A Bayesian analysis of the urban heat effect

T. M. Logana

# Abstract

In this study I investigate the urban heat effect across five US cities. The urban heat affect is important to understand and reduce due to the threat of heat waves. Heat waves are the deadliest of the natural disasters and are very likely (according to the IPCC’s latest assessment report) to be longer and more frequent. High urban temperatures urgently need to be addressed, to reduce our vulnerability to these events. I test a series of hypotheses regarding the effect of greenspace, water bodies, and impervious surfaces on land surface temperature and train predictive models. Understanding whether the hypotheses are consistent across the cities is important for understanding and adaptation decisions. Capable predictive models allow us to see the effect changing the biophysical will have on land surface temperature.

# Introduction

The urban heat island occurs in urban areas where changes in the built environment result in heat being trapped (1). This can be due to increases in heat storing impervious material, the emission of heat from air conditioning units or automobiles, or fine, heat capturing, particulates which are the result of anthropogenic pollution (1). Normally, higher urban temperatures are simply a nuisance at worst. However, they are a dormant threat due to how they exacerbate heat waves. Heat waves occur when temperature is abnormally elevated for a period of days (1). This can be exacerbated by the urban heat island, where already high temperatures are pushed to the extreme. While not as well publicized as some of the more dramatic natural disasters, they are the deadliest and result in large loss of life. The 1995 heat wave in Chicago killed more than 700 people (2). With our current emission profile, forecasts are anticipating heat waves like the 1995 Chicago inferno to be an annual occurrence by 2080 (2). There is an urgent need to understand and adapt our cities to mitigate this threat.

# Data

The data used in this analysis and sources is described in Appendix A. The data that I calculated from land satellite images (land surface temperature (LST), vegetation index (NVDI), and albedo) were averaged across four satellite images over summer months in 2012-2014. Land surface temperature was normalized by subtracting the mean of the city. Once calculated, the data was processed into a grid of 2000 ft square cells for the city. This data is mostly geostatistical data or raster. This data is processed from discrete locations which have a recorded value, e.g. rainfall or temperature gauges across the city. The values are processed by using the geospatial points within the grid, and calculating the cell’s mean, max, and min for each of the variables. The land use cover data was processed so that an area within the cell for each land use type was reported. Finally, for each of the covariates processed, a spatially lagged value is assigned to the cell. The spatially lagged value is simply the average of the surrounding (queen’s move) cells, and is analogous to the time lagging variables in time-series analysis.

# Methodolgy

## Bayesian Regression

When testing the hypotheses proposed my aim is to calculate the posterior of linear regression coefficients, with which I can base any conclusions. The first hypothesis questions how impervious surfaces affect the land surface temperature. To begin this analysis, I preform Bayesian regression on the entire dataset (ignoring the city of the data). The regression is simply

I use weakly informative priors for the intercept and impervious surface coefficient based on Gelman (3). He recommends using a Cauchy prior (0,2.5) for the slope coefficient, because the effect of covariates is usually small, but occasionally large. A Cauchy distribution has a long tail to capture this. The intercept prior is Cauchy(0,10).

The interest is also on how the cities differ, so I repeated this analysis subsetting the data by city. Instead of using the weakly informative prior, I used the Cauchy distribution with a mean equal to the posterior from all cities. Using this prior is double-counting the data, however if the posteriors of the different cities diverge, this is a strong rebuke to the assumption of city similarity.

This analysis was repeated on the second hypothesis regarding how albedo influences land surface temperature at high impervious surface regions. The data for this analysis includes only the cells where the mean impervious surface percentage exceeds 75%.

## Predictive Modelling

The second part of the analysis required predictive modelling of the land surface temperature using biophysical covariates. A Bayesian Additive Regression Tree (BART (1)) was used to predict the land surface temperature. The BART model is a tree based ensemble model. Regression trees are grown on the data and based on weakly informative priors. The resulting prediction is an average of the result of many trees. A spatially independent BART model was trained on four of the cities and tested on Baltimore.

# Results and Discussion

## Bayesian Regression

The posteriors of the regression coefficients are shown in Figure 1 and Figure 2. Figure 1 shows the distribution of the effect percentage impervious surfaces has on land surface temperature within a cell. There is a clear difference between the affect within the different cities, which this more significant given the prior distribution expected them to be near the posterior for the “all” city effect. Figure 2 shows the effect of albedo on land surface temperature in areas with high percentage impervious surface. This is interesting because the posteriors are distributed around zero. The all city case indicates that the whiter an impervious surface is, the cooler it is. This is a fundamental assumption within the interested community. However, this is not the case given this data and regression approach. Some cities indicate that the whiter a surface is, the hotter it is. To investigate the influence of the prior, Figure 3 (Appendix C) shows the posterior distributions with a weak Cauchy(0,2.5) prior (1) and the conclusions drawn are similar. The trace and convergence plots which demonstrate the model converged are included in Appendix C.

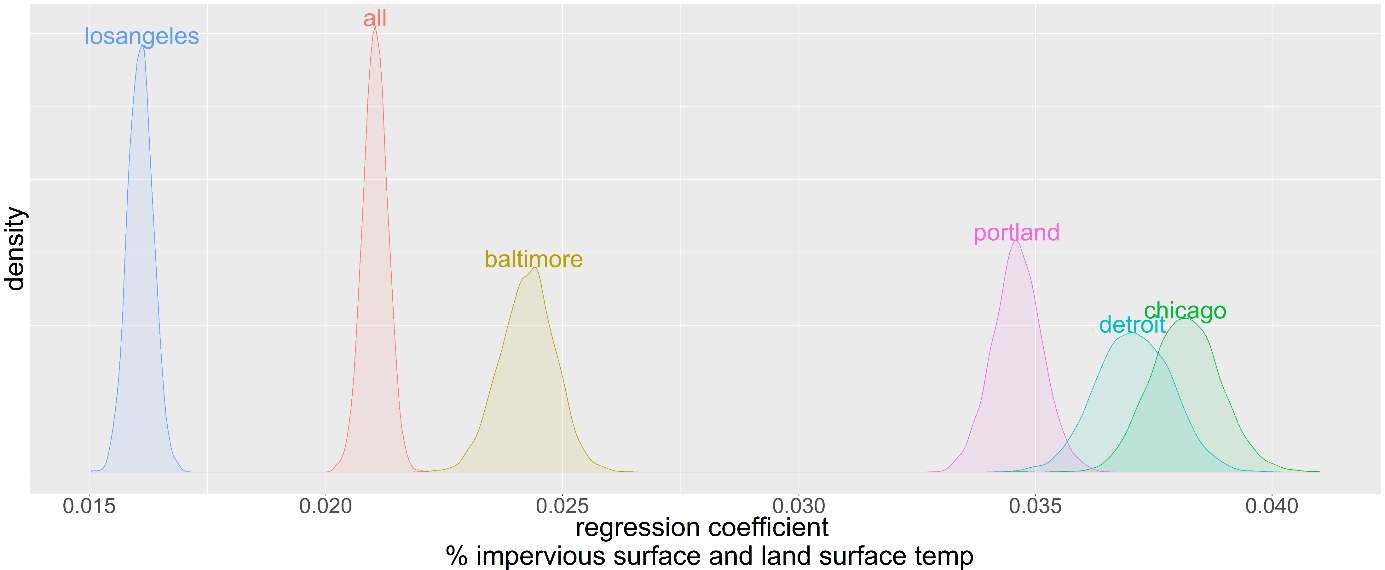


Figure 1. Posterior distributions for the regression coefficient for the effect of percentage impervious surface on land surface temperature. An informative Cauchy prior based on the "all" city posterior was used.

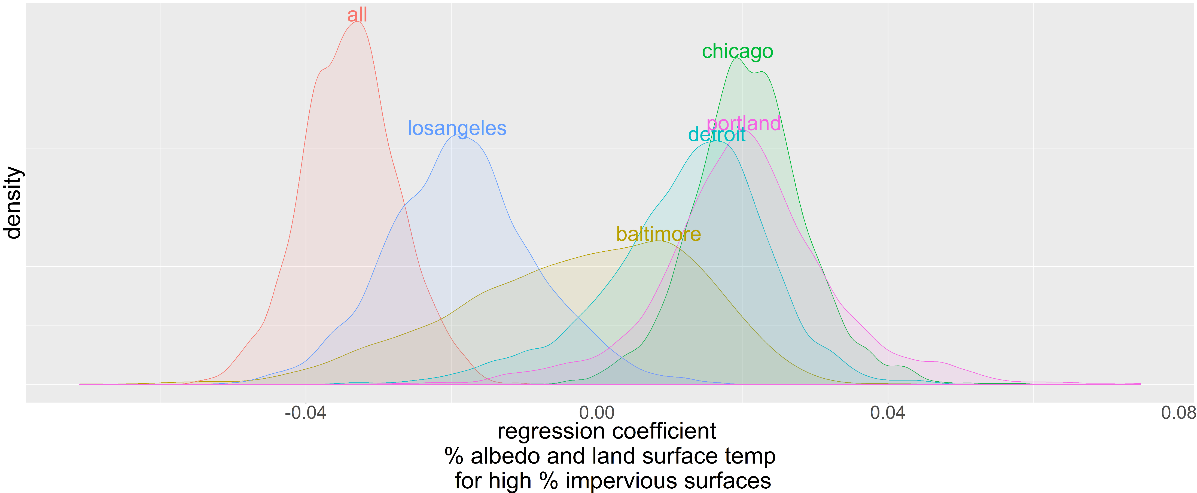


Figure 2. Posterior distributions for the regression coefficient for the effect of albedo on land surface temperature. An informative Cauchy prior based on the "all" city posterior was used.

## Predictive Modelling

The primary goal of this predictive modelling is not to correctly predict the absolute land surface temperature, but rather, accurately predict the land surface temperature relative to its neighbors. This allows the cities to identify areas of risk. The true land surface temperature distribution for Baltimore city is shown in Figure 3a. The predicted land surface temperatures from the BART, spatially independent, is shown in Figure 3b. While the absolute difference in temperatures is incorrect by approximately 1oC, the spatial pattern appears to be captured by the BART. This performance of the models raises concerns about the potential of overfitting. The variable importance plot from a random forest model indicate the model is utilizing information from NVDI, impervious surface, and albedo primarily. NVDI and albedo were calculated using the same landsat images, albeit different bands. I retrained the model without these two variables to see how the predictions were impacted, and they were not. Impervious surface was, in this model, the most important variable. I removed this and retrained the data. The resulting predictive map and variable importance is shown in Figure 4. Surprisingly the predictive accuracy isn’t terrible. The variable importance indicates a reliance on percentage tree canopy (another remotely sensed variable), but also the land cover data.

|  |  |
| --- | --- |
|  |  |
| (a) true | (b) BART predicted |

Figure 3. The true and predicted land surface temperatures for Baltimore City.

|  |  |
| --- | --- |
|  |  |

Figure 4. Predicted land surface temperature in Baltimore City and the variable importance plot with the impervious surface, NVDI, and albedo variables excluded.

# Conclusion and Further Work

This paper is a preliminary report on a journal article which is in progress. Here I present the results of investigating two of the hypotheses I tested using Bayesian data analysis, as well as the early results of the predictive modelling process. The list of hypothesis I will test in the final paper are included in Appendix C.

The results of analyzing the hypotheses support the need for further investigating. We see that the hypothesis reached from one city is not consistent with the behavior in another. This is a potential issue for a field which base much of its work on the generalizability of the results, and focus much of their work on single case-study cities. In addition to investigating the effect of the priors on the existing regressions and analyzing the other hypothesis, I plan to address the issue of the threshold value of high imperviousness. I defined high imperviousness as greater than 75%, but the results are sensitive to this.

The BART predictive model is looking like a capable predictive model; however, it raises questions regarding the use of remote sensing data. Is this simply using proxy data to infer other proxy data? I plan to improve variable selection in the predictive model, as well as test and compare different models, not on error to the actual LST, but to the relative LST, which is what we are interested in.

Further work to understand the urban heat affect and how it is affected by the built environment will hopefully contribute to planners’ ability to reduce it, and mitigate effects of heat waves to our urban communities.

# References

1. Harlan S, Brazel A, Jenerette G, et al. In the shade of affluence: the inequitable distribution of the urban heat island. Research in social [Internet]. 2007; Available from: http://www.emeraldinsight.com/doi/abs/10.1016/S0196-1152(07)15005-5

2. Klinenberg E. Heat wave: A social autopsy of disaster in Chicago. 2015; Available from: https://books.google.co.nz/books?hl=en&lr=&id=LV6zBwAAQBAJ&oi=fnd&pg=PR7&dq=heat+wave+klinenberg&ots=6NgHf21wN7&sig=4BEyUZVv1f9GkgKsCLG1Nad3MPA

3. Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat. 2008 Dec;2(4):1360–1383.

1. Data

|  |  |
| --- | --- |
| Land surface temperature | Calculated from band 8 of the land satellite images of the USGS. |
| Normalized difference vegetation index (NVDI, a measure of living green vegetation) | Calculated from bands 1-4 of the land satellite images. The same images were used here as the ones used to calculate LST. |
| Albedo (a measure of whiteness of a surface) | Calculated from bands 1-4 of the land satellite images. The same images were used here as the ones used to calculate LST and NVDI. |
| Percentage impervious surface | These datasets are also calculated using remote sensing, although are calculated and produced every five years by the Multi-Resolution Land Characteristics (MRLC) consortium (which is a group of US government agencies). |
| Land cover | This data is provided by USGS from surveys |
|  |  |
| Tree canopy cover | The MRLC also produces tree canopy cover from remote sensing data |
| Elevation |  |

1. Bayesian regression diagnostic plots

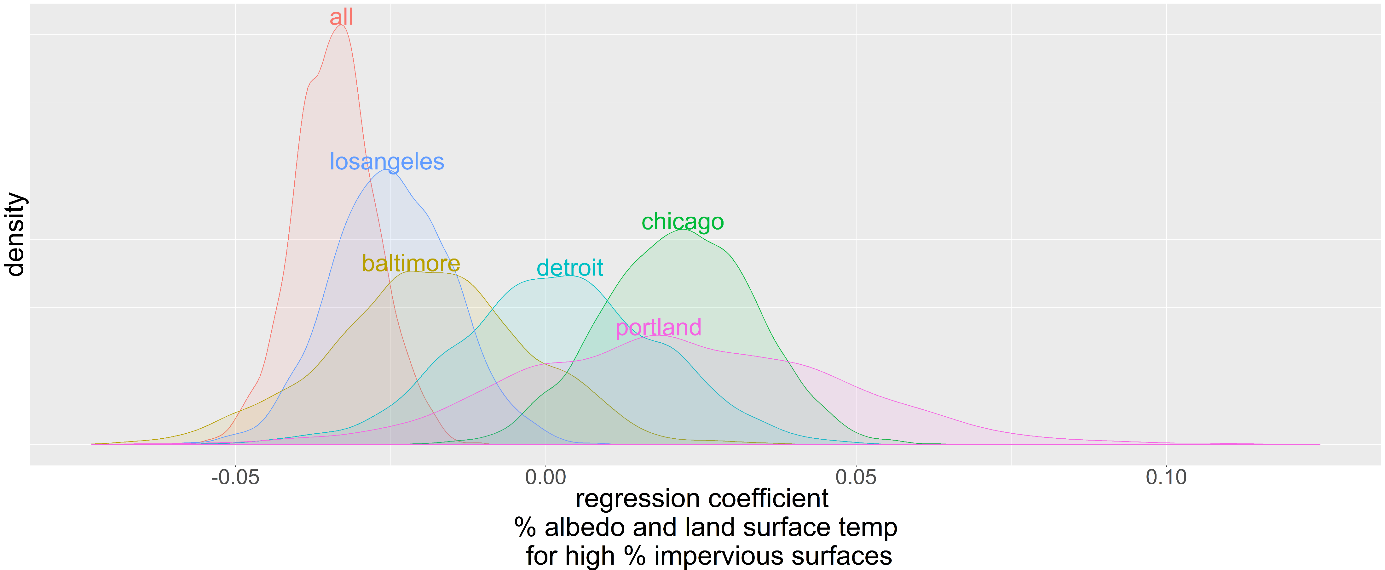


Figure 5. Posterior distributions for the regression coefficient for the effect of albedo on land surface temperature. A weak Cauchy(0,2.5) prior was used here.

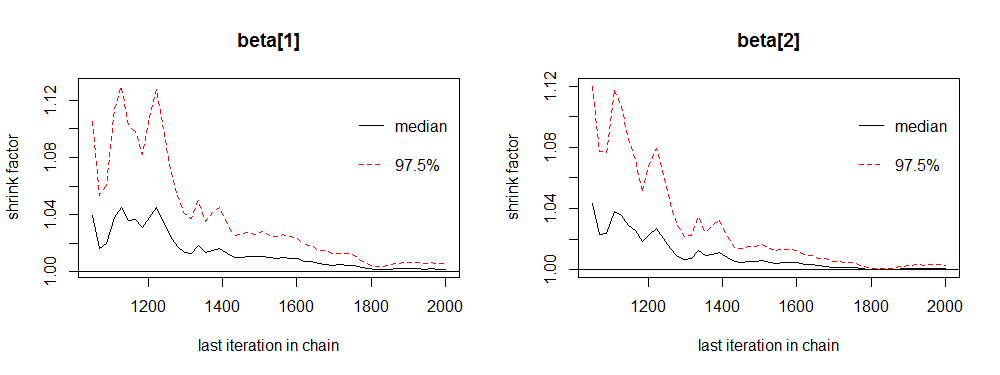


Figure 6. The convergence plot of the intercept and slope for the effect of percentage impervious surface on land surface temperature for all of the data (not subset by city).

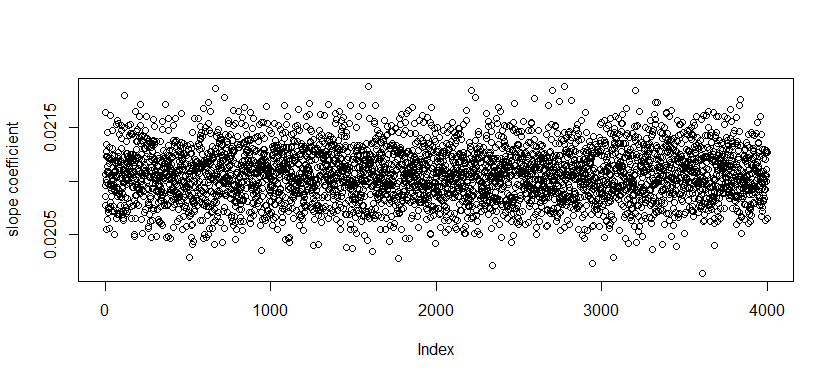


Figure 7. The trace plot for the slope coefficient of percentage impervious surface and its effect on land surface temperature.

1. Hypotheses to test

Descriptive (null is a random distribution):

1. Within Greater Baltimore LST is, to first order, a linear function of percent impervious surface.
2. By LST measures, parks are cooler than non-parks, but there is no significant halo effect and size of park doesn't matter.
3. Within impervious areas, darker surfaces are hotter than brighter surfaces.
4. Poor neighborhoods are more impervious and hotter than wealthy neighborhoods. Controlling for income (need to figure out how to do this best), black neighborhoods have less public greenery, more impervious surface, and higher LST than white neighborhoods.
5. Home value (need parcel level data) correlates with average or minimum LST within 250m of the home, both within and across neighborhoods.
6. LST gradients correlate with home price gradients.

Predictive:

1. At both point and neighborhood scale, LST is a predictable function of impervious surface, elevation, and other components of the physical environment. This allows us to do out of sample predictions of how LST will change as a function of changes to the built environment.

Exploratory:

1. At neighborhood scale, LST can be predicted as a function of social variables. These models will have skill comparable to models based on physical covariates.