SUHII Project Internship Report

Michael Wang

September 2019

1 Objective

The objective of this internship project was to derive the surface urban heat island intensity of cities in the TraK project by implementing the method from Li, et al. 2018 [4] on Google Earth Engine.

2 Data and Methods

The intensity of the surface urban heat island usually refers to the difference in land surface temperature (LST) between non-urban and urban locations. The novelty of Li et al. 2018 [4] lies in the fact that it uses the documented relationship between LST and impervious surface areas (ISA) to fit a line between LST and ISA as a percent (%ISA). The slope of the resulting line would have the equation:

$$f(x) = mx + b$$

$$\Downarrow$$

$$f_{LST}(\%ISA) = m(\%ISA) + ISA_{0\%}$$

Where m is the slope of the line.

Since the intensity of the urban heat island (SUHII) can be defined as the surface temperature difference between urban and non-urban areas:

$$SUHII = f_{LST}(100) - f_{LST}(0)$$

$$\Downarrow$$

$$SUHII = (100m + ISA_{0\%}) - (0m + ISA_{0\%})$$

$$\Downarrow$$

$$SUHII = 100m$$

Thus SUHII is simply the slope of the fitted line multiplied by a constant of 100. The following data was used to derive this value for each city:

- A vector file of the region of interest (ROI)

- The World Settlement Footprint (WSF) 2015 data from DLR

- Landsat-derived Land Surface Temperature data from Parastatidis [3]

- JRC Global Surface Water Mapping Layers v1.1 from the European Commission/Google [2]

The ISA data was first processed using the kernel density estimation (KDE) technique, which was applied to account for the radiance effect of impervious surfaces, where the heat of an impervious pixel is not entire contained within, as it also radiates outward into other pixels. The temperature derived from the satellite LST data, would reflect effect, thus KDE was used to account for it. The specific radius of used for the KDE will affect the results (see section 4), but a radius of 3000 meters was used in line with the original paper. Afterwards, the entire image is normalized, such that 100% ISA refers to the most impervious areas, and 0% the least:

$$normalized = \frac{pixelvalue - minvalue}{maxvalue - minvalue}$$

The LST data is averaged at the 1km level and stacked with the max extent binary water data which was aggregated at the 1km level for a percentage. The ISA layer is then averaged at the 1km level and stacked with the other two layers. The entire stack is then clipped to the ROI, and pixels where there is more than 25% water is removed from further analysis due to the particular thermodynamics of water.

The pixels are then grouped using the ISA layer at 2% intervals, creating 50 "buckets" of data. This particular application did this by dividing each percentage by 2 and applying a floor function, before returning to the original range by multiplying by 2 again. For each bucket, the associated LSTs are averaged. Then a CSV with two columns is exported, with one with %ISA at 2% intervals, the other being the average LST for that particular interval.

The slope of the first order linear regression of these two variables would give the SUHII per the original method. For each city, the a different SUHII was derived for the summer months (May-September), winter months (October-April), and annually.

3 Results

The resulting files are then analyzed and visualized with python. A selection of results are shown here:

In general, the surface urban heat island effect is less intense in winter than in summer, with the annual being somewhere in the middle. The coefficient of determination (R^2) was also higher in summer than in winter, meaning the variances are better explained by the fitted linear regression in summer than in winter.

The SUHII value of each season for each city is appended to the original shape file as a new column through the use of geopandas.
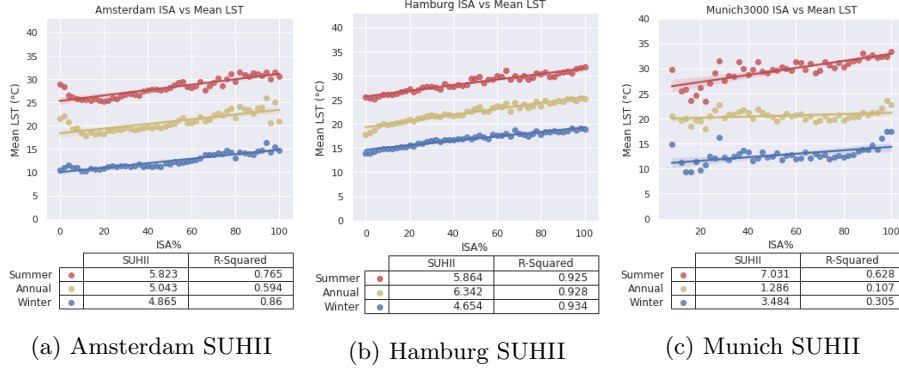
		SUHII	R-Squared
Summer	●	5.823	0.765
Annual	●	5.043	0.594
Winter	●	4.865	0.86

(a) Amsterdam SUHII

		SUHII	R-Squared
Summer	●	5.864	0.925
Annual	●	6.342	0.928
Winter	●	4.654	0.934

(b) Hamburg SUHII

		SUHII	R-Squared
Summer	●	7.031	0.628
Annual	●	1.286	0.107
Winter	●	3.484	0.305

(c) Munich SUHII

Figure 1: A Selection of European City and the plotted %ISA and LST

4 Discussion

The R^2 of the fitted line varies greatly between cities, probably due to the heterogeneous and idiosyncratic nature of the urban landscape. There are a few things of note which will be elaborated on in the following sections.

4.1 R^2 as a Sensitivity Test

The authors of the original method noted the radiant effect of impervious surfaces, and applied the kernel density estimation method to avvount for it. However, this method takes one user defined variable, in the form of the radius. As the author also notes, this will affect the results of the intensity analysis. In the original method, a sensitivity test was done, testing every radius from 1000m to 4000m at 500m intervals, and the radius which produced the results with the highest R^2 was chosen as the final product.

One issue with this is that R^2 is not a perfect a goodness of fit measure (some might say it's not one at all [1]). As seen in this particular data:

		SUHII	R-Squared
Summer	●	10.328	0.897
Annual	●	7.034	0.658
Winter	●	0.325	0.025
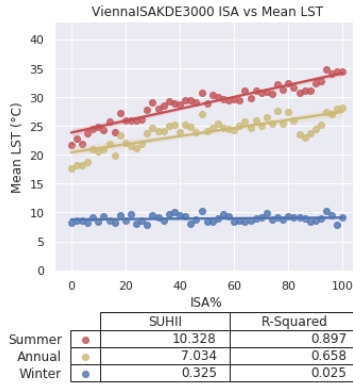
Despite looking like a somewhat reasonable line of fit, the R^2 is abysmally low. One way to think about R^2 is to see it how much better the model does at fitting the data, compared to a naive model (i.e. a straight line) as a percentage. Thus when there's a lot of variance within a data plot that's very flat, the first order linear model does not do much better than a naive straight line model, thus the really low R^2. Despite that, the line fitted is not as horrible as the R^2 would suggest. Its usefullness as the central variable in the radius sensitibity test is thus questionable.

4.2   Bias in Predictions

Another shortcoming of R^2 is its inability to say anything meaningful about any biases in predictions. It is still possible for an invalid model (e.g. a third order relationship modeled as a first ordered one) to produce high R^2 values.
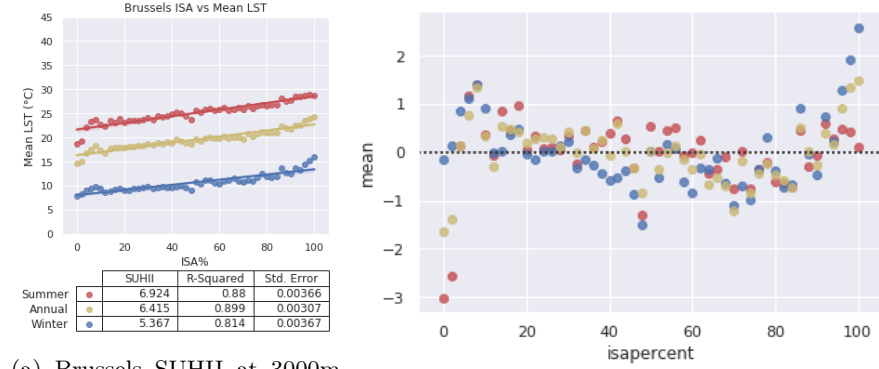
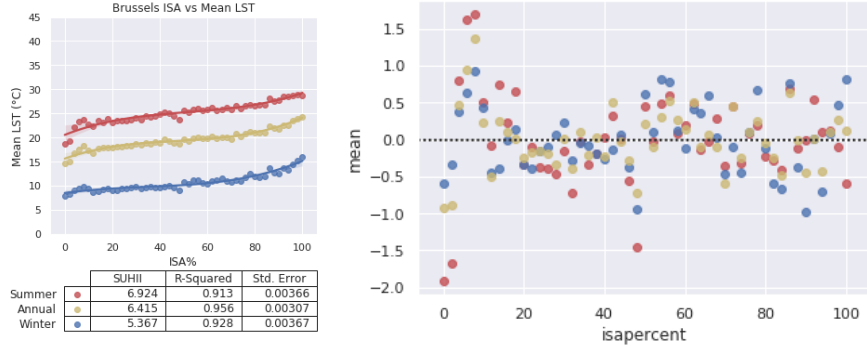(a) Brussels SUHII at 3000m KDE Radius

(b) Residual plot of the fitted linear regression

Seen above, the analysis using the city of Brussels produced fitted lines with very high R^2 values. However, as the residual plot shows, the data deviates from the model in a very regular pattern. Such a residual plot suggests that there's plenty of room for improvement for the model, and indeed, fitting third order polynomial linear regression instead creates much more sensible residual plots.

4.3   Further Outlook

None of this analysis invalidates this method's to obtain a surface heat island intensity, as the method is still a very reasonable way to produce the difference between urban and non-urban surface temperatures using solely remote sensing satellite data.

However, it is also clear that there is more interesting trends and effects at work. The precise relationship between impervious surface and land surface temperature detected isn't necessarily a clear linear relationship, nor should that be expected. The heterogenous of the urban landscape combined with the

(a) Brussels SUHII with 3rd order polynomial

		SUHII	R-Squared	Std. Error
Summer	●	6.924	0.913	0.00366
Annual	●	6.415	0.956	0.00307
Winter	●	5.367	0.928	0.00367

(b) Residual plot of the fitted polynomial linear regression

radiance effect of all kinds of materials makes the relationship very complicated, with even begining to consider the third dimension that's also a primary driver of the urban landscape.

Despite that, the results we got from the method on a variety of different (primarily European) cities still showed a clear positive trend (also negative in some cases), and that trend deviates from a fitted first order linear line in a clear and consistent way, across KDE radiuses and seasons. The precise underlying mechanics, in terms of both thermodynamics as well as the detection thereof, is not completely understood, but it could provide variable and intriguing information about the urban temperature landscape, as well as its management.

For those reasons, this topic can be further investigated into as a Master's thesis topic. The main points of entry for research are:

- Up-to-date research on the Urban Heat Island effect

- Basic mechanics of how LST data is derived

- The radiance effect used in this particular method

- Applicable statistical analyses for these types of data

    - i.e. going beyond R^2

The current expected outcomes are:

- Find the most reasonable and applicable kernel radius

- Extract more meaningful information from the trend between %ISA and KDE

    - i.e. going beyond the ΔT
    - Ideally, to be able to classify cities based on their surface thermo-characteristics

- Creating a model which can more accurately show the relationship between %ISA and LST

5   Description of Delivered Data

The shapefiles of the following cities are appended with three additional columns of data named *SummerSUHII*, *WinterSUHII*, and *AnnualSUHII*:

- Africa

  - Addis Ababa
  - Casablanca
  - Johannesburg
  - Nairobi
  - Tshwane

- Americas

  - Chicago
  - Medellin
  - Montreal
  - Pheonix
  - Portland
  - Santiago Di Chile
  - Vancouver

- Asia

  - Ankara
  - Beijing
  - Delhi
  - Dubai
  - Hong Kong
  - Izmir
  - Jerusalem
  - Niigata
  - Seoul
  - Shizuoka
  - Singapore
  - Taipei

- – Tehran

- Australia

  - – Melbourne
  - – Sydney

- Europe

  - – Amsterdam
  - – Athens
  - – Berlin
  - – Birmingham
  - – Brussels
  - – Budapest
  - – Copenhagen
  - – Dublin
  - – Geneva
  - – Glasgow
  - – Gothenburg
  - – Hamburg
  - – Helsinki
  - – Lisbon
  - – London
  - – Madrid
  - – Milan
  - – Munich
  - – Oslo
  - – Paris
  - – Prague
  - – Rome
  - – Strasburg
  - – Tallinn
  - – Turin
  - – Vienna
  - – Warsaw
  - – Zuerich

References

1. Shalizi, C. *Lecture 10: F-Tests, R2, and Other Distractions* Online. Oct. 2015. `http://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/10/lecture-10.pdf` (2019).

2. Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540,** 418–422. ISSN: 0028-0836, 1476-4687. `http://www.nature.com/articles/nature20584` (2019) (Dec. 2016).

3. Parastatidis, D., Mitraka, Z., Chrysoulakis, N. & Abrams, M. Online Global Land Surface Temperature Estimation from Landsat. *Remote Sensing* **9,** 1208. ISSN: 2072-4292. `http://www.mdpi.com/2072-4292/9/12/1208` (2019) (Nov. 23, 2017).

4. Li, H. *et al.* A new method to quantify surface urban heat island intensity. *Science of The Total Environment* **624,** 262–272. ISSN: 00489697. `https://linkinghub.elsevier.com/retrieve/pii/S0048969717334186` (2019) (May 2018).