Towards High Resolution Urban Heat Analysis: Incorporating Thermal Drones to Enhance Satellite Based Urban Heatmaps

Bryan Rickens
Navid Hashemi Tonekaboni **
rickensbj@g.cofc.edu
hashemin@cofc.edu
Department of Computer Science, College of Charleston, Charleston, South Carolina, USA

ABSTRACT

As remote-sensing becomes more actively utilized in the environmental sciences, our research continues the efforts in adapting smart cities by using civilian UAVs and drones for land surface temperature (LST) analysis. Given the increased spatial resolution that this technology provides as compared to standard satellite measurements, we sought to further study the urban heat island (UHI) effect – specifically when it comes to heterogeneous and dynamic landscapes such as the Charleston peninsula. Furthermore, we sought to develop a method to enhance the spatial resolution of publicly available LST temperature data (such as those measured from the Landsat satellites) by building a machine learning model utilizing remote-sensed data from drones. While we found a high correlation and an accurate degree of prediction for areas of open water and vegetation (respectively), our model struggled when it came to areas containing highly impervious surfaces. We believe, however, that these findings further illustrate the discrepancy between high and medium spatial resolutions, and demonstrate how urban environments specifically are prone to inaccurate LST measurements and are uniquely in need of an industry pursuit of higher spatial resolution for hyperlocal environmental sciences and urban analysis.

KEYWORDS

Remote sensing, smart cities, urban heat island, land surface temperature, Landsat, UAV, drone, high spatial resolution.

1 INTRODUCTION

As climate change becomes an increasingly pressing issue in our society, the need for higher resolution measurements in environmental studies has become paramount [18]. High resolution, as defined in the industry, are any measurements with a spatial resolution within 5 meters per pixel (that is, measurements where each individual pixel covers an area no larger than 5^2 meters) [5]. Urban Heat Islands (UHIs) [2], especially those with dynamic and heterogeneous landscapes such as the one conducted in this research, are especially in need of such higher resolution data [17, 22]. Contemporary measurements gathered in the United States with spatial resolutions at 30 meters covering an area of 900^2 (such as Landsat satellites [11][12] and NLCD land cover classifications [4]) may be too wide to accurately condense the more dynamic landscapes in the country.

Consumer drones and unmanned aircraft vehicles (UAVs) have proven to be an effective solution to the high resolution problem due to their close proximity to the earth's surface, and have seen increased use lately in the environmental sciences. Until publicly available high resolution measurements become industry standard, it is our hope to further demonstrate the efficacy of UAV measurements and use them to propose a solution to the high dimension problem in the meantime. This paper details our attempts at using machine learning to demonstrate the correlation between medium resolution Landsat satellite data and higher resolution UAV measurements, and ultimately to demonstrate the construction of models that can reliably predict the more accurate higher resolution LST measured from UAVs using publicly available Landsat data. To fully realize the potential of smart cities, high spatial resolutions such these are paramount when it comes to analyzing the effects of climate change.

2 RELATED WORK

The need for accurate and readily available LST measurements has been a well discussed topic in the environmental sciences. The method of obtaining them via satellites from space continues to be the industry standard, although this comes with the caveat of emissivity and atmospheric corrections that threaten to be ill-poised [9]. In an effort to circumvent this issue, many recent studies have begun using remote-sensing via civilian UAVs and drones in their research (specifically in measuring LST). Recent research has proven the efficacy of comparing drone and Landsat satellite measurements [1] and has further illuminated how the higher spatial resolution of drone measurements are better equipped at analyzing smaller areas more precisely as compared to satellites [16]. Most comparable to our research, others have similarly found that certain land cover types can have an extremely wide margin of difference in heterogeneous landscapes between the two methods of measurement [7]. While some studies have developed more robust methods of downsampling spatial resolutions for satellite data [15], we contribute to this research by developing further methods of machine learning to enhance satellite spacial resolution with the use of civilian UAV measurements. Since land cover classification has been documented to have a significant contribution on the urban heat island effect [21], we hope to introduce publicly available land cover classification data to increase model efficiency.

3 STUDY AREA AND MATERIALS

3.1 Study Area and Time Period

Harbor Walk is a heterogeneous area within the city of Charleston, South Carolina that presents a unique opportunity to study the implications of high spatial resolution in urban heat analysis. Within a short distance it consists of highly impervious city structures

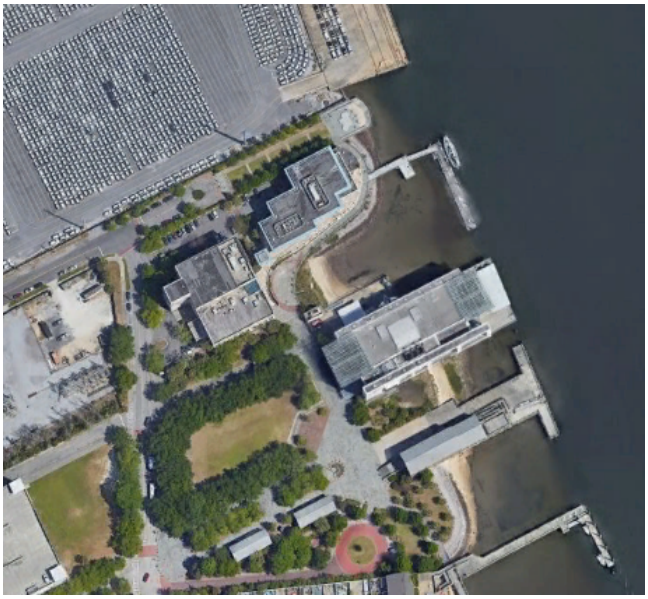

Figure 1: Harbor Walk, City of Charleston
Map Data ©2022 Google

like concrete buildings and streets, parks with open spaces and vegetation, and the harbor itself with open water adjacent to the Cooper River. While still exhibiting the greater urban heat island effect, it is a dynamic area with many various land cover types contained within a short distance that may not be as accurately analyzed with traditional medium-to-large spatial resolutions such as 30, 60, or 100 squared meters.

Data was taken starting from December of 2021, with intermittent samples gathered ranging until April of 2022. Unfortunately, numerous days had to be dropped from the data set due to numerous errors such as cloud cover over the study area and user error. The final dates included in the model are: December 14 2021, March 20 2022, and March 28 2022.

3.2 Satellite Data

Both Landsat 8 and Landsat 9 images were used in our comparative analysis of Harbor Walk. All samples were gathered from the USGS website. The thermal bands used during the measurements (10 and 11) have a spatial resolution of around 100 meters, and are then resampled during the process of generating the LST heatmap. Both Landsat satellites have a spatial resolution of 30 meters after this resampling process [11][12].

3.3 Drone (UAV) Data

The drone used in our research was a DJI Mavic 2 Enterprise UAV. It is a lightweight civilian drone capable of taking both thermal and non-thermal imagery. Below are the full specs for the thermal camera, including its resolution capability.

Figure 2: Drone Flight Path

Figure 3: Thermal Image Taken During Fig. 2's Flight Path (Before Orthorectification)

Table 1: M2E Camera Specs - DJI Mavic 2 Enterprise

Sensor	Uncooled VOx Microbolometer
Lens	HFOV: 57° Aperture: f/1.1
Sensor Resolution	160×120
Pixel Pitch	12 μm
Spectral Band	8-14 μm

The drone flights for each day were closely matched to the Landsat 8 and 9 acquisition schedules. Each flight began roughly before noon and lasted a total of 4 minutes each (2 minutes before and after the exact time when the satellite images were taken).

3.4 Land Use Land Cover (LULC) Data

Land Cover (LC) classification used in the research came from the NLCD 2019 Land Cover data set [4]. The LC types present in our data set, as defined by the NLCD legend, are:

Table 2: NLCD Legend

LULC Classification	Description
High Intensity	80% to 100% impervious surfaces
Medium Intensity	50% to 79% impervious surfaces
Water Intensity	Less than 25% cover of vegetation or soil

One caveat of including this specific land cover data set must be stated however: since the NLCD uses the 30 meter Landsat measurements as its base in LC classification, the high resolution problem is still inherently present in LC classification as well. When comparing LST measurements for differing LC types between Landsat and UAV data, it is highly effective in demonstrating the need for higher resolution analysis; however, its inclusion in our prediction model could prove to be ineffective. Ideally, an LC classification set at a higher resolution would have been preferable in predictive analyses.

4 METHODS

4.1 Satellite Preprocessing

LST measurements of the Harbor Walk area were derived using the Landsat 8-9 Collection 2 Level 2 [19] datasets. Both versions of Landsat satellites use the same TIRS sensors, meaning they both measure using the same thermal bands. This ensures that the measurements from both Landsat satellites are comparable and can ban be used together in the analysis. A small scaling formula [20] was required to take each Band 10 image capturing LST (measured in Kelvin, displayed in unsigned 16-bit integers) to Celsius:

$$[(UnsignedInt * 0.00341802) + 149.0] - 273.15$$

Each day was assessed for any possible cloud cover obstruction. Two days proved to be affected by cloud cover – March 4th, 2022 and April 13th, 2022 – and had to be removed from further analysis.

Rasterio [8], a library in Python specializing in geospatial raster data, was used in identifying the cells covering the Harbor Walk area for each specific day.

As Figure 4 demonstrates, each "cell" demonstrates each individual pixel from a Landsat raster image (a total area of $900m^2$ meters or 30 meters squared). This was found by locating the geospatial boundary for each corner of every pixel, which then allowed us to visualize the partition of area covered in a matrix. Cells are given arbitrary names (e.g., Cell 1) for the sake of analysis.

4.2 Drone (UAV) Preprocessing

An unorthodox method was required to take the temperature matrix given for each individual drone image to be further used in other mapping programs to produce a reliable orthomosaic. At the time of research, recent DJI models (such as the one used) required the use of third-party APIs to decode thermal metadata. The Python library Thermal Base [3] was used in decoding the DJI drone metadata. The temperature matrix output using this library (a matrix of each pixel's temperature value) was then saved as a .tiff file before finally matching each new image with their original image's metadata (using Exotiff [14]). Using the newly decoded images with the appropriate metadata, ArcGIS Pro [6] was used in constructing orthomosaics for each day.

Rasterio was used again to correctly match the area in each orthomosaic to the same 30x30 meter area captured in each Landsat image. This was to ensure that the exact same geographical area

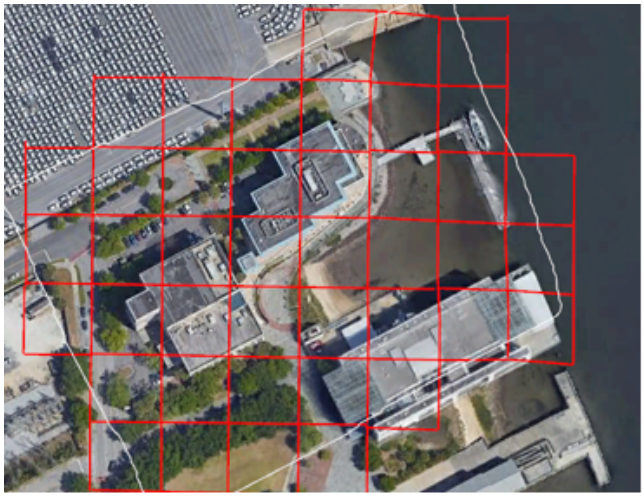

Figure 4: Cell Partition of Landsat Pixels Map Data ©2022 Google

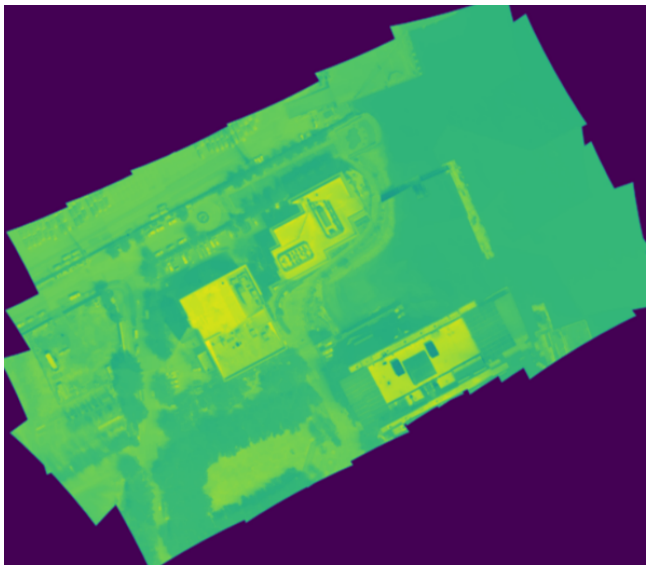

Figure 5: Orthomosaic of Drone Data

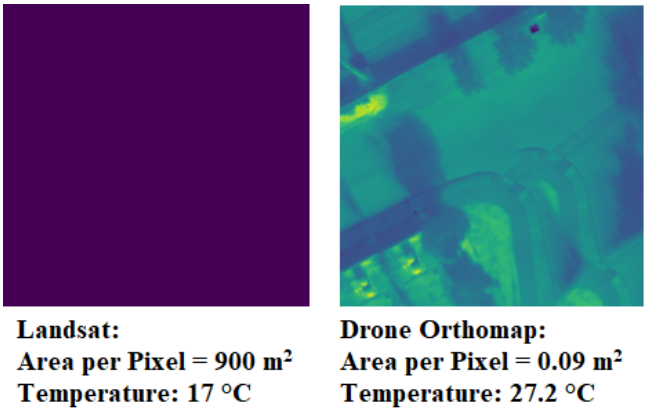

Figure 6: Area Coverage Comparison between Landsat and Drone Data

was analyzed between Landsat and UAV measurements. As Figure 6 demons rates, this allows us to take the same one-to-one area partition of a Landsat cell and a drone orthomosaic cell to then further analyze the LST between them.

4.3 Land Use Land Cover (LULC) Preprocessing

ArcGIS Pro was used in constructing the smaller LULC map of the Harbor Walk area (downgraded to the target area from the original NLCD dataset). Rasterio was then used applying the same method used on the drone orthomosaic to ensure the same cell area coverage was consistent for all sources. Each cell captured from the NLCD data was then encoded with its LC classification and added as a feature.

This, however, proved to be the greatest demonstration of the problem of using a medium spatial resolution for dynamic areas such as Harbor Walk. As compared to both the drone images and other satellite imagery, the LC classifications on the map do not align with the physical surfaces in actuality for numerous areas. This error was most commonly seen at the edge of LC transitions – for example, the waterfront where a High Intensity cell borders an Open Water one. In the case of the former, this would cause areas that are mostly Open Water in reality to be labeled as High Intensity. Another instance demonstrated general misclassification, where a building was labeled as Open Space. This problem is inherently an issue with the map provided by the NLCD, a consequence of the medium spatial resolution used in Landsat measurements. It should be noted that this same dataset is used in various research studies in the field, even despite the discrepancies noted above. Unfortunately, to deal with this issue, multiple cells used in the analysis were dropped to account for the inaccuracy in comparison to high resolution analysis. Figure 8 shows a map of the cells dropped.

4.4 Model Construction

Model construction began with the splitting of the data into a training set – data from 12/14/2021 and 03/28/2022 – and a testing set – data from 03/20/2022. The split was done this way to ensure

Figure 7: Map of Land Cover Classification for Harbor Walk

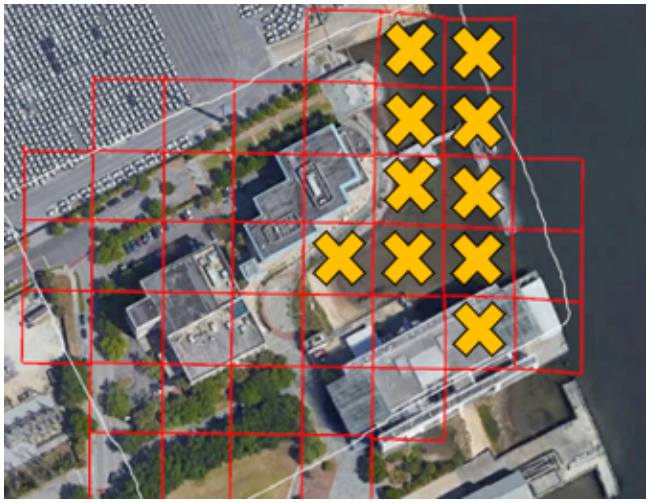

Figure 8: Removed Cells due to NLCD Misclassification Map Data ©2022 Google

the model would not suffer from extrapolation (considering the final day of measurements had the highest temperatures).

Due to the small number of available data and the shape of the correlation, a standard ordinary least squares (OLS) linear regression model was used as the basis for our prediction (using the scikit-learn [13] library in Python);however, multiple algorithms were tested overall. Using Drone LST as our target variable with Landsat LST and LULC classification as predictors, we hoped to demonstrate the potential practicality of using available civilian UAV LST measurements to increase the resolution of (i.e., more

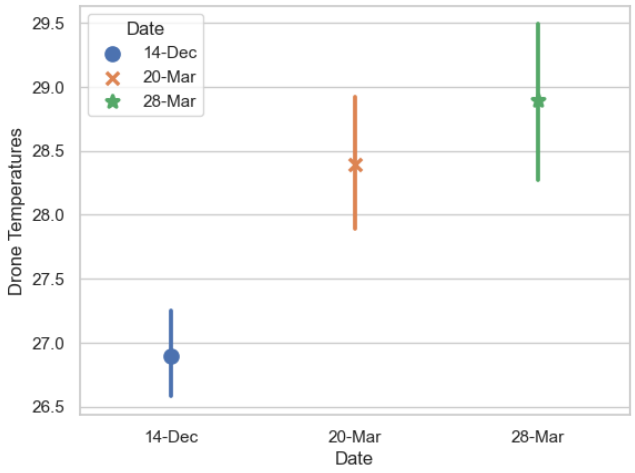

Figure 9: Drone Temperatures by Date

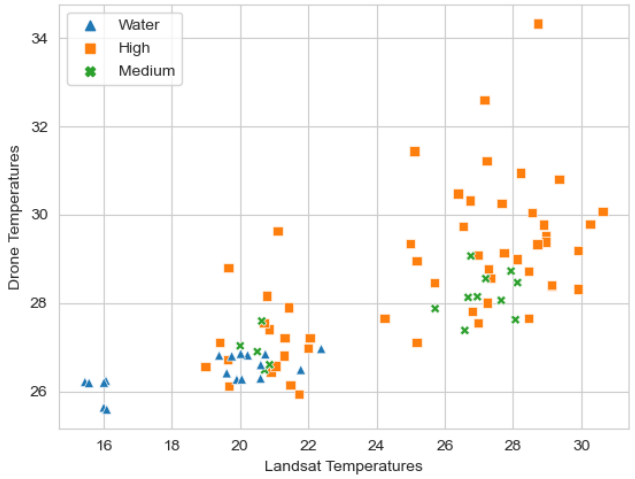

Figure 10: Landsat Temperatures by Drone Temperatures (Grouped by LC)

accurately predict) publicly available Landsat data sets. LULC classification was one-hot encoded with Medium Intensity dropped to avoid multicolinearity.

5 RESULTS

5.1 LST Correlation

Figure 10 shows a scatter plot of Landsat LST by Drone LST. While the land covers for Water and Medium Intensity exhibit a close linear relationship, High Intensity land covers have a much wider range of error when comparing measurements.

This variation between LULC is especially apparent when analyzing the mean temperatures of each cell area across all days. Of particular note are the High Intensity labeled cells 8 and 18, which show significant difference with a number of other cells as seen in Figure 11. After closer inspection and as seen in Figure 12 , these

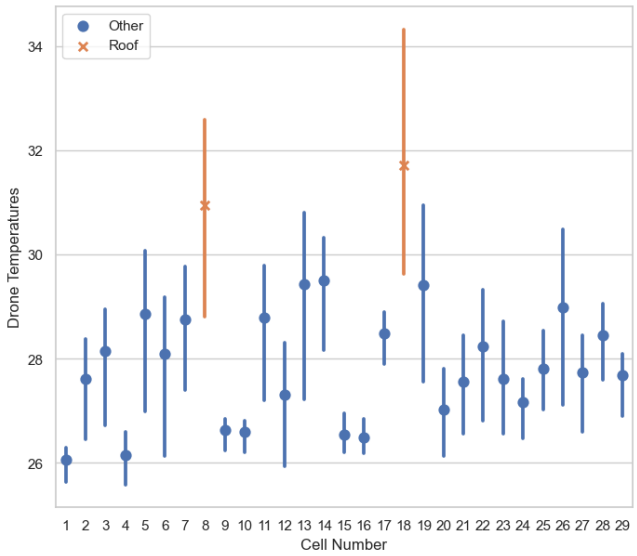

Figure 11: Error Plot of Drone Temperatures per Cell

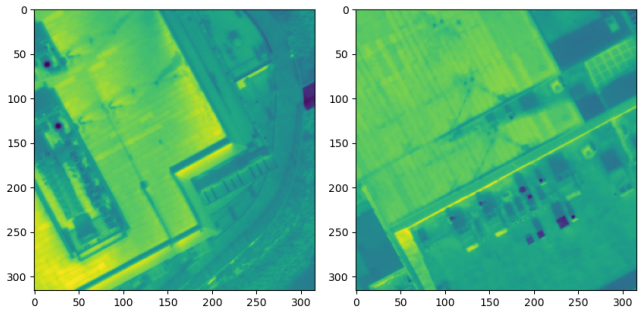

Figure 12: Cells 8 and 18 (Rooftops)

cells are actually the only areas that solely cover a rooftop area (which tracks why they would have such a higher LST than the rest due to the highly impervious surface area).

5.2 Model Prediction

The linear model resulted in an explained variance of 0.49 with a root mean squared error of 1.01 ° C. As seen in the table below, this model outperformed all other algorithms. We believe this to be a result of the shape of the data, which allows linear-based models to easily interpolate between ranges inside the training set. Particularly, interaction-based models such as Random Forest and XGBoost had the poorest results, which could be due to the low sample size of our data set. Noting the obvious linear relationship in the data, however, a linear-based model was the final one used in our efforts to enhance satellite resolution.

Using Shapley values (with the SHAP library in Python [10]) to assess feature importance, it is clear that Landsat Temp had the dominating influence on our linear model. In fact, removing LULC as a variable only resulted in a marginal difference (with an R^2 of 0.42 and an RMSE of 1.07).

Table 3: Model Results

Model Type	RMSE
Ordinary Least Squares	1.01
KNN	1.05
SVM (Linear Kernel)	1.04
Random Forest	1.20
XGBoost	1.57

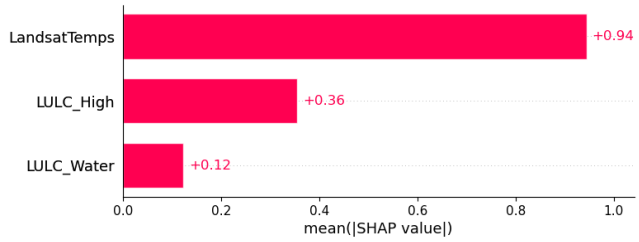

Figure 13: Shapley values for Model Variables

Following what was seen previously, residuals of cells of Open Water and Medium Intensity were much smaller than those of High Intensity.

Table 4: RMSE by LULC

LULC	RMSE
High	1.22
Medium	0.54
Water	0.49

With the validity of the model in question, Figures 14, 15, and 16 demonstrate key errors undermining model performance. Figure14 demonstrates the Actual vs Predicted values, and we see that High Intensity predictions follow no linear trend and exhibit high margins of error. Similarly, Figure 15 highlights the heteroscedasticity of High Intensity predictions. Finally, examining the residuals compared to the actual test values in Figure 16 demonstrates a persistent error of High Intensity predictions across a specific temperature threshold (with High Intensity values below 29 ° C undervalued, with values above this threshold being overvalued).

As discussed further below, while the predictive nature of our model is unreliable and ultimately cannot be used for its initial purpose, we believe it is demonstrative of the problem at hand and further highlights the need for higher spatial resolution measurements.

6 DISCUSSION

Especially for heterogeneous landscapes such as Charleston, Figure 10 is one of the many indications of the need for higher resolution measurements: while the locations with the labels Open Water and Medium Intensity within our data set had uniform geographical distribution suitable for 30x30 meter measurements, the majority

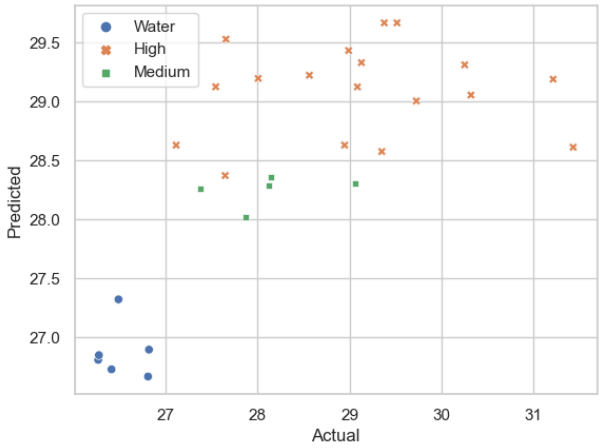

Figure 14: Actual vs Predicted Model Values

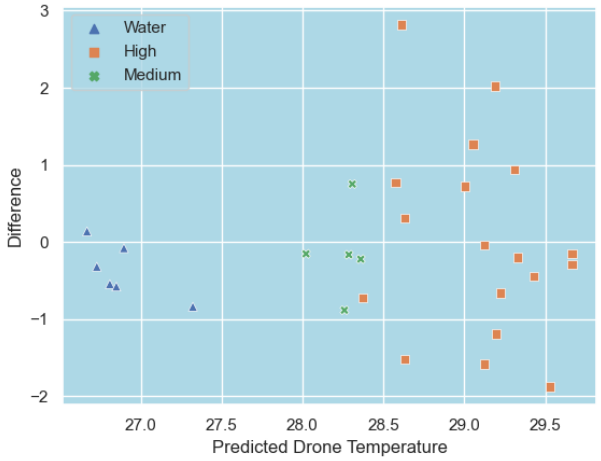

Figure 15: Residuals of Predicted Values

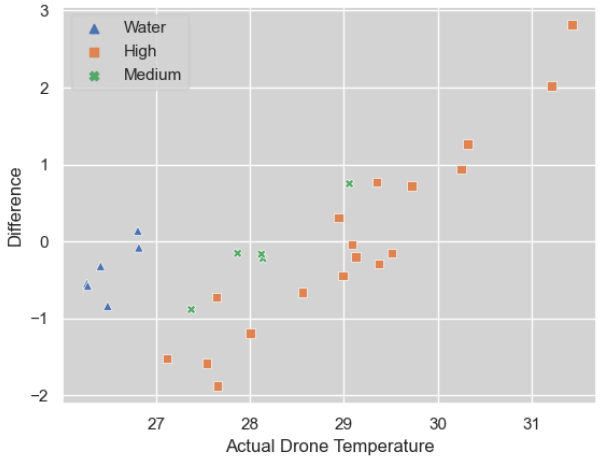

Figure 16: Residuals of Actual Values

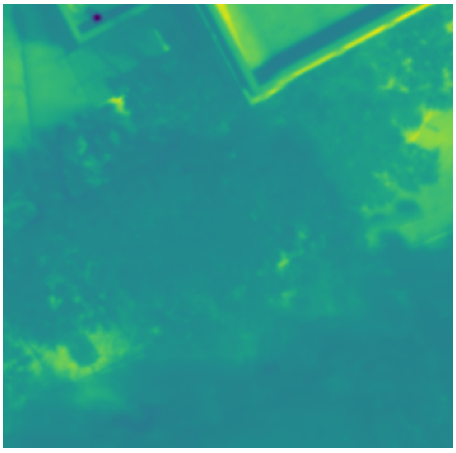

Figure 17: Cell 23 - High Intensity (Containing Heavy Vegetation)

of the cells conducted in our study (labeled as High Intensity) had varying degrees of impervious surfaces contained within, and were shown to have conflicting measurements when taken at a closer altitude with drones in comparison to Landsat satellites. Especially considering that publicly available land cover sets such as the NLCD were also classified using the same Landsat measurements, it is not unsurprising to see so many areas share a High Intensity classification. Various High Intensity cells, when analyzed with drone imagery, exhibited varying ratios of shared land covers within (such as cells that were purely impervious surfaces, to those that had a sizeable portion of vegetation or open water contained within). We believe this is demonstrative of how highly dynamic landscapes are not as accurately measured when using industry standard spatial resolutions.

Figures 17, 18, and 19 demonstrate cells of each land cover classification (arbitrarily numbered for reference) with some of the highest residuals – each exposing the presence of other land covers, most notably the edge where another land cover classification begins. Cell 23 of High Intensity included the border of a group of buildings and a street, but also contained a significant amount of vegetation and tree cover. Cell 24 of Medium Intensity contained mostly tree cover and an open park, but also included a small road at the edge. And lastly, Cell 15 of Open Water included the corner of a building in the edge along with the open harbor. Altogether, proven visually and with the high residuals, we believe this is demonstrative on the consequences of using a medium spatial resolution when classifying land cover of dynamic areas.

Further evidence of this can be seen in our attempts of enhancing Landsat resolution using predictive modeling. While higher resolution LST temperatures of open water and vegetation captured via drones share a stronger correlation with those of satellites, areas with a high degree of impervious surfaces exhibit more variance between the two and struggle the most in prediction. Figure 15 demonstrates the heightened error when dealing with these surfaces.

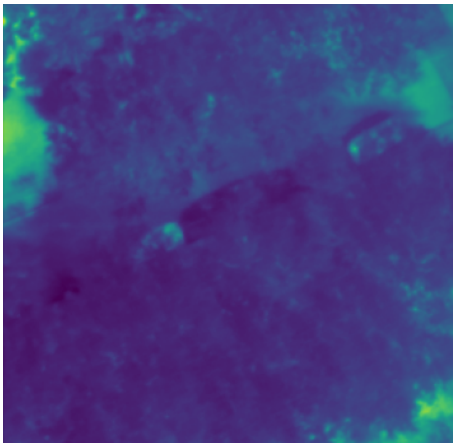

Figure 18: Cell 24 - Medium Intensity (Containing a Street in the Corner)

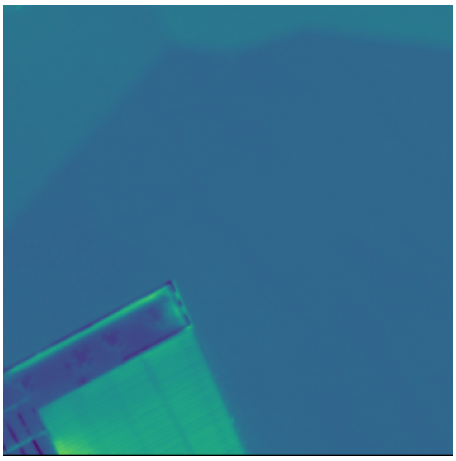

Figure 19: Cell 15 - Open Water (Containing the Edge of a Building)

Table 5: Dynamic Cells

LULC	Cell	Residual
High	23	1.58
Medium	24	0.88
Water	15	0.84

As discussed previously, the use of NLCD classifications of land covers using the same medium spatial resolutions as Landsat measurements further highlighted the disparity between satellite and drone resolutions. While land covers classified with open water and significant vegetation can be accurately enhanced, those classified with high degrees of impervious surfaces cannot. This limitation can have significant implications for monitoring and managing changes in urban land use. While ineffective for our model, we believe these results underline the repercussions of contemporary LULC classifications based on medium spatial resolutions.

7 FUTURE WORK/APPLICATIONS

For future work, we would like to conduct drone flights year round of the Harbor Walk area (both to acquire more data generally as well as to better represent an entire seasonal range of temperatures not seen in our research). With more data, we hope to improve our predictive models and to better represent an entire range of LST measurements. It is our ultimate goal to test the concept of this predictive model on an unseen area similar in geography to our research area, and to conclusively assess the practicality of using civilian UAVs and drones to enhance the resolution of Landsat-derived LST measurements. Also, to further research the unique problem of areas containing highly impervious surfaces exhibiting high variance between the two forms of measurement.

Most crucially, however, further research would benefit from alternative sources of land cover classifications (preferably those captured with a spatial resolution comparable to the one used in remote-sensing).

8 CONCLUSION

As seen in our results, highly dynamic and heterogeneous areas such as those in the Charleston peninsula exhibit varying correlation with publicly available LST measurements taken from a standard medium resolution. While areas with a uniform distribution of surface area (be it water or vegetation) exhibited a stronger linear relationship between drone and Landsat measurements, dynamic areas consisting of a range of land cover types within the same distance exposed increased variation between medium and high resolution measurements. Areas containing a large number of impervious surfaces have been shown to contain the most error when compared to higher resolution measurements captured with drones.

To better analyze hyperlocal urban heat island effects and their implications on the environment, government and other publicly available evaluations of land surface temperature need to adapt to higher resolutions with a smaller spatial areas per-pixel. Given our results indicating increased disparity specifically with areas containing mostly impervious structures, urban environments are uniquely error prone using traditional land surface temperature estimates. While we tried to enhance this resolution in the meantime with predictive modeling techniques, we hope to have shed further light on the issue and demonstrated the implications of the problem at large.

ACKNOWLEDGMENTS

We would like to thank Barbara Szwabowski for providing the foundation of this research (through collecting both Landsat and drone data as well as providing code used in preprocessing).

Landsat data was acquired through the USGS website (earthexplorer.usgs.gov). The NLCD data set of land cover classification was similarly acquired through the USGS website. Google Earth was used for visual representation of the study area and ArcGIS Pro was used in constructing the orthomosaics from the thermal images captured via drone. A special thank you is given to to Detect Technologies and Phil Harvey for their software (Thermal Base and Exiftool, respectively) used in circumventing the DJI decoding process for our specific drone model.

REFERENCES

[1] Awais, M., Li, W., Hussain, S., Cheema, M. J. M., Li, W., Song, R., and Liu, C. Comparative evaluation of land surface temperature images from unmanned aerial vehicle and satellite observation for agricultural areas using in situ data. *Agriculture 12*, 2 (2022), 184.

[2] Deilami, K., Kamruzzaman, M., and Liu, Y. Urban heat island effect: A systematic review of spatio-temporal factors, data, methods, and mitigation measures. *International Journal of Applied Earth Observation and Geoinformation 67* (2018), 30–42.

[3] Detect Technologies. Thermal basel, 2022.

[4] Dewitz, J. National land cover database (nlcd) 2019 products. Tech. rep., U.S. Geological Survey, 2021. https://doi.org/10.5066/P9KZCM54.

[5] EarthData, N. Spatial resolution, 2022.

[6] Esri Inc. Arcgis pro version, 2022.

[7] Garcia-Santos, V., Cuxart, J., Jiménez, M. A., Martinez-Villagrasa, D., Simo, G., Picos, R., and Caselles, V. Study of temperature heterogeneities at sub-kilometric scales and influence on surface–atmosphere energy interactions. *IEEE Transactions on Geoscience and Remote Sensing 57*, 2 (2018), 640–654.

[8] Gillies, S., et al. Rasterio: geospatial raster i/o for Python programmers, 2013–.

[9] Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I. F., and Sobrino, J. A. Satellite-derived land surface temperature: Current status and perspectives. *Remote sensing of environment 131* (2013), 14–37.

[10] Lundberg, S. M., and Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774.

[11] NASA. Landsat 8, 2022.

[12] NASA. Landsat 9, 2022.

[13] 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., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.

[14] Phil Harvey. Exiftool, 2022.

[15] Sánchez, J. M., Galve, J. M., González-Piqueras, J., López-Urrea, R., Niclòs, R., and Calera, A. Monitoring 10-m lst from the combination modis/sentinel-2, validation in a high contrast semi-arid agroecosystem. *Remote Sensing 12*, 9 (2020), 1453.

[16] Silvestri, M., Marotta, E., Buongiorno, M. F., Avvisati, G., Belviso, P., Bellucci Sessa, E., Caputo, T., Longo, V., De Leo, V., and Teggi, S. Monitoring of surface temperature on parco delle biancane (italian geothermal area) using optical satellite data, uav and field campaigns. *Remote Sensing 12*, 12 (2020), 2018.

[17] Tonekaboni, N. H., Kulkarni, S., and Ramaswamy, L. Edge-based anomalous sensor placement detection for participatory sensing of urban heat islands. In *2018 IEEE International Smart Cities Conference (ISC2)* (2018), IEEE, pp. 1–8.

[18] Tonekaboni, N. H., Ramaswamy, L., Mishra, D., Grundstein, A., Kulkarni, S., and Yin, Y. Scouts: A smart community centric urban heat monitoring framework. In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Advances on Resilient and Intelligent Cities* (2018), pp. 27–30.

[19] USGS. Earth explorer, 2022.

[20] USGS. Scale factor, 2022.

[21] Wang, W., Liu, K., Tang, R., and Wang, S. Remote sensing image-based analysis of the urban heat island effect in shenzhen, china. *Physics and Chemistry of the Earth, Parts a/b/c 110* (2019), 168–175.

[22] Yin, Y., Tonekaboni, N. H., Grundstein, A., Mishra, D. R., Ramaswamy, L., and Dowd, J. Urban ambient air temperature estimation using hyperlocal data from smart vehicle-borne sensors. *Computers, Environment and Urban Systems 84* (2020), 101538.