Estimating Heat Storage in Urban Areas Using Multispectral Satellite Data and Machine Learning

Joshua Hrisko [1,*], Prathap Ramamurthy [1], Jorge E. Gonzalez[1]

[1] *Department of Mechanical Engineering and NOAA-CESSRST center, City College of New York, New York, NY, 10031, USA*

[*] *jhrisko000@citymail.cuny.edu*

Abstract

A satellite-derived hysteresis model is presented for estimate heat storage in urban areas. Storage heat flux, one of the dominant terms in the urban surface energy budget (USEB), is largely unknown despite its critical relationship to various urban environmental processes. This study introduces a novel technique for quantifying heat storage by relating multispectral satellite radiances and geophysical properties to ground-truth residual heat storage computed with flux instruments. Gradient-boosted regression trees serve as the method of maximizing the relationship between satellite data and flux measurements. Several flux networks are used to train and validate the model over varying land cover types, which strengthens the robustness of the model. The model performs well under variable weather conditions such as cloudy rainy days. In comparison with other studies, the RMSE and MAE values were found to be lower than some ground-to-ground studies, and is one of few satellite-derived methods that computes direct comparison over a range of different land cover types.

Keywords: Heat Storage, GOES-16, Machine Learning, Radiance, Heat

Flux, Urban, Satellite Remote Sensing, GBRT

1. Introduction and Background

Heat storage has been cited as a major contributor to the urban heat island phenomenon due to increased thermal conductivity and heat capacity of impervious surfaces in cities (Grimmond et al., 1991; Oke, 1988; Ramamurthy and Bou-Zeid, 2017; Roth and Oke, 1994; Taha, 1999). Several studies conclude that heat storage (ΔQ_s) is one of the dominant terms in the urban surface energy budget, in some cases amounting to 40% or more of the net radiation (Bonacquisti et al., 2006; Coutts et al., 2007; Grimmond and Oke, 2002; Offerle et al., 2006; Oke et al., 1999). Heat storage is also significant as a proxy for other fluxes. For example, anthropogenic heat may be difficult to measure, but it can be derived using energy balance closure if each of the other terms has been measured or calculated (Nitis et al., 2017; Offerle et al., 2005; Olivo et al., 2017; Wilson et al., 2002). Despite its significance, there is no standard for calculating ΔQ_s; instead, five common methods can be found scattered throughout the literature: the energy balance residual method (RES), the objective hysteresis model (OHM), the thermal mass scheme (TMS), the town energy balance (TEB), and the element surface temperature method (ESTM) (Chrysoulakis et al., 2018; Grimmond and Oke, 1999; Kerschgens and Kraus, 1990; Lindberg et al., 2020; Lipson et al., 2017; Masson, 2000; Oke and Cleugh, 1987). There are varying degrees of agreement between the different methods, as outlined in the Roberts et al. (2006) article, which reinforces the claims that there is no standard routine for measuring or calculating heat storage.

Much of the progress associated with urban heat storage has been limited to sparsely distributed eddy covariance instruments mounted on flux towers (Nordbo et al., 2012). Flux towers are great tools for measuring accurate fluxes within a given footprint, but they also give rise to incomplete spatial representations over heterogeneous areas due to the separation between sites (Feigenwinter et al., 2018; Kanda et al., 2006; Ramamurthy and Pardyjak, 2011). Some satellite methods remedy this by combining satellite-derived surface temperatures with NDVI-based relationships (Kato and Yamaguchi, 2005a; Parlow, 2003) that are capable of representing the urban form, however, nearly all are time-restricted by satellite overpass periods (Kato and Yamaguchi, 2007; Rigo and Parlow, 2007; Tsuang, 2005). This results in poor statistical significance and sparse diurnal distribution of data points, giving an incomplete picture of spatially-distributed urban heat storage. Fortunately, with the release of two new state-of-the-art geostationary satellites (GOES-16 and GOES-17) from the National Oceanic and Atmospheric Administration (NOAA), the time between satellite observations has finally become competitive with ground-based instruments (Schmit and Gunshor, 2020).

In the present study, two methods are used to calculate heat storage: the residual method (RES) and a satellite-derived hysteresis model. The residual method acts as a ground-truth training and validation tool for the hysteresis model, something that is ordinarily done for the objective hysteresis model (OHM) proposed in the literature (Arnfield and Grimmond, 1998; Meyn and Oke, 2009; Pearlmutter et al., 2005). The GOES-16 satellite radiance data serves as the primary input variable to the model, while land cover and geography-specific properties act as peripheral inputs to characterize each satellite pixel. The inclusion of satellite radiance avoids many of the shortcomings associated with applying the traditional objective hysteresis model (OHM) to satellite data, such as the non-Lambertianity of urban materials and the difficulty in quantifying the temporal hysteresis of net radiation (Herold et al., 2004; Roberts et al., 2012).

Gradient-boosted regression trees (GBRTs) are are used to statistically train and validate the satellite and geographic inputs against ground residual heat storage. Similar machine learning algorithms have been broadly demonstrated as approaches to correlating multivariate systems in Earth-atmosphere interactions (Camps-Valls, 2009; DeFries and Chan, 2000; Lary et al., 2016; Novack et al., 2011; Yoo et al., 2018). Similar methods have been implemented for ground-to-satellite relationships involving aerosols (Just et al., 2018), shortwave radiation (Yang et al., 2018b), water vapor (Just et al., 2019; Lee et al., 2019), soil moisture (Wei et al., 2019), among others.

One of the major challenges facing the development of satellite energy balance estimates in urban areas is the lack of validation points. As a way to combat this, the NYS Mesonet (Mesonet, 2020), National Ecological Observatory Network (NEON) (Network, 2020b), and Ameriflux (Ameriflux, 2020) networks all serve as ground-based instruments used for training and validating the satellite hysteresis algorithm. The GBRT model uses the wide range of geophysical properties from ground stations to develop an accurate and versatile characterization of heat storage in cities. The proposed method overcomes many of the issues plaguing satellite algorithms such as dropped pixels during cloudy periods, inaccurate material properties of urban surfaces, and statistically insignificant analyses. And it does so by approaching the problem from a multi-network, statistical viewpoint. The benefits of this method will be discussed in great detail during the presentation of results.

First, this paper outlines the methods used in quantifying heat storage in urban areas using satellite data. This includes description of heat storage and how it is traditionally measured and modeled with the hysteresis approach, presentation of the GBRT algorithm, and methods for downscaling the resulting product. Then, New York City is introduced as the test area for the summer of 2019 , where summertime was chosen as a particular point of interest because of the potential for heat storage to provide information about extreme heat events and urban heat island phenomena (Golden, 2004; Sailor, 2014; Zhou and Shepherd, 2010). By the end of this study, a diurnal urban heat storage product will be posited, implemented, and validated. As a consequence of using all 16 bands of the GOES-16 satellite (wavelengths from 0.47µm - 13.3µm), all-weather periods are also captured, including clouds and precipitation - which are often caveats when developing satellite-based algorithms (Chrysoulakis et al., 2018; Middel et al., 2012).

Several methods proposed here are new and novel, particularly in relation to the temporal resolution of the satellite radiance data. The multispectral satellite hysteresis model challenges the status quo in surface energy budget estimates by narrowing the temporal capabilities of a widely recognized deficiency in the research area, thus, resulting in an application of heat storage that has potential implications in weather modeling, predictions of energy use, and the partitioning of energy across heat fluxes in urban environments.

2. Methodology

2.1. Satellite Hysteresis Model

The temporal hysteresis between net radiation and heat storage has been widely cited (Anandakumar, 1999; Grimmond et al., 1991; Järvi et al., 2014; Roth and Oke, 1994; Sun et al., 2013; Wang, 2014). Unfortunately, nearly all studies that employ the objective hysteresis model (OHM) use ground-based correlations to approximate heat storage. This leads to issues in spatial resolution, as many towers are located far from one another. For the few studies that employ the OHM with satellite data, they first derive net radiation and then implement land cover-based coefficients from the literature (Rigo and Parlow, 2007). The direct implementation of the OHM to satellite data is highly inaccurate due to uncertainty associated with quantifying net radiation. Satellite-derived net radiation can carry errors greater than the largest errors given in heat storage studies, making the direct implementation of the traditional OHM to satellite data unfavorable. The proposed multispectral method instead uses satellite radiances to avoid amounting errors associated with first deriving net radiation and OHM coefficients.

The traditional objective hysteresis model (OHM) was first developed by Camuffo and Bernardi (1982) and is often attributed to Grimmond et al. (1991), which is stated below for a given surface type:

$$\Delta Q_s = a_1 Q^* + a_2 \frac{dQ^*}{dt} + a_3 \quad (1)$$

where a_1, a_2, a_3 represent coefficients relating to different land cover types (i.e. urban, forest, crop), Q^* is net radiation, and the derivative is typically computed on an hourly basis.

The satellite-derived hysteresis model is proposed below, where spectral radiances are used as the dominant input variables in place of net radiation. Land cover and geophysical properties are also used directly in the model, which leads to the following multi-variate satellite hysteresis relationship:

$$\Delta Q_s = f(t_h, L_i, dL_i/dt, g_j, e, \phi, \lambda) \tag{2}$$

where L_i designates a spectral radiance band of the GOES-16 satellite, replacing Q^* as the hysteresis variable. The index i represents each satellite band, spanning 1-16. The variable t_h is the local time in hours, from 0-23. The land cover input, labeled g_j, is determined using the most recent iteration of the National Land Cover Database (NLCD 2016)- a static database consisting of 20 land cover types, only 16 of which are present outside of Alaska and used in this study. The NLCD parameter g_j ranges from $0 - 1$ and represents the fraction of each land cover class within a corresponding satellite pixel. The sum of all NLCD components over subscript j must sum to 1. The elevation, latitude, and longitude are also included in the model and are labeled e, ϕ, and λ, respectively. The variables that are input to the model are: 16 radiance bands, 16 radiance time derivatives, 20 NLCD classes, latitude, longitude, elevation, and hour of day. This amounts to 56 variables used for training residual heat storage flux from ground stations.

A process flow diagram for the satellite hysteresis model is given in Fig. 1. The diagram represents each required data source and how it is used in the hysteresis model. The input data and process flow mimic the general sequence used in machine learning algorithms - where known parameters and variables are used as inputs, and desired variables are used for training and

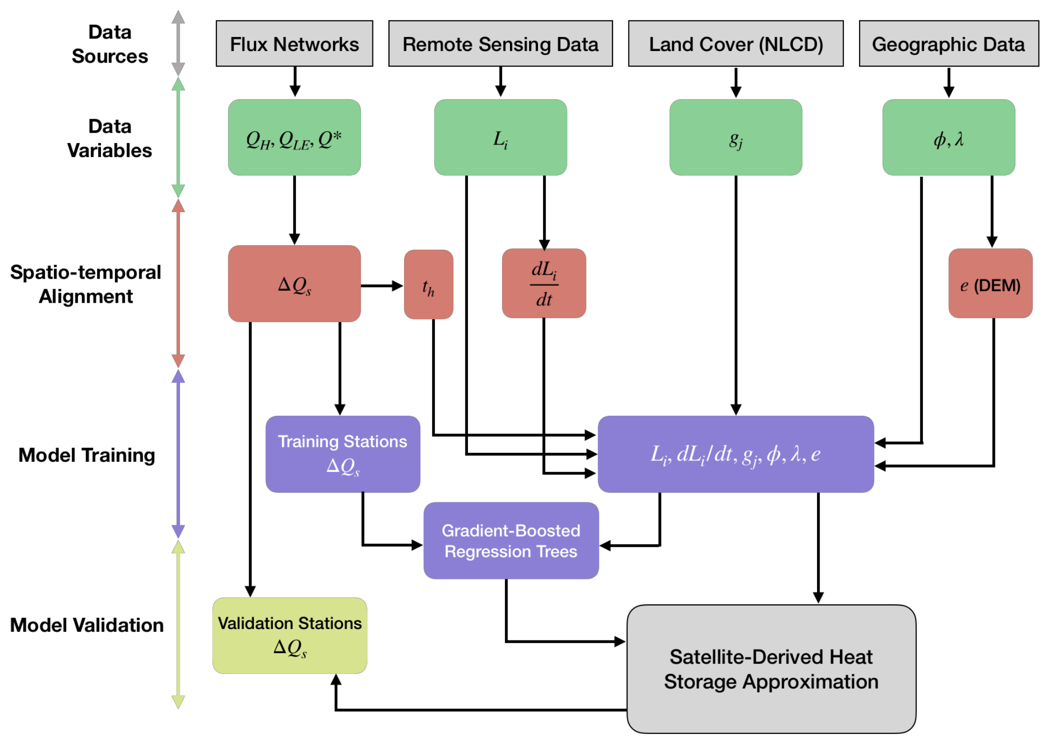


Figure 1: Schematic diagram of the data sources, the derived variables, and how the process of developing and validating the model is carried out

validation. The inputs used specifically in the hysteresis model are both stationary (land cover, latitude, longitude, elevation) and non-stationary (satellite radiance). This combination of stationary and non-stationary information will aid in the extrapolation to areas where ground stations do not exist.

2.2. Residual Heat Storage

The surface energy budget can be used to solve for heat storage as a residual between the heat sources and sinks under energy balance closure assumptions (Oke, 1988; Piringer et al., 2002; Sun et al., 2013, 2017b):

$$\Delta Q_s = Q^* - (Q_H + Q_{LE}) \tag{3}$$

where ΔQ_s denotes the storage heat flux, Q^* is the all-wave net radiation, and Q_H and Q_{LE} represent the sensible and latent heat fluxes, respectively. The residual heat storage derived above is a straightforward method and is often used when radiometers and eddy covariance instruments are available to measure the remaining fluxes (Ferreira et al., 2013; Roberts et al., 2020). The anthropogenic heat flux (Q_F) is often omitted under the assumption that either the error associated with models for Q_F are larger than its contribution to the energy budget (Parlow et al., 2014; Sun et al., 2017a), or the eddy covariance instruments are believed to capture most of the radiative, conductive, and convective components of the anthropogenic release. (Grimmond and Oke, 2002).

The sensible and latent heat fluxes are observed using closed-path eddy covariance systems fitted with gas analyzers and 3-D ultrasonic anemometers (Balogun et al., 2009). Net radiation is measured using net radiometers, taking components of incoming and outgoing shortwave and longwave radiation in balance (Ando and Ueyama, 2017). Two networks were used, the NYS Mesonet and National Ecological Observatory Network (NEON); thus, varying models of instrumentation can be found across all sites.

The flux data was acquired every half hour for comparison with each corresponding satellite pixel, which was time-aligned with the flux measurement down to a 2.5 minute window (resulting from the 5-minute satellite interval). The satellite pixels for each of the 16 bands were available regardless of weather impacts, whereas the flux stations would remove periods of extreme wind or rain automatically.

2.3. Gradient Boosted Regression Trees (GBRT)

Gradient-boosted regression trees (GBRTs) were selected based on their performance with multivariate systems and their ability to handle nonlinearities without overfitting (Kedem et al., 2012). The GBRT algorithm used here is similar to the method developed by Ke et al. (2017), wherein the difference between variable and observation is calculated as a 'loss function' and broken into parts, called trees. The number of trees is determined by the increase in accuracy for subsequent added trees. For example, if the increase in number of trees decreases the error down by a certain amount, then another tree is added, and the partitioning continues. If an asymptote in accuracy is reached, then the adding of trees ceases (Friedman, 2001; Mason et al., 2000). The accuracy for GBRTs here is calculated using the least-squares method. For the gradient boosting aspect, pseudo-residuals are computed as the gradient of the loss function and used at each time step to increase the prediction capabilities of the model (Friedman, 2002).

Using GBRTs, the 16 satellite radiances, the 16 satellite radiance time derivatives, and land cover and geography-specific properties will be trained with the residual heat storage, as stated in Eqn. 2. The goal is to create a robust algorithm that uses not only the radiance data relayed from the satellite, but also the contribution of land cover and other phenomena. By using land cover-specific parameters, the hysteresis algorithm will uncover potential relationships between satellite data and ground properties, making it easier to use the algorithm in uncalibrated areas. GBRTs are sensitive to overfitting, and as a result, independent stations will be used to assess the true performance of the model. Training size will also be varied as a way to explore the time-series dependence of the model and asses the peak performance of the model (Robinzonov et al., 2012; Schonlau, 2005).

Python's Scikit-learn library (Pedregosa et al., 2011) is used to implement the gradient-boosted regression tree method outlined above (Prettenhofer and Louppe, 2014). For training and validation, the residual heat storage flux is partitioned in time and across 34 flux stations over the summer of 2019. The stations are then separated into testing and validation groups, where some stations are reserved for independent verification of the model. The data was partitioned in time as a percentage of total data points for each station to be inputted into the GBRT algorithm. For example, if a station has 1000 valid data points, a training size of 80% uses 800 data points for training and 200 data points for validation. Training periods are partitioned sequentially as a way to optimize the algorithm for real-world implementation. Using the sequential partition allows the data to be used in a similar method for an operational product, where the training can be done continuously with data as it is updated in time. This may enable real-time calibration and perhaps improve the overall accuracy of the model in the long-term. The stations are randomly shuffled when trained, meaning the errors can vary from training to training. This helps diversify the model with different inputs and test the stability and overfitting artifacts, if any, present in the model.

2.4. Downscaling Routine

A simple downscaling routine is proposed that requires no further tools or training to the machine learning algorithm. The downscaling takes the available parameters: NLCD, geography, satellite radiances; and uses the higher spatial resolution parameters to create higher resolution maps of heat storage. The proposed scaling takes the 2-km native algorithm (chosen based on the dominant resolution of 12 of the 16 radiance bands) and outputs a 320-m product.

The NLCD and digital elevation model both have the advantage of 30-m spatial resolution. The 30-m resolution, therefore, dictates the smallest resolution possible for the algorithm, assuming the downscaling procedure functions accurately and linearly, without introducing considerably large errors. The downscaling proposed here follows a similar method used in various satellite algorithms relating to meteorology (Busch et al., 2012; Mascaro et al., 2010; Ranney et al., 2015). Three types of algorithms can be found for downscaling satellite products: satellite-to-satellite methods, methods that use geoinformation data, and model-based models (Mitraka et al., 2015; Peng et al., 2017). One or a combination of the aforementioned methods can be seen across the literature for accurately downscaling satellite data.

The downscaling presented here hinges on the hypothesis that the heat storage relies heavily on land cover fraction, meaning the machine learning model will respond to the higher resolution inputs without change to its accuracy. And since there is no comparative satellite with higher spatial resolution *and* similar temporal or spectral resolution - the satellite-to-satellite method is not employable here. Of course, a higher resolution satellite could be used while available over its specific overpass times, and this is perhaps the most commond method of downscaling, however, since the model used here is a multi-spectral approach, the satellite comparison would require downscaling first in the spectral domain, which is outside the scope of this particular study. The model-based method is also difficult to employ in this particular case due to lack of high-resolution ground networks for model training and the lack of a standard for comparison with numerical models. Consequently, geoinformation data is the method that will be used to downscale the model.

These geostatistical assumptions may break down when compared with flux towers, due to the flux tower footprint in urban areas measuring from 0.5km - 2km across the literature (Bergeron and Strachan, 2011; Feigenwinter et al., 2017; Kotthaus and Grimmond, 2012, 2014; Velasco et al., 2005, 2009), however, this will be assessed during the presentation of results in later sections.

2.5. Error Metrics

The hysteresis model given in the forthcoming analysis uses standard statistical methods to assess its performance against ground station residual flux data. The following metrics are given in relation to the comparison between model storage heat flux and ground-truth station heat storage flux derived as a residual (Laurent et al., 1998; Singh and Irmak, 2009; Şahin, 2012):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\Delta Q_{s,i,model} - \Delta Q_{s,i,station})^2} \tag{4}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\Delta Q_{s,i,model} - \Delta Q_{s,i,station}| \tag{5}$$

$$\mathrm{MBE} = \frac{1}{N}\sum_{i=1}^{N}(\Delta Q_{s,i,model} - \Delta Q_{s,i,station}) \tag{6}$$

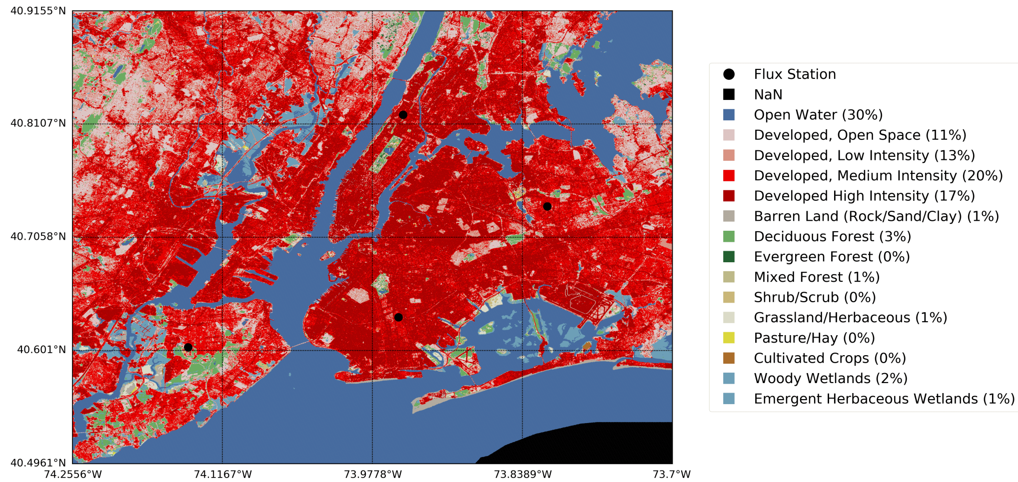


Figure 2: Flux station locations and land cover map for New York City mapped from the National Land Cover Database (NLCD).

$$R^2 = 1 - \frac{\sum_i (\Delta Q_{s,i,model} - \Delta Q_{s,i,station})^2}{\sum_i (\Delta Q_{s,i,station} - \overline{\Delta Q_{station}})^2} \tag{7}$$

where RMSE represents the root-mean-square error, MAE is the mean-absolute error, MBE is the mean bias error, and R^2 is the coefficient of determination, sometimes called the model efficiency. These four metrics were chosen as a way of normalizing the comparisons in the literature - which use a varying amount of the relationships given above.

3. Geography and Data Selection

3.1. Study Area

The study area contains a grid of 16x24 GOES-16 satellite pixels at the 2-km scale, resulting in a total of 384 pixels in the NYC region for the native satellite-derived algorithm. A total of 34 stations were used for the analysis:

20 from NEON, 10 from Ameriflux, 3 from the NYS Mesonet, and 1 from the City College of New York. The observation period spanned June - August 2019, and the geographic spread of the stations was limited to the bounds of the CONUS (continental United States). All of the stations were trained for different percentages of the available data, whereas the particular urban analytics are focused on the four urban sites located in New York City. Figure 2 shows the four NYC flux stations plotted atop the NLCD map of NYC. The study area is dominated by open water and developed land cover, both of which can be observed in Fig. 2.

3.2. Surface Flux Stations

Three networks are used for analysis: the National Ecological Observatory Network (NEON), the Ameriflux network, and the New York State Mesonet. For each of the networks, fluxes are derived using the eddy-covariance method (Network, 2020a). NEON sites use Campbell Scientific CSAT-3 sonic anemometers and Li-Cor LI-7200 gas analyzers mounted atop vertical towers. The raw data are used to generate 30-minute turbulent flux data products for sensible and latent heat fluxes. Net radiation is derived using components of incoming and outgoing shortwave and longwave radiation, acquired with Hukseflux NR01 net radiometers. A total of 21 NEON sites are, for the most part, non-urban and will help decouple vegetative components of land cover in urban sites produced by the NYS Mesonet.

Ameriflux core sites are used in conjunction with the NEON sites and employ flux towers and gas analyzers similarly from Campbell Scientific and Li-Cor (Ameriflux, 2020). Nine stations are used from the Ameriflux network, most of which are non-urban. The Ameriflux sites were introduced as a way of diversifying the training and validation of the satellite algorithm. Similar to the NEON network, sensible and latent heat fluxes, along with net radiation were acquired at 30-minute intervals to produce the heat storage residuals.

The final ground network is the NYS Mesonet. NYS Mesonet stations use Kipp & Zonen CNR4 net radiometers and Campbell Scientific CSAT3A 3D ultrasonic anemometers and EC155 gas analyzers. The instrumentation located on the NYS Mesonet sites also record sensible, latent, and 4-component radiation every thirty minutes. Three NYS Mesonet sites are used, all of which are specifically urban and are located within New York City. An additional flux tower is located in Manhattan at the City College of New York (CCNY), which is identical to the NYS Mesonet instruments but is not maintained by the NYS Mesonet. This results in a total of 34 ground flux stations for the analysis. The urban sites will serve as the test bed for the satellite routine and will be used as performance benchmarks for the machine learning routine.

3.3. GOES-16 Satellite Data

The Geostationary Operational Environmental Satellite-R Series (GOES-R) renamed GOES-16 upon reaching its operational orbit, is used as the weather satellite for comparison with ground-based residual heat storage fluxes acquired from the flux networks NEON and NYS Mesonet. The raw spectral radiance data is acquired from the Advanced Baseline Imager (ABI) in a data product called L1b (Level 1b), which are openly available to anyone on the Google BigQuery database.

The units associated with L1b spectral radiances are [$\mathrm{W{\cdot}m^{-2}sr^{-1}\mu m^{-1}}$]. GOES-16 scan mode 3 is used and results in one observation of the continental United States (CONUS) every five minutes, for each of the 16 bands. This allows a temporal alignment accuracy of 2.5 minutes between ground-based flux stations and corresponding satellite pixels. The spatial resolution between neighboring pixels depends on the chosen band, but vary roughly 0.5 km - 2.0 km (Group and Program, 2017).

3.4. Land Cover and Digital Elevation Model

The U.S. Geological Survey recently published its fifth National Land Cover Database (NLCD), designated the NLCD 2016. The contiguous U.S. (CONUS) NLCD 2016 product is used here, which is produced at 30-m spatial resolution and contains 16 land cover classes (Jin et al., 2019; Wickham et al., 2014; Yang et al., 2018a). The classes are divided into the following categories: open water; perennial ice/snow; developed: open space, low intensity, medium intensity, and high intensity; barren land (rock/sand/clay); forest: deciduous, evergreen, mixed; shrub/scrub; grasslands/herbaceous; pasture/hay; cultivated crops; wetlands: woody and emergent herbaceous. Refer to Fig. 2 for the NLCD breakdown in New York City. The NLCD 2016 incorporates four urban categories (developed classes) and will serve as the determination of urbanization for given satellite pixels.

Along with the 16 NLCD classes, a digital elevation model (DEM) will be added as part of the classification of each satellite pixel. The Shuttle Radar Topography Mission (SRTM) is run by the U.S. Geological Survey and publishes a freely available, 30-m resolution, elevation product that spans the entire contiguous U.S. (Elkhrachy, 2018). Latitude and longitude coordinates dictate the elevation across a given satellite pixel, and is used in the machine learning algorithm to capture the sensitivity of heat storage to changes in elevation and as well as land class. This is commonly done for satellite-based assessments of evapotranspiration or thermodynamic processes at the Earth's surface (Cheng et al., 2011; Semmens et al., 2016; Xian and Crane, 2006; Zhou et al., 2014). Both the DEM and NLCD are at much higher resolution than the satellite, which will aid in the downscaling of the final satellite algorithm.

3.5. Relationship Between Satellite Bands and Residual Heat Storage

The hypothesis of this research hinges on the correlation between satellite radiance and ground station residual heat storage. If net radiation is derived using satellite radiances throughout the literature (Bisht and Bras, 2010; Carmona et al., 2015; Hou et al., 2014; Jin et al., 2011), then an application of raw radiances, without the intermediary routine for predicting net radiation, may suffice for approximating of heat storage directly. Particularly, with the high temporal resolution of the 16 satellite bands covering the visible, near-infrared, and infrared wavelengths (Schmit et al., 2018) - the correlation between satellite radiances and heat storage should be high. Figure 3 demonstrates the correlation between ground station and nearest satellite pixel for an urban area (Brooklyn, NY), where the correlation between variables is defined as (Benesty et al., 2009; Inglada, 2002):

$$\text{Corr} = \frac{\sum_{k=1}^{N}(\Delta Q_{s,k} - \overline{\Delta Q_s}) \cdot (L_{\lambda,k} - \overline{L_\lambda})}{\sqrt{\sum_{k=1}^{N}(\Delta Q_{s,k} - \overline{\Delta Q_s})^2} \cdot \sqrt{\sum_{k=1}^{N}(L_{\lambda,k} - \overline{L_\lambda})^2}} \tag{8}$$

The shortwave bands can be seen to negatively correlate with the heat storage flux during the daytime, which is expected due to the influence of direct solar irradiation. During the nighttime, the shortwave bands have

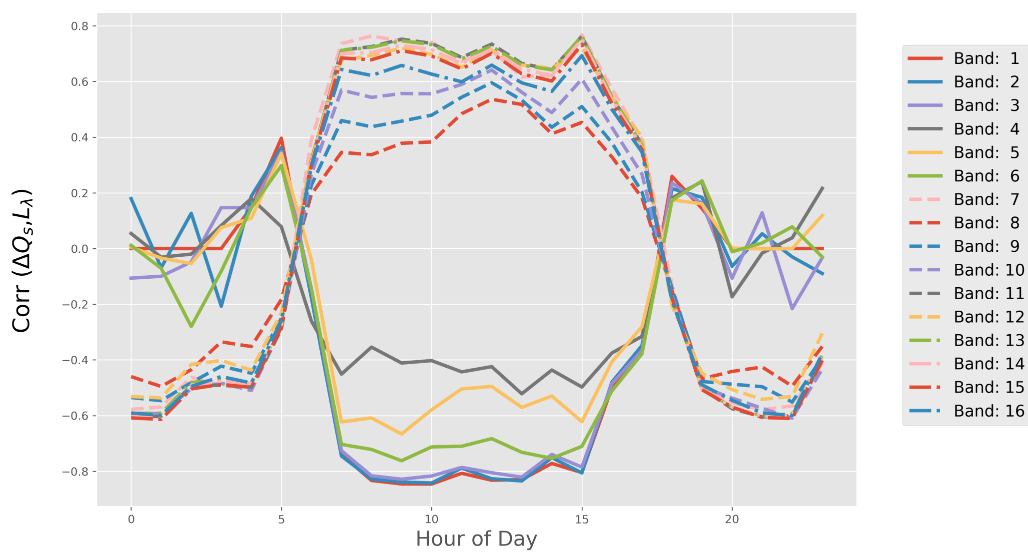


Figure 3: Correlation between ground station residual heat storage flux and nearest satellite pixel for each GOES-16 band. The shortwave bands are negatively correlated to the heat storage flux during the daytime, and minimally correlated during the nighttime; whereas the longwave bands are positvely correlated to the storage flux during the daytime and negatively correlated during the nighttime. These correlations are essential to the hypothesis that radiance bands can be used to calculate heat storage flux.

almost no correlation to the heat storage, as expected, due to the opposite reasoning proposed in the previous sentence. For longwave bands, there is high correlation during the nighttime and daytime. Longwave bands are positvely correlated to the storage flux during the daytime and negatively correlated during the nighttime. These correlations reinforce the original hypothesis that the GOES-16 radiance bands can be used to calculate heat storage flux - the primary motivator going forward in this research.

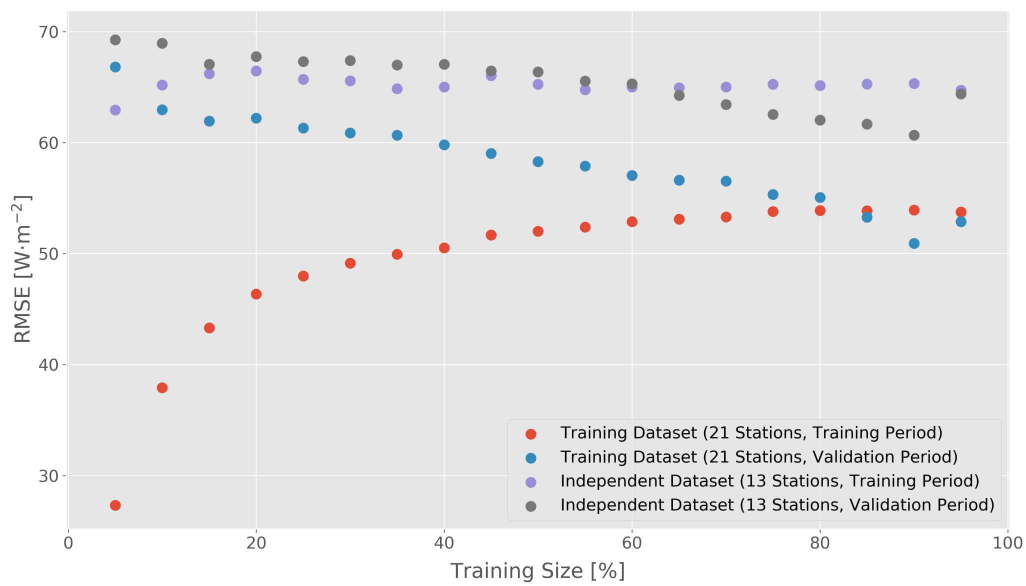


Figure 4: RMSE as a function of training size for the training dataset and independent dataset. The training dataset and independent dataset have also been divided into the training period and validation period as well. We see the decrease in RMSE as a function of increased training size for the validation periods, as expected.

4. Results and Discussion

4.1. Training and Validation

Neither training nor validation were uniform across the swath of available stations due to the variability of local meteorological conditions (high winds, heavy rain) and complications with instrumentation, both of which cause drops in data. This results in differing amounts of points per station over the full testing period. Figure 4 shows the average RMSE for training and validation periods, where the training dataset is a unique set of 21 ground stations taken from all three networks, while the independent dataset is a separate unique set of 13 ground stations from each of the three networks.

Table 1: Division of training and validation data for the analysis over summer 2019.

Dataset	Period	# Stations	Mean # Points	RMSE
Training	Training	21	2927	55.8
Training	Validation	21	326	52.6
Independent	Training	13	2594	63.7
Independent	Validation	13	289	60.5

The profiles given in Fig. 4 exhibit expected behavior as the training size changes. The training dataset carries the lowest error during the training period, which is most likely due to overfitting - a common artifact of GBRT algorithms (Kedem et al., 2012). Similarly, the error in the validation period is also lower than that of the independent dataset. The independent dataset does, however, demonstrate consistent error across the range of training sizes, indicating that the training is somewhat stable in training size, with a slight decrease in error as the training size increases.

Considering the results of Fig. 4, 90% of the data was selected for training and 10% was reserved for validation. It would be valid to select any training size over 80%, as that point marks the approximate asymptote in training and validation error. The independent dataset experiences similar phenomena, with less variability, indicating a more accurate prediction of the model's true performance in relation independent ground stations (or satellite pixels).

Statistics relating to the satellite-derived GBRT model for heat storage at 90% training are given in Table 1. A resulting argument can be made regarding the performance of the satellite hysteresis model , stating that the 2-km spatial model carries errors on the order of 60-65 W$\cdot$m^{-2}, when

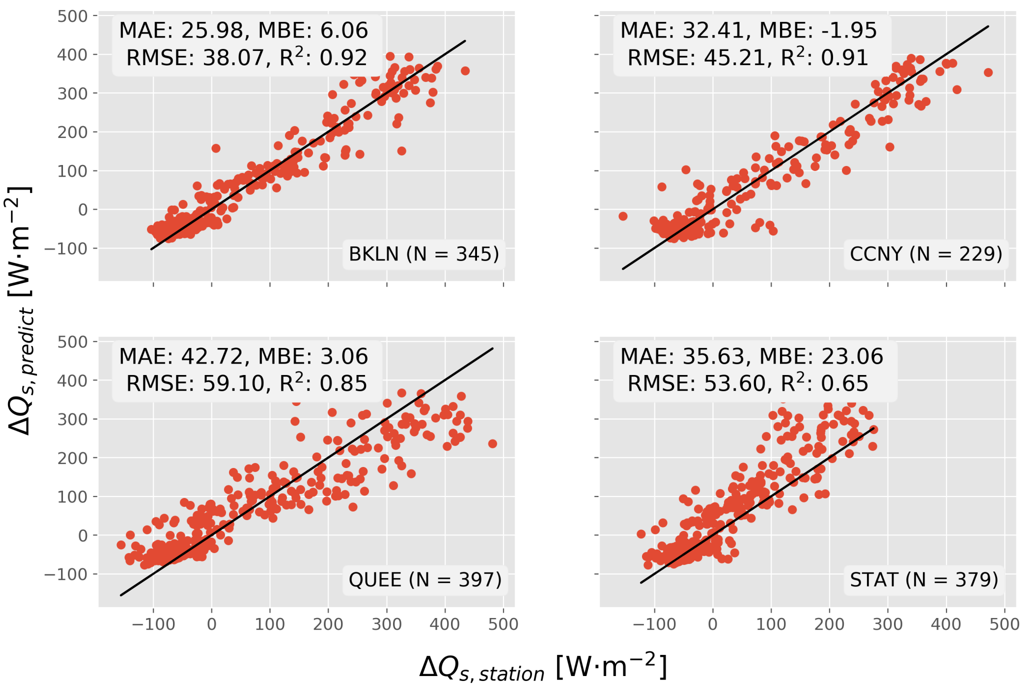


Figure 5: Scatter for NYC flux stations and the model performance using GOES-16 satellite.

compared with ground-based residual heat flux. This, of course, has only been posited for the summer of 2019, and the validity of this statement is contingent upon further validation across multiple seasons.

4.2. The Case Study of New York City

In Fig. 5, four scatter plots are given for each of the urban stations in New York City, exclusively for the validation period. Two of the urban stations were used during training (BLKN and CCNY), while the other two were not (QUEE and STAT). For all four stations, the RMSE values were below 60 W·m^{-2} and the MAE were below 43 W·m^{-2}. The average value for both RMSE and MAE in the urban region were 49 and 34 W·m^{-2}, respectively.

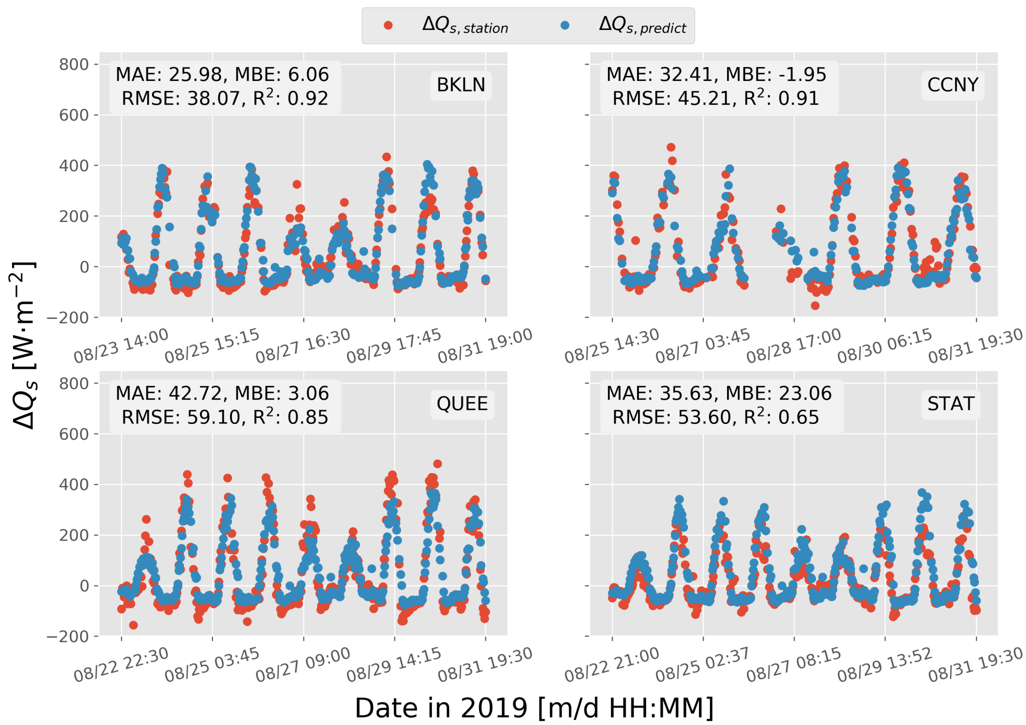


Figure 6: Temporal reconstruction of ΔQ_s using the satellite hysteresis model.

The average MBE was 7.6 W·m^{-2}, and the average R^2 was 0.83. All four performance metrics are well within the ranges cited across the literature (Roberts et al., 2006).

Investigation of Fig. 6 shows the temporal reconstruction of the same set of validation data given in Fig. 5. The array of temporal reconstructions is capable of tracking the diurnal profile of ΔQ_s quite well, something that has not been demonstrated in the contemporary satellite research (Chrysoulakis et al., 2018; Kato and Yamaguchi, 2007; Parlow, 2003; Rigo and Parlow, 2007). The satellite hysteresis model is also capable of approximating heat storage during rainy and cloudy periods, another capability lacking in the satellite literature. On the 22nd, 23rd, and 28th of August, historical weather records for the NYC area show rainy and cloudy conditions, which is exhibited in Fig. 6 by the lower amplitudes of heat storage.

For most of the diurnal cycle, the model is able to recreate the trend of heat storage for each of the sites. What varies most with the model is its ability to capture the full-scale amplitude of both the daytime maximums and the nighttime minimums. For the Queens station (QUEE), the daytime satellite heat storage peaks are smaller than the station residuals, and during the nighttime the opposite can be observed. For the Staten Island (STAT) station, the daytime peaks are somewhat over-predicted by the satellite, which can also be concluded by looking at the statistical bias.

Since both QUEE and STAT were omitted from the training dataset, we can hypothesize that the satellite-derived ΔQ_s may have a positive bias for less urban stations (STAT island is only about 60% developed in land cover class), and has a more tempered response in amplitude both during the daytime and nighttime for more urban areas (observed for QUEE, where the land cover class is nearly 100% developed). However, more urban stations are needed to fully verify this claim.

Something to note is that both training stations, BKLN and CCNY, are each nearly 100% urban, meaning that each of the stations and satellite pixels have different responses despite similar classifications in land cover. This is an important observation and one that increases the confidence of the model's ability to capture varying responses over urban environments where ground stations are unavailable for comparison. Although the error associated with QUEE is higher than the other three stations, it is well within the range of errors cited across the literature for ground-based OHM methods, indicating that the satellite routine is a viable approach to quantifying heat storage. The errors will also be explored in subsequent sections, once the downscaling routine has been introduced.

4.3. Downscaling from 2-km to 320-m

The ubiquity of water surrounding the land masses of New York City results in a particular obstacle for development of the satellite algorithm. The issue arises when training the algorithm, where the lack of ground stations over water creates a weakness for pixels that contain certain amounts of water. As a consequence, pixels with a water fraction greater than 0.05 (5%) were dropped. And with the satellite-derived heat storage outputting data at a native 2-km resolution , many of the pixels in the study area were dropped.

As a way to both increase the number of satellite pixels in the small window of NYC and avoid dropping so many pixels due to water content - a downscaling routine was developed. The downscaling takes the satellite algorithm from 2-km to 320-m, which results in fewer dropped pixels and a better representation of the distribution of heat storage in cities. This is specifically important for coastal cities where water surrounds highly urban areas.

As explained in Section 2.4, the native resolution of the satellite-derived heat storage is 2-km, set by the dominant spatial resolution of the radiance bands. Here, the downscaling is set to 320-m, which increases the resolution of the NYC grid to 100x150. The implementation of the downscaling algorithm is validated for QUEE and STAT, with a decrease in RMSE of 5.4 $W{\cdot}m^{-2}$ for QUEE and 0.3 $W{\cdot}m^{-2}$ for STAT. This is shown in Fig. 7.

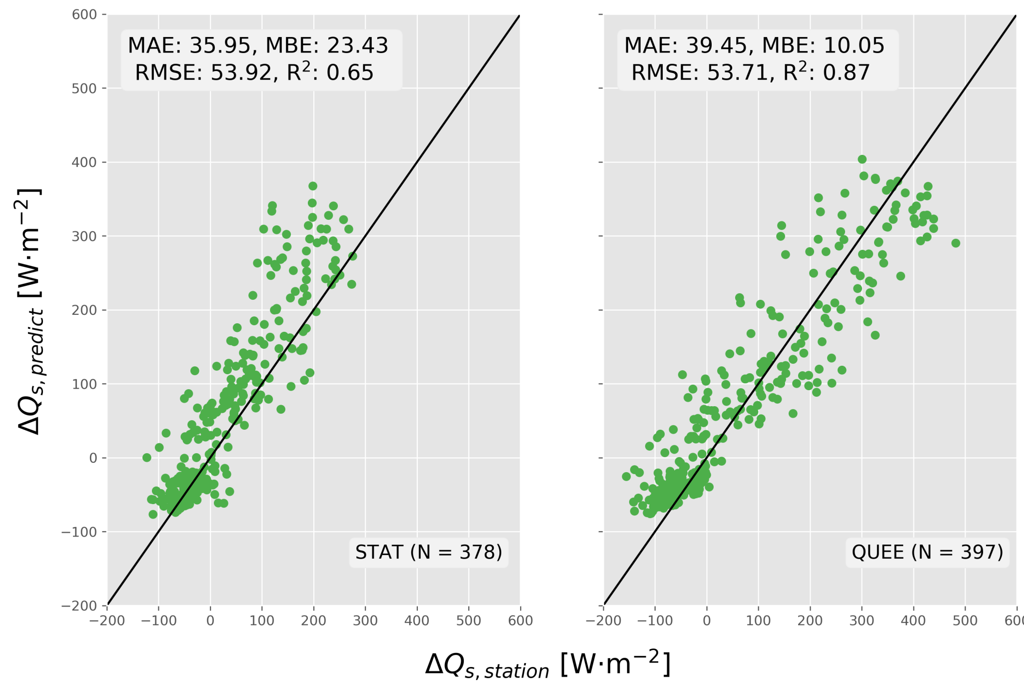


Figure 7: Downscaling performance for STAT (left) and QUEE (right) sites. The downscaling from 2-km to 320-m resulted in a decrease in RMSE (5.5W $\cdot$m^{-2}) for the QUEE pixel and no change in RMSE for the STAT pixel, partially validating the accuracy of the downscaling procedure.

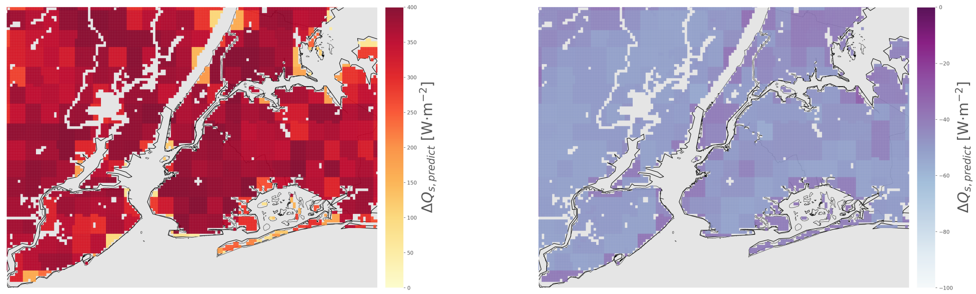
The same was observed for both BKLN and CCNY stations as well, where their errors never deviated more than 2 W·m^{-2} when comparing the 2-km pixel to the 320-m downscaled value. A sensitivity analysis was done for neighboring pixels surrounding each ground station, and similar results were found for the adjacent pixels, where marginal variability was found due to land cover changes. The results presented henceforth will be on the 320-m satellite-derived heat storage using the GBRT method and the downscaling presented here.

4.4. Spatial Representation of Heat Storage

A spatial representation of heat storage using the satellite hysteresis model is given in Fig. 8 for both midday and midnight periods on August 24th, 2019. The data has been downscaled as per the routine given in the previous section, resulting in a resolution of 320-m, rather than 2-km. A spatial filter has also been implemented based on water content: any pixel with more than 5% water is omitted. Additionally, any pixel that exists outside the range [-200,600] W$\cdot$m^{-2} is also omitted. This is based on the average maxima and minima observed over several standard deviations.

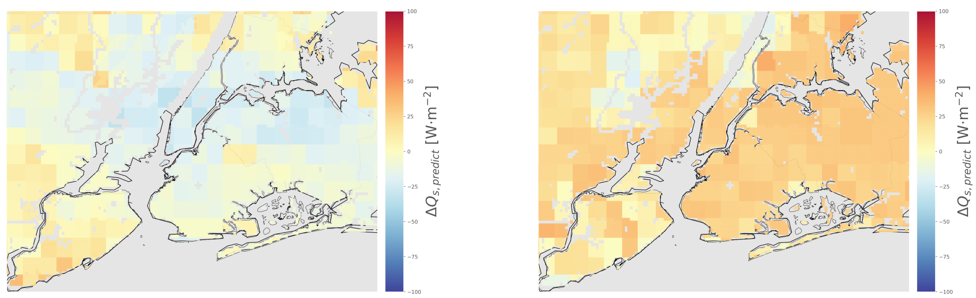
The first and perhaps most obvious inference is that the native 2-km pixels have overwhelmingly influenced each downscaled pixel. The distribution of color in the spatial domain seems to be dominanted by each underlying 2-km pixel, giving the squared-off artifact in both plots in Fig. 8. This is likely due to the priority of each variable in the GBRT model, i.e., time and radiance bands take precedent over land cover and elevation. This, simply put, indicates that the model is more sensitive to macro changes in geography, rather than local changes in geography, perhaps due to viewing angle of the satellite or large changes in elevation.

The transition between positive and negative fluxes is also visible in the spatial representations of heat storage. Sunrise, for example, is captured on the 27th of August in Fig. 9a, where the storage is largely varied throughout the city. The sunrise on that day was observed at 6:18 a.m., indicating a possible delayed effect in heat storage for certain areas. The same can be observed a few hours after peak heating at 5:30 p.m., in Fig. 9b, where the release of heat (negative storage) is seen for some pixels, while positive



(a) (b)

Figure 8: (a) Daytime representation of heat storage, ΔQ_s on Aug. 24 at 12:00 EDT.. (b) Nighttime representation of heat storage on Aug. 24 at 01:30 EDT.



(a) (b)

Figure 9: (a) Heat storage after sunrise (07:30 EDT on Aug. 27) capturing the variability of the city's response to heating of its surface. (b) The inversion of heat storage (zero-crossing point) is shown just after sunset on Aug. 27 at 18:30 EDT.

storage is seen for others. This indicates another possible spatial delay.

The standard deviation of ΔQ_s across valid pixels is also observed to vary throughout the day. The midday spatial distribution of heat storage varies up to $\pm$65 W·m^{-2}, and the nighttime down to $\pm$ W·m^{-2}. These values, on average, amount to 15-20% of the relative amplitudes of ΔQ_s. The largest variability occurs during the transitional periods given in Fig. 9, at sunrise and a few hours after peak heating.

4.5. Comparisons with Literature

A statistical comparison between satellite-derived heat storage algorithms is nearly impossible as no identifiable studies have used an approach that validates their model directly against ground station values. Therefore, no satellite-to-satellite performance comparison is possible. This is likely due to the lack of measurement standards for heat storage, and moreover, a consequence of ground station sparseness in urban areas. One thing that is investigated in the literature is the fraction of net radiation occupied by heat storage. This is not valid in our case as net radiation is a component in the residual method and, accordingly, used to train the model.

One definite advantage of the GBRT-based method is that it avoids a common pitfall associated with satellite algorithms in urban areas, namely, the handling of non-Lambertian surfaces. This is likely due to the inclusion of multispectral data, which may handle some of the urban heterogeneity issues in reflected radiation (de Almeida Castanho et al., 2007; Hamedianfar and Shafri, 2015). Moreover, the temporal resolution of the satellite also permitted a large statistical comparison between model and ground stations, which also contributes to the uniqueness and stability of the algorithm. MODIS comparisons, for example, are limited to two comparison points per day along the diurnal cycle.

A more appropriate evaluation of the satellite hysteresis model is through comparison with studies that calculate statistics based on ground-to-ground measurements. As mentioned in the introduction, there are four methods used for calculating heat storage in the relevant literature: the residual method (RES), the objective hysteresis model (OHM), the thermal mass

Table 2: Comparison between satellite-derived hysteresis model and the objective hysteresis model derived from ground net radiation data for various cities. Results from this study are in bold and are taken from the downscaled comparison with the nearest NYC ground station. The table is ordered by increasing RMSE.

Site/Description	# Points	R^2	RMSE
Los Angeles, CA (suburban)[a]	424	0.92	29.0
Mexico City (city center)[a]	61	0.96	33.6
Brooklyn, NY (urban)	**345**	**0.92**	**39.5**
Manhattan, NY (urban)	**229**	**0.91**	**45.0**
Vancouver, Canada (industrial)[a]	312	0.88	48.9
Queens, NY (urban)	**397**	**0.87**	**53.7**
Staten Island, NY (suburban)	**379**	**0.65**	**53.9**
Miami, FL (suburban)[a]	204	0.79	61.9
Vancouver, Canada (suburban)[a]	464	0.67	62.9
Sacramento, CA (suburban)[a]	222	0.56	66.0
São Paulo city, Brazil (suburban)[a]	353	0.69	74.1
Chicago, IL (suburban)[a]	163	0.56	83.3
Marseille, France (city center)[b]	192	0.70	94.8
Tucson, AZ (suburban)[a]	75	0.75	107.4

[a] Grimmond and Oke (1999)

[b] Roberts et al. (2006)

scheme (TMS), and the town energy balance (TEB). One widely recognized paper by Roberts et al. (2006) compiles a comparison of all four methods into one study, and includes an error analysis for the OHM, TMS, and TEB against the RES method. It also agglomerates other studies in an effort to corroborate its statistics. Those errors are used here as a guide, in part, to assess the results produced by the satellite hysteresis model.

Table 2 shows the comparison between 14 different heat storage calculations by hysteresis model, four of which contain the satellite-derived results from NYC. Overall, the New York City stations: Queens (QUEE), Brooklyn (BKLN), Manhattan (CCNY), and Staten Island (STAT), all outperform 7 of the 10 stations in root-mean-square error. This is quite remarkable, particularly for the independently verified stations STAT and QUEE. All four NYC stations are also validated with more data points than 6 out of 10 stations. These statistics are an indication that the satellite-derived heat storage is a viable model against the ground-based OHM.

As for the town energy balance (TEB) method, the same study uses Marseille, France as a test site. The TEB estimated a mean hourly RMSE between TEB and RES of 79 $W{\cdot}m^{-2}$. Another evaluation of the TEB was done by Masson et al. (2002) for Mexico City and Vancouver, which managed mean RMSE values of 39 and 87 $W{\cdot}m^{-2}$, respectively. A third experiment carried out in the city of Basel, Switzerland compared two different implementation schemes for the TEB and found RMSE values of 64 and 70 $W{\cdot}m^{-2}$. Resolutely, it is fair to say that the satellite-derived hysteresis model outperforms the town energy balance.

For the previous study in Marseille, the final calculative method for heat storage is examined: the thermal mass scheme (TMS). The RMSE between RES and TMS was measured to be 109 $W{\cdot}m^{-2}$ - quite a large error when compared to the satellite hysteresis model. Heat storage derived by thermal mass scheme and objective hysteresis model are most suited for satellite data due to incorporation of thermal properties rather than aerodynamic properties. Thus, the dilemma described above regarding inability to compare satellite methods arises again for the TMS. As a result, very few studies have errors associated with the RES method.

Building information is often used as an input to urban storage models, such as the town energy balance (TEB) or element surface temperature method (ESTM). Building height and building area fraction information were deliberately excluded in the proposed satellite hysteresis model, due to the limitations in training, where the majority of the ground stations either do not contain buildings or lack building information. Thus, training with building information would be very limited and likely result in further overfitting in the GBRT model. This is where the national land cover database's four urban categories become important, as they account for the range of urbanization in the ground networks ranging from open space to high intensity urban development.

Another important aspect of quantifying heat storage, either from the satellite or surface perspective, is the accurate accounting of anthropogenic heat flux. A wide range of studies have developed algorithms for determining anthropogenic heat flux using traffic information, population density, fuel economies, among others (Sailor and Lu, 2004). For the proposed multispectral hysteresis model, it is assumed that the anthropogenic heat flux is inherent in the measurement of conductive, convective, and radiative fluxes captured by eddy covariance instruments placed in the urban areas (Grimmond and Oke, 2002). This is based on observations made in several cities that argue that the majority of the anthropogenic influence is outputted as sensible heat flux (Kato and Yamaguchi, 2005b; Olivo et al., 2017; Sailor, 2011). This assumption may result in under-prediction of heat storage, but it is difficult to quantify by how much, and is designated as an area for future research.

The lack of studies comparing satellite-derived heat storage to ground stations is a motivating factor for other future works relating to satellite-derived heat storage. It is likely that in the future, the thermal mass scheme will be implemented using satellite data, notably due to the temporal resolution advancements of the GOES-16 and GOES-17 satellites. For now, the comparison between satellite-derived heat storage and ground stations remains chiefly neglected.

5. Conclusion

A multispectral hysteresis model was introduced as a way to predict heat storage flux in urban areas using land cover and geographic properties contained within satellite pixels. The model bridges the divide between single-point ground measurements and spatially-distributed satellite approximations, with direct validation - something nonexistent in the peered literature. A gradient-boosted regression tree (GBRT) method was used to train input variables against a series of ground flux stations. The satellite hysteresis model outperformed many of the ground-to-ground hysteresis models, indicating that the satellite method may be an improved, more robust method for calculating heat storage flux.

The error associated with the urban satellite hysteresis model was among the best inhen compared with other studies in the research field. For all four urban stations, the independently validated, 320-m spatially downscaled, average RMSE value was found to be 48.0 W·m^{-2}, the average mean-absolute error (MAE) was found to be 33.8 W·m^{-2}, the average mean bias error (MBE) was 9.3 W·m^{-2}, and the mean R^2, 0.84.

The satellite-derived heat storage is also able to recreate spatial patterns in heat storage that were not previously possible, specifically in relation to the full diurnal cycle. Because the algorithm was independently trained and validated, its accuracy and ability to recreate hourly approximations of heat storage is noteworthy and unprecedented.

Another accomplishment of the satellite hysteresis model is its ability to capture all-weather profiles. We saw for several periods that the satellite radiances were able to capture cloudy and rainy days, which permitted the calculation of tempered heat storage despite limited solar irradiance. The captured heat storage under rainy weather is something that has not been observed in the literature. One hypothesis is that the machine learning algorithm is able to relate the diminishment of radiance amplitude to the decrease in heat storage. It could also be true that some of the radiance bands are capturing minimal radiation beyond the clouds, however, this has not been verified or studied at length, and remains a topic for future exploration.

Lastly, the maintained accuracy of the approximation of heat storage under downscaled conditions proves that the algorithm is capable of component analysis and needs to be explored to the fullest extend. In summary, accurate quantification of the spatial distribution of heat storage in built environments has been an open question that, if resolved, will open a wide range of opportunities in closing the surface energy balance. Temporal, spectral, and spatial resolutions of new generation geostationary satellites are making this quantification a closer reality. This may enable significant advances in weather and climate modeling for accurate prediction of UHI, for urban planning, and for relationships between thermal responses of urban environments and energy demands. This contribution is a first major step in this direction.

Acknowledgements

This study is supported and monitored by The National Oceanic and Atmospheric Administration – Cooperative Science Center for Earth System Sciences and Remote Sensing Technologies (NOAA-CESSRST) under the Cooperative Agreement Grant #: NA16SEC4810008. The authors would like to thank The City College of New York, NOAA-CESSRST (aka CREST) program and NOAA Office of Education, Educational Partnership Program for full fellowship support for Joshua Hrisko. The statements contained within the manuscript/research article are not the opinions of the funding agency or the U.S. government, but reflect the author's opinions. The research is also funded by The Department of Defense Army Research Office Grant # W911NF-18-1-0371. This research is also made possible by the New York State (NYS) Mesonet. Original funding for the NYS Mesonet was provided by Federal Emergency Management Agency grant FEMA-4085-DR-NY, with the continued support of the NYS Division of Homeland Security & Emergency Services; the state of New York; the Research Foundation for the State University of New York (SUNY); the University at Albany, SUNY; the Atmospheric Sciences Research Center (ASRC) at SUNY Albany; and the Department of Atmospheric and Environmental Sciences (DAES) at SUNY Albany. Funding for AmeriFlux core site data was provided by the U.S. Department of Energy's Office of Science.

References

de Almeida Castanho, A., Prinn, R., Martins, V., Herold, M., Ichoku, C., Molina, L., 2007. Urban visible/SWIR surface reflectance ratios from satellite and sun photometer measurements in Mexico City, in: Atmospheric Chemistry and Physics Discussions, European Geosciences Union, pp. 8113–8139.

Ameriflux, 2020. Core flux sites. https://ameriflux.lbl.gov/. lawrence Berkeley National Laboratory.

Anandakumar, K., 1999. A study on the partition of net radiation into heat fluxes on a dry asphalt surface. Atmospheric Environment 33, 3911 – 3918.

Ando, T., Ueyama, M., 2017. Surface energy exchange in a dense urban built-up area based on two-year eddy covariance measurements in Sakai, Japan. Urban Climate 19, 155 – 169.

Arnfield, A.J., Grimmond, C., 1998. An urban canyon energy budget model and its application to urban storage heat flux modeling. Energy and Buildings 27, 61 – 68.

Balogun, A.A., Adegoke, J.O., Vezhapparambu, S., Mauder, M., McFadden, J.P., Gallo, K., 2009. Surface energy balance measurements above an exurban residential neighbourhood of Kansas City, Missouri. Boundary-Layer Meteorology 133, 299.

Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Pearson correlation coefficient, in: Noise reduction in speech processing. Springer, pp. 1–4.

Bergeron, O., Strachan, I.B., 2011. CO2 sources and sinks in urban and suburban areas of a northern mid-latitude city. Atmospheric Environment 45, 1564 – 1573.

Bisht, G., Bras, R.L., 2010. Estimation of net radiation from the MODIS data under all sky conditions: Southern Great Plains case study. Remote Sensing of Environment 114, 1522 – 1534.

Bonacquisti, V., Casale, G., Palmieri, S., Siani, A., 2006. A canopy layer model and its application to Rome. Science of The Total Environment 364, 1 – 13.

Busch, F.A., Niemann, J.D., Coleman, M., 2012. Evaluation of an empirical orthogonal function–based method to downscale soil moisture patterns based on topographical attributes. Hydrological Processes 26, 2696–2709.

Camps-Valls, G., 2009. Machine learning in remote sensing data processing, in: 2009 IEEE International Workshop on Machine Learning for Signal Processing, pp. 1–6.

Camuffo, D., Bernardi, A., 1982. An observational study of heat fluxes and their relationships with net radiation. Boundary-Layer Meteorology 23, 359–368.

Carmona, F., Rivas, R., Caselles, V., 2015. Development of a general model to estimate the instantaneous, daily, and daytime net radiation with satellite data on clear-sky days. Remote Sensing of Environment 171, 1 – 13.

Cheng, L., Xu, Z., Wang, D., Cai, X., 2011. Assessing interannual variability of evapotranspiration at the catchment scale using satellite-based evapotranspiration data sets. Water Resources Research 47.

Chrysoulakis, N., Grimmond, C., Feigenwinter, C., Lindberg, F., Gastellu-Etchegorry, J.P., Marconcini, M., Mitraka, Z., Stagakis, S., Crawford, B., Olofson, F., Landier, L., Morrison, W., Parlow, E., 2018. Urban energy exchanges monitoring from space. Scientific Reports 8.

Coutts, A.M., Beringer, J., Tapper, N.J., 2007. Impact of increasing urban density on local climate: Spatial and temporal variations in the surface energy balance in Melbourne, Australia. Journal of Applied Meteorology and Climatology 46, 477–493.

DeFries, R., Chan, J.C.W., 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. Remote Sensing of Environment 74, 503 – 515.

Elkhrachy, I., 2018. Vertical accuracy assessment for SRTM and ASTER digital elevation models: A case study of Najran city, Saudi Arabia. Ain Shams Engineering Journal 9, 1807–1817.

Feigenwinter, C., Parlow, E., Vogt, R., Schmutz, M., Chrysoulakis, N., Lindberg, F., Marconcini, M., Del Frate, F., 2017. Spatial distribution of sensible and latent heat flux in the URBANFLUXES case study city Basel (Switzerland), in: 2017 Joint Urban Remote Sensing Event (JURSE), IEEE. pp. 1–4.

Feigenwinter, C., Vogt, R., Parlow, E., Lindberg, F., Marconcini, M., Frate, F.D., Chrysoulakis, N., 2018. Spatial distribution of sensible and latent heat flux in the city of Basel (Switzerland). IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 2717–2723.

Ferreira, M.J., de Oliveira, A.P., Soares, J., 2013. Diurnal variation in stored energy flux in São Paulo city, Brazil. Urban Climate 5, 36 – 51.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics , 1189–1232.

Friedman, J.H., 2002. Stochastic gradient boosting. Computational statistics & data analysis 38, 367–378.

Golden, J.S., 2004. The built environment induced urban heat island effect in rapidly urbanizing arid regions–a sustainable urban engineering complexity. Environmental Sciences 1, 321–349.

Grimmond, C., Cleugh, H., Oke, T., 1991. An objective urban heat storage model and its comparison with other schemes. Atmospheric Environment. Part B. Urban Atmosphere 25, 311 – 326.

Grimmond, C.S.B., Oke, T.R., 1999. Heat storage in urban areas: Local-scale observations and evaluation of a simple model. Journal of Applied Meteorology 38, 922–940.

Grimmond, C.S.B., Oke, T.R., 2002. Turbulent heat fluxes in urban areas: Observations and a local-scale urban meteorological parameterization scheme (LUMPS). Journal of Applied Meteorology 41, 792–810.

Group, G.C.W., Program, G.S., 2017. NOAA GOES-R series Advanced Baseline Imager (ABI) level 1b radiances. NOAA National Centers for Environmental Information.

Hamedianfar, A., Shafri, H.Z.M., 2015. Detailed intra-urban mapping through transferable OBIA rule sets using WorldView-2 very-high-resolution satellite images. International Journal of Remote Sensing 36, 3380–3396.

Herold, M., Roberts, D.A., Gardner, M.E., Dennison, P.E., 2004. Spectrometry for urban area remote sensing—development and analysis of a spectral library from 350 to 2400 nm. Remote Sensing of Environment 91, 304–319.

Hou, J., Jia, G., Zhao, T., Wang, H., Tang, B., 2014. Satellite-based estimation of daily average net radiation under clear-sky conditions. Advances in Atmospheric Sciences 31.

Inglada, J., 2002. Similarity measures for multisensor remote sensing images, in: IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. 104–106.

Järvi, L., Grimmond, C.S.B., Taka, M., Nordbo, A., Setälä, H., Strachan, I.B., 2014. Development of the surface urban energy and water balance scheme (SUEWS) for cold climate cities. Geoscientific Model Development 7, 1691–1711.

Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., Zhu, Z., Xian, G., Howard, D., 2019. Overall methodology design for the United States National Land Cover Database 2016 products. Remote Sensing 11.

Jin, Y., Randerson, J.T., Goulden, M.L., 2011. Continental-scale net radiation and evapotranspiration estimated using MODIS satellite observations. Remote Sensing of Environment 115, 2302 – 2319.

Just, A.C., De Carli, M.M., Shtein, A., Dorman, M., Lyapustin, A., Kloog, I., 2018. Correcting measurement error in satellite aerosol optical depth with machine learning for modeling PM2.5 in the Northeastern USA. Remote Sensing 10.

Just, A.C., Liu, Y., Sorek-Hamer, M., Rush, J., Dorman, M., Chatfield, R., Wang, Y., Lyapustin, A., Kloog, I., 2019. Gradient boosting machine learning to improve satellite-derived column water vapor measurement error. Atmospheric Measurement Techniques Discussions 2019, 1–22.

Kanda, M., Moriwaki, R., Kasamatsu, F., 2006. Spatial variability of both turbulent fluxes and temperature profiles in an urban roughness layer. Boundary-Layer Meteorology 121, 339–350.

Kato, S., Yamaguchi, Y., 2005a. Analysis of urban heat-island effect using ASTER and ETM+ data: Separation of anthropogenic heat discharge and natural heat radiation from sensible heat flux. Remote Sensing of Environment 99, 44 – 54. Scientific Results from ASTER.

Kato, S., Yamaguchi, Y., 2005b. Analysis of urban heat-island effect using ASTER and ETM+ data: Separation of anthropogenic heat discharge and natural heat radiation from sensible heat flux. Remote Sensing of Environment 99, 44–54.

Kato, S., Yamaguchi, Y., 2007. Estimation of storage heat flux in an urban area using ASTER data. Remote Sensing of Environment 110, 1 – 17.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree, in: Advances in neural information processing systems, pp. 3146–3154.

Kedem, D., Tyree, S., Sha, F., Lanckriet, G.R., Weinberger, K.Q., 2012. Non-linear metric learning, in: Advances in neural information processing systems, pp. 2573–2581.

Kerschgens, M.J., Kraus, H., 1990. On the energetics of the urban canopy layer. Atmospheric Environment. Part B. Urban Atmosphere 24, 321 – 328.

Kotthaus, S., Grimmond, C., 2012. Identification of micro-scale anthropogenic CO2, heat and moisture sources – processing eddy covariance fluxes for a dense urban environment. Atmospheric Environment 57, 301 – 316.

Kotthaus, S., Grimmond, C., 2014. Energy exchange in a dense urban environment – part ii: Impact of spatial heterogeneity of the surface. Urban Climate 10, 281 – 307. ICUC8: The 8th International Conference on Urban Climate and the 10th Symposium on the Urban Environment.

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geoscience Frontiers 7, 3 – 10. Special Issue: Progress of Machine Learning in Geosciences.

Laurent, H., Jobard, I., Toma, A., 1998. Validation of satellite and ground-based estimates of precipitation over the Sahel. Atmospheric Research 47-48, 651 – 670.

Lee, Y., Han, D., Ahn, M.H., Im, J., Lee, S.J., 2019. Retrieval of total precipitable water from himawari-8 ahi data: A comparison of random forest, extreme gradient boosting, and deep neural network. Remote Sensing 11.

Lindberg, F., Olofson, K., Sun, T., Grimmond, C., Feigenwinter, C., 2020. Urban storage heat flux variability explored using satellite, meteorological and geodata. Theoretical and Applied Climatology , 1–14.

Lipson, M.J., Hart, M.A., Thatcher, M., 2017. Efficiently modelling urban heat storage: an interface conduction scheme in an urban land surface model (aTEB v2.0). Geoscientific Model Development 10, 991–1007.

Mascaro, G., Vivoni, E.R., Deidda, R., 2010. Downscaling soil moisture in the southern Great Plains through a calibrated multifractal model for land surface modeling applications. Water Resources Research 46.

Mason, L., Baxter, J., Bartlett, P.L., Frean, M.R., 2000. Boosting algorithms as gradient descent, in: Advances in neural information processing systems, pp. 512–518.

Masson, V., 2000. A physically-based scheme for the urban energy budget in atmospheric models. Boundary-Layer Meteorology 94, 357–397.

Masson, V., Grimmond, C.S.B., Oke, T.R., 2002. Evaluation of the town energy balance (teb) scheme with direct measurements from dry districts in two cities. Journal of Applied Meteorology 41, 1011–1026.

Mesonet, N., 2020. Flux network, downloaded from. http://www.nysmesonet.org/networks/flux.

Meyn, S.K., Oke, T., 2009. Heat fluxes through roofs and their relevance to estimates of urban heat storage. Energy and Buildings 41, 745 – 752.

Middel, A., Brazel, A., Kaplan, S., Myint, S., 2012. Daytime cooling efficiency and diurnal energy balance in Phoenix, Arizona, USA. Climate Research 54, 21–34.

Mitraka, Z., Chrysoulakis, N., Doxani, G., Del Frate, F., Berger, M., 2015. Urban surface temperature time series estimation at the local scale by spatial-spectral unmixing of satellite observations. Remote Sensing 7, 4139–4156.

Network, N.E.O., 2020a. Data product: Dp4.00200.001. provisional data downloaded from. https://www.neonscience.org/. Battelle, Boulder, CO, USA.

Network, N.E.O., 2020b. Provisional data downloaded from. https://www.neonscience.org/.

Nitis, T., Tsegas, G., Moussiopoulos, N., Gounaridis, D., Bliziotis, D., 2017. Satellite data based approach for the estimation of anthropogenic heat flux over urban areas, in: Themistocleous, K., Michaelides, S., Papadavid, G., Ambrosia, V., Schreier, G., Hadjimitsis, D.G. (Eds.), Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017), International Society for Optics and Photonics. SPIE. pp. 531 – 539.

Nordbo, A., Järvi, L., Vesala, T., 2012. Revised eddy covariance flux calculation methodologies – effect on urban energy balance. Tellus B: Chemical and Physical Meteorology 64, 18184.

Novack, T., Esch, T., Kux, H., Stilla, U., 2011. Machine learning comparison between WorldView-2 and QuickBird-2-simulated imagery regarding object-based urban land cover classification. Remote Sensing 3, 2263–2282.

Offerle, B., Grimmond, C., Fortuniak, K., Pawlak, W., 2006. Intraurban differences of surface energy fluxes in a central european city. Journal of applied meteorology and climatology 45, 125–136.

Offerle, B., Grimmond, C.S.B., Fortuniak, K., 2005. Heat storage and anthropogenic heat flux in relation to the energy balance of a central european city centre. International Journal of Climatology 25, 1405–1419.

Oke, T., 1988. The urban energy balance. Progress in Physical Geography: Earth and Environment 12, 471–508.

Oke, T., Cleugh, H., 1987. Urban heat storage derived as energy balance residuals. Boundary-Layer Meteorology 39, 233–245.

Oke, T., Spronken-Smith, R., Jáuregui, E., Grimmond, C., 1999. The energy balance of central Mexico City during the dry season. Atmospheric Environment 33, 3919 – 3930.

Olivo, Y., Hamidi, A., Ramamurthy, P., 2017. Spatiotemporal variability in building energy use in New York City. Energy 141, 1393–1401.

Parlow, E., 2003. The urban heat budget derived from satellite data. Geographica Helvetica 58, 99–111.

Parlow, E., Vogt, R., Feigenwinter, C., 2014. The urban heat island of Basel - seen from different perspectives. Die Erde; Zeitschrift der Gesellschaft für Erdkunde zu Berlin 145, 1–2.

Pearlmutter, D., Berliner, P., Shaviv, E., 2005. Evaluation of urban surface energy fluxes using an open-air scale model. Journal of Applied Meteorology 44, 532–545.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Peng, J., Loew, A., Merlin, O., Verhoest, N.E., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. Reviews of Geophysics 55, 341–366.

Piringer, M., Grimmond, C., Joffre, S., Mestayer, P., Middleton, D., Rotach, M., Baklanov, A., De Ridder, K., Ferreira, J., Guilloteau, E., Karppinen, A., Martilli, A., Masson, V., Tombrou, M., 2002. Investigating the surface energy budget in urban areas - recent advances and future needs. Water, Air, & Soil Pollution: Focus 2, 1–16.

Prettenhofer, P., Louppe, G., 2014. Gradient boosted regression trees in scikit-learn. PyData 2014.

Ramamurthy, P., Bou-Zeid, E., 2017. Heatwaves and urban heat islands: a comparative analysis of multiple cities. Journal of Geophysical Research: Atmospheres 122, 168–178.

Ramamurthy, P., Pardyjak, E.R., 2011. Toward understanding the behavior of carbon dioxide and surface energy fluxes in the urbanized semi-arid Salt Lake Valley, Utah, USA. Atmospheric environment 45, 73–84.

Ranney, K.J., Niemann, J.D., Lehman, B.M., Green, T.R., Jones, A.S., 2015. A method to downscale soil moisture to fine resolutions using topographic, vegetation, and soil data. Advances in Water Resources 76, 81 – 96.

Rigo, G., Parlow, E., 2007. Modelling the ground heat flux of an urban area using remote sensing data. Theoretical and Applied Climatology 90, 185–199.

Roberts, D.A., Quattrochi, D.A., Hulley, G.C., Hook, S.J., Green, R.O., 2012. Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) decadal survey mission. Remote Sensing of Environment 117, 83–101.

Roberts, S., Oke, T., Voogt, J., Grimmond, C., Lemonsu, A., 2020. Energy storage in a european city center. CD Proceedings, 5th International Conference on Urban Climate .

Roberts, S.M., Oke, T.R., Grimmond, C.S.B., Voogt, J.A., 2006. Comparison of four methods to estimate urban heat storage. Journal of Applied Meteorology and Climatology 45, 1766–1781.

Robinzonov, N., Tutz, G., Hothorn, T., 2012. Boosting techniques for nonlinear time series models. AStA Advances in Statistical Analysis 96, 99–122.

Roth, M., Oke, T., 1994. Comparison of modelled and "measured' heat storage in suburban terrain. Beitr. Phys. Atmosph. 67, 149–156.

Sailor, D.J., 2011. A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment. International journal of climatology 31, 189–199.

Sailor, D.J., 2014. Risks of summertime extreme thermal conditions in buildings as a result of climate change and exacerbation of urban heat islands. Building and Environment 78, 81–88.

Sailor, D.J., Lu, L., 2004. A top–down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas. Atmospheric environment 38, 2737–2748.

Schmit, T.J., Gunshor, M.M., 2020. Chapter 4 - ABI imagery from the GOES-R series, in: Goodman, S.J., Schmit, T.J., Daniels, J., Redmon, R.J. (Eds.), The GOES-R Series. Elsevier, pp. 23 – 34.

Schmit, T.J., Lindstrom, S., Gerth, J., Gunshor, M.M., 2018. Applications of the 16 spectral bands on the Advanced Baseline Imager (ABI). J. Operational Meteor. 6, 33–46.

Schonlau, M., 2005. Boosted regression (boosting): An introductory tutorial and a stata plugin. The Stata Journal 5, 330–354.

Semmens, K.A., Anderson, M.C., Kustas, W.P., Gao, F., Alfieri, J.G., McKee, L., Prueger, J.H., Hain, C.R., Cammalleri, C., Yang, Y., et al., 2016. Monitoring daily evapotranspiration over two California vineyards using Landsat 8 in a multi-sensor data fusion approach. Remote Sensing of Environment 185, 155–170.

Singh, R.K., Irmak, A., 2009. Estimation of crop coefficients using satellite remote sensing. Journal of Irrigation and Drainage Engineering 135, 597–608.

Sun, T., Kotthaus, S., Li, D., Ward, H.C., Gao, Z., Ni, G.H., Grimmond, C.S.B., 2017a. Attribution and mitigation of heat wave-induced urban heat storage change. Environmental Research Letters 12, 114007.

Sun, T., Wang, Z.H., Ni, G.H., 2013. Revisiting the hysteresis effect in surface energy budgets. Geophysical Research Letters 40, 1741–1747.

Sun, T., Wang, Z.H., Oechel, W.C., Grimmond, S., 2017b. The Analytical Objective Hysteresis Model (AnOHM v1.0): methodology to determine bulk storage heat flux coefficients. Geoscientific Model Development 10, 2875–2890.

Taha, H., 1999. Modifying a mesoscale meteorological model to better incorporate urban heat storage: A bulk-parameterization approach. Journal of Applied Meteorology 38, 466–473.

Tsuang, B.J., 2005. Ground heat flux determination according to land skin temperature observations from in situ stations and satellites. Journal of Hydrometeorology 6.

Velasco, E., Lamb, B., Pressley, S., Allwine, E., Westberg, H., Jobson, B.T., Alexander, M., Prazeller, P., Molina, L., Molina, M., 2005. Flux measurements of volatile organic compounds from an urban landscape. Geophysical Research Letters 32.

Velasco, E., Pressley, S., Grivicke, R., Allwine, E., Coons, T., Foster, W., Jobson, B.T., Westberg, H., Ramos, R., Hernández, F., Molina, L.T., Lamb, B., 2009. Eddy covariance flux measurements of pollutant gases in urban mexico city. Atmospheric Chemistry and Physics 9, 7325–7342.

Wang, Z.H., 2014. A new perspective of urban–rural differences: The impact of soil water advection. Urban Climate 10, 19 – 34.

Wei, Z., Meng, Y., Zhang, W., Peng, J., Meng, L., 2019. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. Remote Sensing of Environment 225, 30 – 44.

Wickham, J., Homer, C., Vogelmann, J., McKerrow, A., Mueller, R., Herold, N., Coulston, J., 2014. The Multi-Resolution Land Characteristics (MRLC) Consortium — 20 years of development and integration of USA national land cover data. Remote Sensing 6, 7424–7441.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R., Verma, S., 2002. Energy balance closure at FLUXNET sites. Agricultural and Forest Meteorology 113, 223 – 243. FLUXNET 2000 Synthesis.

Xian, G., Crane, M., 2006. An analysis of urban thermal characteristics and associated land cover in Tampa Bay and Las Vegas using Landsat satellite data. Remote Sensing of Environment 104, 147 – 156. Thermal Remote Sensing of Urban Areas.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., Xian, G., 2018a. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. ISPRS Journal of Photogrammetry and Remote Sensing 146, 108 – 123.

Yang, L., Zhang, X., Liang, S., Yao, Y., Jia, K., Jia, A., 2018b. Estimating surface downward shortwave radiation over China based on the gradient boosting decision tree method. Remote Sensing 10.

Yoo, C., Im, J., Park, S., Quackenbush, L.J., 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.

Zhou, W., Qian, Y., Li, X., Li, W., Han, L., 2014. Relationships between land cover and the surface urban heat island: seasonal variability and effects of spatial and thematic resolution of land cover data on predicting land surface temperatures. Landscape ecology 29, 153–167.

Zhou, Y., Shepherd, J.M., 2010. Atlanta's urban heat island under extreme heat conditions and potential mitigation strategies. Natural Hazards 52, 639–668.

Şahin, M., 2012. Modelling of air temperature using remote sensing and artificial neural network in Turkey. Advances in Space Research 50, 973 – 985.

List of Figures