Figure 3.8. Decadal trends in the number of Tmax extreme heat event per summer during the 1970-2010 time period in the Maurer dataset. The symbology regarding color groupings and statistical significance is the same as figure 3.6.

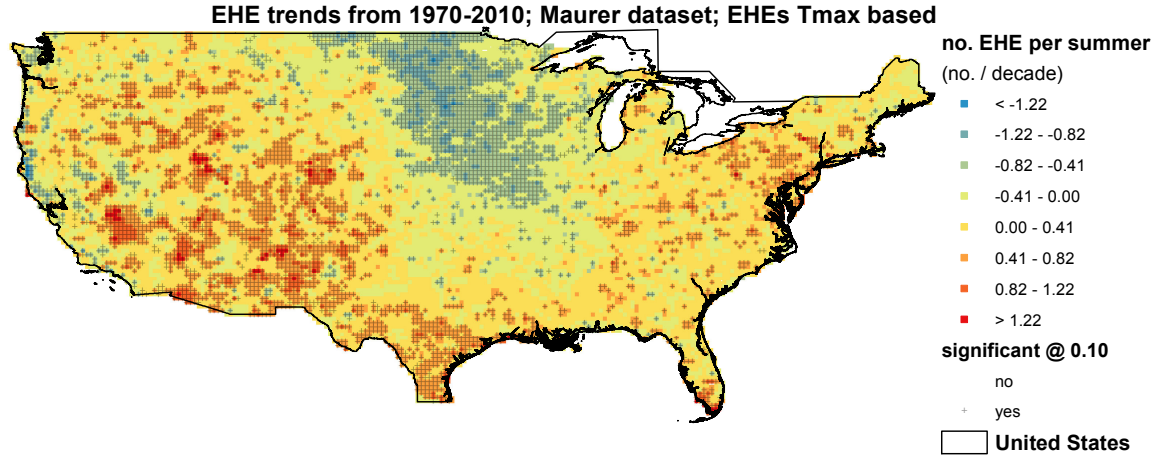


Figure 3.9. Decadal trends in the annual mean duration (in days) of Tmax extreme heat events during the 1970-2010 time period in the Maurer dataset. The symbology regarding color groupings and statistical significance is the same as figure 3.6.

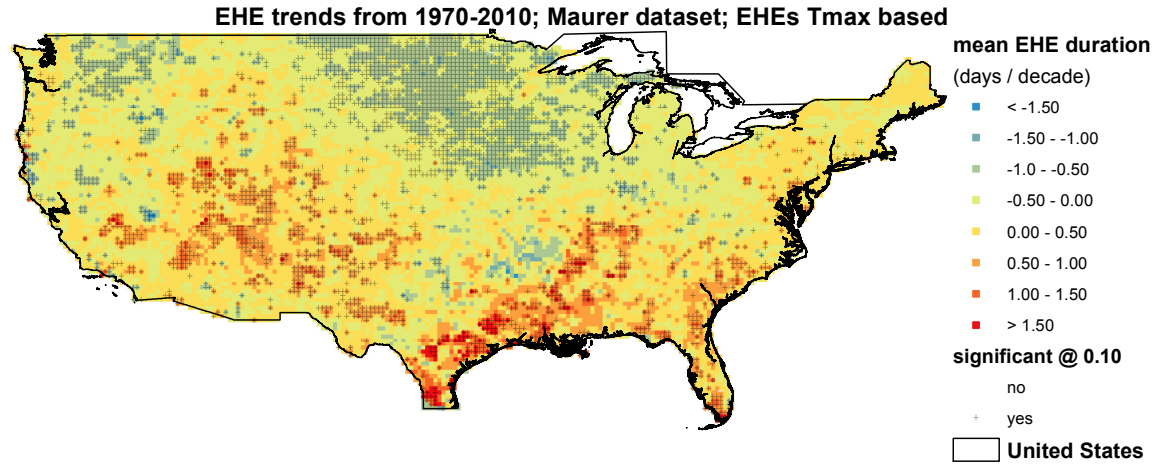


Figure 3.10. Decadal trends in the number of Tmnx extreme heat event per summer during the 1970-2010 time period in the Maurer dataset. The symbology regarding color groupings and statistical significance is the same as figure 3.6.

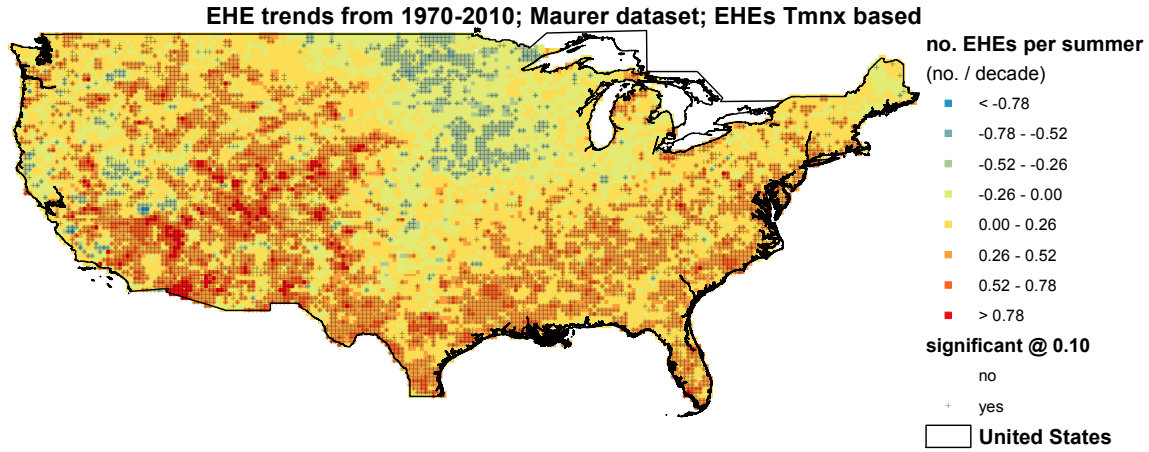
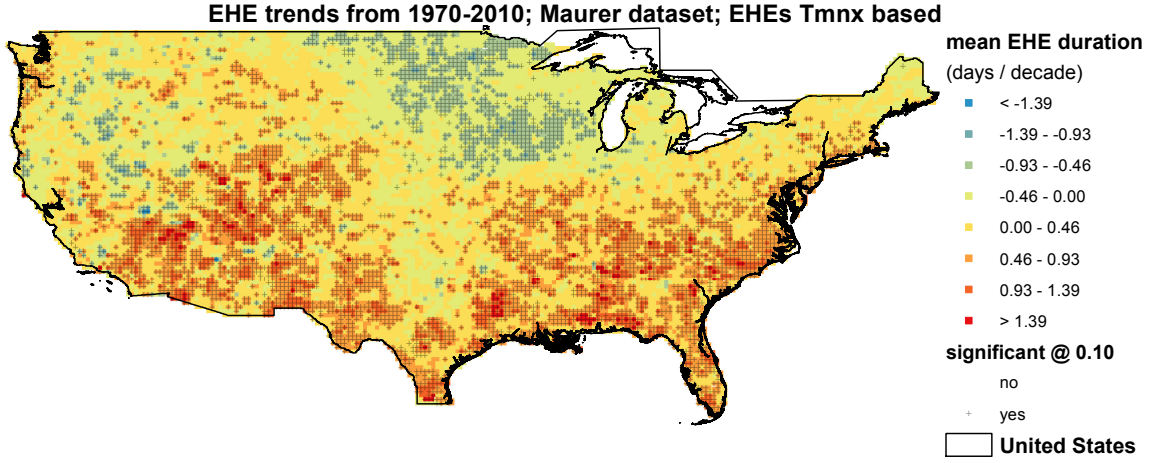


Figure 3.11. Decadal trends in the annual mean duration (in days) of Tmax extreme heat events during the 1970-2010 time period in the Maurer dataset. The symbology regarding color groupings and statistical significance is the same as figure 3.6.



3.12. Tables

Table 3.1. A continental characterization of trend sign and significance. The percent of stations with negative trend values (significant negative) and positive trend values (significant positive). PEC stands for percentile exceedence counts, EHE-days stands for the number of extreme heat event days.

	Maurer	PRISM/DiLuzio	DAYMET
Tmin temperature bias	43(43), 57(56)	61(58), 39(39)	52(47), 48(45)
Tmax temperature bias	32(31), 68(66)	35(32), 65(63)	39(31), 61(58)
Tmin PEC residual trend	66(47), 34(25)	52(29), 48(28)	52(24), 48(21)
Tmax PEC residual trend	76(63), 24(12)	51(28), 49(29)	67(35), 33(12)
EHE-days residual trend	71(43), 29(12)	47(12), 53(20)	65(21), 35(8)

Table 3.2. Continental spatial averages of the metrics of comparison. Specifically the summer average temperature bias (°C), the decadal residual trends in percentile exceedence counts (PEC), the decadal residual trends in the number of extreme heat event days (EHE-days) and the same in both the USHCN and datasets being evaluated. The numbers in parenthesis are the upper and lower 90% confidence bounds as determined by bootstrapping.

	Maurer	PRISM/DiLuzio	DAYMET
Tmin temperature bias	-0.15 (-0.17, -0.13)	-0.12 (-0.13, -0.11)	0.02 (0.01, 0.04)
Tmax temperature bias	0.11 (0.10, 0.13)	0.09 (0.08, 0.11)	0.18 (0.17, 0.20)
Tmin PEC residual trend	0.14 (0.09, 0.19)	0.16 (0.13, 0.20)	-0.22 (-0.28, -0.17)
Tmax PEC residual trend	-0.58 (-0.59, -0.55)	0.06 (0.03, 0.10)	-0.97 (-1.02, -0.91)
EHE-days residual trend	-0.09 (0.07, 0.11)	0.12 (0.10, 0.14)	-0.31 (-0.35, -0.28)
EHE-days USHCN trend	0.75 (0.72, 0.78)	1.04 (1.01, 1.07)	0.84 (0.78, 0.90)
EHE-days Target trend	0.66 (0.62, 0.69)	1.16 (1.13, 1.19)	0.53 (0.46, 0.60)

Table 3.3. Sensitivity of continental spatial average comparison metrics to map making method. IDW representing the inverse distance weighting and OK representing the kriging method. Comparison metrics include the summer average temperature bias, the decadal residual trends in percentile exceedence counts (PEC), the decadal residual trends in the number of extreme heat event days (EHE-days) and the same in both the USHCN and datasets being evaluated. The number in parenthesis are the upper and lower 90% confidence bounds.

	Maurer	PRISM/DiLuzio	DAYMET
Tmin temperature bias (IDW) (°C)	-0.15 (-0.17, -0.13)	-0.12 (-0.13, -0.11)	0.02 (0.01, 0.04)
Tmin temperature bias (OK) (°C)	-0.15 (-0.17, -0.13)	-0.12 (-0.12, -0.11)	0.02 (0.01, 0.03)
Tmax temperature bias (IDW) (°C)	0.11 (0.10, 0.13)	0.09 (0.08, 0.11)	0.18 (0.17, 0.20)
Tmax temperature bias (OK) (°C)	0.12 (0.11, 0.13)	0.10 (0.09, 0.11)	0.18 (0.17, 0.18)
Tmin PEC trend diff (IDW)	0.14 (0.09, 0.19)	0.16 (0.13, 0.20)	-0.22 (-0.28, -0.17)
Tmin PEC trend diff (OK)	0.10 (0.06, 0.14)	0.14 (0.12, 0.16)	-0.22 (-0.25, -0.18)
Tmax PEC trend diff (IDW)	-0.58 (-0.59, -0.55)	0.06 (0.03, 0.10)	-0.97 (-1.02, -0.91)
Tmax PEC trend diff (OK)	-0.59 (-0.60, -0.58)	0.09 (0.07, 0.11)	-0.99 (-1.03, -0.96)
Tmnx EHE-days USHCN (IDW)	0.75 (0.72, 0.78)	1.04 (1.01, 1.07)	0.84 (0.78, 0.90)
Tmnx EHE-days USHCN (OK)	0.76 (0.73, 0.78)	1.02 (1.00, 1.05)	0.82 (0.77, 0.87)
Tmnx EHE-days Target (IDW)	0.66 (0.62, 0.69)	1.16 (1.13, 1.19)	0.53 (0.46, 0.60)
Tmnx EHE-days Target (OK)	0.69 (0.66, 0.72)	1.15 (1.13, 1.18)	0.49 (0.43, 0.55)
Tmnx EHE-days Target-USHCN (IDW)	-0.09 (-0.07, -0.11)	0.12 (0.10, 0.14)	-0.31 (-0.35, -0.28)
Tmnx EHE-days Target-USHCN (OK)	-0.11 (-0.09, -0.12)	0.12 (0.11, 0.13)	-0.32 (-0.34, -0.30)

Table 3.4. Regional mean residual trends as determined. PEC stands for percentile exceedence counts, EHE-days stands for the number of extreme heat event days. The numbers in parenthesis are the upper and lower 90% confidence bounds as determined by bootstrapping.

Tmin PEC	Maurer	PRISM/DiLuzio	DAYMET
Northwest (218)	0.11 (0.05 - 0.16)	0.42 (0.37 - 0.46)	0.23 (0.15 - 0.31)
Southwest (188)	1.40 (1.31 - 1.50)	0.23 (0.18 - 0.29)	-0.65 (-0.76 - -0.51)
Northcentral (123)	-0.36 (-0.41 - -0.31)	0.04 (-0.03 - 0.10)	-0.12 (-0.20 - -0.04)
Southcentral (145)	-0.28 (-0.35 - -0.22)	-0.01 (-0.08 - 0.06)	-0.04 (-0.16 - 0.09)
Northeast (47)	-0.52 (-0.61 - -0.43)	-0.18 (-0.28 - -0.09)	-0.34 (-0.51 - -0.17)
Southeast (98)	-0.72 (-0.83 - -0.59)	0.08 (-0.08 – 0.25)	-0.71 (-1.00 - -0.42)
Tmax PEC			
Northwest (218)	-0.55 (-0.58 - -0.52)	0.18 (0.11 - 0.25)	-1.33 (-1.42 - -1.25)
Southwest (188)	-0.47 (-0.51 - -0.43)	0.11 (0.03 - 0.19)	-1.35 (-1.42 - -1.27)
Northcentral (123)	-0.71 (-0.75 - -0.67)	-0.08 (-0.15 - -0.02)	-0.55 (-0.63 - -0.46)
Southcentral (145)	-0.68 (-0.74 - -0.62)	-0.10 (-0.18 - 0.01)	-0.10 (-0.24 – 0.03)
Northeast (47)	-0.46 (-0.52 - -0.40)	-0.21 (-0.31 - -0.11)	-0.96 (-1.12 - -0.80)
EHE-days			
Northwest (218)	0.10 (0.07 - 0.13)	0.22 (0.21 - 0.24)	0.11 (0.06 - 0.16)
Southwest (188)	-0.40 (-0.44 - -0.37)	0.12 (0.08 - 0.16)	-0.70 (-0.78 - -0.62)
Northcentral (123)	0.35 (0.32 - 0.37)	0.05 (0.02 - 0.08)	-0.42 (-0.46 - -0.38)
Southcentral (145)	0.24 (0.22 - 0.27)	0.05 (0.00 - 0.09)	-0.17 (-0.24 - -0.10)
Northeast (47)	0.37 (0.32 - 0.41)	-0.06 (-0.10 - -0.02)	-0.49 (-0.58 - -0.40)
Southeast (98)	0.41 (0.38 - 0.44)	0.18 (0.11 – 0.24)	-0.43 (-0.52 - -0.35)

Table 3.5. Regional mean residual trends as determined via the kriging map making method. PEC stands for percentile exceedence counts, EHE-days stands for the number of extreme heat event days. The numbers in parenthesis next to the region are the number of grid cells within each region. The number in the parenthesis next to the averages are the upper and lower 90% confidence bounds as determined by a bootstrapping procedure.

Tmin PEC	Maurer	PRISM/DiLuzio	DAYMET
Northwest (218)	0.10 (0.07 - 0.14)	0.45 (0.42 - 0.48)	0.17 (0.12 - 0.23)
Southwest (188)	1.21 (1.14 – 1.27)	0.17 (0.12 - 0.20)	-0.53 (-0.63 - -0.44)
Northcentral (123)	-0.35 (-0.38 - -0.34)	0.06 (0.03 - 0.09)	-0.12 (-0.18 - -0.06)
Southcentral (145)	-0.22 (-0.26 - -0.18)	0.05 (0.01 - 0.08)	-0.15 (-0.21 - -0.09)
Northeast (47)	-0.47 (-0.51 - -0.42)	-0.24 (-0.27 - -0.19)	-0.31 (-0.42 - -0.21)
Southeast (98)	-0.82 (-0.89 - -0.74)	-0.14 (-0.22 - -0.04)	-0.59 (-0.74 - -0.45)
Tmax PEC			
Northwest (218)	-0.51 (-0.53 - -0.49)	0.17 (0.14 - 0.21)	-1.30 (-1.35 - -1.24)
Southwest (188)	-0.57 (-0.60 - -0.55)	0.20 (0.16 - 0.25)	-1.46 (-1.50 - -1.41)
Northcentral (123)	-0.68 (-0.70 - -0.66)	-0.06 (-0.11 - -0.02)	-0.55 (-0.61 - -0.50)
Southcentral (145)	-0.66 (-0.69 - -0.63)	-0.09 (-0.14 - -0.05)	-0.09 (-0.15 - -0.02)
Northeast (47)	-0.40 (-0.43 - -0.38)	-0.14 (-0.20 - -0.10)	-0.95 (-1.10 - -0.80)
EHE-days			
Northwest (218)	0.09 (0.07 – 0.11)	0.21 (0.19 - 0.22)	0.03 (0.00 - 0.06)
Southwest (188)	-0.33 (-0.36 - -0.30)	0.13 (0.10 - 0.16)	-0.63 (-0.68 - -0.57)
Northcentral (123)	0.34 (0.32 – 0.36)	0.05 (0.04 - 0.07)	-0.41 (-0.43 - -0.38)
Southcentral (145)	0.23 (0.22 – 0.25)	0.08 (0.06 - 0.11)	-0.21 (-0.25 - -0.17)
Northeast (47)	0.35 (0.33 – 0.38)	-0.07 (-0.09 - -0.05)	-0.53 (-0.58 - -0.46)
Southeast (98)	0.42 (0.40 – 0.44)	0.14 (0.11 - 0.18)	-0.42 (-0.45 - -0.39)

Table 3.6. Moran's I index analysis of residual trends autocorrelation. The first value is the index value and the second value is the P-value. Percentile exceedence counts (PEC) and the number of extreme heat event days (EHE-days) are both listed for each dataset.

	Maurer	PRISM/DiLuzio	DAYMET
Tmin PEC res. trends	0.080, 0.004	0.069, 0.017	-0.006, 0.999
Tmax PEC res. trends	-0.005, 0.863	0.041, 0.137	0.030, 0.137
EHE-days res. trends	0.033, 0.199	0.028, 0.295	-0.020, 0.586

Table 3.7. Correlation of population growth trends for each dataset and comparison metric. First value listed is the radius of strongest correlation, the values in the parenthesis are the Pearson's correlation coefficient and the leftmost number the P-value. Bold values are statistically significant with 90% confidence.

	Maurer	PRISM/Diluzio	DAYMET
Tmin PEC	**3.61km (+0.16; 0.02)**	3.00km (+0.07; 0.31)	**1.80km (+0.17; 0.04)**
Tmax PEC	3.61km (+0.03; 0.67)	1.12km (+0.07; 0.30)	**1.80km (+0.14; 0.10)**
EHE-days	3.61km (+0.12; 0.10)	**3.00km (+0.16; 0.01)**	1.80km (+0.07; 0.91)

Table 3.8. Comparison of instrument-type grouped residual trends. The median (10th and 90th percentile) of daily minimum (Tmin) and maximum (Tmax) percentile exceedence count (PEC) mean trend residuals by instrument type (liquid in glass thermometer in cotton region shelters (CRS), maximum minimum temperature sensors (MMTS) or hygrometer (Hygro)). Values estimated via 7500 bootstraps of the sample mean. Values in square brackets are the number of stations.

	Maurer	PRISM/DiLuzio	Daymet
CRS Tmin PEC	-0.36 (-0.68, -0.05) [39]	-0.02 (-0.19, 0.16) [54]	0.21 (-0.16, 0.56) [37]
MMTS Tmin PEC	-0.28 (-0.41, -0.13) [124]	0.01 (-0.12, 0.14) [156]	-0.48 (-0.78, -0.20) [94]
Hygro Tmin PEC	0.01 (-0.29, 0.36) [32]	0.39 (-0.09, 0.88) [30]	0.89 (0.20, 1.63) [23]
CRS Tmax PEC	-0.52 (-0.69, -0.36) [39]	0.08 (-0.11, 0.28) [54]	-1.48 (-1.84, -1.12) [37]
MMTS Tmax PEC	-0.66 (-0.77, -0.55) [124]	0.00 (-0.16, 0.16) [156]	-0.57 (-0.84, -0.31) [94]
Hygro Tmax PEC	-0.36 (-0.55, -0.16) [32]	-0.03 (-0.39, 0.40) [30]	-1.05 (-1.70, -0.42) [23]
CRS EHE-days	-0.25 (-0.38, -0.12) [39]	0.03 (-0.05, 0.11) [54]	-0.11 (-0.32, 0.11) [37]
MMTS EHE-days	-0.29 (-0.35, -0.22) [124]	-0.08 (-0.02, 0.14) [156]	-0.57 (-0.74, -0.42) [94]
Hygro EHE-days	-0.02 (-0.11, 0.08) [32]	0.15 (-0.06, 0.35) [30]	0.06 (-0.32, 0.46) [23]

Table 3.9. Correlations with proxies for homogenization levels. Pearson's correlation coefficients are provided between both the trends in percentile exceedence counts (PEC) residual trends and extreme heat event days (EHE-days), and the trends in summer average temperature residuals between time series of USHCN data with different levels of homogenization. The daily temperature extreme corresponding to the PEC (i.e. daily minimum and daily minimum) is only shown, but the first value in the EHE-days is the daily minimum and the second daily maximum.

	Maurer	PRISM/DiLuzio	DAYMET
Tmin PEC (raw-tob)	0.10	0.11	0.03
Tmax PEC (raw-tob)	0.37	0.13	0.20
EHE-days (raw-tob)	0.20/0.33	0.13/0.17	0.14/0.15
Tmin PEC (tob-full)	0.82	0.38	0.52
Tmax PEC (tob-full)	0.69	0.35	0.38
EHE-days (tob-full)	0.62/0.31	0.34/0.16	0.28/0.17
Tmin PEC (raw-full)	0.84	0.40	0.53
Tmax PEC (raw-full)	0.80	0.37	0.45
EHE-days (tob-full)	0.67/0.44	0.36/0.18	0.31/0.22

CHAPTER 4. AN INVESTIGATION INTO THE SPATIAL VARIABILITY OF NEAR-SURFACE AIR TEMPERATURES IN THE DETROIT, MI METROPOLITAN REGION

Full citation of corresponding manuscript

Oswald, E. M., R. B. Rood, K. Zhang, C. J. Gronlund, M. S. O'Neill, J. L. White-Newsome, S. J. Brines and D. G. Brown, 2012: An investigation into the spatial variability of near-surface air temperatures in the Detroit, MI metropolitan region. *Journal of Applied Meteorology and Climatology*, **51**: 1290-1304, doi: 10.1175/JAMC-D-11-0127.1

Abstract

Annually, heat is the chief cause of weather-related deaths in the United States. Understanding the air temperature patterns throughout urban regions is important for locating resources during dangerously hot weather. This study focused on the air temperatures in the Detroit, MI metropolitan region during the summer of 2009. An observational network was established that included 1) monitors sited in the back yards of residential participants, 2) National Weather Service standard observations, and 3) a network of monitors operated by the state of Michigan. Daily maximum and minimum temperatures were analyzed for their spatial pattern, magnitude of spatial variability and relationships with weather conditions. The existence of spatial variability was confirmed specifically during weather considered dangerous to the public's health. The relationships between temperature observations and distance to water, distance to city center, and local percent impervious surface were investigated. The investigation was redone using only dates with weather conditions favorable to spatial variability to assess the differences.

The spatial variability during the daily minimum was typically stronger in magnitude and the spatial pattern more consistent than that during the daily maximum. The existence of spatial variability during dangerously hot weather was confirmed via several methods of classifying dangerous weather, and only a weak inverse relationship existed between daily minimum temperatures and the amount of spatial variability. The largest correlation with land-cover and location attributes was between percent impervious surface values and daily minimum temperatures. Daily maximum temperatures were most correlated with distance to water. Consistent with previous studies on spatial variability in urban environments, the results suggest a need for sensitivity to the spatially variable nature of exposure to heat events in both public health and urban planning. For example these results showed the downtown area experienced elevated temperatures in daily minimums and the eastern portions of Detroit experienced decreased temperatures in daily maximums. When reevaluated only during nights with low wind speeds and light cloud cover, the spatial variability was larger, the statistical models performed about equivalently and the locations of relative hot and cold were the same as those predicted by the model using all 110-days of field observations.

4.1. Introduction

Near surface air temperatures vary spatially over a region because the land cover land use (LCLU) types vary spatially. LCLU types alter the surface energy budget and how the surface interacts with the lower atmosphere. Urban areas are effectively congregations of LCLU types that cause the overnight temperatures to be warmer than the surrounding rural LCLU types; this effect is termed the Urban Heat Island effect (UHI effect) (Bornstein 1968; Oke 1973; Landsberg 1981).

In particular, the fabric of the urban environment has a difficult time converting the incident radiative energy gain into an upward flux of latent heat because of lack of water availability at the surface (Oke 1982; Grimmond 2007). Moreover, the materials that comprise the bulk of the developed LCLU types have a relatively high potential to store and conduct heat throughout the day. Overnight, the geometric characteristics (e.g. mean building height and street width) of developed LCLU types lead to relatively small radiative cooling rates. This is because the view of space (cold) is often reduced (Chen et al. 2010), often referred to as the *urban canyon effect* (Nunez and Oke 1977; Barring and Mattsson 1985). Another physical mechanism for the UHI is the reduction in total turbulent sensible heat flux by reduced low-level winds.

These wind speeds are reduced because of relatively large ground roughness prevents momentum exchange with (faster) higher altitude winds (Grimmond 2007); this mechanism should work to reduce the flux during both nighttime and daytime. Physically this is because the ground roughness dictates the vertical extent of the lowest two layers of the atmosphere. First the urban canopy layer which extends to the ground to just above the buildings and has the exchanges of momentum, heat and moisture. Next the the roughness sublayer extends from the ground to the height where blending action is complete. Both of these are dependent on the roughness length (Oke 2004).

Another physical mechanism is a reduction in shortwave albedo. Interestingly, regardless of its color relative to the rural surroundings - the geometric shape of cities causes them to trap upwards to 20% more shortwave radiation (Aida 1982), because

this radiation comes in and bounces around until it finds an absorptive surface. Lastly, heat is also added into the urban system by anthropogenic means (Hart and Sailor 2009) such as cars engines, industrial and commercial exhaust, and cooling and refrigeration systems.

Conversely during the daytime developed LCLU types can potentially cause relatively cool temperatures, albeit not as consistent nor strongly an effect. One mechanism is when less shortwave radiation is received at the ground level due to increased pollution (e.g. soot). The other mechanism (that applies to semiarid climates) is called the Urban Oasis effect (Oke 1987). It works on the scale of an entire city and on the smaller scale, for example in urban parks (Spronken-Smith et al. 2000). That is because both urban parks and cities can both be local maximums for vegetation and soil moisture, which causes relatively higher latent heat fluxes and thus lower air temperatures.

This illuminates an important point: on a larger scale temperatures over rural landscapes are different from those within an adjacent city, but on a smaller scale (within a city) the temperature varies by neighborhoods with similar types of urban development (Oke 1984). This arises in part from the differences in the aforementioned characteristics (albedo, geometrical shape, moisture availability, etc.) between different “urban” LCLU types. Most cities have clustering of land primarily consisting of industrial and commercial use; regular, compact and block housing; old and modern city cores (Stewart and Oke 2009). These are laid out different in each city. The natural geography also plays a role in the temperature distribution throughout a developed region (Kuttler et al. 1996). For example on the small scale, topographical elevation allows cool air drainage and inversions to occur. On the larger scale sizeable water bodies/coastlines (Childs and Raman 2005; Freitas et al. 2007) and mountain ranges can interact with the city.

Observational studies have not fundamentally changed since the 1980’s, although numerical modeling studies have emerged. Examples of the typical methods of the observational studies include time trends (i.e. the site is becoming more urban) at singular locations (e.g. Tarleton and Katz 1995; Tereshchenko and Filonov 2001), comparative time trends at one (or more) urban stations and one (or

more) rural stations (e.g. Ackerman 1985; Magee et al. 1999; Philandras et al. 1999; Morris and Simmonds 2000; Zhou and Shepherd 2009), networks of fixed stations within and around a city (e.g. Basara et al. 2008, 2010; Camilloni and Barrucand 2010) and transects across an urban area (e.g. Saitoh et al. 1996; Goh and Chang 1999; Bottyán and Unger 2003; Wong and Yu 2005; Yokobori and Ohta 2009). It is common for studies to use more than one method (Kuttler et al. 1996; Montávez et al. 2000; Longxun et al. 2003).

Most urban climate studies focus on the UHI effect on a large scale, or the temperature differences between an urban station and its rural counterparts (e.g. Zhou and Shepherd 2009; Basara et al. 2008; Camilloni and Barucand 2011; Chow and Roth 2006; Morris et al. 2001; Mohsin and Gough 2011; Velazquez-Lozada et al. 2006). That difference is often termed the *magnitude of the UHI, UHI magnitude*, *Intensity of the UHI* or *UHI Intensity*. Its effectiveness as a metric relies on the study's success to characterize the stations as rural or urban. However these metrics have limitations, for example LCLU types categorized as both "urban" and "rural" vary within those categories (e.g. agriculture vs. industrial; forest vs. commercial; barren vs. residential; etc).

It is rare for studies to investigate the smaller scale variability of temperatures within an urban region and/or city. Some of those studies demonstrated the existence of, and the amount/sign, variability within cities (e.g. Saaroni et al. 2000; Basara et al. 2008; Gaffin et al. 2008); others attempted to map the spatial structure (Montávez et al. 2000; Longxun et al. 2003; Bottyán and Unger 2003; Kim and Baik 2005; Hart and Sailor 2009). More involved studies showed the variability was linked to the local scale spatial attributes (Kuttler et al. 1996; Goh and Chang 1999; Bottyán and Unger 2003; Hart and Sailor 2009; Buttstädt et al. 2010) and land cover types (Wong and Yu 2005; Yokobori and Ohta 2009). Spatial variables found to be good predictors were building height, building density, building area, building height-to-street width ratios, building use type, canopy cover, sky view factor, water surface ratio, amount of impervious surface and ground elevation.

Studies have shown the maximum UHI magnitude resides after sunset but before the daily minimum temperature (Tmin) (e.g. Oke 1982, Morris and Simmonds

2000) and some have shown when Tmin occurs can even be at different times depending on LCLU type (Morris and Simmonds 2000). The time of strongest signal is the time of day typical UHI studies focus on because they are focusing on the physical impacts of the urban landscape. However epidemiological literature more commonly has linked the daily temperature extremes to mortality data (for whatever reason). Summertime daily maximum temperature (Tmax) is associated with both heat-related medical dispatches (Dolney and Sheridan 2006; Golden et al. 2008) and mortality rates (Baccinni et al. 2008; Anderson and Bell 2009; Basu 2009; Gosling et al. 2009). Summertime Tmin is also associated with mortality rates (Kalkstein and Smoyer 1993; Hajat et al. 2002; Grize et al. 2005; Schwartz 2005; Hajat et al. 2006; Fouillet et al. 2007; Basu et al. 2008). It is not clear which daily temperature extreme is most important, however it likely varies with region (Kalkstein 1989). Consequently this study focused on the spatial variability specifically on Tmin and Tmax. Thus while our analysis does not focus on the strongest urban landscape signal it compensates by being more useful scientists in the heat-health field.

To focus the scope of this study the goal was to answer the following set of questions. First, did the daily temperature extremes vary spatially across the metropolitan domain? Was the day-to-day magnitude of that variability diagnosable by the large-scale weather? Was the spatial pattern predictable by the land attributes? Lastly, did the spatial variability exist during hot/oppressive weather?

4.2.Methods

4.2.1. Observational network

4.2.1.1. Descriptions of networks

For this study it was chosen to integrate multiple networks into one observing network. While it was not the focus of this study, future studies might also construct such networks. In this section, first the networks and data used are described, then comparability between the networks and uncertainty in using the integrated-network are explained.

NWS network

The hourly NWS observations used to anchor our integrated network were observed at the local airports by the NWS's Automated Weather Observing Station (AWOS) and Automated Surface Observing Station (ASOS) networks (Mannarano 1998). Temperature measurements are observed at a 1.5-meter height. These are the standard operational products used by the weather service and are subject to calibration standards (Federal Aviation Administration 1998). Five airports were used in the network, located from well within to the outskirts of the city (Figures 4.1, 4.2) and on a sizeable island (Grosse Isle). Acquired from the National Climatic Data Center (NCDC) was the hourly temperature, cloud cover, and wind speed (taken at 10 m above ground level) data during the observational period.

The Michigan Department of Environmental Quality network

The Michigan Department of Environmental Quality (MDEQ) operates an air pollution-monitoring network as mandated by the U.S. Environmental Protection Agency for monitoring compliance with National Ambient Air Quality Standards under the Clean Air Act (Department of Natural Resources and Environment 2010). This network provided hourly means of air temperatures five meters above ground level (i.e. still within the urban canopy layer) and wind characteristics measured at ten meters above ground level. Six of their monitoring locations were suitable for this study (two locations were used for co-location) and the aforementioned data were acquired through contact with the MDEQ staff. Figures 4.1 and figures 4.3-4.5 show locations and some photographs of observing sites.

The temporarily established network

For the custom study network, 21 HOBO Pro V2 External Temp/RH data loggers (Onset Computer Corporation 2010) (Onset Computer Corp., Pocasset/MA) were used, herein referred to as HOBOs. Previous studies have demonstrated that these monitors are capable of similarly detailed studies of the spatial structure of temperature fields (Whiteman et al. 2000). The HOBO monitors consist of temperature and relative humidity sensors inside naturally aspirated radiation shields and fastened to thin wooden stakes, along with a data logger, at a height of 1.5

meters above ground level. Detroit-area residents, and some businesses, were approached for permission to place loggers outside on their property, which ranged in location from the city limits to the old downtown district; near and far from the two lakes (Erie and St. Clair) and the Detroit River; in densely populated suburbs to the outskirts of more distant suburban communities and in the countryside (Figure 4.1). To the east of the Detroit River lies the city of Windsor, Ontario, Canada; it was assumed that Windsor's close proximity did not affect Detroit's temperature pattern because it is typically downwind of Detroit and separated by a river up to 4 km in width. The HOBOs were placed primarily in back yards (or grassy areas at businesses) of participants, sampling at a 10-minute frequency, and sited to minimize microclimate impacts (e.g. over grassy groundcover, ample sky view, at least 15 meters away from heat sources) (Figures 4.5-4.6). Lastly, it was assumed the turbulent mixing produced by flow around barriers was adequate to blend the observed atmospheric temperatures so as to allow measurement of local-scale (e.g. neighborhoods) average temperatures.

4.3.1.2. Evaluation of network comparability and uncertainty

A primary advantage of the HOBO network was the ability to calibrate its measurements with those of existing networks. In general, the method was for a given calibration period, to a) locate a HOBO monitor in as close proximity as possible to at least one monitor from an existing network, b) quantify the relationship, or *bias*, between the co-observed daily temperature extremes, c) if necessary, apply bias corrections or simple modeling in order to make observations as similar as possible, then d) re-quantify the mean difference between HOBO and existing network observations over the calibration experiment, referred to as the *uncertainty* (Table 4.1) in using the existing network with the HOBO network.

Complications arose when integrating multiple networks; each network had unique monitoring uncertainties and also further uncertainties when using them in concert. The details of how the uncertainties associated with each network's observations are described in this section. A metric of comparison was desired that

would allow for comparison of values across the network but also ensured that any spatial variability found would exceed the unique uncertainties of the networks. This new formed metric was referred to as the "anomaly" and mathematically was

$$Anomaly_{i,j,d} \equiv (T_{i,j,d} - \overline{T_{j,d}}) \pm uncert_{i,j} \begin{cases} +, (T_{i,j,d} - \overline{T_{j,d}}) < 0 \\ -, (T_{i,j,d} - \overline{T_{j,d}}) > 0 \end{cases}$$

with *d* being the day, *j* signifying either Tmax or Tmin and *i* indicating location. First, the differences between each day's value of the Tmax (or Tmin), at each location ($T_{i,j,d}$), and the network mean (i.e. across all locations) of that particular Tmax or Tmin ($\overline{T_{j,d}}$) were calculated. Then depending on the sign of the first term ($T_{i,j,d}$ - $\overline{T_{j,d}}$) the network-specific uncertainty value ($uncert_{i,j}$) was either subtracted or added in such a manner that always worked to reduce the absolute value of the first term.

First the accuracy of the HOBO monitors relative to one another was empirically determined. Pre and Post-deployment tests were both undertaken but the later was bigger. Post-deployment tests were chosen because they would capture any drift in accuracy of the monitors during the field study. For those tests, all monitors were placed in a temperature-controlled room during testing. With 3097 observations per monitor (for the post -deployment test), the range across the monitors was calculated for each observation-time and then the mean was calculated at 0.44 °C. This was effectively the typical disparity in temperatures observed across the HOBO monitors when they should have been monitoring the same temperature. This was referred to as the "inherent-relative uncertainty" of the HOBO network (Table 4.1) and it represents an amount of uncertainty in our observations.

Then the differences between the HOBO network and the standard NWS observations from the airports were sought. Although it was impossible to gain permission to collocate monitors on airport property, a monitor was located 2.7 km away from one of the airport monitoring locations (KDTW) and the differences were consecutively observed for two weeks. This large separation and small sample size is an admitted limitation to the network calibration. In order to lessen that limitation, the siting and sampling algorithm differences were separately accounted for. This

HOBO siting was typical of the other HOBO monitor sitings as opposed to the very open airport monitor sitings. Using this nearby HOBO monitor's observations, which sampled at a higher frequency (every two minutes) than the typical HOBO monitor (every ten minutes), the KDTW observations (i.e. mean value within five minutes prior to hour) were estimated, and the mean bias for Tmin and Tmax was calculated at +0.14 °C and +0.49 °C, respectively. Once the HOBO observations were adjusted for these biases, the mean differences between the HOBO and the KDTW observations for both Tmax and Tmin were calculated at values less than the prescribed uncertainty of the ASOS/AWOS instrumentation (0.5 °C); therefore the "HOBO siting uncertainty" was taken as 0.0 °C (Table 4.1). While this seems counterintuitive, it was hard to assign an uncertainty when the mean difference was less than the uncertainty of the ASOS/AWOS instrumentation; however in response we took a more conservative approach when determining the total HOBO uncertainty. In regards to the NWS AWOS/ASOS observations in the network, the only uncertainty it was assumed they had was the 0.5°C "instrumentation uncertainty" (Mannarano 1998) (Table 4.1).

For this observing network Tmin at each station was taken as the minimum temperature recorded between 0100 and 0700 EDT, and Tmax was taken as the maximum temperature recorded between 1300 and 1900 EDT. Initially this convention was chosen over a 24-hour maximum/minimum temperature selection to filter out ensure the Tmax time was later in the day than the Tmin time. However, this method also was ensured to represent the temperatures while humans slept (morning) and were active (afternoon). Moreover, it allows better relationships to weather conditions, as it is guaranteed that temperature observations and the weather observations happened at roughly the same time of day. However because this study focused on the daily temperature extremes, instead of the time of day (i.e. usually more in the middle of the overnight cooling process) with the most robust influence from the landscape (i.e. largest MTR), our results should be considered to be underestimates of the maximum amount of spatial variability that metropolitan Detroit experiences over the course of a night.

The temporal sampling varies across the three networks. The HOBO monitors were programmable but the typical configuration was a 10-minute sampling frequency with no additional averaging. The MDEQ network reports hourly means every hour, and the airport network reports five-minute means every hour. To test the impact of the temporal sampling differences, a HOBO was set to sample on a one-minute sampling frequency over the duration of the study and used to replicate the various network products each day. This was used to estimate the biases and uncertainties associated with the different temporal sampling algorithms among the networks.

As mentioned previously, the HOBO-AWOS/ASOS collocation was only two weeks long, and so it was elected to explicitly account for the difference in sampling algorithms. This bias between AWOS/ASOS and HOBO observations was calculated at -0.48 °C for Tmax, and +0.16 °C for Tmin. The mean difference, after these bias corrections were applied, was calculated at 0.28 °C for Tmax and 0.11 °C for Tmin, and this was referred to as the "HOBO sampling algorithm uncertainty" (Table 4.1). Since the HOBO-MDEQ collocation experiment occurred over a large sample, we decided to incorporate the sampling algorithm difference between the two networks in the collocation experiment itself. However using the HOBO-AWOS/ASOS sampling algorithm bias and uncertainty values was still required, since the airport sampling – not the HOBO sampling - was the integrated network baseline.

To integrate data from the MDEQ network, quantification was needed of the differences in observations that were due to both the differences in height, instrumentation and sampling algorithm between the network's measurements and HOBO observations. To quantify the sum impact of these differences two 1.5-meter above ground level HOBO monitors were collocated at two MDEQ monitoring sites for two months (Figures 4.4-4.5). On site MDEQ wind-speed and nearby NWS AWOS/ASOS sky-cover observations were useful in predicting the daily temperature extreme HOBO observations from MDEQ observations. The relationships determined were designed to be non-site specific by pooling the data from both collocation sites for the two months duration (128 days) and then building the relationships. The linear multiple regression equations used for this prediction are

briefly provided (Table 4.1), and analysis of fit was performed (not shown). The mean differences between MDEQ-statistical-model predicted and HOBO observed Tmin and Tmax values were adopted as the "MDEQ collocation uncertainty" (Table 4.1) and calculated at 0.43 °C and 0.52 °C, for Tmin and Tmax, respectively. The measurement uncertainty stated in the MDEQ instrumentation specifications was 0.3°C, but this was less than the aforementioned determined uncertainties and it was assumed to be included in those values. Thus the uncertainties were summed related to the collocation and sampling algorithm differences. Summing these uncertainties resulted in an uncertainty in Tmin of 0.54 °C, and 0.80 °C in Tmax, when using the MDEQ stations observation within the larger network.

Additionally, even with careful siting standards, inconsistencies between HOBO observations exists, simply because they monitored in yards with different microclimates. It was decided not to include uncertainty arising from microclimate variability across either the six MDEQ stations or five AWOS/ASOS stations because the level of monitoring standardization and siting strictness was better and assumed adequate. To assess the magnitude of this uncertainty within the HOBO network, data from two similarly located and sited monitors (separated by ~300 meters) were analyzed to determine the yard-to-yard differences, indicated in table 4.1 as the "HOBO microclimate uncertainty." First the mean difference between the two monitors was calculated, over the duration of the study, per extreme, and showed a Tmax value of 0.76 °C, and Tmin value of 0.33 °C. For Tmin the inherent-relative uncertainty of the HOBO network was larger than this difference, so the microclimate uncertainty was assumed to be 0 °C. However, for Tmax the microclimate uncertainty exceeded the inherent-relative uncertainty by 0.32 °C, and thus the (Tmax) microclimate uncertainty was taken as 0.32 °C. Therefore the final HOBO uncertainty calculation all four uncertainties (inherent-relative, microclimate, siting and sampling algorithm difference) were summed, per daily temperature extreme. We recognize this was a conservative approach because some cancelling or overlap of the uncertainties likely occurs. This resulted in an assumed uncertainty in Tmin of 0.55 °C, and 1.04 °C in Tmax when using HOBO values in the integrated network.

Lastly, Tmax measurements from two HOBO stations were excluded from the study. These stations were first identified as having possibly compromised sitings, more specifically, a built or cement surface within three meters of the monitor. These observations were subsequently compared to nearby observations to assess their quality (not shown). The daily minimum temperature was devoid of obvious siting impacts, consistent with preliminary tests (not shown) indicating this type of undesirable microclimate to have a notably lesser impacts on Tmin than Tmax (e.g. building receiving shortwave radiation and warming, then radiating heat into immediate area). These stations were the Southgate and West Detroit #2 HOBO monitors (Figure 4.1). All other locations passed scrutiny of potentially compromised siting, so that the integrated network utilized a total of 30 and 28 stations for Tmin and Tmax, respectively. One station was located in a park roughly 1 km across but we retained it within the network because we felt it was very near the edge of the park.

4.2.2. Analysis of observations

In order to characterize the weather during the study, relative to long-term summer weather data, the 2009 KDTW observations were converted to percentiles relative to historical observations (Figure 4.1). The KDTW observing station's data from 1979-2009 was acquired online at the National Climatic Data Center. The Tmax (Tmin) was again considered the highest (lowest) temperature each day observed between one and seven pm (am) (EDT or GMT-4). The 1979-2008 period was used as a climate base period and a sub sampling method within that base period of a 19-day window centered on the calendar date (Figure 4.7). The percentiles were determined by calculating the empirical cumulative distribution function (i.e. no assumptions on the underlying distribution) and bi-linear interpolation of those percentiles (the same procedure as described in section 2.2.2). Subsequently, the percentile for each day during the 2009 field study was determined, for both daily temperature extremes, with respect to 570 (19x30) observations of temperatures experienced at KDTW around that calendar date. The time series, during the summer of 2009, of both Tmax and Tmin percentiles was then provided along with

the 110-day mean (110DM) percentiles. 110 days was the length of this field study and thus 110DM values are meant to represent the typical conditions during the observational study.

Initially, the basic statistics of the magnitude of spatial variability across the integrated network were calculated. The magnitude of spatial variability was taken - each day and for both daily temperature extremes - as the *range* in the anomalies across all sensors, and was referred to as the *metropolitan temperature range* (MTR). This was done for each of the 110 Tmax and Tmin of the field experiment. First the 110DM, standard deviation and maximum MTR were calculated, for both daily temperature extremes. Histograms were constructed for the MTR, in both daily temperature extremes over the field experiment duration. Lastly a Student's t-test was used to determine statistical significance (at the 0.05 significance level) in the summer mean MTR.

Next a characterization of the spatial structure of anomalies was sought. Foremost the 110DM anomalies of both daily temperature extremes were calculated and plotted on maps. Subsequent investigation into the stability (i.e. consistency) of that spatial pattern was undertaken by comparing the range in the 110DM anomalies to the 110DM MTR. The 110DM anomalies require no consistency and the 110DM MTR is reduced by inconsistency, and thus we compared the ratio of the two values in both Tmax and Tmin. We know of no other studies using this method of inference. In order to combine the temporal and spatial characteristics of the anomalies, figures were created to compare the signal-to-noise across all stations. This method of assessing the spatial consistency makes the assumption that the disparity in consistency was not limited to only the hottest and coolest stations but rather indicative of the consistency of the spatial pattern of anomalies. To eliminate uncertainty arising from the sample size (n=110), a bootstrapping procedure with 10000 resamples was used to better estimate the true means and standard deviations of the 110DM anomalies.

4.2.2.1. Intersection with dangerously hot weather

The importance of the spatial variability in temperature to the heat-health discussion depends on whether it exists during dangerously hot weather. The definition of "dangerous" is not universal, thus we chose to evaluate the intersection of the MTR with days labeled dangerous by various methods. Methods were generally chosen because they were familiar to heat health warning system operators in most U.S. cities, and can apply outside of the region. These methods are more fully described below.

The first method used the air-mass-based Spatial Synoptic Classification 2 method (SSC2) (Sheridan 2002; Kalkstein and Sheridan 2003) to diagnose days. Air masses are large homogeneous bodies of air, which when present, exhibit distinctive meteorological characteristics (Kalkstein et al. 2008). Then historical mortality data is examined for substantial empirical relationships with air mass observations, along with cofounders (e.g. duration, timing within season, etc.). It was a great way to evaluate the MTR because most modern heat health warning system operators are familiar with it and even forecast it on an operational basis (R. Pollman 2009, personal communication). Also the results are useful to locations outside of Detroit since the air masses are not specific to Detroit. In Detroit "Dry Tropical", "Moist Tropical +" and "Moist Tropical ++" are the dangerous air masses in the SSC/SSC2. The mean MTR was calculated as a function of all air masses observed, as well as a group of the oppressive air masses. This was done to further our characterization of the MTR with respect to weather conditions. Calendar data provided by Dr. Scott Sheridan's website (Sheridan 2010) was used for this purpose.

The second method was that of the NWS heat advisory. Local NWS forecast offices call heat advisories based on weather and heat stress related index forecasts, with familiarity of their local climate and recent weather conditions (R. Pollman 2009, personal communication). All major cities within the U.S. have NWS forecast offices available to them, and thus this method to evaluate the MTR in some ways should relate to all cities. Tmax and Tmin MTR were selected that occurred during period of advisories as indicated from the NWS's *Non-Precipitation Warnings, Watches, Advisories* bulletins (NCDC HDSS Access System reference). The mean MTR was

calculated for the Tmax on all dates with heat advisories spanning the afternoon and for the Tmin on all dates with heat advisories overnight.

The last method was to investigate the linear relationships between temperature and MTR; this method was advantageous because it's simple and direct. Specifically, ordinary least squares regression between the normalized observed MTR and the temperature percentile values was used to determine the magnitude of trends and their significance. Normalized refers to dividing all the values by the 110-day maximum value. The temperature percentile is always an important driver of any heat health warning system. Calculation details regarding temperature percentile values were in section 4.2.2. Scatter plots and the 95 percent confidence intervals of the linear trends were calculated as well as the y-intercepts at the extrapolated 100th percentiles. The correlation coefficient (R), percent of variance explained (R^2) and model P-value were also calculated.

4.2.2.2. Spatial variability's relationship with weather

Large day-to-day variability of the MTR led to investigation into whether the variability could be diagnosed, or predicted, by the city-wide weather conditions. The elected weather variables were chosen because they theoretically could impact the physical processes that lead to some areas being warmer than others (e.g. cloud cover as an proxy for radiative heating or cooling rates). The Tmin MTR was evaluated against three variables: previous afternoon average cloud-cover percentage, morning average cloud-cover percentage and morning average wind speed. The relationship with the previous afternoon cloud cover had not been previously inspected. The Tmax MTR was tested against two variables: average afternoon cloud-cover percentage and average afternoon wind speed.

These weather observations were spatially averaged over three surrounding NWS ASOS/AWOS locations (KVLL, KDET and KDTW) (Figure 4.1), and temporally averaged for the morning between four and eight AM, and the afternoon between two and six PM (all EDT, or GMT-4). Once the relationships with the NWS ASOS/AWOS observations were determined the investigation was repeated with corresponding

variables from a reanalysis dataset. Reanalysis datasets are blended products of observations and weather forecasting models, and thus variables available in those models are available in the reanalysis datasets. The relationships were rebuilt with reanalysis data in order to examine the differences in predicative abilities between in-situ observational data and model-observation blended products. The North American Regional Reanalysis (NARR) dataset (Mesinger et al. 2006) was used and similarly spatially averaged (e.g. grid points that the three airports were located within) and temporally averaged (e.g. only the times within the time ranges previously used). The wind speed values, for both afternoon and nighttime, were calculated from NARR's *3-hourly value of U-Component at 10-meters* and the analogous *V-Component* variable. To represent cloud cover the *3-hourly forecast of Total Cloud Cover of the entire atmosphere* was used for the overnight periods and the *3-hourly average Downward Shortwave Radiation Flux at the surface* for the afternoon periods. Then for both predictands stepwise multiple regression, via the backwards elimination method, was performed to select the appropriate predictands.

Evaluation of the regression coefficients determined from scaled inputs, was done to determine each predictor's relative contributions to the models. To determine the goodness of those regression models, statistics such as P-values, root mean squared error (RMSE) and R^2 were compared to both the mean and maximum MTR. Multiple regression models that passed initial scrutiny were further tested for their functionality by evaluating their predictive ability in a cross-validation manner. This was done by breaking the sample into five even number groups by counting by fives throughout the 110-night sample (i.e. classic k-folds method). Then five times over, a model could be trained using four of the groups and then have its performance quantified by the left-out group. Then the mean RMSE and average absolute difference were calculated across all five groups and compared to the maximum and mean MTR of the 110-night evaluation sample. The *average absolute difference* was calculated to give confidence to the RMSE values, and allow others an additional way to compare our models performance to their own models. Additionally, the confidence in the determined regression coefficients was evaluated using a bootstrapping technique to evaluate regression coefficients. This was

accomplished by creating 10000 bootstrap samples, by sampling with replacement from the 110-sample, and then determining an estimate of the regression coefficients each time. In addition to calculating the median coefficient (for the regression coefficient estimator), the standard deviation of those 10000 coefficients was calculated as a measure of variability.

4.2.2.3. Mean anomalies relationships with land attributes

Next, investigation into the relationships between the spatial attributes and observed temporally averaged anomalies was undertaken. These relationships were built and assessed (for performance) using the observing stations, and since spatial attributes are continuous they can be used to estimate the temperatures in the spaces between the observing stations. These processes are described below.

We created three spatial attributes that we will refer to as “distance to city center” (D2CC), “distance to water” (D2H2O) and “percent impervious surface” (PIS). The values of these variables were extracted at each monitoring location using Esri ArcGIS software. The Spearman Rank (SR) correlation coefficients between each different spatial variable were calculated, and then investigation into their relationships with the 110DM anomalies was undertaken. SR correlations were used because the spatial data values looked non-normal (not shown). Pearson’s correlation coefficients were also calculated (not shown) and did not disagree with the SR correlation coefficients. The spatial variables are better described below.

The first variable, PIS is an indicator of the built environment (Oke 1973; Arnold and Gibbons 1996). Impervious surfaces are the hard constructed surfaces that cover buildings, roadways, parking lots, etc. Areas surrounded by a larger PIS may be more likely to store heat, and then release that heat overnight. The spatial dataset used was the National Land Cover Data 2001 Imperviousness dataset, which was derived from Landsat imagery taken in 2001 (U.S. Geological Survey 2008). Maps were constructed with both the sites of the monitoring and a superimposed data layer of per-pixel (30m resolution) estimates of imperviousness. Using Esri ArcGIS software, the percent of surrounding surface indicated as impervious was

calculated for each station. Imperviousness does not take into account the 3-dimensional geometrical factors or distance from city center. Initially investigated was the sensitivity between the temporally averaged mean anomalies, both Tmax and Tmin, for each station and the PIS values within various circular radii from 0.15 to 3.0 km. SR correlation coefficients were calculated between each radius' PIS values (at each station) and the 110DM anomalies. This was done in order to determine the strongest relationship between local percent impervious values and both daily temperature extremes. Subsequent assessment of the relationship with PIS was performed only with the radius with the strongest relationship.

The second spatial variable, D2H2O was a logical driving mechanism of temperature variability. If the thermal inertial of a water body could dampen the diurnal cycle, then it would impact both daily temperature extremes of near-shore locations. Specifically with regards to Tmax, both small-scale 'lake breezes' and meso-scale cold-air advection could potentially cool near-shore regions. With regards to Tmin, lower-level atmospheric water vapor content at near-shore locals might decrease long wave radiative cooling rates at the surface and subsequently affect shelter height temperatures. Relevant water bodies were taken here as the Detroit River, Lake St. Clair and Lake Erie and the straight-line distances were drawn from the land-water boundary nearest to each location.

Next, distance from city center was examined as a driving mechanism of the temperature variability. The traditional, and crude, model of the UHI phenomenon consists of concentric isotherms surrounding a city (coolest on the outside). The city center might be warmer because of contributions from increased anthropogenic heat flux and roughness length nearer the city center. Roughness length retards the rate of heat loss (upwards) within the urban canopy layer and roughness sub layer because it stretches them vertically. Additionally, latent heat flux and short-wave albedo should be lowest in a city's interior (due to moisture availability and short-wave radiation trapping). However during Tmax, one can make some counter points in Detroit, with the water bodies' proximity to downtown. Conversely, the relationship between Tmin and distance from city center is uncomplicated, considering the urban-canyon effect (e.g. long-wave radiation trapping) and volumetric heat capacity are

expected to be largest at the city center, and the proximity to water is not expected cool near-shore locations at night.

We then evaluated the ability to predict the 110DM anomalies from the aforementioned spatial attributes. The backwards elimination stepwise regression method was employed with all three variables as predictors. In order to evaluate the determined regression equations, the RMSE, R^2 and full-model P-value were calculated. A brief evaluation within each equation (e.g. strength of the standardized correlation coefficients) and between equations (e.g. RMSE, R^2) was undertaken. To test the utility of the regression equations that were deemed successful, we predicted each station's temporally averaged temperature anomalies using the other stations, in a leave-one-out cross-validation manner (i.e. the k-folds method). Equation goodness was then evaluated by calculating the RMSE and average absolute error and comparing them to the range in the anomalies being predicted and the RMSE from the model built on all stations**.** The confidence in the regression coefficients was evaluated with a bootstrapping technique where 10000 resamples of the stations values were made and the regression coefficients calculated for each resample. Then the median regression coefficient (of those 10000 coefficients) was calculated as well as the standard deviation of the 10000 coefficients (as a measure of variability of those coefficients). Lastly, using the spatially continuous data sets and the regression model in ArcGIS software a spatially continuous map of the bootstrap derived model output was created. An analogous map was created using the statistical model created without bootstrapping, and compared it with the aforementioned map.

4.2.2.4. Anomalies relationship during conducive conditions

The last matter investigated was how the characterizations changed from the experiment average to during weather conditions favorable to large MTRs. Thus some of the analysis was repeated, with observations used only during days where a large MTR was predicted, and compared to the summer average analysis. The bootstrapped derived regression equation relating the MTR to the weather conditions

was used to choose the dates. It was decided to select days in which the equation predicted MTR was larger than the median value; thus the new sample size was 55 days. For comparison purposes, we recalculated the signal-to-noise ratio, the mean MTR, the range in mean anomalies and then the ratio between the two. Lastly, the spatial distribution of mean anomalies was again plotted atop a map of Detroit. All these were then compared to the analogous products for the temporal average.

Evaluation was then undertaken of the ability to predict these subselected mean anomalies from the spatial variables. This was undertaken only for Tmin. The backwards elimination stepwise regression method was used and indicated that all three variables were appropriate predictors for a regression equation of the anomalies. Again the RMSE, R^2 and full-model P-value were calculated for performance evaluation purposes. The coefficients calculated using normalized predictor variables were used to quantify the influence each variable had. To further test the ability of the spatial attributes to predict the mean anomalies we estimated each station's anomaly in the leave-one-out cross-validation manner previously described; then calculated the RMSE and average absolute error in predicting the true values. Again, the same bootstrapping method as described for the 110DM anomalies to estimate the regression coefficients and quantify the variability was used to subsequently assess the same things here using the subselected 55 day sample.

Lastly using this regression equation and the spatially continuous landscape attributes, a map was created and compared to the analogous map created using the regression model based on the entire 110-day sample. This assessed the model output's spatial pattern sensitivity between summer average and conducive-weather.

4.3. Results

When calculating the percentiles w.r.t the past 30 years worth of summer (M, J, J, A, S) data at the KDTW location, there existed ample data (e.g. only 0.0042% of the temperature data was missing). The 110DM Tmin and Tmax percentiles were both slightly cool, calculated at the 47.0^{th} and 43.5^{rd} percentiles, respectively.

Inspection of the time series (Figure 4.8) of the 110 days during the field study revealed that there was a wide range of percentiles experienced with episodes of warmer and cooler temperatures, in both Tmax and Tmin.

Analysis of the spatial variability across the network, via the MTR, showed that Tmin had more variability than Tmax. The mean (standard deviation) over the study period in Tmin was 2.8 °C (1.5 °C) and 1.4 °C (0.8 °C) in Tmax. The maximum MTR observed in Tmin was 6.3 °C and in Tmax was 4.8 °C. The histograms for both Tmin and Tmax (Figure 4.9) showed that while mild MTRs were frequent in the Tmax, large MTRs were more prevalent in Tmin. This was expected because the nighttime mechanisms (e.g. radiative cooling differences, heat storage, etc.) work in concert to create large spatial variability in temperature in Tmin. Conversely, with sometimes conflicting physical mechanisms in Tmax, large spatial variability was harder to accomplish. Lastly, for both daily temperature extremes a Student's t-test rejected the hypothesis that the true mean MTR could be zero. This simply means over the duration of the summer there was statistically significant spatial variability in both daily temperature extremes.

Next, the anomalies' spatial patterns were characterized. The 110DM anomalies, per station, were calculated and displayed on maps of the Detroit Metropolitan region (Figure 4.10). These indicated that warmer Tmin generally occurred in the downtown area (Figure 4.1) and near the waterfront while warmer Tmax generally occur in the western suburbs. The range in Tmin 110DM anomalies was calculated at 2.0 °C, and at 0.6 °C for Tmax. The ratios between those values and the corresponding 110DM MTR values (2.8 °C and 1.4 °C, respectively) were 0.72 and 0.42. These ratios imply Tmin has more consistency in which stations were the hottest and coolest, than Tmax. Physically these results make sense because wind speeds are stronger during the day, leading to wind direction and cold air advection (from the water bodies) playing bigger roles in the resultant spatial pattern. Moreover, lake breezes play a big role (restricted to daytime) and vary in extent and direction. Looking at maps of each date's spatial anomalies fully supported these results (not shown).

When the stations' mean anomalies were plotted with their respective standard deviations (Figure 4.11), it was clear that both of the signal-to-noise ratios were noise heavy – especially Tmax. These figures imply large day-to-day variability. These results were physically sensible: on some dates the spatial variability was small and on others it was large; there was large variability at each station relative to the temperature differences between station means, particularly in Tmax.

4.3.1. Intersection with dangerously hot weather

Next the existence of spatial variability across the urban region during dangerously hot weather was examined. The first method of classifying dates as stressful was the air mass-based SSC2, which indicated only three days (June 24th, Aug. 9th and Sept. 22nd) were in air masses considered oppressive to Detroit, and they were not consecutive (i.e. an extreme heat event). These days were classified as "moist tropical plus" airmass days, and no "dry tropical" days were observed. The mean MTR observed during those three dates was 2.0 °C in Tmin, and 0.9 °C in the Tmax. There was not enough data for a Student's t-test to be meaningful. Also the mean MTR was calculated for each air mass type (Table 4.2); the results indicated larger Tmin MTRs during dry air masses and did not indicate bias with warmer (as opposed to colder) air masses. This was physically sensible as atmospheric humidity is associated with clouds and consequently low radiative heat loss (and thus low disparity in heat loss rates) at the surface.

The second method of selecting dangerous dates, the region's NWS heat advisories, only had two days (June 24th and Aug. 9th) during which the afternoon was under alert, and no dates with heat advisories at nighttime. The mean MTR was 0.7 °C in Tmax during those dates.

The last method of exploring the relationship with dangerous weather was to examine the relationships between temperature percentiles and the MTR. Both the Tmin and Tmax MTRs were normalized and thus the units of slope here were percentile MTR per temperature percentile. A scatter plot (Figure 4.12) indicated that Tmin MTR exhibited a weak inverse relationship with temperature. Figure 4.12

suggests Tmax was not a function of temperature percentile. The regression coefficients and confidence intervals confirmed that the there was no significant relationship with Tmax percentiles and only a weak, albeit significant, relationship with Tmin percentiles (Table 4.8). The R for Tmin was 0.41 (R^2=0.17) and 0.05 (R^2=0.002) for Tmax; the P-value for the two regression models was < 0.001 for Tmin and 0.62 for Tmax. Further analysis indicated that cloud cover was somewhat to blame for the inverse correlation (not shown).

4.3.2. Spatial variability relationship with weather

The ability to predict the amount of spatial variability (i.e. MTR) from the larger scale weather conditions was investigated. Preliminary scatter plots indicated linear relationships (Figure 4.13). The multiple stepwise regression method indicated all three predictors, of Tmin MTR, should be included. For Tmax, the multiple stepwise regression method indicted only the wind speed would be an appropriate predictor. Speculating on physical mechanisms to explain why cloud cover wasn't included; cloud cover percentage is potentially a weak predictor of downwelling shortwave radiation (due to scattering and lack of information about cloud height or type) or possibly during cloudy afternoons large spatial variability in temperature can arise due to isolated precipitation (e.g. thunderstorms). The Tmin MTR multiple regression coefficients, with the predictor variables normalized, indicated that overnight cloud cover was the most important (-0.72) variable and of lesser importance were the previous afternoon cloud cover (-0.39) and overnight wind speed (-0.28) variables. These results suggested that the driver of the spatial variability of temperature during the daily low in Detroit was dominated by factors related to radiative cooling (e.g. sky view), as opposed to those related to mixing (e.g. wind).

The Tmin regression equation performed relatively well with a RMSE of 1.0 °C, a R^2 of 0.55 and P-value of effectively zero compared to a 2.8 °C mean and 6.3 °C maximum MTR. The Tmax regression equation was less effective with a RMSE of 0.8 °C, a R^2 of 0.11 and P-value on the order of 10^{-4} compared to a 1.4 °C mean and 4.8 °C maximum MTR. Such results for Tmax suggested that the Tmax spatial

variability was influenced by variables not considered in this study. It is the author's inclination that both the synoptic wind direction and temperature difference between the city and the adjacent lakes were possibly such variables.

Further investigation seemed warranted into the expected fit of the observational-based model predicting the Tmin MTR. This was accomplished via cross validation by separating model training and testing samples, and then averaging the equation performance measures (i.e. RMSE, average absolute difference). The equations built on these samples (n=88) had similar coefficients for the individuation terms to those coefficients built on the entire sample (n=110) (not shown). The mean (std) RMSE was 1.0 °C (0.2 °C) and the average absolute difference between predicted and observed values was 0.8 °C (0.2 °C) and these can be compared with the mean MTR of 2.8 °C and maximum of 6.3 °C. The lack of performance decline suggests that this model was useful and the performance was not overly dependant on the data that was used to train the model. Lastly, we used a bootstrapping technique to estimate the regression coefficients. The mean regression coefficients were calculated and compared well with the coefficients calculated from the 110-night sample (Table 4.4). The standard deviations suggested we had confidence in these coefficients, but also that the confidence in the overnight cloud cover coefficient was largest and the confidence was lowest in the coefficient of the previous day's cloud cover. Possible physical explanations for the performance of previous day's cloud cover as a reliable predictor were the roles of diffuse radiation and precipitation, as well as the amount of cloud cover earlier than 1pm (cloud cover was calculated from the observations from 2 to 6pm).

The appeared ability to diagnose the MTR from city-wide weather observations led to investigation into whether a reanalysis product could be used instead. The SR correlation coefficient (R=0.33) indicated substantial differences between the NARR derived afternoon cloud cover percentages and the corresponding observational percentages. Due to the spatial scale of precipitation and cloud formation processes, disconnects between a model-observation blended product and the observations are not uncommon. Subsequently, in its place a possibly superior NARR variable *afternoon downward shortwave radiation flux at the*

surface was employed instead of afternoon cloud cover percentage. Such a variable could be superior to cloud cover as it likely incorporates factors such as cloud height, scattering and atmospheric humidity, but in general there was low confidence in such a variable due to it being based on cloud cover (it is difficult for models to accurately diagnose).

Multiple stepwise regression with respect to Tmin MTRs indicated that all three predictors were appropriate to use for regression. Conversely, for Tmax the multiple stepwise regression method indicted again that only the wind speed would be an appropriate predictor. The Tmin MTR multiple regression coefficients, with the predictor variables normalized, indicated overnight cloud cover was the most important (-0.72) and previous afternoon radiation flux (0.36) and overnight wind speed (-0.20) being of lesser importance. Besides the strength of the wind speed coefficient being reduced these coefficients were very similar to those in the model based on the in-situ observations. Overall the performance decline was small when using reanalysis variables relative to in-situ observations models. The equation had a RMSE of 1.2 °C, a R^2 of 0.42 and P-value of effectively zero. The Tmax regression model was also slightly less effective than the corresponding model with observational inputs. That model's performance was not as good as the Tmin model's performance and had a RMSE of 0.8 °C, a R^2 of 0.08 and P-value on the order of 10^{-3}.

Likewise, it also seemed warranted to further investigate the usefulness of the reanalysis-based Tmin model predicting MTRs. A cross validation technique was employed to test the dependency of the training sample. The equations built here were similar to the one built using the entire sample, and performed similarly to the 110-day trained model with a RMSE of 1.2 °C and an average absolute difference of 1.0 °C (Table 4.5). Again, we used bootstrapping to estimate confidence in the regression coefficients and to more accurately determine the mean. These estimates were very comparable to those from the 110-day sample. They also indicated a similar confidence in the coefficients as was indicated in the model with in-situ observations as input (Table 4.4). The only difference was a lesser confidence (e.g.

larger standard deviation to mean ratio) in the wind speed term coefficient from the reanalysis data.

4.3.3. Mean anomalies relationship with land attributes

Prior to predicting each station's 110DM anomaly the spatial attributes inherent to those stations themselves were briefly investigated. First, the radius of PIS best correlated to the 110DM anomalies was calculated. Correlations with Tmin seemed to decrease with larger distance and ranged from 0.68 (P-value=10^{-5}) associated with the 0.20 km radius to 0.27 associated with the 1.95 km radius (Figure 4.14). The daily maximum correlations were less systematic, but ranged from 0.39 (P-value=0.04) at 1.95 km to -0.06 at 2.65 km. These results suggested that Tmin was preferentially influenced by smaller scale factors (e.g. moisture availability and heat storage characteristics) as opposed to Tmax that were hardly scale size dependant. This makes sense because the heat released from the urban fabric overnight would affect smaller areas due to advection/mixing being driven by winds, which are typically weaker during the daytime. In addition, the Pearson's correlation coefficients did not disagree with these findings (Figure 4.14). Moving forward, the two aforementioned radii of strongest correlation were used to calculate the PIS attributes.

The correlations between the spatial variables were then investigated. The SR correlation coefficient between 0.20 km PIS and D2H2O was calculated at -0.30 (P-value=0.11), and between the 1.95 km PIS and the D2H2O it was calculated at -0.12 (P-value=0.54). The correlation coefficient between 0.20 km PIS and D2CC values was calculated at -0.60 (P-value=10^{-4}), and similarly for the 1.95 km values was calculated at -0.65 (P-value=10^{-4}). Lastly, the correlation coefficient between the D2CC and D2H2O attributes was calculated at 0.74 (P-value=10^{-6}). These results were anticipated since the spatial patterns of D2H2O and D2CC variables were similar in many regards and much less similar to the PIS variable. Certainly the correlation between D2CC and PIS will vary from city to city. Detroit's lack of correlation allowed evaluation of which variable was more influential.

Next we evaluated the ability to predict each station's 110DM anomalies from the aforementioned spatial attributes. The stepwise multiple regression method indicated for Tmin that all three variables were appropriate predictors. The RMSE for the Tmin model was calculated at 0.3 °C (the range in 110DM anomalies was 2.0 °C), the R^2 was 0.62 and the P-value was 10^{-5}. For the Tmax equation, the stepwise multiple regression indicated that D2H2O and D2CC variables only should be used. This equation led to a RMSE of 0.1 °C (the range in 110DM anomalies was 0.6 °C), a R^2 of 0.33 and a P-value on the order of 10^{-3}. The superior performance of the Tmin model, compared to the Tmax model, was not a surprise as previous results alluded to possible important factors (to explain Tmax) not being accounted for. Conversely, explaining more than half of the variability with only the three variables was a pleasant surprise, as the model did not incorporate urban geometry variability (e.g. sky view factor, height to width ratios).

Examination of the Tmin correlation coefficients of the normalized predictors showed PIS was most influential (0.38), and D2H2O (-0.30) and D2CC (0.27) were secondary (Table 4.7). It was expected that PIS would outperform D2CC, but the substantial contribution from D2H2O was unexpected. Physically D2CC is simply an ideal model of land development, thus theoretically PIS should outperform it. The physical mechanism behind D2H2O is more ambiguous, however. It has already been stated that locations near the water might have more water vapor in the atmosphere, which could reduce radiative cooling rates overnight (thus raising the temperatures). However another hypothesis was as follows: on nights with a large temperature differential between land and water land breeze circulations should develop, and thus locations near shore would have wind directed out towards the water. That air is originating either from a location just inland (often of denser development than the onshore locations) or from sinking air aloft. A similar analysis for Tmax indicated that the D2H2O variable was dominant (0.12) over the D2CC (-0.08) variable. These values for coefficients were expected inasmuch as the D2H2O variable being the dominant driver of Tmax.

The Tmax model explained a small portion of a spatial variability that did not vary much, and thus it was subsequently decided that only the Tmin equation would

be tested further. Through cross-validation each station's 110DM anomaly was predicted with an equation built on the other 29 stations, and these predicted values were compared to the actual values to give a mean RMSE of 0.4 °C and absolute difference of 0.3 °C.

Cross validation is primarily designed to test for *overfitting*, or the drop in performance when tested on a location outside of the equation-training sample. Said another way, it tests how well the relationship holds in spaces (or times) where Y and X_n are not available to build the relationship(s). The lack of sizeable performance decline suggests the model can confidently be used to describe the space between the stations. By design our training stations robustly spanned the predicator variables. Lastly, bootstrapping was employed to more accurately estimate the regression coefficients. These bootstrap estimates of the coefficients compared well to the coefficients estimated from the original sample (Table 4.6) but not as well as the coefficients of the regression equation used to predict the MTR (Table 4.4). This was anticipated since the training sample size was 28, instead of the 110 in the model to predict the MTR. The estimated regression coefficients and their standard deviations suggested larger confidence in the PIS variable than the other two variables, and the least amount of confidence in the D2CC variable.

The map created from the traditional regression method (Figure 4.15) had a spatial pattern highly similar to that from the bootstrapped method (Figure 4.16). The only noticeable difference was a modestly larger range of temperature predictions from the traditional regression model. These maps showed in general the area to the north west was cool and most of the downriver (south) region and certain east Detroit areas were warm, along with the expected city center (downtown) area.

4.3.4. Relationship during conducive weather

To further explore the behavior of Detroit's urban climate, it was reevaluated with respect to only days expected to have larger than normal spatial variability in temperatures. Selecting those days was accomplished by using the regression equation derived from bootstrapping (Table 4.4) to predict the Tmin MTR. Since the

regression equation for predicting Tmax was not effective, days were retained instead because they had less than the median windspeed (8.6 m s^{-1}).

The means and standard deviations were calculated for the different locations and the subsequent signal-to-noise ratios were larger in both daily temperature extremes (Figure 4.17). The mean MTRs during the 55 selected dates were larger than the respective 110DM anomalies, calculated at 3.7 °C (Tmin) and 1.8 °C (Tmax). The ranges of the aforementioned mean anomalies were also larger than the corresponding 110-day ranges, calculated at 3.1 °C for Tmin and 1.0 °C for Tmax. Thus the ratio of the range in mean anomalies to the mean MTR was 0.82 in Tmin and 0.55 in Tmax; these were again larger than the corresponding 110-day ratios (0.72 and 0.42). Larger ratios physically suggest that the spatial patterns were more consistent during the 55 selected dates than during the entire 110-day sample. Physically sensible, as the 55 selected dates had fewer dates with small Tmin-MTR values (dates not selected were preferentially biased towards small MTR because the selection was based on prediction of the MTR equation, from the weather variables, being above the median value), and during Tmax less influence from the wind. Lastly, the mean anomalies were plotted over a map of Detroit (Figure 4.18), and showed a similar-yet-stronger pattern to that of the 110DM anomalies.

The stepwise regression method indicated that all three spatial variable predictors were appropriate for predicting the 55-selected-dates mean anomalies at each station. This model explained slightly less variability compared to the equation used to predict the 110DM anomalies; the RMSE was 0.5 °C (the range in anomalies was 3.1 °C), the R^2 was 0.60 and the P-value was on the order of 10^{-5}. This was physically sensible because when the weather conditions were conducive to spatial variability in temperature, the temperature pattern was more sensitive to differences in landscape (e.g. urban geography, soil moisture, vegetation, etc.); thus more variability existed and the statistical model explained a smaller portion of it (because we did not add predictor variables). The ratio of the RMSE to range in mean anomalies was slightly smaller than that of the corresponding 110-day model. The most influential predictor was again the PIS (normalized coefficient of 0.56) with D2H2O being second (-0.43) and D2CC slightly less influential (0.40). All three non-

normalized coefficients were larger for the 55-selected-dates regression model because the variability they were explaining was larger (Figure 4.17).

Cross validation efforts showed the RMSE was 0.6 °C and average absolute difference 0.5 °C. The lack of large performance decline suggests this model was independent of the training stations and thus would hold true in the space between the stations. The bootstrapping method indicated again that the MTR models (Table 4.4, 4.5) did a better job of estimating the regression coefficients; likely due to training sample size (Table 4.7). The estimated regression coefficients and their corresponding standard deviations also indicated slightly less confidence in the estimates for this model compared to the one of the 110-day period. Similarly to the 110DM anomalies model, the results indicated the least (most) confidence in the D2CC (PIS) coefficient. The map created using this bootstrap derived regression model (Figure 4.19) was nearly identical (albeit larger range, more distinctive features) to those of modeling the 110DM anomalies (Figures 4.15-4.16).

4.4. Conclusions

While spatial variability in temperatures existed for both the Tmax and Tmin, the variability in Tmin was larger and the pattern more consistent than in Tmax. Results such as the histograms of the MTR and the maps of the averaged anomalies indicated larger spatial variability. The ratios between the mean temperature range and range of mean anomalies also suggested such spatial pattern coherency. This was similar to other recent studies that found (similar) indices such as the “UHI magnitude” (Runnalls and Oke 2000; Wilby 2003; Erell and Williamson 2007; Gaffin et al. 2008; Basara et al. 2008; Camilloni and Barrucand 2010) and “intra-urban spatial variability” (Wilby 2003; Erell and Williamson 2007; Gaffin et al. 2008; Basara et al. 2008) to be largest at nighttime. The existence of spatial variability during Tmax was confirmed similar to other studies (e.g. Oklahoma City in 2003 (Basara et al. 2008)).

The amount of spatial variability was comparable during hot weather as it was during typical weather. The mean MTRs calculated during subsets of hot weather

being comparable to the temporal averages supported this conclusion. Furthermore, a lack of substantial relationships between temperature percentiles and MTR shored up this conclusion. Very few studies have investigated this, but our results seemed to confirm the findings of a previous study that examined, and found, both the intra-urban and intra-suburban temperature variability during a heat event (Basara et al. 2010).

The amount of spatial variability during Tmin depends on the large-scale weather conditions, and there were indications that forecasting that variability may be possible. The strong performance of the statistical models using both reanalysis and in-situ observational data led to these conclusions. Furthermore, bootstrapping and cross validation methods suggested confidence/utility in these statistical models. This was consistent with other studies that indicated both the nighttime intra-urban spatial variability (Eliasson 1996, Wilby 2003, Kim and Baik 2005, Erell and Williamson 2007) and UHI magnitude (Runnalls and Oke 2000, Morris et al. 2001, Gedzelman et al. 2003, Wilby 2003, Camilloni and Barrucand 2010) to be largest during low wind speed and clear sky weather conditions. However, our results build on these findings, first by suggesting an explicit relationship between Tmin spatial variability and the previous afternoon's average cloud cover, which was not found in the literature. Secondly, previous studies relating the spatial variability to weather variables derived from a reanalysis dataset were not found.

The spatial patterns of temperature were more closely associated to land characteristics during the Tmin than during the Tmax. The stronger performance of the Tmin statistical models supported this conclusion. Interestingly, Buttstädt et al. (2010) found stronger relationships between geospatial variables and temperature during the afternoon than early morning. Possible explanations for this discrepancy were the definition of time periods (i.e. Tmin versus "before 7:30am (local time)"), differences in climate (Detroit versus western Germany) or differences in geospatial variables (urban geometry and LCLU classification variables were primarily used by Buttstädt et al. (2010)). Conversely Kuttler et al. (1996) found stronger relationships during the nighttime than daytime. Our model of the spatial pattern of temperature was primarily driven by local imperviousness and distance to water. The Kuttler et al.

(1996) study also found local imperviousness to be a relatively strong predictor of overnight temperatures. Other similar studies using statistical models to explain the nighttime metropolitan spatial pattern of temperature (Kuttler et al. 1996, Bottyán and Unger 2003, Buttstädt et al. 2010) confirmed it could be done adequately (R^2 between 0.50-0.85) with only a small number of variables (2-4). Other studies (Gaffin et al. 2008) however, failed to find land-cover variables driving temperature but did not use the same variables this study did. No previously studies were found that used distance to water as a predictor for Tmin temperatures.

The signal-to-noise of the spatial variability was small over the temporal average but larger during weather conditions linked to large MTR values. Comparisons of the station anomaly signal-to-noise ratios between date-samples (110 vs 55) support these conclusions. The implications were that the probability of expected temperature patterns is higher during certain weather conditions (e.g. anticyclonic weather). We know of no other study that explicitly investigated the signal-to-noise ratio as a function of weather condition.

Lastly, relationships with the mean temperatures built during weather conditions linked to large MTR values performed well but explained relatively less of the total variability (which was larger than during the temporal average). This physically suggests that the spatial variability of urban climate temperatures is highly reduced by wind speed, and so when wind speed is held low you have more factors meaningfully influencing the temperatures. The performance of both the statistical models and the maps of mean anomalies during a subset of conducive conditions supported these conclusions. Other studies (Kuttler et al. 1996; Buttstädt et al. 2010) showed marginally improved performance of regression models during dates selected based on weather conditions.

4.5. Discussion

The conclusions concerning the existence of spatial variability in temperature during dangerously hot weather would benefit from exploration into whether this is true for climates other than that in Detroit and Oklahoma City (Barasra et al. 2008)).

Such conclusions inform heat-health decision makers whether they should consider exposure disparities by location within the city, and thus its role in heat stress vulnerability mapping. Daytime temperatures being unusually hot during the early portion of the summer likely has a stronger relationship with temperature variability through the physical mechanisms of land-lake breezes, than they do during the later portion of the summer. This is because in the late summer the water temperatures are likely the highest of the year, which reduces the land – water temperature gradient.

Atmospheric moisture is associated with both warmer overnight temperatures, and less spatial variability. Specifically air masses with higher atmospheric moisture exhibited lower spatial variability in temperature; thus predictions of a future climate with more (less) occurrences of relatively moist air masses would suggest a future decrease (increase) in attention to the spatial variability in temperatures will be required. Additionally with respect to climate forecasts, a future climate with stronger (weaker) winds and/or more clouds would require less (more) attention to spatial variability in temperature in urban areas.

The conclusions regarding the relationships between variables representing the spatial variability in temperatures during Tmin and the preceding afternoon's shortwave radiation received would profit from exploration into whether this is a universal relationship and/or how its influence changes depending on climate or latitude. Furthermore, such conclusions suggest that reducing (daytime) shortwave radiation absorption can mitigate overnight spatial variability in temperature over an urban region – supporting the legitimacy of albedo-alteration as a method of urban geoengineering towards mitigating the urban heat island.

Furthermore, our conclusion concerning the temporal variability of spatial variability conveys to heat-health decision makers that optimally the amount of spatial variability would be forecasted for upcoming hot weather episodes. Heat exposure is only one spatially variable factor in heat vulnerability mapping, and if the amount of variability is going to be small then placing it in an area focusing on other variables (e.g. population age, income) may be preferable. Conversely, if the amount of spatial variability is deemed high, resources should be deployed in areas with high heat

stress risk. Since the models built during average weather suggest the same locations as the models built during weather with above average spatial variability, statistical models based on only conducive weather conditions were unnecessary for spatial indicating those locations of high heat.

The implications of this study still hold outside of the Detroit region. First the way the intersection with dangerously hot weather was explored (e.g. NWS advisories, relationship with temperature percentile, SSC2 method) applies outside of the Detroit region. Next these results demonstrate that when modeling spatial patterns of temperature both the geospatial variables intrinsic to urban development (e.g. percent impervious surface) and variables intrinsic to the natural terrain (e.g. distance to water) are important in modeling spatial patterns of temperature. Lastly, the proof of concept for exploring the spatial variability in temperature can be applied to other cities/regions and existing observing networks. Specifically, similar integrated observing networks and statistical models temporally and spatially forecasting the variability could be used to inform the HHWS designers in any city.

4.6. Acknowledgements

The research was sponsored by the Graham Environmental Sustainability Institute at the University of Michigan, the U.S. Environmental Protection Agency Science to Achieve Results (STAR) under grant R832752010 and the Centers for Disease Control and Prevention under grant R18EH000348. I also appreciate Craig Fitzner (MDEQ) for scientific data support. Additionally, the generosity is acknowledged of all those who granted site access permission for installation and operation of observing monitors. Finally, I must acknowledge the hard work by Drs. Marie S. O'Neill, Kai Zhang, Jalonne White-Newsome, Carina Gronlund and Daniel Brown, as well as Mr. Shannon Brines as they are coauthors on the published manuscript.

4.8. References

Ackerman B., 1985: Temporal march of the Chicago heat island. *Journal of Climatology and Applied Meteorology*, **24**: 547-554, doi: 10.1175/1520-0450(1985)024<0547:TMOTCH>2.0.CO;2

Aida, M., 1982: Urban albedo as a function of the urban structure – a model experiment. *Boundary-layer Meteorology*. **23**: 405-413, doi: 10.1007/BF00116269.

Barring, L., J. O. Mattsson, S. Lindqvist, 1985: Canyon geometry, street temperatures and urban heat island in Malmö, Sweden. *Journal of Climatology*, **5**: 433-444, doi: 10.1002/joc.3370050410

Basara, J. B., P. K. Hally Jr., A. J. Schroeder, B. G. Illston, K. L. Nemunaitis, 2008: Diurnal cycle of the Oklahoma City urban heat island. *Journal of Geophysical Research*, **113**, doi:10.1029/2008JD010311.

Basara, J. B., H. G. Basara, B. G. Illston, K. C. Crawford, 2010: The impact of the urban heat island during an intense heat wave in Oklahoma City. *Advances in Meteorology*, **2010**, doi:10.1155/2010/230365.

Bornstein, R. D., 1968: Observations of the urban heat island effect in New York City. *Journal of Applied Meteorology*, **7**: 575-582, doi: 10.1175/1520-0450.

Bottyán, Z. J. Unger, 2003: A multiple linear statistical model for estimating the mean maximum urban heat island. *Theor. Appl. Climatol.*, **75**, 233-243, doi: 10.1007/s00704-003-0735-7.

Buttstädt, M., T. Sachsen, G. Ketzler, H. Merbitz, C. Schneider: Urban temperature distribution and detection of influencing factors in urban structure. In: Int. seminar on Urban Form. Hamburg (2010).

Camilloni, I. and M. Barrucand, 2010: Temporal variability of the Buenos Aires, Argentina, urban heat island. *Theoretical and Applied Climatology*, doi: 10.1007/s00704-011-0459-z.

Chen, L., E. Ng, X. An, C. Ren, M. Lee, U. Wang, Z. He, 2012: Sky view factor analysis of street canyons and its implications for daytime intra-urban air temperature differentials in high-rise, high-density urban areas of Hong Kong: a GIS-based simulation approach. *International Journal of Climatology*, **32**: 121-136, doi: 10.1002/joc.2243.

Childs, P. P. and S. Raman, 2005: Observations and numerical simulations of urban heat island and sea breeze circulations over New York City. *Pure and Applied Geophysics*, **162:** 1955-1980, doi: 10.1007/s00024-005-2700-0.

Erell, E. and T. Williamson, 2007: Intra-urban differences in canopy layer air temperature at a mid-latitude city. *Int. Journal of Climatology*, **27**, 1243-1255, doi: 10.1002/joc.1469.

Eliasson, I., 1996: Intra-urban nocturnal temperature differences: a multivariate approach. *Climate Research*, **7** 21-30, doi: 10.1002/joc.1469.

Freitas, E. D., C. M. Rozoff, W. R. Cotton, P. L. Silva Dias, 2007: Interactions of an

urban heat island and sea-breeze circulations during winter over the metropolitan area of São Paulo, Brazil. *Boundary-Layer Meteorology*, **122**:43-65, doi: 10.1007/s10546-006-9091-3.

Gaffin, S. R., and coauthors, 2008: Variations in New York city's urban heat island strength over time and space. *Theor. Appl. Climatol.*, **94**, 1–11, doi: 10.1007/s00704-007-0368-3.

Gedzelman, S. D., S. Austin, R. Cermak, N. Stefano, S. Partridge, S. Quesenberry, D. A. Robinson, 2003: Mesoscale aspects of the urban heat island around New York City. *Theor. Appl. Climatol.,* **75**, 29-42, doi: 10.1007/s00704-002-0724-2.

Georgescu, M., M. Moustaoui, A. Mahalov, J. Dudhia, 2011: An alternative explaination of the semiarid urban area "oasis effect". *Journal of Geophysical Research*, **116**: D24113, doi: 10.1029/2011JD016720.

Grimmond, S., 2007: Urbanization and global environmental change: local effects of urban warming. *The Geographic Journal*, **173**: 83-88, doi: 10.1111/j.1475.2007.232_3.x.

Hart, M. A. and D. J. Sailor, 2009: Quantifying the influence of land-use and surface characteristics on spatial variability in the urban heat island. *Theoretical and Applied Climatology*, **95**: 397-406, doi:10.1007/s00704-008-0017-5.

Kim, Y. H. and J. J. Baik, 2005: Spatial and temporal structure of the urban heat island in Seoul. *J. Appl. Meteor.,* **44**, 591-605, doi: 10.1175/JAM2226.1.

Kuttler W., A-B Barlag, F. Roβmann, 1996: Study of the thermal structure of a town in a narrow valley. *Atmospheric Environment,* **30**: 365-378, doi: 10.1016/1352-2310(94)00271-1.

Landsberg, H. E., 1981: The Urban Climate, 278 pp., Academic, San Diego, Calif.

Magee N., J. Curtis, G. Wendler, 1999: The urban heat island effect at Fairbanks, Alaska. *Theoretical and Applied Climatology,* **64**: 39-47, doi: 10.1007/s007040050109.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. J and J. Woolen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, W. Shi, 2006: North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, **87**: 343–360, doi: 10.1175/BAMS-87-3-343

Morris C. J. G and I. Simmonds, 2000: Associations between varying magnitudes of the urban heat island and the synoptic climatology in Melbourne, Australia. *International Journal of Climatology*, **20**: 1931-1954, doi: 10.1002/1097-0088(200012)20:15<1931::AID-JOC578>3.0.CO;2-D.

Morris, C. J. G., I. Simmonds, N. Plummer, 2001: Quantification of the influences of wind and cloud on the nocturnal urban heat island of a large city. *J. Appl. Meteor.*, **40,** 169-182, doi: 10.1175/1520-0450(2001)040<0169:QOTIOW>2.0.CO;2.

Nunez, M. and T. R. Oke, 1977: The energy balance of an urban canyon. *Journal of Applied Meteorology*. **16:** 11-19, doi: 10.1175/1520-

0450(1977)016<0011:TEBOAU>2.0.CO;2

Oke, T. R., 1973: City size and the urban heat island. *Atmospheric Environment*. **7**: 769-779, doi: 10.1016/0004-6981(73)90140-6.

Oke, T. R., 1982: The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society.* **108**, 1-24, doi: 10.1002/qj.49710845502.

Oke, T. R., 1987: Boundary Layer Climates, 2nd ed., 435 pp., Methuen, London.

Oke, T. R., 1984: Methods in urban climatology. *Applied Climatology*, Zürcher Geographische Schriften, **14**, 19-29.

Oke, T. R., 2004: Initial guidance to obtain representative meteorological observations at urban sites. World Meteorological Organization report 81. 47 pp.

Philandras C. M., D.A. Metaxas, P.T. Nastos, 1999: Climate variability and urbanization in Athens. *Theoretical and Applied Climatology,* **64**: 65-72, doi: 10.1007/s007040050092.

Runnals, K. E. and T. R. Oke, 2000: Dynamics and controls of the near-surface heat island of Vancouver, British Columbia. *Physical Geography*, **21,** 283-304,

Spronken-Smith, R. A., T. R. Oke, W. P. Lowry, 2000: Advection and the surface energy balance across an irrigated urban park. *International Journal of Climatology*, **20**: 1033-1047, doi: 10.1002/1097-0088(200007)20:9<1033::AID-JOC508>3.0.CO;2-U.

Stewart, I. and T. R. Oke, 2009: Newly developed "thermal climate zones" for defining and measuring urban heat island "magnitude" in the canopy layer. Extended Abstracts, The 89th American Meteorological Society Annual Meeting, J8.2A, Phoenix, AZ.

Tarleton, L.F. and R.W. Katz, 1995: Statistical explanation for trends in extreme summer temperatures at Phoenix, Arizona. *Journal of Climate*, **8**: 1704-1708, doi: 10.1175/1520-0442(1995)008<1704:SEFTIE>2.0.CO;2.

Tereshchenko I. E. and A.E. Filonov, 2001: Air temperature fluctuations in Guadalajara, Mexico, from 1926 to 1994 in relation to urban growth. *International Journal of Climatology,* **21**:483-494, doi: 10.1002/joc.602.

Whiteman, C. D., J. M. Hubbe, and W. J. Shaw, 2000: Evaluation of an inexpensive temperature datalogger for meteorological applications. *J. Atmos. Oceanic Technol.*, **17**, 77–81.

Wilby, R. L., 2003: Past and projected trends in London's urban heat island. *Weather,* **58**, 251-260, doi: 10.1256/wea.183.02.

Zhang, K, R. B. Rood, G. Michailidis, E. M. Oswald, J. D. Schwartz, A. Zanobetti, K. L. Ebi, M. S. O'Neill, 2012: Comparing exposure metrics for classifying 'dangerous heat' in heat wave and health warning systems, Environment International, **46**: 23-29, doi: 10.1016/j.envint.2012.05.001.

Zhou Y. and J. M. Shepherd, 2009: Atlanta's urban heat island under extreme heat conditions and potential mitigation strategies. *Natural Hazards*, **52**: 639-668, doi: 10.1007/s11069-009-9406-z.

4.8. Figures

Figure 4.1: Map of the observing sites within the network used to monitor the Detroit Metropolitan region. Airport refers to FAA observing locations, HOBO refers to the temporary network and MDEQ refers to the locations of Michigan Department of Environmental Quality observing sites. A layer of imperviousness is provided for reference and four locations referenced within the text labeled.

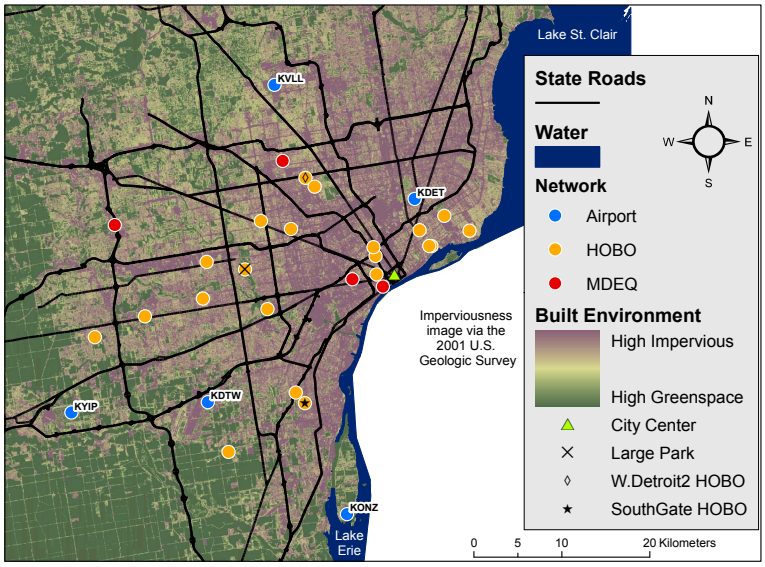


Figure 4.2: A photograph taken of the Coleman Young Municipal City airport observing station.

Figure 4.3: A photograph taken of the MDEQ observing station in Oakpark (north west of downtown).

Figure 4.4: A photograph taken of the Joy road observing site. This location was a site of MDEQ and HOBO co-location. The MDEQ monitor is circled in red and the HOBO monitor circled in yellow.

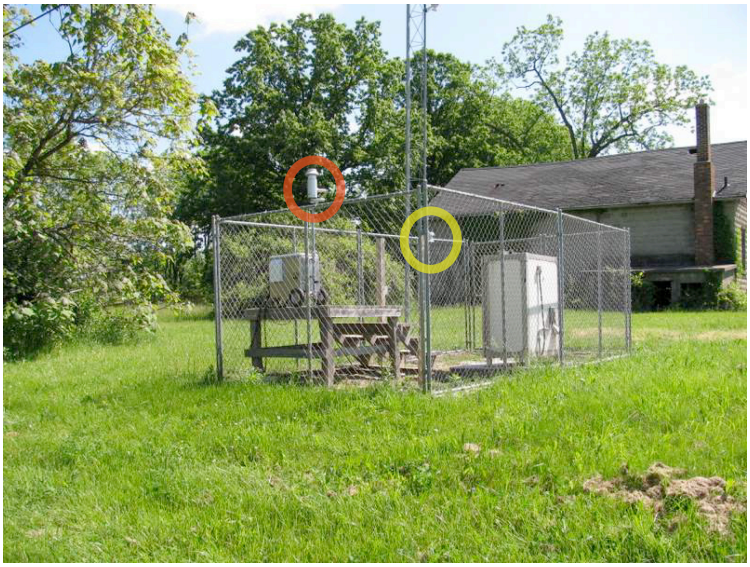


Figure 4.5: A photograph of the Allen Park observing site. This location was a site of MDEQ and HOBO co-location. The MDEQ monitor is not visible in this photograph.

Figure 4.6: A photograph of a HOBO monitoring site in Redford.

Figure 4.7: Schematic of the calendar date-dynamic base climate sample used to calculate the percentiles at each station. Thirty years worth of the dates with red (yellow, green) on them are used to determine the percentiles on the date marked with a red (yellow, green) X on it.

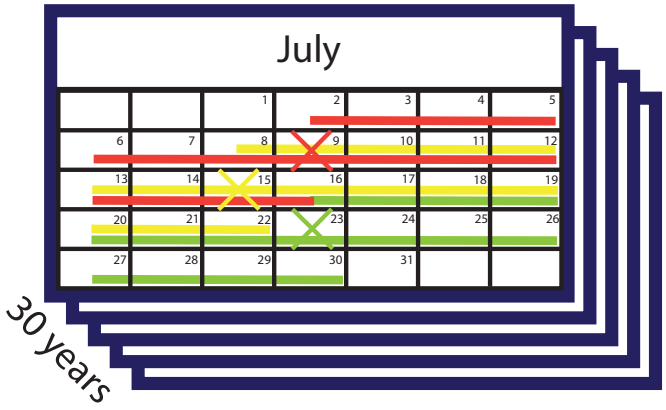


Figure 4.8: The time series of daily maximum and minimum temperature percentiles as observed by the KDTW airport observing station during the observational period. Tmax provided in the top panel and Tmin in the bottom. In each panel both the 5-day running average and the un-averaged time series is provided.

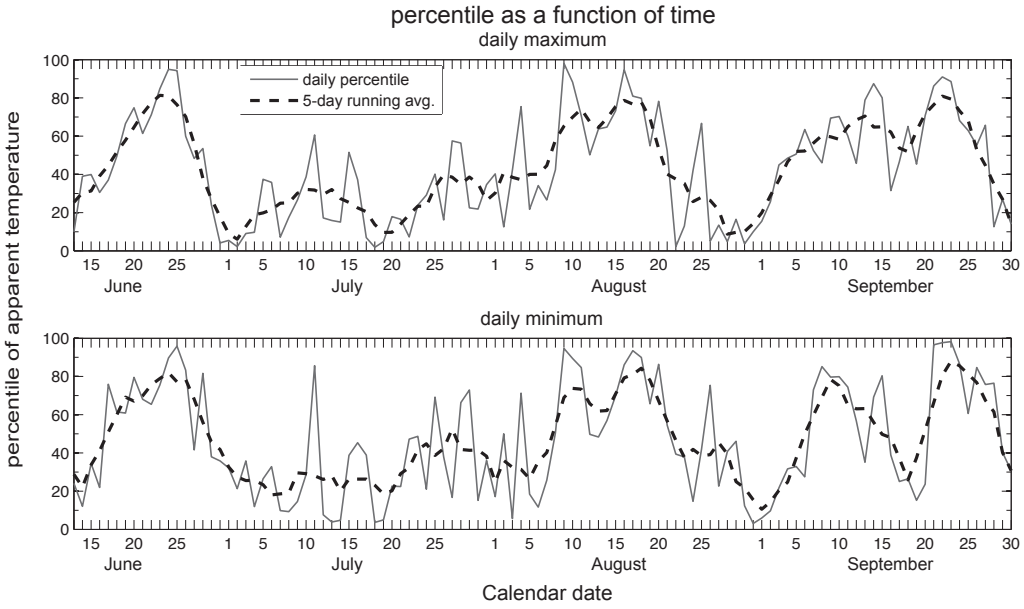


Figure 4.9: The histograms of the observed Metropolitan temperature range over the 110-day experiment. Both Tmin (a) and Tmax (b) histograms are provided.

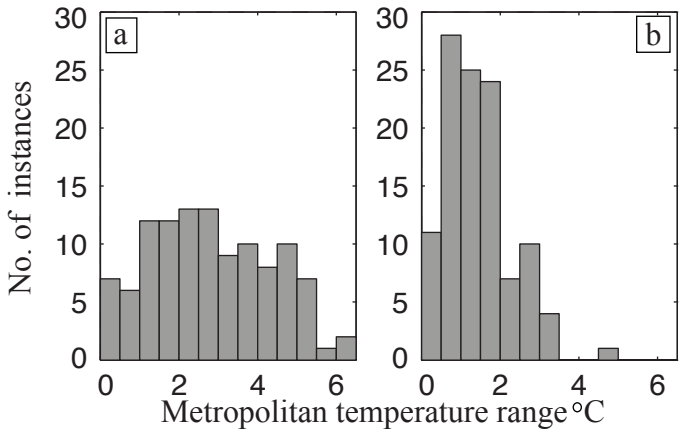


Figure 4.10: The 110-day average anomaly values at each station. Both Tmin (left panel) and Tmax (right panel) are provided. The six symbol groups for each figure are equal interval categories moving away from the zero value.

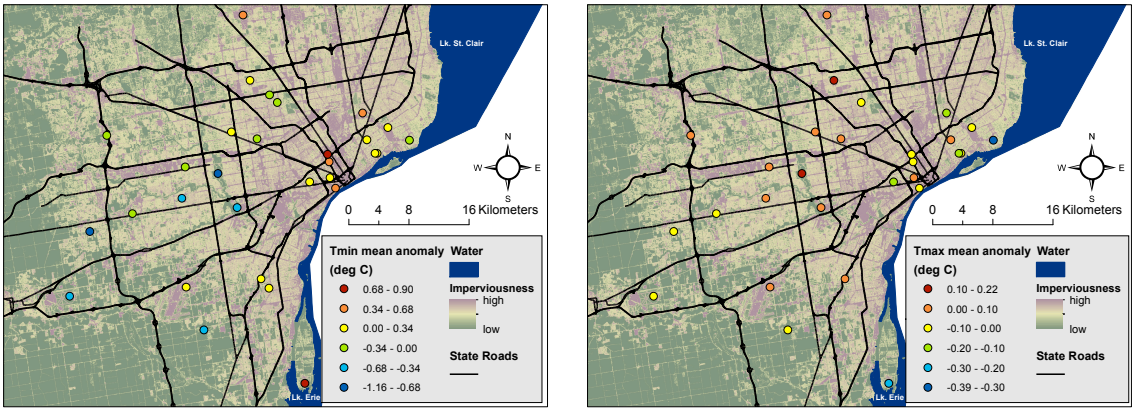


Figure 4.11: The mean and standard deviations of the anomalies over the 110-day experiment. Each station's anomaly is ordered by lowest to highest mean. Both Tmin (top panel) and Tmax (bottom panel) are provided.

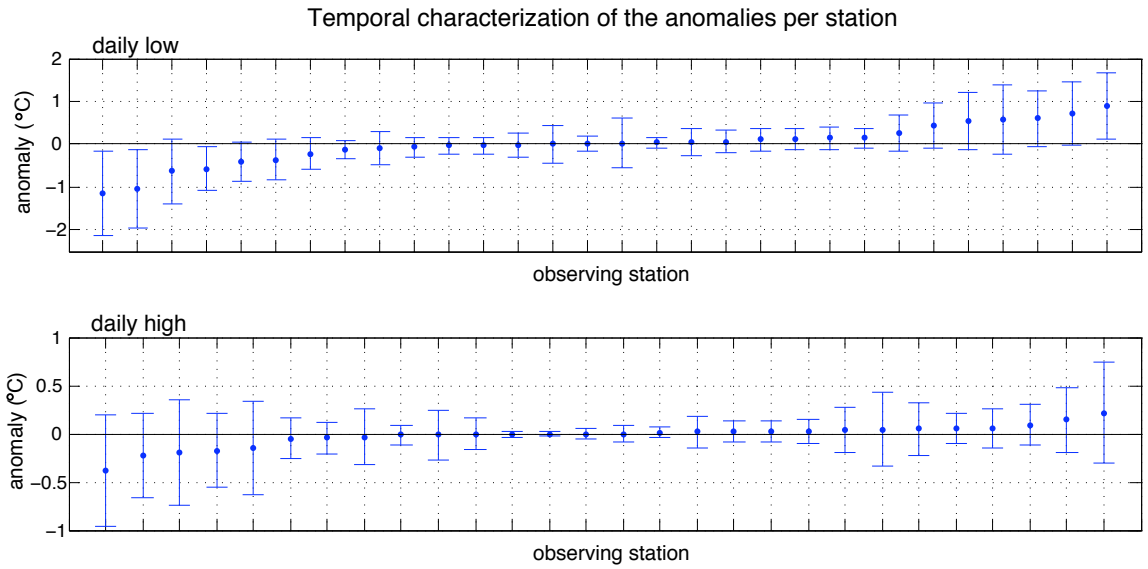


Figure 4.12: The metropolitan temperature range as a function of temperature percentile observed over the 110-day period. Both Tmax (red) and Tmin (blue) are shown, as well as the ordinary least squares (OLS) linear fit to both sets. The Y-axis has been normalized so that 1 (0) is the maximum (minimum) metropolitan temperature range.

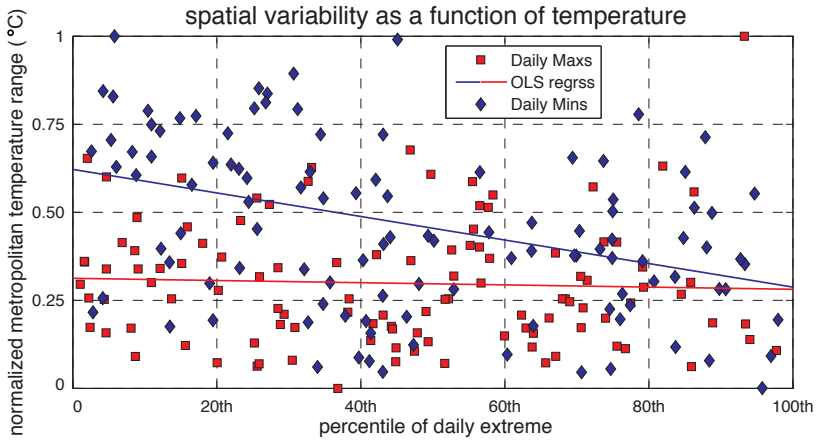


Figure 4.13: Scatterplots of the metropolitan temperature range as a function of city-average weather observations. Morning (afternoon) wind speeds and cloud cover related to Tmin (Tmax) temperature ranges.

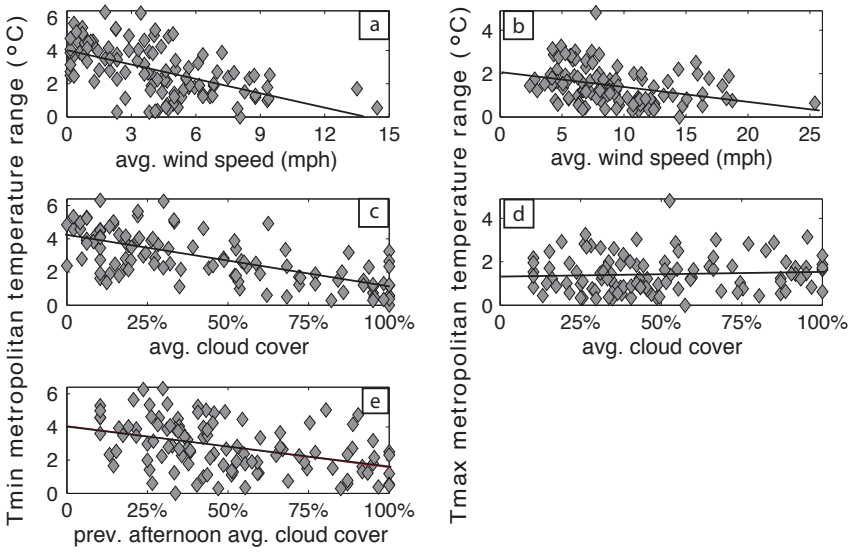


Figure 4.14. Correlation coefficients between 110-day mean anomalies and PIS values, as a function of radius of PIS. Both Tmin (a) and Tmax (b) are provided. Both Spearman rank and Pearsons correlation coefficients provided.

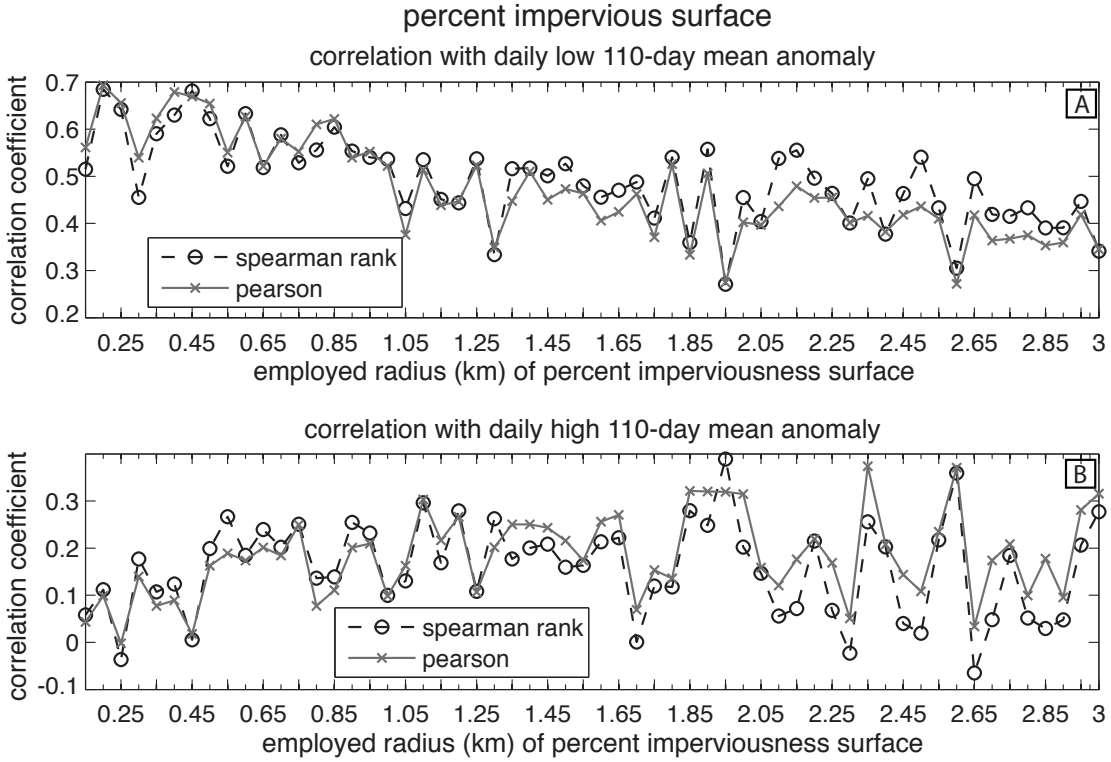


Figure 4.15: Spatially continuous model output from the statistical model built to predict the 110DM anomaly values at each observing station. The 110DM anomaly observations are also displayed (circles) and are categorized by half-standard deviations away from the zero value.

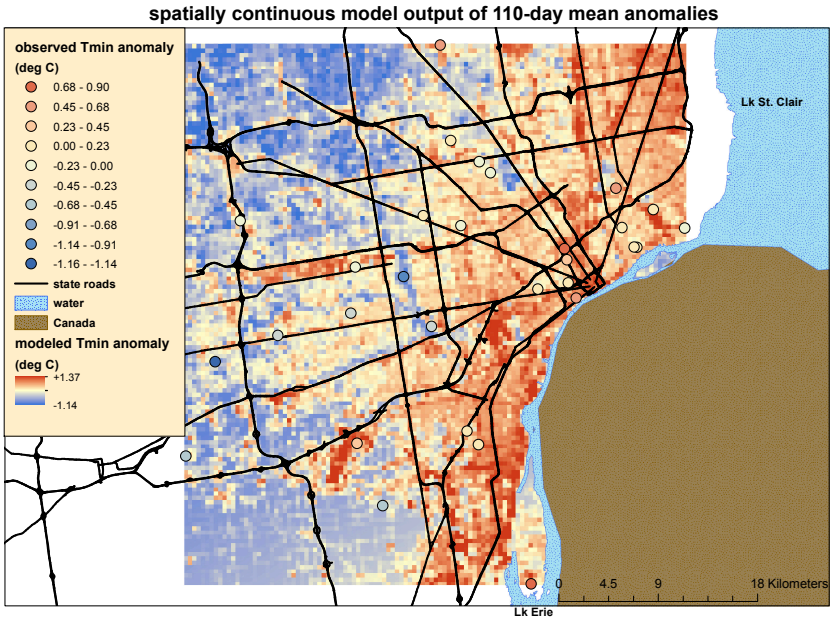


Figure 4.16: Continuous model output from the statistical model built to predict the 110DM anomaly values at each observing station using bootstraps to estimate the model. The 110DM anomaly observations are also displayed (circles) and are categorized by half-standard deviations away from the zero value.

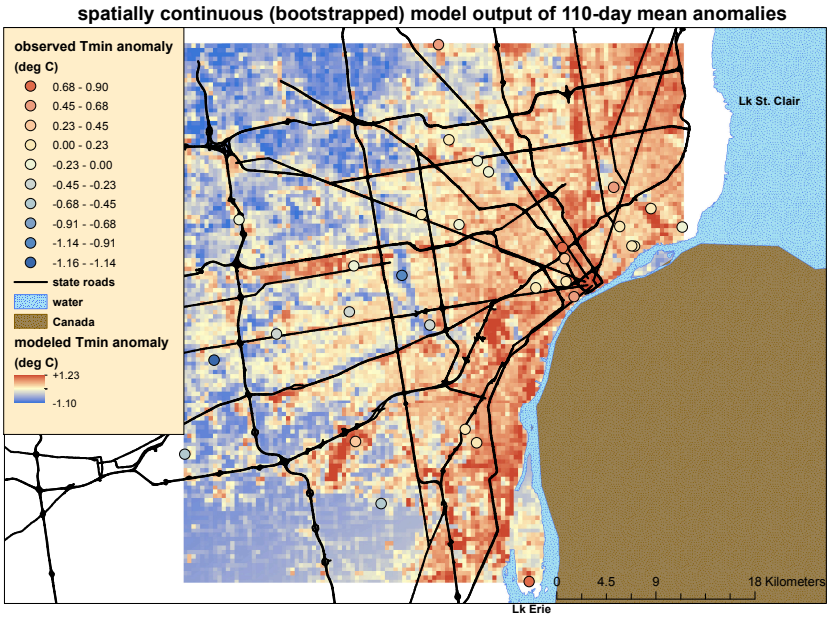


Figure 4.17: The mean and standard deviation of anomalies over the 55-day subselected sample. Each station is listed and ordered from lowest to highest mean anomaly. Both Tmin (top panel) and Tmax (bottom panel) are provided.

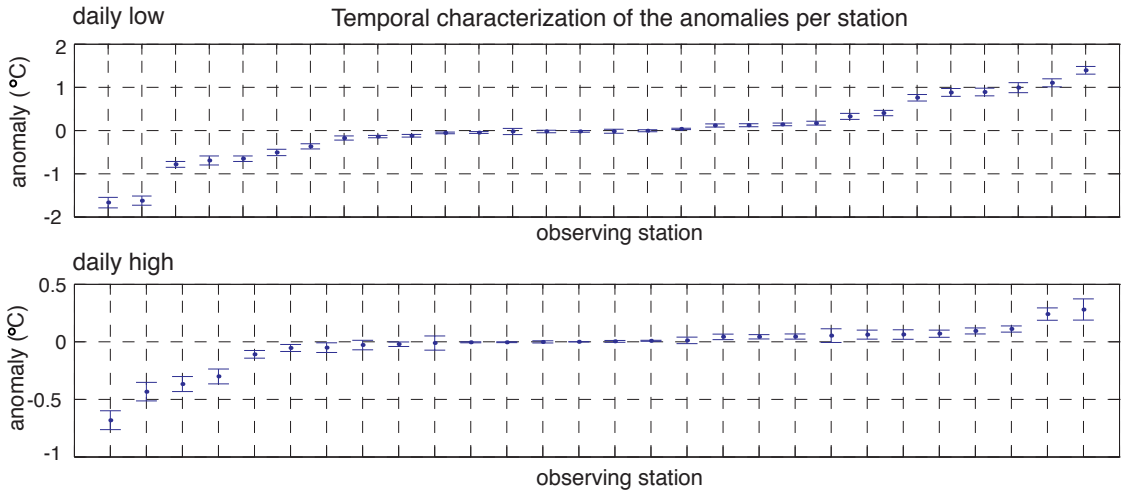


Figure 4.18: The mapped 55-day subselected sample average anomaly values. Both Tmin anomalies (left panel) and Tmax anomalies (right panel) are provided for each station. The six symbol groups (for each map) are equal interval categories moving away from the zero value.

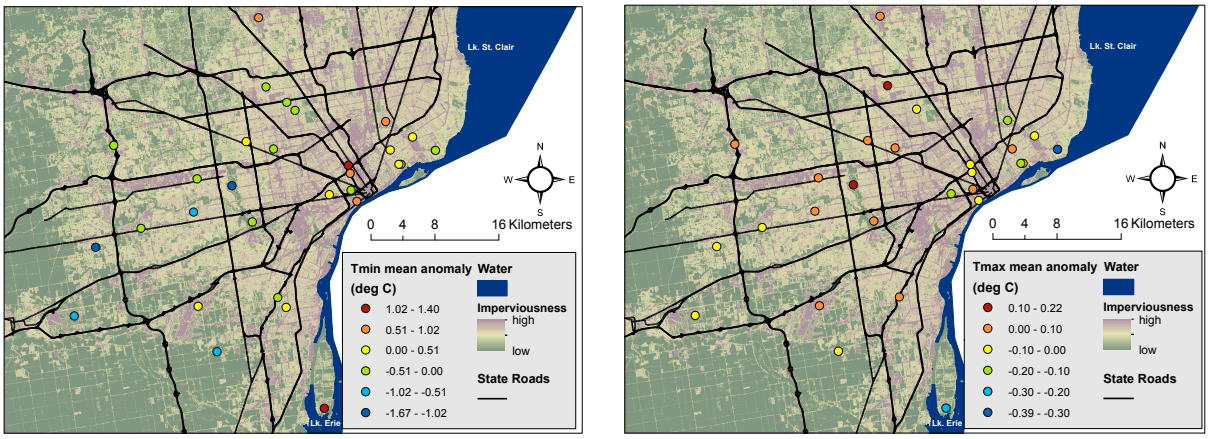
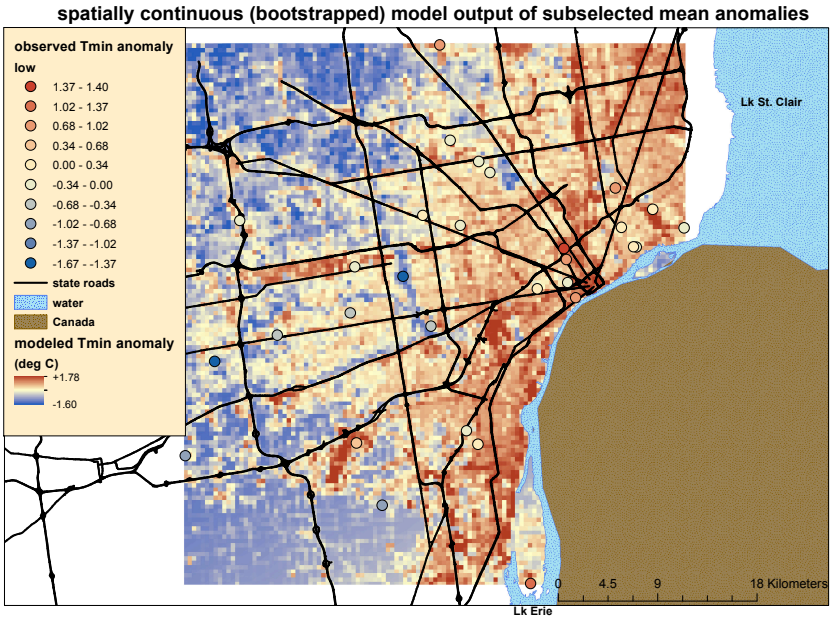


Figure 4.19: Continuous model output from the statistical model built to predict the subselected mean anomalvalues at each observing station using bootstraps to estimate the model. The subselected mean anomaly observations are also displayed (circles) and are categorized by half-standard deviations.



4.9. Tables

Table 4.1. Determined biases and uncertainties between networks. All biases (mean difference) and uncertainties (mean absolute difference after bias correction) values are in degrees Celsius. HOBO_t denotes the HOBO temperature extreme observations in degrees Celsius, MDEQ_t denotes the MDEQ temperature extreme observations in degrees Celsius, avg_WS denotes MDEQ stations mean observed wind speed in meters per second between 0300-0800 EDT or 1400-1800 EDT and CC denotes the mean airport observed cloud-cover percentage between 0400-0800 EDT or 1300-1800 EDT.

	Network	Difference	Bias	Uncertainity
Tmin	Airport	instrument	n/a	0.50
	HOBO	inherent/relative	n/a	0.44
		microclimate	n/a	0.00
		siting	0.14	0.00
		sampling algorithm	0.16	0.11
	MDEQ	co-location	*	0.43
		sampling algorithm	0.16	0.11
Tmax	Airport	instrument	n/a	0.50
	HOBO	inherent/relative	n/a	0.44
		microclimate	n/a	0.32
		siting	0.49	0.00
		sampling algorithm	-0.48	0.28
	MDEQ	co-location	**	0.52
		sampling algorithm	-0.48	0.28
*HOBO_t ≈ -4.2+1.1*MDEQ_t+0.13*avg_WS-2.0*(1-CC%)				
**HOBO_t ≈ 5.8 + 0.97 × MDEQ_t - 0.16 × avg_WS + 1.3 × (1-CC%)				

Table 4.2. Characterization of average MTR as a function of SSC2 airmass types. Percent refers to the portion of the field experiment the airmass was observed.

Air Mass Type	Days	Percent	Tmin MTR	Tmax MTR
Dry Polar	9	8.2%	3.59°C	1.42°C
Dry Mod.	41	37.3%	3.48°C	1.47°C
Dry Trop.	0	0.0%	n/a	n/a
Moist Polar	5	4.5%	1.23°C	1.41°C
Moist Mod.	21	19.1%	1.96°C	1.70°C
Moist Trop.	22	20.0%	2.34°C	1.31°C
Moist Trop.+/++	3	2.7%	2.01°C	0.89°C

Table 4.3. Regression coefficients between the observed normalized MTR and temperature percentiles. xMax Y value refers to the normalized MTR value corresponding to the far extreme (100th percentile) of percentiles

	Y-intercept	Percentile coeff.	xMax Y value
Tmin	61.1 (52.9, 69.2)	-0.34 (-0.49, -0.20)	26.74
Tmax	31.2 (24.7, 37.8)	-0.03 (-0.16, 0.09)	28.13

Table 4.4. Regression coefficients of multiple linear regression model predicting the Tmin MTR from three observered weather variables. The top row depicts results of an equation built on the observations; the bottom row depicts the results of an equation built using bootstrapping to better estimate the true coefficients. The median coefficients are listed while the corresponding standard deviations are inside parentheses. CC % denotes cloud cover percentage.

	Y-intercept	Windspeed (mph)	Overnight CC %	Prev. afternoon CC %
110-day	4.883	-0.136	-2.157	-1.082
10,000 bootstrap	4.890 (0.217)	-0.137 (0.037)	-2.157 (0.390)	-1.086 (0.397)

Table 4.5. Regression coefficients of multiple linear regression model predicting the Tmin MTR from three weather variables derived from reanalysis data. The top row depicts results of an equation built on the observations; the bottom row depicts the results of an equation built using bootstrapping to better estimate the true coefficients. The median coefficients are listed while the corresponding standard deviations are inside parentheses. CC % denotes cloud cover percentage and SW denotes “short wave”.

	Y-intercept	Windspeed (mph)	Overnight CC %	Prev. afternoon downwelling SW radiation (W/m^2)*1000
110-day	2.966	-0.059	-2.177	2.875
10,000 bootstrap	2.961 (0.584)	-0.059 (0.032)	-2.176 (0.378)	2.884 (0.964)

Table 4.6. Regression coefficients of multiple linear regression between the 110-day mean anomalies and three spatial variables. The top row depicts results of an equation built on the observations; the bottom row depicts the results of an equation built using bootstrapping to better estimate the true coefficients. The median coefficients are listed while the corresponding standard deviations are inside parentheses. 200m PIS denotes the percent impervious surface predictor, D2H2O denotes the distance-to-water predictor and D2CC denotes the distance-to-city center predictor.

	Y-intercept	200m PIS	D2H2O (km)	D2CC (km)
30-station	-0.914	0.018	-0.031	0.026
10,000 bootstrap	-0.866 (0.263)	0.017 (0.004)	-0.027 (0.015)	0.021 (0.016)

Table 4.7. Regression coefficients of multiple linear regression between the 55-day subslected sample mean anomalies and three spatial variables. The top row depicts results of an equation built on the observations; the bottom row depicts the results of an equation built using bootstrapping to better estimate the true coefficients. The median coefficients are listed while the corresponding standard deviations are inside parentheses. 200m PIS denotes the percent impervious surface predictor, D2H2O denotes the distance-to-water predictor and D2CC denotes the distance-to-city center predictor.

	Y-intercept	200m PIS	D2H2O (km)	D2CC (km)
30-station	-1.355	0.026	-0.045	0.038
10,000 bootstrap	-1.280 (0.438)	0.025 (0.006)	-0.038 (0.023)	0.030 (0.024)

CHAPTER 5. THESIS CONCLUSIONS

This work used observational data to examine extreme temperature trends and how those trends were portrayed in new high-resolution datasets. This work also investigated the variability of air temperatures within an urban region. These investigations led to learning in regards to how temperature extremes have been changing and how that change has related to changes in other parts of the climate system. As well as learning in regards to how to characterize the uncertainty present in the gridded datasets and how they might be corrected, and in regards to what drives spatial variability in temperatures across urban regions and why some locations were hotter or colder. This learning developed take away messages such as extreme heat being sensitive to many factors including region, the daily extreme they are based on and the timing within the summer. Also messages such as there are some applications for which the gridded datasets are not appropriate for, but applications that focus on small-scale variability will have larger uncertainty associated with them. Lastly, the spatial variability in urban regions occurs during hot weather and is relatively well linked to weather conditions and land attributes. This chapter will reflect on what was learned and how what questions could be explored next.

5.1. Regarding the EHE trends analysis

Described below was how this EHE trend analysis contributed to the overall understanding of the topic of EHEs. Trends in EHEs were unexpectedly sensitive to which daily temperature extreme met the requirements of the EHE definition. This implied that EHEs based on different daily temperature extremes were quite different, which was party a reflection of the daily maximum and minimum temperatures having changed differently and partly due to differences between summer mean temperature trends and EHE trends. Future studies of EHEs should appreciate and further study physical mechanisms that work oppositely on the different daily temperature extremes (e.g. humidity, certain kinds of LCLU change, urbanization). A study looking closer at the relationship between the temporal coupling of elevated

temperature, dubbed the "Tmax-Tmin coupling" (Section 2.5; Gershunov et al. 2009), would also improve the understanding of EHEs. Essentially, do extreme daily maximum temperatures beget extreme daily minimum temperatures in the following morning more so than extreme daily minimum temperatures lead to extreme daily maximum temperatures in the ensuing afternoon (i.e. which side of the Tmax-Tmin coupling is stronger)? How does this mechanism compare to mechanisms that work oppositely on both daily temperature extremes? This conclusion suggests EHE trend analyses should separately quantify EHEs with daily minimum and maximum temperatures above their chosen threshold(s). Moreover they should also quantify EHEs with both daily temperature extremes over the threshold(s). As is, these conclusions should aid the informed use of past and future EHE trend analyses as guidance for heat-health scientists. Specifically that the trends of EHEs based on different daily temperature extremes are not interchangeable.

Trends in EHEs during the earlier portion of the summer can differ from those of the later portion of the summer. This implied that the tendencies of the atmosphere over the CONUS in the early and later parts of the summer could change independently of one another. A future study better quantifying the past changes in EHEs as a function of time within the summer season might be very beneficial to the field. Moreover if a reanalysis dataset was used, the changes in the physical system (e.g. North American Jet Stream positioning, speed of progression of upper-level waves) could be easily attributed to the changes (in EHEs). More simply, moving forward it might be appropriate for EHE trend analyses to explicitly quantify early summer trends in EHEs, along with full summer trends. Early season EHEs are particularly important to both human-health (Kalkstein 1990) and physical systems like agriculture. As for heat-health scientists these conclusions add to the informed use of existing and future EHE trend analyses. Specifically it adds uncertainty to what to expect in the early part of the summer, based on results likely describing the summer as a whole.

There exist differences between the trends in summer mean daily temperature extremes and the corresponding EHEs. Those differences existed primarily at the regional level, but also at the continental level the summer average temperatures

seemed to have differences (e.g. more positive bias trends in the 1930-2010 period, overall stronger trends). This implies that the different portions of the statistical distribution (e.g. very cold, cool, normal, warm, very hot) were changing differently. Shen et al. (2011) has already shown the various statistical moments (i.e. mean, variability, skew, kurtosis) have changed differently in the CONUS single-day temperature observations. Subsequent studies in this field should explicitly examine what roles the trends in the different statistical moments (e.g. mean, variability, etc.) play in the trends in the EHEs. Do those relationships vary regionally, and are they different for different EHE characteristics (e.g. duration, frequency, intensity) or EHE types? What do climate forecasts of the different statistical moments then indicate about future EHEs? Also these conclusions should encourage climatologists to retire the concept that the probability distribution function is simply shifting towards warmer temperatures, and begin thinking about the different physical mechanisms that correspond to the different statistical moments (e.g. upper-level Rossby waves and synoptic variability in temperatures at the surface, GHG concentrations and the mean temperature, etc.). As for heat-health scientists these conclusions add to the informed use of existing and future EHE trend analyses. Specifically it adds uncertainty as in what to expect of EHEs from trend analyses of the summer mean temperatures, which are much more common.

5.2.Regarding the evaluation of gridded climate datasets

Described below was how this evaluation of gridded climate datasets contributed to the overall confidence of this type of climate information product, and pointed to future work in this area. On large scales the datasets had statistically significant differences in both trends and temporal averages. This was contrary to prior beliefs, and implies that the effects of the non-climatic biases do not average out over time and space. To put it differently, while the uncertainty this brings to the utility of the datasets depends on the application, it is hoped the existence of uncertainty across scales will encourage correction of the causes. Our results suggest the corrections should focus on the homogenization of the underlying data network. On the short term, climatologists must recognize the implications on

downscaling products, as the bulk of the uncertainty in the gridded datasets manifests itself at same scale (small scale). Specifically, downscaling methods use gridded datasets to spatially refine the larger-scale model output, but this evaluation demonstrated differences in temperature over the life of the datasets (i.e. the temporal average) that operate heavily at the small scale. Heat-health scientists should use (for decision making guidance) the current versions of these datasets with some caution, particularly at the small scale and in trends.

Non-climatic biases that occurred at the stations were not as strongly related to the differences between the differences the datasets had (with the reference dataset) at those locations, as expected. This was because the errors within the underlying network (due to lack of homogenization) were propagated spatially away from those stations and merged throughout the dataset by the interpolation processes used to create the grids. This means each grid point is plagued by more individual discontinuities than a typical station within the underlying network. The next study in this line of evaluation should be designed to explicitly test the different interpolation methods in these datasets for their impact on the propagation of these errors. Which interpolation method handles the non-climatic biases the best (e.g. is weighting the stations based on quality possible)? Until better versions of these datasets are made available, these conclusions suggest that climatologists could opt to use ungridded station data (e.g. NWS Co-Op stations or USHCN station) instead of the gridded data. Scientists who are not comfortable acquiring their own station data can still use the datasets but should be aware that the data has frequent discontinuities related to non-climatic biases and this is true at every gridpoint.

5.3.Regarding the temperature variability across an urban region

Explained below is how this evaluation of an urban region contributed tools available for studying and to the overall knowledge in urban climate. First, it was concluded that reanalysis data was able to force a statistical model of the spatial variability across the region nearly as well as in-situ observations. This insinuates not only are city-wide meteorological variables linked to the spatial variability, but

their representation within model-observation blended products can be used. This is important because there might be variables linked to the spatial variability that are hard to observe via observing stations such as cloud height or lower level atmospheric moisture. Future work in this area should examine the relationships between a large suite of meteorological variables and the amount of spatial variability in anticipation of finding a set of variables with excellent diagnostic capabilities. Does that set of variables change with region or time of summer? Then an analysis should be done of the ability to forecast (into the future) the spatial variability using the knowledge of which variables are optimal, and which forecast models are the most capable of doing so. For now, climatologists can infer how the tendencies of spatial variability will change in the future from those relationships (between the city-wide weather variables and the amount of spatial variability) and the changes in those variables as predicted by climate forecast models.

The previous afternoon's radiative characteristics were a useful predictor of the following overnight spatial variability in temperatures. This suggests that physical mechanisms linked to afternoon downwelling shortwave radiation (albedo, heat storage ability within and on the surface, moisture availability, etc.) are influential in overnight temperatures. Future observational studies ought to explicitly examine the roles these three land attributes play in the amount of spatial variability overnight. This can be done specifically by observing at locations spanning those variables, building statistical models predicting the spatial variability amount and examining their performance as a function of previous afternoon cloud cover. Urban planners and climatologists can use knowledge from this conclusion immediately by discouraging the creation of neighborhoods with large heat storage (e.g. lots of cement, wood), low albedo (e.g. dark surfaces, tall buildings with narrow streets) and low moisture availability (e.g. no vegetation, too much cement coverage).

This study represents a framework to observationally study the urban climate. This unique observing network was an example of an urban measurement network called for in the Grimmond et al. (2010) assessment of climate information for city planners. The statistical models predicting the amount of variability and locations of warm and cool temperatures were easy to make and performed reasonably. Moving

forward with this framework, it is felt it could potentially be used to realize the potential the nation-wide “weather underground” (Weather Underground, Ann Arbor MI) observing network has in studying urban climate. More specifically the Weather Underground network uses several different types of weather stations and thus currently is of use on a qualitative, not quantitative, level due to the uncertainty introduced by observing with different weather station types. So if the biases and uncertainties can be quantified, similar to in our observing network (or in a more sophisticated manner), the Weather Underground network could be integrated so that it would be more appropriate assessing spatial variability. It then could be used to build the statistical models. There exists available large-scale weather and land attribute information that can be quantified and used to build the relationships. More short term, this framework can be used to study other cities as well for both its use as a tool for learning the physical environment and informing a heat-health vulnerability analysis (i.e. when and where is it hottest).

The occurrence of the spatial variability was specifically confirmed during hot weather. This speaks to the relevancy of the spatial variability to periods of hot weather and suggests a lack of dependence on air temperatures. Future studies exploring the relationships between hot weather and spatial variability in temperatures (across urban regions) should use more metrics of hot weather (e.g. heat index, other biometeorological indices) and focus on regions outside of the Midwest (e.g. dry hot climates, marine dominated climates). This knowledge is also immediately usable to scientists as it suggests a need for sensitivity to spatial variability in temperatures during hot weather as well as it justifies attempts at mapping the variability of spatial for use within human health vulnerability analyses (Wilhelmi and Hayden 2010).

5.4. References

Grimmond, S., 2007: Urbanization and global environmental change: local effects of urban warming. *The Geographic Journal*, **173**: 83-88, doi: 10.1111/j.1475.2007.232_3.x.

Kalkstein, L. S.,1990: Climate change and public health: what do we know and where are we going? *Environmental Impact Assessments and Reviews*, **10**: 383-392.

Shen, S. S. P., A. Gurung, H. Oh, T. Shu, D. R. Easterling, 2011: The twentieth century contiguous US temperature changes indicated by daily data and higher statistical moments. *Climate Change*, **109**: 287-317. doi: 10.1007/s10584-011-0033-9

Wilhelmi, O. V. and M. H. Hayden, 2010: Connecting people and place: a new framework for reducing urban vulnerability to extreme heat. *Environmental research letters*, **5**: 014021. doi: 10.1088/1748-9326/5/1/014021