad), AM (Ad), NDVI			Linear Regression	East Africa	0.17–0.39	-
Recondo et al. (2013)	LST (Td), NDVI, Julien day, Elev			Robust Regression	Spain	0.40–0.51	-
Ho et al. (2016)*	LST (Landsat), PWW (Td), NDWI, Dist. ocean, Solar radiation, Sky view factor			Random Forest	Greater Toronto	0.87	9.0°F
Li and Zha (2018)	Red, Green, Blue reflectance (MODIS), EVI, Land cover (MODIS – categorical), Dist. water, Dist. road, Dist. city, Elev Night-time lights			Random Forest	China	0.70	-
Liao et al. (2020)	LST (Td, Tn), AT (Td, Tn), AM (Td, Tn), PWW (Td, Tn)			Linear Regression	US	0.28–0.50	15.3–17.0%
Ramirez-Beltrán et al. 2020	LST (Ad, An, Td, Tn), LST (GOES), PWW (Ad, Td), Brightness (GOES), Albedo (GOES), NDVI			Autoregressive Model	Mesoamerica and Caribbean	0.56–0.62	8.5–10.2%

Ad, An, Td, Tn refer to Aqua Day, Aqua Night, Terra Day, and Terra Night, respectively, LST = Land Surface Temperature (MOD11A1/MYD11A1), PWW = Precipitable Water Vapour (MOD05/MYD05), AM = Atmospheric Moisture (MOD07/MYD07), AT = Atmospheric Temperature (MOD07/MYD07), NDVI = Normalized Difference Vegetation Index, EVI = Enhanced Vegetation Index, NDWI = Normalized Difference Water Index, NDBI = Normalized Difference Built-Up Index, (by land cover) indicates separate models were fit for each land cover type, *denotes Ho et al. (2016) estimated HUMINDEX.

products and (2) assess which factors are most important in this estimation. Specifically, daily maximum 1-km HI was modelled across the CONUS for 2012 using MODIS LST, PWV, and near-infrared (NIR) indices, as well as land cover, topographic, and locational factors.

This study represents the first attempt to estimate daily HI from remotely sensed products. The approach has the potential to provide a significant improvement in spatial resolution over traditional HI computations, which are derived from in-situ weather stations. As seen in Figure 1, these weather stations are not adequate to confidently cover the entire conterminous U.S. and capture the associated HI variability, especially in rural areas. While researchers have used satellite images to predict high-to-moderate spatial resolution Tmax and RH, with the exception of Ho et al. (2016), none have directly computed an index related to apparent temperature (and this only done for a single region). Moreover, HI, the metric used by meteorologists, policy makers, and health professionals in the US, has yet to be derived via remote sensing techniques.

2. Methodology

2.1. Study area

The CONUS was identified as an ideal study area due to its wide range of topographic and climatic conditions, in addition to the availability of auxiliary data (Figure 1). The time-frame for this analysis was defined by the 5-month period from 1 May to 30 September of 2012. These months capture the warmest part of the year and provide a several-week buffer for late/early extreme temperatures. In addition, 2012 was an unusually hot year in

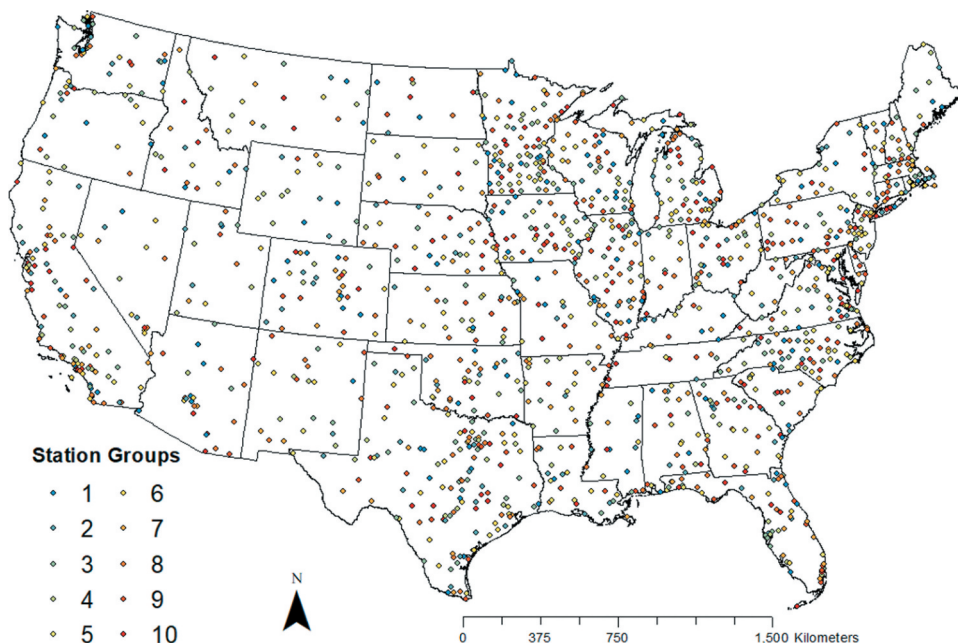


Figure 1. The CONUS study area and location of the 1,395 utilized Local Climatology Dataset (LCD) stations. the station grouping scheme used for the modified 10-fold cross-validation (discussed in Section 2.3) is displayed as well.

which most of the CONUS was adversely affected by prolonged and excessive heat waves (Rippey 2015). However, there were also some cooler periods, resulting in a broad and normally distributed range in HI values (see Figure A1 in Appendix A for more information on HI variability in the utilized dataset).

2.2. Variable identification and justification

2.2.1. Response variable: Daily Heat Index (HI) from weather stations

Daily HI was available through the US Local Climatology Dataset (LCD). This dataset consists of hourly meteorology parameters, including air temperature and RH (US National Ocean and Atmospheric Administration (US NOAA) (2019)). Similar to Ho et al. (2016) and Zhang et al. (2014), daily HI was calculated at the time of maximum air temperature to reflect the apparent temperature during the hottest part of the day. The polynomial formula for computing HI is described by NWS (US NWS 2014). Figure 1 shows the location of the 1,395 CONUS stations with valid data for 2012. Valid refers to having hourly observations from 11:00 AM to 6:00 PM (local time) for at least 145 days (95%) from 1 May to 31 September.

2.2.2. Predictor variables

Independent variables used for HI estimation are summarized in Table 2 and described below in greater detail. The MODIS sensor was the source for daily satellite observations, including LST, PWV, red, NIR, and mid-infrared (MIR) products. Other static variables were employed as well. Land cover information was obtained from the 2011 National Land Cover Dataset (NLCD); topographic factors were derived from the Shuttle Radar Topography Mission (SRTM).

2.2.3. Land surface temperature (LST) variables

LST is the most widely utilized variable for estimating Tmax and has also been used to estimate RH (see Table 1). Daily, 1-km daytime LST from the Aqua satellite was employed (MYD11A1) (Wan 2015), since Aqua's daytime overpass is closer to solar noon than Terra's (1:00 PM vs. 10:00 AM for the CONUS). While some have used atmospheric temperature (AT) from MODIS Atmospheric Profile products (MOD07/MYD07) to estimate RH, we used LST, since it was shown to be closer to radiant ground temperatures (Wang, Liang and Meyers 2008). LST from Terra (MOD11A1) was also used to assess potential benefits of alternatively using morning observations.

2.2.4. Precipitable Water Vapor(PWV)

With the exception of Li and Zha (2018), PWV or atmospheric moisture was used by all studies that estimated RH and Lin et al. (2012) to estimate Tmax (see Table 1). Daily, 1-km PWV from the Aqua satellite was available through the MODIS Total Precipitable Water product (MYD05) (MODIS Science Team 2014). While some have used AM from the MODIS Atmospheric Profile products (MOD07/MYD07), we used PWV, since it was shown to be closer to ground-level humidity (Wong et al. 2015). PWV from Terra (MOD05) was also used to assess potential benefits of alternatively using morning observations.

Table 2. Overview of predictor variables.

Variable	Definition	Type	Res. (m)	Source
LST _{AD}	Afternoon (Aqua daytime) LST	Daily	1,000	MYD11A1
LST _{TD}	Morning (Terra daytime) LST	Daily	1,000	MOD11A1
PWV _{AD}	Afternoon (Aqua daytime) precipitable water vapour (PWV)	Daily	1,000	MYD05
PWV _{TD}	Morning (Terra daytime) PWV	Daily	1,000	MOD05
Δ LST _M	Delta Morning LST: Difference between afternoon and morning LST	Daily	1,000	MYD/MOD11A1
Δ LST _N	Delta Night LST: Difference between afternoon and night-time LST	Daily	1,000	MOD11A1
Δ LST _{DB}	Delta Day Before LST: Difference between afternoon LST and afternoon LST from the day before	Daily	1,000	MOD11A1
Δ PWV _M	Delta Morning PWV: Difference between afternoon and morning PWV	Daily	1,000	MYD/MOD05
Δ PWV _{DB}	Delta Day Before PWV: Difference between afternoon PWV and afternoon PWV from the day before	Daily	1,000	MYD05
NDVI	Normalized Difference Vegetation Index: $(\text{NIR} - \text{Red})/(\text{NIR} + \text{Red})$	Daily	500	MYD09 GA
NDWI	Normalized Difference Water Index: $(\text{NIR} - \text{MIR})/(\text{NIR} + \text{MIR})$	Daily	500	MYD09 GA
%Imp	Percent impervious cover (NLCD classes 21–24)	Static	30	NLCD
%Ag	Percent agriculture cover (NLCD classes 81–82)	Static	30	NLCD
%For	Percent forest cover (NLCD 41–43)	Static	30	NLCD
Elev	Elevation (m)	Static	90	SRTM
Slope	Slope (as percent)	Static	90	SRTM
Lat	Latitude (in decimal degrees – NAD83)	Static	-	-
DC	Distance to nearest coast (includes ocean and Great Lakes)	Static	-	USGS (USGS 2014)

Res. = Spatial resolution of the source data, NIR = near-infrared reflectance (Band 2), Red = red reflectance (Band 1), MIR = mid-infrared reflectance (Band 7), NLCD = 2011 National Land Cover Database, SRTM = Shuttle Radar Topography Mission.

2.2.5. Δ LST(s)

To account for temperature trends over the course of a given day, the difference (or Δ) between afternoon LST and morning (Terra day, MOD11A1), night (Aqua night, MYD11A1), and day before LST (Aqua day-1, MYD11A1) was included. Several authors have used both day and night LST images for T_{max} estimation; others have included observations from both the Terra and Aqua satellites or fit a diurnal curve (see Table 1). Although some have compared different day/night and Aqua/Terra combinations, findings regarding the best group of predictors are inconsistent (Zhang et al. 2011; Zeng et al. 2015a; Phan et al. 2019). Yoo et al. (2018) noted that LST from the day before was critical for estimating T_{max}.

2.2.6. Δ PWV(s)

Similar to LST, the difference between afternoon and morning (Terra day, MOD05) and day before PWV (Aqua day-1, MYD05) was used to account for humidity trends over the course of a day. Night-time PWV was not available, since this product is based on NIR bands.

2.2.7. Normalized Difference Vegetation Index (NDVI)

To account for differences in vegetative cover and evaporation rates across study areas, several authors have included Normalized Difference Vegetation Index (NDVI) or the similar Enhanced Vegetation Index (EVI) for T_{max} and RH estimation (see Table 1). Daily NDVI was computed with the red and NIR bands (1 and 2, respectively) from the Aqua reflectance product (Vermote and W 2015).

2.2.8. Normalized Difference Water Index (NDWI)

Authors also used Normalized Difference Water Index (NDWI) for estimating Tmax and RH, as it tends to be more related to plant water content than NDVI (see [Table 1](#)). We computed daily NDWI with the NIR and MIR bands (2 and 7, respectively) from the Aqua reflectance product (MYD09 GA).

2.2.9. Percentland cover (impervious, forest, agriculture)

As air temperature and humidity are highly dependent on land cover, authors have included percent land cover variables in their models or proximity to certain urban features, such as roads or city centres (see [Table 1](#)). Some fit separate models for each land cover type (Zhang et al. 2011; Rhee and Im 2014; Zeng et al. 2015a). To account for varying land cover, percent impervious (%Imp), forest (%For), and agriculture (%Ag) were computed with the 2011 NLCD (USGS 2017). These cover types are the most commonly considered for Tmax and RH estimation (Zhang et al. 2011; Rhee and Im 2014; Zeng et al. 2015a).

2.2.10. Additional topographic and locational factors

As Tmax and RH vary with respect to elevation, many authors utilized elevation from a digital elevation model (DEM) for estimation (see [Table 1](#)). Some additionally considered slope (Rosenfeld et al. 2017; Serra et al. 2020). To reflect the effects of topography, we incorporated elevation and slope from the Shuttle Radar Topography Mission (SRTM) (Kautz 2017). Latitude (Lat) was used to adjust for warmer temperatures closer to the equator and distance to nearest coast (DC) was included to account for greater humidity closer to large bodies of water. We used USGS's definition for the CONUS coast (USGS 2014), which includes the shorelines of the Great Lakes and oceans. To assign a higher weight to closer stations, natural log of distance was taken.

2.3. Variable geoprocessing and grid alignment

The MODIS 1-km LST grid defined the spatial reference for the HI model. All other variables were resampled to match. After the independent variables were aligned, values for pixels that contained a LCD station were extracted. To account for registration issues, an inverse distance weighted spatial average of valid values was taken using a 3 by 3 window around each station. Invalid values within the window were ignored. However, if the centre pixel was invalid, the average was set to invalid. This was done for LST, PWV, all Δ parameters, Lat, Elev, and Slope. To reflect a larger area of environmental conditions and the landscape around each station, NDVI, NDWI, and land cover variables were calculated with a 5 by 5 unweighted average of valid values. The unweighted average assigned equal weights to all values within a 25-km² area around each station. This approach was based on authors who employed the TVX method with a sliding window size of 5–7 pixels (Zhu, Lú and Jia 2013; Kitsara et al. 2018; Misslin et al. 2018). As with the 3 by 3 window, an average was taken only if the centre pixel was valid; invalid pixels were not used to compute the average.

Observations were then pooled across stations and days (1,395 stations X 153 days = 213,435 observations). Each station had 153 different values for daily variables, which in some cases were invalid due to cloud cover or emissivity error. For static variables (e.g. elevation), stations were assigned the same value for each day.

2.4. HI model development, validation, and selection

Due to the large number of observations and nonlinearity of the HI equation, random forest regression was identified as an appropriate method for model development. Random forest regression is a nonparametric machine learning technique that uses a set of regression trees, each trained with a subset of training data, with a random subset of available predictors used to split the data into each node of each tree (Ho et al. 2016; Yoo et al. 2018). Refer to Breiman (2001) for further details.

Due to its flexibility in areas with complicated and heterogeneous landscapes (Noi, Degener and Kappas 2017; Yoo et al. 2018), random forest regression has been extensively used to estimate air temperature (Xu, Knudby and Ho 2014; Meyer et al. 2016; Zhang et al. 2016; Noi, Degener and Kappas 2017; Sanikhani et al. 2018; Yoo et al. 2018; Zhu et al. 2019; Otgonbayar and Avirmed 2019; Hough et al. 2020; Shen et al. 2020), humidity (Li and Zha 2018), and HUMIDEX (Ho et al. 2016) from MODIS products. Several authors have found random forest to serve as a better predictor of air temperature than ordinary least squares regression (Xu, Knudby and Ho 2014; Meyer et al. 2016; Noi, Degener and Kappas 2017) and other machine learning techniques (Zhang et al. 2016; Zhu et al. 2019; Shen et al. 2020).

To evaluate relative improvements in HI estimation from adding additional parameters, a series of simple to complex models were developed. Starting with LST, variables were included in order of highest expected importance to least, based on previous findings for Tmax, RH, and HUMIDEX models (see Table 1). NIR indices (NDVI and NDWI) were added last, since these variables were the most data-intensive to derive and may already be captured by land cover. To ensure an unbiased comparison, the same observations were used to train and validate each model. These represent instances in which all MODIS parameters were valid ($n = 52,464$).

To avoid using observations from the same station for both training and validation, a modified 10-fold cross-validation approach was employed, such that stations were randomly assigned to 10 groups (see Figure 1). When a group was used for validation, the other nine groups were used for training. The optimal parameters for random forest regression were identified by iteratively testing values and finding the combination that produced the lowest mean square error (MSE). These parameters included the learn rate, maximum number of splits, and minimum number of leaves. Further information on the selected optimal random forest parameters is available in Table A1 of Appendix A.

Model performance was assessed on the basis of the resulting coefficient of determination (R^2) and root mean square error (RMSE). Multicollinearity was evaluated with variable inflation factors (VIFs). The best model was selected on having an ideal trade-off between performance and loss of observations. 'Best', in this context, refers to the most ideal model across a restrictive set of predictor variables (described in Table 2) and does not imply that all possible parameters or combination of parameters were tested. Once the best model was selected, variable importance was further evaluated using the percent increase in mean square error (%IncMSE), similar to Xu, Knudby and Ho

(2014), Ho et al. (2016), Didari and Zand-Parsa (2018), Li and Zha (2018), and Yoo et al. (2018). This metric quantifies the relative increase in model error when each variable is removed.

3. Results and discussion

3.1. Model selection: performance assessment and tradeoffs

To assess the utility of LST for estimating HI and improvements derived from incorporating additional variables, 14 models were fit via random forest regression. The results are summarized in Table 3. For consistency, models were trained and validated on the exact same observations, which had valid values for all MODIS-derived variables ($n = 52,464$). This is different than the number of potential observations (n_p) reported in Table 3, which had no bearing on model development/assessment. The number of potential observations varies by model due to the incorporation of different variables; models with more variables tend to have a lower n_p , since there is a greater chance for clouds to invalidate at least one of the variables.

Adding PWV resulted in the most significant improvement in model performance (Model 2). Including all of the other parameters only increased the R^2 by 0.14 and decreased the RMSE by 1.6°F (Model 14). While HI estimation can be enhanced by incorporating these factors, using just afternoon LST and PWV for a simplified model may be sufficient, depending on the accuracy required for a given application. Missing afternoon observations from the Aqua sensor could also be substituted with morning observations from the Terra sensor to reduce the frequency of invalid data (as indicated by the similar performance of Models 2 and 3).

An important finding from this analysis is that LST alone is a poor predictor of apparent temperature. However, researchers commonly utilize MODIS LST to examine air temperature in the context of human health (Stathopoulou et al. 2005; Klein Rosenthal, Kinney and Metzger 2014; Bao, Li and Yu 2015; Ho, Knudby and Huang 2015; Morabito et al. 2015; Declat-Barreto et al. 2016; Lehoczy et al. 2017; Xu et al. 2017; Chen et al. 2018; Karimi et al. 2018; Méndez-Lázaro et al. 2018; Mushore et al. 2018; Song and Wu 2018; Sun et al. 2018; Valmassoi et al. 2018). Future work in this field could be significantly improved by incorporating our proposed models.

A major disadvantage of using multiple Δ parameters is the loss of usable observations, especially since Model 7 performed only slightly better than the Δ_{morning} (Model 4), Δ_{night} model (Model 5), and $\Delta_{\text{day-before}}$ (Model 6) models. Thus, we recommend that researchers use either Δ_{morning} LST/PWV or Δ_{night} LST variables. Adding land cover, topography, locational, and NIR factors resulted in only a marginal improvement in HI estimation. Considering the extensive data processing required to incorporate daily NDWI and NDVI, these variables can potentially be omitted.

It is important to note that if the primary goal is HI estimation, multicollinearity would not necessarily be an issue. In fact, several researchers developed T_{max} models that utilized multiple LST products with a high degree of correlation (Noi, Kappas and Degener 2016; Didari and Zand-Parsa 2018; Yoo et al. 2018; Zhang et al. 2018; Phan et al. 2019). However, Model 13 was selected as the best model for subsequent error and variable importance assessment, since it had a maximum VIF smaller than 5 and provided

Table 3. Overview of HI model results. Listed are the R², RMSE, maximum VIF, number of potential observations (n_p), and included variables.

Model	R ²	RMSE (°F)	VIF _{max}	n _p	LST _{AD}	LST _{TD}	PWV _{AD}	PWV _{TD}	ΔLST _M	ΔLST _N	ΔLST _{DB}	ΔPWV _M	ΔPWV _{DB}	%Imp	%For	%Ag	Elev	Slope	DC	Lat	NDVI	NDWI
1	0.34	8.4	1.0	117,269	X																	
2	0.70	5.7	1.0	117,237	X		X															
3	0.71	5.6	1.0	120,086		X		X														
4	0.75	5.1	1.3	94,932	X		X		X			X										
5	0.78	4.8	4.5	82,109	X		X			X												
6	0.74	5.3	1.2	76,594	X		X				X		X									
7	0.80	4.6	6.0	52,483	X		X		X	X		X										
8	0.79	4.7	1.8	94,932	X		X		X			X		X	X	X						
9	0.81	4.5	4.9	82,109	X		X			X				X	X	X						
10	0.81	4.5	1.9	94,932	X		X		X			X		X	X	X	X	X	X	X		
11	0.82	4.4	5.5	82,109	X		X			X				X	X	X	X	X	X	X		
12	0.84	4.1	8.8	82,080	X		X			X				X	X	X	X	X	X	X	X	
13	0.83	4.3	3.4	94,891	X		X		X			X		X	X	X	X	X	X	X	X	
14	0.84	4.1	5.7	81,097	X		X			X				X	X	X	X	X	X	X	X	

While the total number of potential variables is listed, each model was trained and validated with the same 52,464 observations. Refer to Table 2 for variable definitions. The total number of potential observations was 213,435.

a good trade-off between performance and loss of observations. In this context, 'best' refers to the most ideal across all 14 models that were tested. The results discussed in the following sections (3.2 and 3.3) were derived from retraining and validating this model with all 94,891 observations available for regression. Refer to Section A3 in Appendix A for a detailed discussion and comparison of the different models.

3.2. Error assessment for the selected model

When the selected model (Model 13) was retrained and validated using all available 94,891 observations via the modified 10-fold cross validation approach, the resulting R^2 and RMSE were 0.83 and 4.4°F, respectively; the maximum VIF was 3.1. Note that these values differ from those specified in Table 3, since all potential observations were used (94,891 vs 52,464). Figure 2 shows a point density plot of estimated versus actual HI values; there is a strong degree of linear association between the two. For extreme HI (<50°F and >110°F), there were some instances of over and under estimation, most likely due to the limited number of observations available for model development in this range. However, the vast majority of observations follow the one-to-one relationship.

Based on the R^2 , Model 13 compared favourably to previous Tmax models and outperformed most RH models (see Table 1). The RMSE was lower than the RMSE for Ho et al. (2016) HUMIDEX model (9°F). While some caution is advised when interpreting direct comparisons, as these analyses utilized different dependent variables and study areas, this indicates that our HI model performs reasonably well, especially when considering that it was validated across a wide range of conditions over the entire CONUS.

Model residuals were normally distributed, demonstrating that there was no bias towards over or under estimation (Figure 3). Note that residuals were derived with estimates from the modified 10-fold cross validation approach. The maximum and

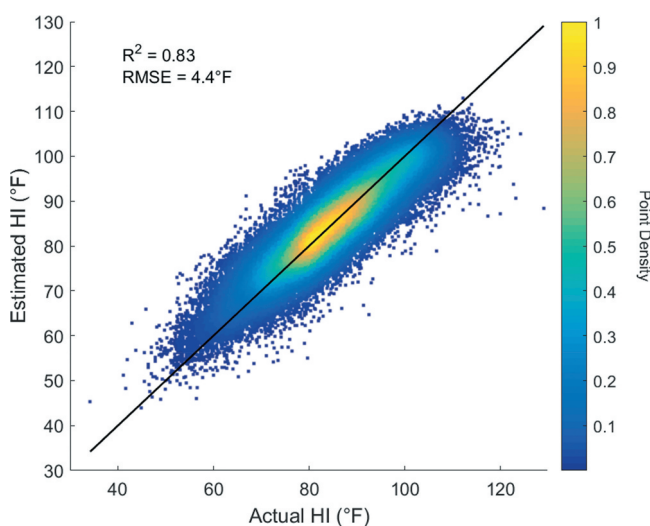


Figure 2. Point density plot of estimated vs. actual HI for the selected model (Model 13). Note: this plot reflects the 94,891 observations that were used to train and validate the model; plot made using function from Henson (2022).

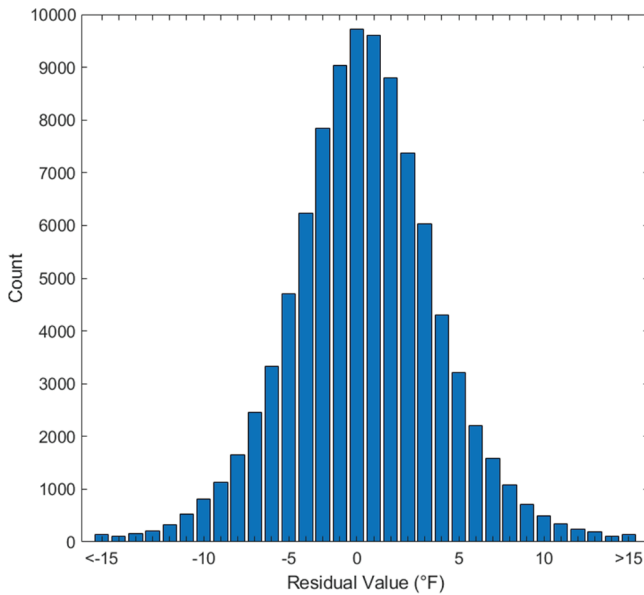


Figure 3. Histogram of residuals for the selected model (Model 13) with all 94,891 potential observations.

minimum residuals were 40.7°F and −24.3°F, respectively. However, the vast majority (77.6%) were within $\pm 5^\circ\text{F}$; <0.1% were outside of $\pm 20^\circ\text{F}$. These unusually high residuals did not appear to be concentrated among a small group of stations, as no station had more than 2. In fact, only 8 stations had more than 1. In addition, they did not appear to be caused by abnormal HI or predictor values. While the model performed very poorly for these observations, they represented a negligible portion of the data ($n = 45$) and are not a cause of major concern.

The residual plot indicated that model error was homogenous (Figure 4). There did not appear to be greater error for higher or lower HI.

RMSEs were computed for each station (across all observations for a given station). Most station RMSEs were $< 5^\circ\text{F}$ (73.5%). There were no major spatial trends in station error across the CONUS (i.e. increasing error north to south, east to west, or closer to the coast) (Figure 5). However, several stations with a RMSE $> 10^\circ\text{F}$ were concentrated in Southern Coastal California. This may be due to the region's later peak in maximum annual temperatures, which occurs in mid-September (US National Ocean and Atmospheric Administration (US NOAA) 2019). Hashimoto et al. (2008) produced a vapour pressure deficit model from MODIS data and similarly found that this area of the CONUS to have a high prediction error.

When comparing the RMSE map of the selected model (Figure 5) to the RMSE map of the LST only model (Figure 6), it is evident that including additional factors considerably improved HI estimation across the CONUS. The improved performance was most apparent for stations along the eastern coast of the US (south of Pennsylvania) and Gulf of Mexico. This was expected though, as these areas are hotter and more humid than the rest of the CONUS. Interestingly, the LST only model performed better in the western portion of states bisected by the 100° Meridian, especially Texas, likely due to the dryer

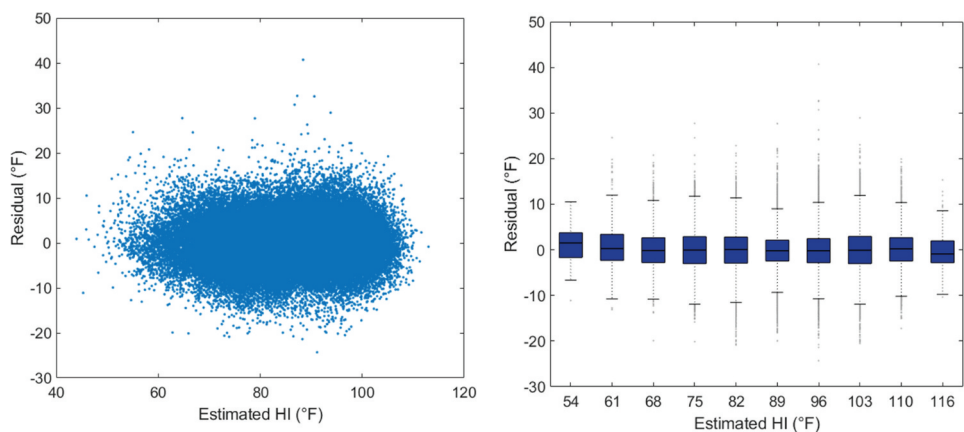


Figure 4. Residual plot (left) and corresponding box plot (right) for the selected model (Model 13). Note: x-axis for the boxplot lists the midpoint for each estimated HI range.

climate. While this analysis did not consider Longitude, subsequent models may benefit by incorporating it to account for drier conditions west for the 100° Meridian. Moreover, it appears that including additional variables to account for atmospheric moisture, land cover, and topography is most important for hot and humid regions of the CONUS.

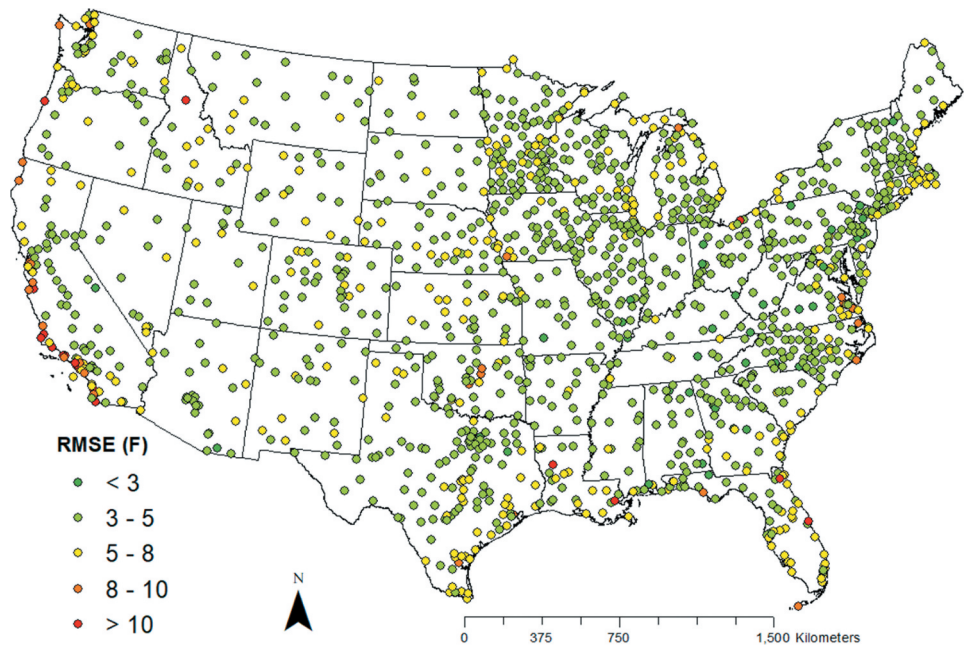


Figure 5. RMSE by station (in °F) for the selected model (Model 13).

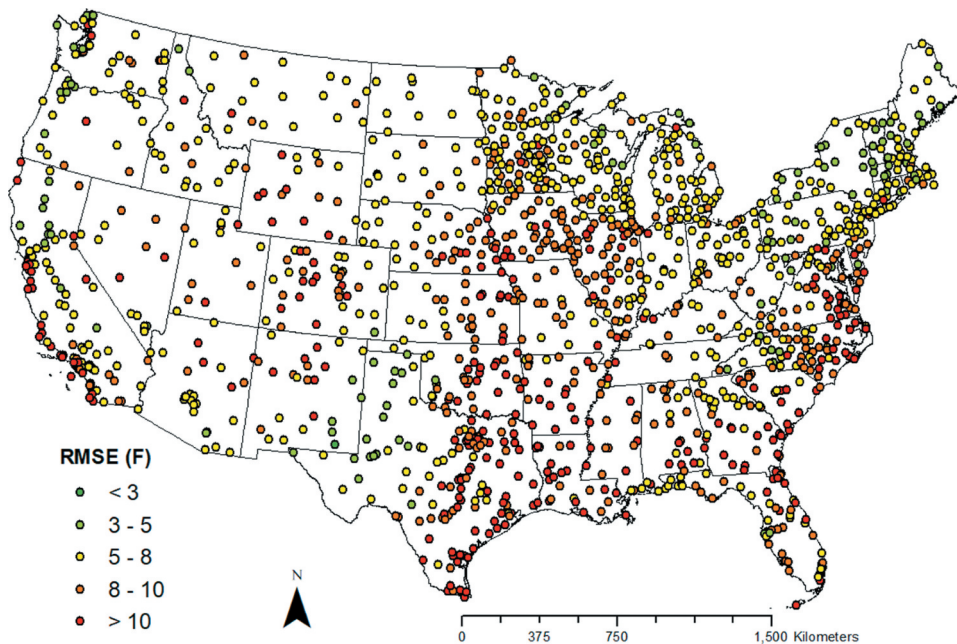


Figure 6. RMSE by station (in °F) for the LST only model (Model 1).

Much of the justification for this paper applies to metropolitan areas, since urban areas encompass >80% of the US population, exacerbate heat waves, and have sharp changes in temperature over relatively short distances due to the urban heat island effect (making spatially explicit temperature datasets especially important). However, the model could be applied to rural locations as well. HI estimates would be of great use to rural communities, since weather stations in these areas are highly sparse. Furthermore, the model could be used to better examine the impact of heat stress on ecosystems. Understanding and predicting biological responses to extreme heating events is critical for effective ecosystem modelling (Jentsch, Kreyling and Beierkuhnlein 2007; Smale and Wernberg 2013). These HI estimates offer complementary insight to climatic pressures exhibited on terrestrial ecosystems, thus leading to improved modelling opportunities.

Another point to make with respect to the random forest modelling capabilities is that when Model 13 was replaced with an ordinary least squares regression, results were close to those derived from random forest ($R^2 = 0.79$ vs. 0.83 , $RMSE = 4.9^\circ F$ vs. $4.4^\circ F$). This is consistent with Meyer et al. (2016), who found only a slight improvement of random forest over ordinary least squares regression when modelling air temperature. Thus, the use of least squares may be sufficient for HI prediction.

3.3. Variable importance assessment for the selected model

The percent increase in mean square error (%IncMSE) was used to further assess the importance of variables included in the selected model (Model 13) (Table 4, first column). LST and PWV were, by far, the most important variables for HI estimation. This corresponds to previous authors who determined that these factors had the highest variable

importance in their respective Tmax (Lin et al. 2012; Xu, Knudby and Ho 2014; Recondo et al. 2013; Noi, Kappas and Degener 2016; Yoo et al. 2018; Hough et al. 2020), RH (Lin et al. 2013; Recondo et al. 2013), and HUMIDEX (Ho et al. 2016) model. The %IncMSE for LST was about 2/3rds larger than the %IncMSE for PWV. However, the inclusion of both LST and PWV is essential, as model error roughly doubled when either was removed (2.3X greater for LST, 1.8X greater for PWV).

NDWI had the next highest importance score, with a %IncMSE just above Δ LST and Δ PWV. Authors have similarly found NIR indices to be the second most important factor behind LST and/or PWV (Kim and Han 2013; Lin et al. 2012; Xu, Knudby and Ho 2014; Recondo et al. 2013; Li and Zha 2018; Yoo et al. 2018); others have found them to have little to no influence on estimation (Lin et al. 2013; Ho et al. 2016; Noi, Kappas and Degener 2016; Didari and Zand-Parsa 2018). Shen et al. (2020) state that LST and NDVI explain the vast majority of variability in Tmax. Considering the number of Tmax models that included multiple LST predictors (see Table 1), it was surprising to find that Δ LST and Δ PWV did not have a higher %IncMSE.

With the exception of elevation (Elev) and percent impervious (%Imp), removing land cover, topographic, and locational factors resulted in virtually no increase in model error. In prior work, these variables were commonly found to have moderate to low importance scores (Lin et al. 2012; Kim and Han 2013; Recondo et al. 2013; Xu, Knudby and Ho 2014; Ho et al. 2016; Noi, Kappas and Degener 2016; Li and Zha 2018; Yoo et al. 2018). Percent impervious (%Imp) was the most important land cover variable, perhaps due to differences in temperature and humidity across urbanization levels (Yang, Ren and Hou 2017; Lokoshchenko 2017; Hao et al. 2018). However, the importance score for all land cover variables was relatively low. In contrast to authors that used a smaller study area (Xu, Knudby and Ho 2014; Ho et al. 2016; Rosenfeld et al. 2017; Li and Zha 2018), distance to nearest coast (DC) did not appear to have any influence on estimation at a continental scale.

Table 4. The percent increase in mean square error (%incmse) of each variable, R^2 , and RMSE for the selected HI model and comparison to similar Tmax and RH models.

	HI	Tmax	RH
R²	0.83	0.83	0.74
RMSE	4.4 (°F)	4.3 (°F)	7.4%
%IncMSE			
LST	126.3	172.9	73.1
Δ LST	11.7	12.9	6.2
PWV	76.0	32.8	44.8
Δ PWV	11.1	9.1	2.1
%Imp	2.0	3.1	4.1
%For	0.9	1.7	1.1
%Ag	0.0	0.0	1.0
Elev	8.6	8.9	25.7
Slope	0.7	1.1	0.0
DC	0.0	0.0	0.0
Lat	0.2	0.5	0.0
NDWI	13.3	13.4	7.3

In general, quantifying %IncMSE for predictors produced results that were consistent with [Section 3.1](#). The most important factors were LST and PWV, Δ parameters and NDWI were moderately important, and land cover, topographic, and locational factors had little to no impact on HI estimation.

To further compare our results to previous studies, maximum air temperature (Tmax) and relative humidity (RH) models were fit with the same parameters via random forest regression ([Table 4](#), last two columns). Consistent with [Recondo et al. \(2013\)](#), the RH model performed worse than the Tmax model. As expected, LST was the most important variable in the Tmax model. However, PWV had the second highest factor, suggesting that the incorporation of PWV is also important for Tmax estimation. This is a significant finding, as only [Lin et al. \(2012\)](#) included PWV in their Tmax model. A major distinction between HI and Tmax is that PWV had a much larger importance factor for HI. When PWV was removed from the Tmax model, the estimation error increased by only 33%. When PWV was removed from the HI model, however, the error nearly doubled. In fact, PWV was just as important for HI estimation as it was for RH. Thus, failing to include PWV will result in model estimates that poorly reflect apparent temperature.

3.4. Limitations

It is important to note several limitations and constraints for the proposed models. As mentioned by [Lin et al. \(2012\)](#), weather stations are not randomly distributed and tend to spatially correlate with population. There was, however, a sufficient spread of both urban and rural stations across the CONUS (see [Figure 1](#)). Most had 10–35% impervious cover and there was decent representation for all levels from 0% to 100% (see [Figure A2](#) in [Appendix A](#)). The large number of utilized stations ($n = 1,395$) encompassed a wide range of conditions for model development.

In its current state, our model can only produce HI estimates with cloud-free observations, which represented 46% of values in the utilized dataset. There are several daily LST interpolation methods that can be used to derive a temporally and spatially continuous dataset ([Neteler 2010](#); [Maffei, Alfieri and Menenti 2012](#); [Alfieri, De Lorenzi and Menenti 2013](#); [Metz, Rocchini and Neteler 2014](#); [Fan et al. 2014](#); [Yu et al. 2015](#); [Zeng et al. 2015b](#); [Shwetha and Kumar 2016](#); [Kang et al. 2018](#)); some were specifically designed for CONUS urban areas ([Li et al. 2018](#); [Long et al. 2020](#)). [Liao et al. \(2020\)](#) estimated RH for cloud-contaminated pixels using MODIS atmospheric profile datasets. To our knowledge, no such method exists for NDWI. An interpolation approach for LST, PWV, and NDWI that is applicable to the entire CONUS could be used to produce a gap-free HI product.

Our HI model is also limited by its 1-km spatial resolution. While this scale is sufficient to examine sub-regional heating trends, aligning HI estimates with socioeconomic data at the Census block-group level may be challenging. A growing body of work has focused on downscaling MODIS LST to a 60-m resolution using Landsat imagery ([Bindhu, Narasimhan and Sudheer 2013](#); [Yoo et al. 2020](#)). As [Ren et al. \(2015\)](#) demonstrated how PWV could be derived from Landsat 8, it may be possible to downscale MODIS PWV and NDWI as well.

While applying interpolation and downscaling methods to the entire CONUS would be data-intensive, a gap-free, daily 60-m HI product could be highly beneficial to future researchers.

4. Conclusions

This analysis aimed to estimate daily maximum 1-km HI across the CONUS using MODIS products in conjunction with ancillary spatial datasets and determine which factors were most important for estimation. We produced a model that was capable of estimating HI with a satisfactory level of error ($R^2 = 0.83$, $RMSE = 4.4^\circ F$). Stations with a relatively high level of error were concentrated in Southern Coastal California, perhaps due to the late peak in maximum summer temperatures. However, the vast majority of observations were estimated within $5^\circ F$. LST and PWV were, by far, the most important variables for HI estimation. The incorporation of additional NIR indices, land cover, topographic, and locational factors resulted in further improvements in model performance ($R^2: +0.14$, $RMSE: -1.6^\circ F$). Applying a simplified model that includes just LST and PWV may be sufficient, depending on the required accuracy. Prior to this analysis, research regarding satellite-based apparent temperature indices had been limited. In comparison to previous Tmax and RH models, our derived HI product is unique in that it was produced at a continental scale and is applicable across a wide range of conditions. An important finding from this study is that the use of LST alone is a poor predictor of apparent temperature. However, LST is widely utilized to examine urban heat island effects in the context of human health. Future work on this topic could be greatly improved by the incorporation of PWV.

Although the scope of this paper was to propose a novel method for estimating HI, our overarching goal is to create a seamless grid product for the entire CONUS in future work. All employed datasets are widely and freely available, thus making product creation feasible. We hope that researchers in multiple study areas, for example in epidemiology, building energy demand, and environmental justice, can use this analytical framework to assess sub-regional heating patterns at a much greater spatial resolution than previously possible. Further work to interpolate cloud-contaminated values and downscale estimates to a 60-m resolution would considerably increase the utility of this HI product. A follow-up study could build off these findings to create a seamless, daily 60-m HI dataset across the entire CONUS from 2001 to 2021.

Acknowledgements

This study was supported by NASA's Land Cover Land Use Change Program (grant #: NNX15AD42G) and a Graduate Assistantship awarded through Environmental Resources Engineering Department at the SUNY College of Environmental Science and Forestry.

Data availability statement

All data utilized for this analysis is publicly available. MODIS products can be downloaded from NASA (<https://modis.gsfc.nasa.gov/data/>). The 2011 National Land Cover Dataset (NLCD) can be obtained from the Multi-Resolution Land Characteristics Consortium (<https://www.mrlc.gov/data>). Elevation

data from the Shuttle Radar Topography Mission (SRTM) can be downloaded from the Earth Resource Observation and Science (EROS) Centre (https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1?qt-science_center_objects=0#qt-science_center_objects). You may contact the authors for additional information.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Giorgos Mountrakis  <http://orcid.org/0000-0001-5958-8134>

References

- Adab, H., K. D. Kanniah, K. Solaimani, and K. P. Tan. (2013). Estimating Atmospheric Humidity Using MODIS Cloud-Free Data in a Temperate Humid Region. *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS Melbourne*, Australia, 1827–1830.
- Ahmed, S. 2018. "Assessment of Urban Heat Islands and Impact of Climate Change on Socioeconomic Over Suez Governorate Using Remote Sensing and GIS Techniques." *The Egyptian Journal of Remote Sensing and Space Science* 21 (1): 15–25. doi:10.1016/j.ejrs.2017.08.001.
- Alfieri, S. M., F. De Lorenzi, and M. Menenti. 2013. "Mapping Air Temperature Using Time Series Analysis of LST: The SINTESI Approach." *Nonlinear Processes in Geophysics* 20 (4): 513–527.
- Auffhammer, M. and E. T. Mansur. 2014. "Measuring Climatic Impacts on Energy Consumption: A Review of the Empirical Literature." *Energy Economics* 46: 522–530. doi:10.1016/j.eneco.2014.04.017.
- Bao, J., X. Li, and C. Yu. 2015. "The Construction and Validation of the Heat Vulnerability Index, a Review." *International Journal of Environmental Research and Public Health* 12 (7): 7220–7234. doi:10.3390/ijerph120707220.
- Benali, A., A. C. Carvalho, J. P. Nunes, N. Carvalhais, and A. Santos. 2012. "Estimating Air Surface Temperature in Portugal Using MODIS LST Data." *Remote Sensing of Environment* 124: 108–121. doi:10.1016/j.rse.2012.04.024.
- Bindhu, V. M., B. Narasimhan, and K. P. Sudheer. 2013. "Development and Verification of a Non-Linear Disaggregation Method (NL-DisTrad) to Downscale MODIS Land Surface Temperature to the Spatial Scale of Landsat Thermal Data to Estimate Evapotranspiration." *Remote Sensing of Environment* 135: 118–129. doi:10.1016/j.rse.2013.03.023.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.
- Chen, Q., M. Ding, X. Yang, K. Hu, and J. Qi. 2018. "Spatially Explicit Assessment of Heat Health Risk by Using Multi-Sensor Remote Sensing Images and Socioeconomic Data in Yangtze River Delta, China." *International Journal of Health Geographics* 17 (1). doi:10.1186/s12942-018-0135-y.
- Chung, J., Y. Lee, W. Jang, S. Lee, and S. Kim. 2020. "Correlation Analysis Between Air Temperature and MODIS Land Surface Temperature and Prediction of Air Temperature Using TensorFlow Long Short-Term Memory for the Period of Occurrence of Cold and Heat Waves." *Remote Sensing* 12 (19): 3231. doi:10.3390/rs12193231.
- Declet-Barreto, J., K. Knowlton, G. D. Jenerette, and A. Buyantuev. 2016. "Effects of Urban Vegetation on Mitigating Exposure of Vulnerable Populations to Excessive Heat in Cleveland, Ohio." *Weather, Climate, and Society* 8 (4): 507–524. doi:10.1175/WCAS-D-15-0026.1.
- Didari, S. and S. Zand-Parsa. 2018. "Enhancing Estimation Accuracy of Daily Maximum, Minimum, and Mean Air Temperature Using Spatio-Temporal Ground-Based and Remote-Sensing Data in Southern Iran." *International Journal of Remote Sensing* 39 (19): 6316–6339. doi:10.1080/01431161.2018.1460500.

- dos Santos, R. S. 2020. "Estimating Spatio-Temporal Air Temperature in London (UK) Using Machine Learning and Earth Observation Satellite Data." *International Journal of Applied Earth Observation and Geoinformation* 88: 102066.
- Emamifar, S., A. Rahimikhoob, and A. A. Noroozi. 2013. "Daily Mean Air Temperature Estimation from MODIS Land Surface Temperature Products Based on M5 Model Tree: Daily Mean Air Temperature Estimation from MODIS." *International Journal of Climatology* 33 (15): 3174–3181. doi:10.1002/joc.3655.
- Fan, X.-M., H.-G. Liu, G.-H. Liu, and S.-B. Li. 2014. "Reconstruction of MODIS Land-Surface Temperature in a Flat Terrain and Fragmented Landscape." *International Journal of Remote Sensing* 35 (23): 7857–7877. doi:10.1080/01431161.2014.978036.
- Hao, L., X. Huang, M. Qin, Y. Liu, W. Li, and G. Sun. 2018. "Ecohydrological Processes Explain Urban Dry Island Effects in a Wet Region, Southern China." *Water Resources Research* 54 (9): 6757–6771. doi:10.1029/2018WR023002.
- Hashimoto, H., J. Dungan, M. White, F. Yang, A. Michaelis, S. Running, and R. Nemani. 2008. "Satellite-Based Estimation of Surface Vapor Pressure Deficits Using MODIS Land Surface Temperature Data." *Remote Sensing of Environment* 112 (1): 142–155. doi:10.1016/j.rse.2007.04.016.
- Henson, R. (2022). "Flow Cytometry Data Reader and Visualization" MATLAB Central File Exchange. Retrieved March 25 2022 from. <https://www.mathworks.com/matlabcentral/fileexchange/8430-flow-cytometry-data-reader-and-visualization>
- Ho, H., A. Knudby, and W. Huang. 2015. "A Spatial Framework to Map Heat Health Risks at Multiple Scales." *International Journal of Environmental Research and Public Health* 12 (12): 16110–16123. doi:10.3390/ijerph121215046.
- Ho, H. C., A. Knudby, Y. Xu, M. Hodul, and M. Aminipouri. 2016. "A Comparison of Urban Heat Islands Mapped Using Skin Temperature, Air Temperature, and Apparent Temperature (Humidex), for the Greater Vancouver Area." *The Science of the Total Environment* 544: 929–938. doi:10.1016/j.scitotenv.2015.12.021.
- Hough, I., A. C. Just, B. Zhou, M. Dorman, J. Lepeule, and I. Kloog. 2020. "A Multi-Resolution Air Temperature Model for France from MODIS and Landsat Thermal Data." *Environmental Research* 183: 109244. doi:10.1016/j.envres.2020.109244.
- Jentsch, A., J. Kreyling, and C. Beierkuhnlein. 2007. "A New Generation of Climate-Change Experiments: Events, Not Trends." *Frontiers in Ecology and the Environment* 5 365–374.
- Kang, J., J. Tan, R. Jin, X. Li, and Y. Zhang. 2018. "Reconstruction of MODIS Land Surface Temperature Products Based on Multi-Temporal Information." *Remote Sensing* 10 (7): 1112. doi:10.3390/rs10071112.
- Karimi, M., R. Nazari, D. Dutova, R. Khanbilvardi, and M. Ghandehari. 2018. "A Conceptual Framework for Environmental Risk and Social Vulnerability Assessment in Complex Urban Settings." *Urban Climate* 26: 161–173. doi:10.1016/j.uclim.2018.08.005.
- Kautz, S. 2017. "Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global [Data Set]" doi:10.5066/F7PR7TFT
- Kim, D.-Y. and K.-S. Han. 2013. "Remotely Sensed Retrieval of Midday Air Temperature Considering Atmospheric and Surface Moisture Conditions." *International Journal of Remote Sensing* 34 (1): 247–263. doi:10.1080/01431161.2012.712235.
- Kitsara, G., G. Papaioannou, A. Retalis, D. Paronis, and P. Kerkides. 2018. "Estimation of Air Temperature and Reference Evapotranspiration Using MODIS Land Surface Temperature Over Greece." *International Journal of Remote Sensing* 39 (3): 924–948. doi:10.1080/01431161.2017.1395965.
- Klein Rosenthal, J., P. L. Kinney, and K. B. Metzger. 2014. "Intra-Urban Vulnerability to Heat-Related Mortality in New York City, 1997–2006." *Health & Place* 30: 45–60. doi:10.1016/j.healthplace.2014.07.014.
- Lehoczy, A., J. Sobrino, D. Skoković, and E. Aguilar. 2017. "The Urban Heat Island Effect in the City of Valencia: A Case Study for Hot Summer Days." *Urban Science* 1 (1): 9. doi:10.3390/urbansci1010009.

- Li, L. and Y. Zha. 2018. "Mapping Relative Humidity, Average and Extreme Temperature in Hot Summer Over China." *The Science of the Total Environment* 615: 875–881. doi:[10.1016/j.scitotenv.2017.10.022](https://doi.org/10.1016/j.scitotenv.2017.10.022).
- Li, X., Y. Zhou, G. R. Asrar, and Z. Zhu. 2018. "Creating a Seamless 1 Km Resolution Daily Land Surface Temperature Dataset for Urban and Surrounding Areas in the Conterminous United States." *Remote Sensing of Environment* 206 (1): 84–97. doi:[10.1016/j.rse.2017.12.010](https://doi.org/10.1016/j.rse.2017.12.010).
- Liao, Q.-Y., P. Leng, Z.-L. Li, C. Ren, Y.-Y. Sun, M.-F. Gao, S.-B. Duan, and G.-F. Shang (2020). A Method for Deriving Relative Humidity from MODIS Data Under All-Sky Conditions *IEEE Transactions on Geoscience and Remote Sensing* 59 . , 8992–9006 doi:[10.1109/TGRS.2020.3036248](https://doi.org/10.1109/TGRS.2020.3036248).
- Lim, J. and G. Yun. 2017. "Cooling Energy Implications of Occupant Factor in Buildings Under Climate Change." *Sustainability* 9 (11): 2039. doi:[10.3390/su9112039](https://doi.org/10.3390/su9112039).
- Lin, S., N. Moore, J. Messina, M. DeVisser, and J. Xu. 2012. "Evaluation of Estimating Daily Maximum and Minimum Air Temperature with MODIS Data in East Africa." *International Journal of Applied Earth Observation and Geoinformation* 18: 128–140.
- Lin, S., N. J. Moore, J. P. Messina, and J. Wu. 2013. "Evaluation of MODIS Surrogates for Meteorological Humidity Data in East Africa." *International Journal of Remote Sensing* 34 (13): 4669–4679. doi:[10.1080/01431161.2013.781288](https://doi.org/10.1080/01431161.2013.781288).
- Lokoshchenko, M. A. 2017. "Urban Heat Island and Urban Dry Island in Moscow and Their Centennial Changes." *Journal of Applied Meteorology and Climatology* 56 (10): 2729–2745. doi:[10.1175/JAMC-D-16-0383.1](https://doi.org/10.1175/JAMC-D-16-0383.1).
- Long, D., L. Yan, L. Bai, C. Zhang, X. Li, H. Lei, H. Yang, et al. 2020. "Generation of MODIS-Like Land Surface Temperatures Under All-Weather Conditions Based on a Data Fusion Approach." *Remote Sensing of Environment* 246: 111863. doi:[10.1016/j.rse.2020.111863](https://doi.org/10.1016/j.rse.2020.111863).
- Maffei, C., S. Alfieri, and M. Menenti (2012). Characterizing Fire Hazard from Temporal Sequences of Thermal Infrared MODIS Measurements. *1st EARSel Workshop on Temporal Analysis of Satellite Images* Mykonos, Greece.
- Maia-Silva, D., R. Kumar, and R. Nateghi. 2020. "The Critical Role of Humidity in Modeling Summer Electricity Demand Across the United States." *Nature Communications* 11 (1). doi:[10.1038/s41467-020-15393-8](https://doi.org/10.1038/s41467-020-15393-8).
- Meehl, G. A. 2004. "More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century." *Science* 305 (5686): 994–997.
- Méndez-Lázaro, P., F. E. Muller-Karger, D. Otis, M. J. McCarthy, and E. Rodríguez. 2018. "A Heat Vulnerability Index to Improve Urban Public Health Management in San Juan, Puerto Rico." *International Journal of Biometeorology* 62 (5): 709–722. doi:[10.1007/s00484-017-1319-z](https://doi.org/10.1007/s00484-017-1319-z).
- Metz, M., D. Rocchini, and M. Neteler. 2014. "Surface Temperatures at the Continental Scale: Tracking Changes with Remote Sensing at Unprecedented Detail." *Remote Sensing* 6 (5): 3822–3840. doi:[10.3390/rs6053822](https://doi.org/10.3390/rs6053822).
- Meyer, H., M. Katurji, T. Appelhans, M. Müller, T. Naus, P. Roudier, and P. Zawar-Reza. 2016. "Mapping Daily Air Temperature for Antarctica Based on MODIS LST." *Remote Sensing* 8 (9): 732. doi:[10.3390/rs8090732](https://doi.org/10.3390/rs8090732).
- Misslin, R., Y. Vaguet, A. Vaguet, and É. Daudé. 2018. "Estimating Air Temperature Using MODIS Surface Temperature Images for Assessing Aedes Aegypti Thermal Niche in Bangkok, Thailand." *Environmental Monitoring and Assessment* 190 (9). doi:[10.1007/s10661-018-6875-0](https://doi.org/10.1007/s10661-018-6875-0).
- Mitchell, B. C. and J. Chakraborty. 2018. "Exploring the Relationship Between Residential Segregation and Thermal Inequity in 20 U.S. Cities." *Local Environment* 23 (8): 796–813.
- MODIS Science Team. (2014). "MYD05_L2 Modis/aqua Total Precipitable Water Vapor 5-Min L2 Swath 1km and 5km [Data Set]" [10.5067/MODIS/MYD05_L2.006](https://doi.org/10.5067/MODIS/MYD05_L2.006)
- Morabito, M., A. Crisci, B. Gioli, G. Gualtieri, P. Toscano, V. Di Stefano, and G. F. Gensini. 2015. "Urban-Hazard Risk Analysis: Mapping of Heat-Related Risks in the Elderly in Major Italian Cities." *Plos One* 10 (5): e0127277.
- Mushore, T. D., J. Odindi, T. Dube, and O. Mutanga. 2018. "Outdoor Thermal Discomfort Analysis in Harare, Zimbabwe in Southern Africa." *South African Geographical Journal* 100 (2): 162–179. doi:[10.1080/03736245.2017.1339630](https://doi.org/10.1080/03736245.2017.1339630).

- National Oceanic and Atmospheric Administration (NOAA). (2019). "Warmest Day of the Year Map" Accessed on May 2 2019 from. <https://www.ncdc.noaa.gov/file/us-warmest-day-year-map.jpg>
- Neteler, M. 2010. "Estimating Daily Land Surface Temperatures in Mountainous Environments by Reconstructed MODIS LST Data." *Remote Sensing* 2 (1): 333–351. doi:10.3390/rs1020333.
- Noi, P., M. Kappas, and J. Degener. 2016. "Estimating Daily Maximum and Minimum Land Air Surface Temperature Using MODIS Land Surface Temperature Data and Ground Truth Data in Northern Vietnam." *Remote Sensing* 8 (12): 1002. doi:10.3390/rs8121002.
- Noi, P., J. Degener, and M. Kappas. 2017. "Comparison of Multiple Linear Regression, Cubist Regression, and Random Forest Algorithms to Estimate Daily Air Surface Temperature from Dynamic Combinations of MODIS LST Data." *Remote Sensing* 9 (5): 398. doi:10.3390/rs9050398.
- Ortiz, L., J. E. González, and W. Lin. 2018. "Climate Change Impacts on Peak Building Cooling Energy Demand in a Coastal Megacity." *Environmental Research Letters* 13(9): 094008. doi:10.1088/1748-9326/aad8d0.
- Otgonbayar, A. M. and E. Avirmed. 2019. "Estimation of Climatologies of Average Monthly Air Temperature Over Mongolia Using MODIS Land Surface Temperature (LST) Time Series and Machine Learning Techniques." *Remote Sensing* 11 (21): 2588.
- Peng, G., J. Li, Y. Chen, A. P. Norizan, and L. Tay. 2006. "High-Resolution Surface Relative Humidity Computation Using MODIS Image in Peninsular Malaysia." *Chinese Geographical Science* 16 (3): 260–264. doi:10.1007/s11769-006-0260-6.
- Phan, T. N., M. Kappas, K. T. Nguyen, T. P. Tran, Q. V. Tran, and A. R. Emam. 2019. "Evaluation of MODIS Land Surface Temperature Products for Daily Air Surface Temperature Estimation in Northwest Vietnam." *International Journal of Remote Sensing* 40(14) 1–19 doi:10.1080/01431161.2019.1580789.
- Ramírez-Beltrán, N. D., C. M. Salazar, J. M. Castro Sánchez, and J. E. González. 2019. "A Satellite Algorithm for Estimating Relative Humidity, Based on GOES and MODIS Satellite Data." *International Journal of Remote Sensing* 40 (24): 9237–9259. doi:10.1080/01431161.2019.1629715.
- Recondo, C., J. J. Peón, E. Zapico, and E. Pendás. 2013. "Empirical Models for Estimating Daily Surface Water Vapour Pressure, Air Temperature, and Humidity Using MODIS and Spatiotemporal Variables. Applications to Peninsular Spain." *International Journal of Remote Sensing* 34 (22): 8051–8080. doi:10.1080/01431161.2013.828185.
- Ren, H., C. Du, R. Liu, Q. Qin, G. Yan, Z.-L. Li, and J. Meng. 2015. "Atmospheric Water Vapor Retrieval from Landsat 8 Thermal Infrared Images: Water Vapor Estimate from L8." *Journal of Geophysical Research: Atmospheres* 120 (5): 1723–1738. doi:10.1002/2014JD022619.
- Rhee, J. and J. Im. 2014. "Estimating High Spatial Resolution Air Temperature for Regions with Limited in situ Data Using MODIS Products." *Remote Sensing* 6 (8): 7360–7378. doi:10.3390/rs6087360.
- Rippey, B. R. 2015. "The U.S. Drought of 2012." *Weather and Climate Extremes* 10: 57–64. doi:10.1016/j.wace.2015.10.004.
- Rosenfeld, A., M. Dorman, J. Schwartz, V. Novack, A. C. Just, and I. Kloog. 2017. "Estimating Daily Minimum, Maximum, and Mean Near Surface Air Temperature Using Hybrid Satellite Models Across Israel." *Environmental Research* 159: 297–312. doi:10.1016/j.envres.2017.08.017.
- Rothsfuz, L. P. (1990). National Weather Service. the Heat Index Equation: Or, More Than You Ever Wanted to Know About the Heat Index Equation. Western Technical Attachment, 9024. Fort Worth, TX.
- Sabrin, S., M. Karimi, and R. Nazari. 2020. "Developing Vulnerability Index to Quantify Urban Heat Islands Effects Coupled with Air Pollution: A Case Study of Camden, NJ." *ISPRS International Journal of Geo-Information* 9 (6): 349. doi:10.3390/ijgi9060349.
- Sanikhani, H., R. C. Deo, P. Samui, O. Kisi, C. Mert, R. Mirabbasi . . . , and Z. M. Yaseen. 2018. "Survey of Different Data-Intelligent Modeling Strategies for Forecasting Air Temperature Using Geographic Information as Model Predictors." *Computers and Electronics in Agriculture* 152: 242–260.
- Serra, C., X. Lana, M. D. Martínez, J. Roca, B. Arellano, R. Biere, M. Moix, and A. Burgueño. 2020. "Air Temperature in Barcelona Metropolitan Region from MODIS Satellite and GIS Data." *Theoretical and Applied Climatology* 139 (1–2): 473–492. doi:10.1007/s00704-019-02973-y.

- Shen, H., Y. Jiang, T. Li, Q. Cheng, C. Zeng, and L. Zhang. 2020. "Deep Learning-Based Air Temperature Mapping by Fusing Remote Sensing, Station, Simulation and Socioeconomic Data." *Remote Sensing of Environment* 240: 111692. doi:10.1016/j.rse.2020.111692.
- Shi, Y., Z. Jiang, L. Dong, and S. Shen. 2017. "Statistical Estimation of High-Resolution Surface Air Temperature from MODIS Over the Yangtze River Delta, China." *Journal of Meteorological Research* 31 (2): 448–454. doi:10.1007/s13351-017-6073-y.
- Shwetha, H. R. and D. N. Kumar. 2016. "Prediction of High Spatio-Temporal Resolution Land Surface Temperature Under Cloudy Conditions Using Microwave Vegetation Index and ANN." *ISPRS Journal of Photogrammetry and Remote Sensing* 117: 40–55. doi:10.1016/j.isprsjprs.2016.03.011.
- Smale, D. A. and T. Wernberg (2013). Extreme Climatic Event Drives Range Contraction of a Habitat-Forming Species Proceedings of the Royal Society B . , 280(1754), 20122829 doi:10.1098/rspb.2012.2829.
- Sofan, P., T. Sugiharto, and G. Hasnaeni (2010). Relative Humidity Estimation Based on Modis Precipitable Water for Supporting Spatial Information Over Java Island. *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, 4(1).
- Song, Y. and C. Wu. 2018. "Examining Human Heat Stress with Remote Sensing Technology." *GIScience & Remote Sensing* 55 (1): 19–37. doi:10.1080/15481603.2017.1354804.
- Stathopoulou, M. I., C. Cartalis, I. Keramitsoglou, and M. Santamouris (2005 Proceedings of SPIE - The International Society for Optical Engineering). *Thermal Remote Sensing of Thom's Discomfort Index (DI): Comparison with in-Situ Measurements* 5983 131–139 doi:10.1117/12.627541 . (edited by M. Ehlers and U. Michel).
- Steadman, R. G. 1979. "The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science." *Journal of Applied Meteorology* 18 (7): 861–873. doi:10.1175/1520-0450(1979)018<0861:TAOSPI>2.0.CO;2.
- Sun, Y., C. Gao, J. Li, W. Li, and R. Ma. 2018. "Examining Urban Thermal Environment Dynamics and Relations to Biophysical Composition and Configuration and Socio-Economic Factors: A Case Study of the Shanghai Metropolitan Region." *Sustainable Cities and Society* 40: 284–295. doi:10.1016/j.scs.2017.12.004.
- US Centre for Disease Control and Prevention (CDC). (2013). "Heat-Related Illness" Accessed on March 25 2019 from. https://www.cdc.gov/pictureofamerica/pdfs/picture_of_america_heat-related_illness.pdf
- US Geological Survey (USGS). (2014). *USGS Small-scale Dataset - 1:1,000,000-Scale Coastline of the United States 201403 Shapefile*. Accessed on January 10 2019 from. <https://www.sciencebase.gov/catalog/item/581d051de4b08da350d523be>
- US Geological Survey (USGS). (2017). "National Land Cover Database 2011, Product Legend" Accessed on March 2nd 2017 from. https://www.mrlc.gov/nlcd11_leg.php
- US National Ocean and Atmospheric Administration (US NOAA). (2019). "National Centers for Environmental Information" *Local Climatology Dataset (LCD)*. Accessed on March 25th from. <https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/quality-controlled-local-climatological-data-qclcd>
- US National Weather Service (US NWS). (2014). "Weather Prediction Center. the Heat Index Equation" Accessed on March 25 2019 from. https://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml
- US National Weather Service (US NWS). (2019). "What is the Heat Index?" Accessed on March 25 2019 from. <https://www.weather.gov/ama/heatindex>
- Valmassoi, A., S. Gharbia, S. Di Sabatino, P. Kumar, and F. Pilla. 2018. "Future Impacts of the Reforestation Policy on the Atmospheric Parameters in Ireland: A Sensitivity Study Including Heat Discomfort Impacts on Humans and Livestock." *Personal and Ubiquitous Computing* 23 .
- Vermote, E. and R. W. 2015. "MYD09GA Modis/aqua Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V006." doi:10.5067/MODIS/MYD09GA.006.
- Wan, Z. (2015). *MYD11A1 MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006*.

- Wang, W., S. Liang, and T. Meyers. 2008. "Validating MODIS Land Surface Temperature Products Using Long-Term Nighttime Ground Measurements." *Remote Sensing of Environment* 112 (3): 623–635. doi:[10.1016/j.rse.2007.05.024](https://doi.org/10.1016/j.rse.2007.05.024).
- Wong, M. S., X. Jin, Z. Liu, J. Nichol, and P. W. Chan. 2015. "Multi-Sensors Study of Precipitable Water Vapour Over Mainland China." *International Journal of Climatology* 35 (10): 3146–3159. doi:[10.1002/joc.4199](https://doi.org/10.1002/joc.4199).
- Xu, Y., A. Knudby, and H. C. Ho. 2014. "Estimating Daily Maximum Air Temperature from MODIS in British Columbia, Canada." *International Journal of Remote Sensing* 35 (24): 8108–8121. doi:[10.1080/01431161.2014.978957](https://doi.org/10.1080/01431161.2014.978957).
- Xu, H., X. Hu, H. Guan, and G. He. 2017. "Development of a Fine-Scale Discomfort Index Map and Its Application in Measuring Living Environments Using Remotely Sensed Thermal Infrared Imagery." *Energy and Buildings* 150: 598–607.
- Yang, P., G. Ren, and W. Hou. 2017. "Temporal-spatial Patterns of Relative Humidity and the Urban Dryness Island Effect in Beijing City." *Journal of Applied Meteorology and Climatology* 56 (8): 2221–2237. doi:[10.1175/JAMC-D-16-0338.1](https://doi.org/10.1175/JAMC-D-16-0338.1).
- Yoo, C., J. Im, S. Park, and L. J. Quackenbush. 2018. "Estimation of Daily Maximum and Minimum Air Temperatures in Urban Landscapes Using MODIS Time Series Satellite Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 137: 149–162. doi:[10.1016/j.isprsjprs.2018.01.018](https://doi.org/10.1016/j.isprsjprs.2018.01.018).
- Yoo, C., J. Im, S. Park, and D. Cho. 2020. "Spatial Downscaling of MODIS Land Surface Temperature: Recent Research Trends, Challenges, and Future Directions." *대한원격탐사학회지* 36 (4): 609–626.
- Yu, W., Z. Nan, Z. Wang, H. Chen, T. Wu, and L. Zhao. 2015. "An Effective Interpolation Method for MODIS Land Surface Temperature on the Qinghai-tibet Plateau." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (9): 4539–4550. doi:[10.1109/JSTARS.2015.2464094](https://doi.org/10.1109/JSTARS.2015.2464094).
- Zeng, L., B. Wardlow, T. Tadesse, J. Shan, M. Hayes, D. Li, and D. Xiang. 2015a. "Estimation of Daily Air Temperature Based on MODIS Land Surface Temperature Products Over the Corn Belt in the US." *Remote Sensing* 7 (1): 951–970.
- Zeng, C., H. Shen, M. Zhong, L. Zhang, and P. Wu. 2015b. "Reconstructing MODIS LST Based on Multitemporal Classification and Robust Regression." *IEEE Geoscience and Remote Sensing Letters* 12 (3): 512–516. doi:[10.1109/LGRS.2014.2348651](https://doi.org/10.1109/LGRS.2014.2348651).
- Zhang, W., Y. Huang, Y. Yu, and W. Sun. 2011. "Empirical Models for Estimating Daily Maximum, Minimum and Mean Air Temperatures with MODIS Land Surface Temperatures." *International Journal of Remote Sensing* 32 (24): 9415–9440. doi:[10.1080/01431161.2011.560622](https://doi.org/10.1080/01431161.2011.560622).
- Zhang, K., Y.-H. Chen, J. D. Schwartz, R. B. Rood, and M. S. O'Neill. 2014. "Using Forecast and Observed Weather Data to Assess Performance of Forecast Products in Identifying Heat Waves and Estimating Heat Wave Effects on Mortality." *Environmental Health Perspectives* 122 (9): 912–918. doi:[10.1289/ehp.1306858](https://doi.org/10.1289/ehp.1306858).
- Zhang, H., F. Zhang, M. Ye, T. Che, and G. Zhang. 2016. "Estimating Daily Air Temperatures Over the Tibetan Plateau by Dynamically Integrating MODIS LST Data." *Journal of Geophysical Research: Atmospheres* 121 (19): 11–425.
- Zhang, H., F. Zhang, G. Zhang, Y. Ma, K. Yang, and M. Ye. 2018. "Daily Air Temperature Estimation on Glacier Surfaces in the Tibetan Plateau Using MODIS LST Data." *Journal of Glaciology* 64 (243): 132–147. doi:[10.1017/jog.2018.6](https://doi.org/10.1017/jog.2018.6).
- Zhu, W., A. Lü, and S. Jia. 2013. "Estimation of Daily Maximum and Minimum Air Temperature Using MODIS Land Surface Temperature Products." *Remote Sensing of Environment* 130: 62–73. doi:[10.1016/j.rse.2012.10.034](https://doi.org/10.1016/j.rse.2012.10.034).
- Zhu, X., Q. Zhang, C.-Y. Xu, P. Sun, and P. Hu. 2019. "Reconstruction of High Spatial Resolution Surface Air Temperature Data Across China: A New Geo-Intelligent Multisource Data-Based Machine Learning Technique." *Science of the Total Environment* 665: 300–313.