Spatial Influences of Heat Exposures and Social Vulnerability on the Temperature-Mortality Relationship: A Case Study in the Greater Vancouver Area

by

Hung Chak Ho

M.S., Mississippi State University, 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Geography
Faculty of Environment

© Hung Chak Ho 2016

SIMON FRASER UNIVERSITY

Spring 2016

All rights reserved.
However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name:	**Hung Chak Ho**
Degree:	**Doctor of Philosophy**
Title:	***Spatial Influences of Heat Exposures and Social Vulnerability to the Temperature-Mortality Relationship: A Case Study in the Greater Vancouver Area***
Examining Committee:	**Chair: Meg Holden,** Associate Professor

Nadine Schuurman
Senior Supervisor
Professor

Anders Knudby
Co-Supervisor
Assistant Professor
Department of Geography, Environment and Geomatics, University of Ottawa

Sarah B. Henderson
Supervisor
Assistant Professor
School of Population and Public Health, University of British Columbia

Kirsten Zickfeld
Supervisor
Associate Professor

Margaret Schmidt
Internal Examiner
Associate Professor

Yelena Ogneva-Himmelberger
External Examiner
Associate Professor
Department of International Development, Community, and Environment, Clark University/

Date Defended/Approved: **April 4, 2016**

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Climate change has increased the frequency and intensity of extremely hot weather. The health risks are not uniform across affected areas due to variability in heat exposure and social vulnerability, but these differences are challenging to map with precision. Therefore, this PhD research evaluated the uses of linear and nonlinear statistical models to map the intra-urban difference of temperature distribution, compared the temperature distributions mapped from land surface temperature, air temperature, and apparent temperature (Humidex), and developed a spatially- and temporally-stratified case-crossover approach for delineation of areas with higher and lower risk of mortality on extremely hot days, and applied this framework in greater Vancouver, Canada. A digital elevation model, Landsat 5 TM, Landsat 7, and MODIS water vapor products were used to map the air temperature and Humidex. Records of all deaths with an extremely hot day as a case day or a control day were extracted from an administrative vital statistics database spanning the years of 1998-2014. Three heat exposure and eleven social vulnerability variables were assigned at the residential location of each decedent. Conditional logistic regression was used to estimates the odds ratio for a 1°C increase in daily mean temperature from a fixed site for subsets of the data above and below different values of the spatial variables. An important result was that intra-urban variability in air temperature and Humidex could be mapped using a Random Forest model with good accuracy (RMSE = 2.31°C and 2.04°C respectively). This project also found that Humidex could better demonstrate the spatial distributions of heat exposure when it was compared with the temperature maps from land surface temperature and air temperature. Also, the heat exposure and social vulnerability variables with the strongest spatially-stratified results were Humidex and labour nonparticipation rate. Areas at higher risk had values ≥3.7°C warmer than Vancouver International Airport, the reference site, and ≥60% of the population neither employed nor looking for work. These variables were combined in a composite index to quantify their interaction and enhance visualization of high-risk areas. In conclusion, methods from this PhD research provided a data-driven framework for spatial delineation of the temperature-mortality relationship by heat exposure and social vulnerability. Results can be used to map and target the most vulnerable areas for public health intervention, and the methodology is directly transferable to other cities in Canada and abroad.

Keywords: Extreme Heat; Urban Climate; Remote Sensing; Temperature Mortality; Environmental Health

Dedication

Many thanks to the support from my parents and family so I can be here. Many thanks for all my friends and colleagues to support me to survive in this city from the beginning to nowadays.

Acknowledgements

I acknowledge the Pacific Institute for Climate Solutions and Simon Fraser University for partial funding of this project. I hereby thank Dr. Anders Knudby, Dr. Sarah Henderson, Dr. Kirsten Zickfeld, and Dr. Nadine Schuurman for their supervision and their support to this PhD research. I also thank Dr. Yongming Xu, Mehdi Aminipouri, Paul Sirovyak, Matus Hodul, Dr. Tom Kosatsky, Angela Yao, Blake Walker, Kaylee Girard, Spencer Behn, and Melvin Pan for their support to this project in order to lead this research to succeed.

Table of Contents

Approval ii
Ethics Statement iii
Abstract iv
Dedication vi
Acknowledgements vii
Table of Contents viii
List of Tables x
List of Figures xi
List of Abbreviations xii

Chapter 1. Introduction, objectives, and study area 1

Chapter 2. Evaluating the uses of linear and nonlinear statistical models to map intra-urban difference of temperature distributions 7
2.1 Data and methods 7
2.1.1 Satellite data 7
2.1.2 Satellite-derived predictors 8
2.1.3 Calibration/validation data 10
2.1.4 Regression models 11
2.1.5 Accuracy assessment 12
2.1.6 Variable importance analysis 13
2.2 Results and discussion 13
2.2.1 Model performance 13
2.2.2 Spatial distribution of relative maximum air temperature 15
2.2.3 Variable importance 17
2.2.4 Temperature pattern variability 19
2.2.5 Weather underground data 21
2.2.6 Complementary data source 22

Chapter 3. Comparison of the intra-urban difference of temperature distributions mapped with different temperature metrics 23
3.1 Data and methods 23
3.1.1 Humidex field observations 23
3.1.2 Spatial data layers 24
3.1.3 Regression modelling 25
3.1.4 Accuracy Assessment 26
3.1.5 Variable importance analysis 26
3.1.6 Production of temperature maps 26
3.2 Results and discussion 27
3.2.1 Predictions of (Humidex) 27
3.2.2 Variable importance (Humidex) 28
3.2.3 Spatial distribution of the three temperature measures 29

3.2.4 Humidex patterns and variability 32
3.2.5 Comparison and validation of temperature distributions 34

Chapter 4. Delineation of the spatial variability of the temperature-mortality relationship 36
4.1 Data and methods 36
4.1.1 Mortality data 36
4.1.2 Time-varying temperature data 37
4.1.3 Spatial heat exposure data 37
4.1.4 Spatial social vulnerability data 37
4.1.5 Spatial delineation of the temperature-mortality relationship using case-crossover analyses 39
4.2 Results and discussion 40
4.2.1 Data summary 40
4.2.2 Spatial variablitiy in the temperature-mortality relationship by heat exposure 42
4.2.3 Spatial variablitiy in the temperature-mortality relationship by social vulnerability 44
4.2.4 Sensitivities analyses 44
4.2.5 Combined heat exposure and social vulnerability index 45
4.2.6 Discussion and limitations 47

Chapter 5. Conclusion 51

References 53

List of Tables

Table 1 Landsat images used for this study 8

Table 2 Maximum air temperature (Tmax) error estimates based on cross-validation, for each model type. 13

Table 3 Results from variable importance analyses (air temperature) 18

Table 4 Results from the variable importance analysis (Humidex) 29

Table 5 Summary statistics for the mortality analytic dataset 40

List of Figures

Figure 1 Map of Study Area 6
Figure 2 Scatter Plot of Validation Results from Air Temperature Mapping 15
Figure 3 Spatial Distribution of Air Temperature (Random Forest Model) 17
Figure 4 Boxplot of Air Temperature Variation of Each Weather Station 20
Figure 5 Scatter Plot of Validation Results from Humidex Mapping 28
Figure 6 Spatial Distributions of Temperature from Three Metrics 30
Figure 7 Comparison of Air Temperature and Humidex Maps 32
Figure 8 Boxplot of Humidex Variation of Each Weather Station 33
Figure 9 Delineation of Temperature-Mortality Relationship 43
Figure 10 Heat Risk Map from Heat Exposure and Social Vulnerability 46

List of Abbreviations

AVHRR – Advanced very-high-resolution radiometer

BC – British Columbia

BCCDC – BC Centre for Disease Control

DEM – Digital Elevation Model

EC – Environment Canada

ICD-10 – 10th Revision of the International Classification of Diseases

k_1 – thermal band calibration constant

k_2 – thermal band calibration constant

$L^{\uparrow}$ – Upwelling Atmospheric Radiance

$L^{\downarrow}$ – Downwelling Atmospheric Radiance

L_{sen} – at-senor radiance

LST – Land Surface Temperature

MAE – Mean Absolute Error

ME – Mean Error

MOD05 – MODIS Precipitable Water product

MODIS – Moderate Resolution Imaging Spectroradiometer

NAALSED – NASA's Atmospheric Correction Parameter Calculator

NDVI – Normalized Difference Vegetation Index

NDWI – Normalized Difference Water Index

OR – Odds ratio

RMSE – Root Mean Squared Error

SD – Standard Deviation

SEVIRI – Spinning Enhanced Visible and Infrared Imager

SLC – Scan-Line Corrector

SVF – Skyview Factor

SVM – Support Vector Machine

Tmax – Maximum Air Temperature

UHI – Urban Heat Island

UTM – Universal Transverse Mercator

VANDIX – Vancouver Area Neighborhood Deprivation Index

VGI – volunteered geographic information

WU – Weather Underground

YVR – Vancouver International Airport

ρNIR – near-infrared band

ρMIR – mid-infrared band

τ – Atmospheric Transmittance

ε – emissivity values

%IncMSE – Percentage Increase in Mean Squared Error

Chapter 1.

Introduction

Climate change has increased the severity, intensity, and frequency of extreme hot weather events over the past century (Meehl and Tebaldi 2004; Reid et al. 2009). Many of these events have been associated with excess mortality in urban areas, as documented in the United States (Semenza et al. 1996; Kaiser et al. 2001; Curriero et al. 2002), Europe (Filleul et al. 2006; Schifano et al. 2009), China (Huang et al. 2010), and Russia (Trenberth and Fasullo; 2012). Less severe impacts have also been documented in the Canadian cities of Montreal (Smargiassi et al. 2009), Toronto (Pengelly et al. 2007), and Vancouver (Kosatsky et al. 2012), the latter being the setting for our study. Cities are at particular risk during extreme hot weather events due to the urban heat island (UHI) effect, which leads to intra-urban variability in heat exposure (Hart and Sailor 2009; Hawkins et al. 2004; Roth et al. 1989).

The temperature-mortality relationship for any given city is typically evaluated with time series models that quantify the exposure of the entire population using near-surface air temperatures measured at a single reference station, such as an airport (Basu 2009, Hajat et al. 2010). The most common indicators of heat exposure are air temperature and apparent temperature, where the latter combines measures of temperature and humidity (Steadman 1979; Basu 2009; Massetti et al. 2014). Either of these can be quantified as the daily minimum, mean, maximum, or as multi-day averages of these values. Air temperature and apparent temperature are highly correlated (Basu 2009; Zhang et al. 2014), but apparent temperature has been more closely associated with mortality (Almeida et al. 2010; Barnett et al. 2010; Zhang et al. 2014). Furthermore, apparent temperature is a better predictor of indoor temperatures than air temperature (Nguyen et al. 2014), which is important for industrialized countries where people spend >90% of their time indoors (Höppe and Martinac 1998).

While time-series studies using single reference station measurements are straightforward and methodologically consistent, they do not facilitate quantification of intra-urban spatial variability in the temperature-mortality relationship (Basu 2009). These spatial differences are likely driven by two overarching mechanisms: spatial variability in the exposure and spatial variability in the distribution of vulnerable populations. Previous efforts to evaluate the effect of spatial variability in exposure have typically characterized UHI impacts using satellite-derived estimates of land surface temperature (LST) (Smargiassi et al. 2009; Hondula et al. 2012; Laaidi et al. 2012). Other studies have demonstrated that living in a relatively hot part of a city is a significant risk factor during extreme hot weather events (Smargiassi et al. 2009; Laaidi et al. 2012), though the utility of LST for delineating these risks has been inconsistent (INVS 2004). One possible reason is that LST may not be the most appropriate proxy for human heat exposure, but it is often used because, unlike air temperature, it is directly measured by remote sensing instruments. Another possible reason is that many previous studies have used LST from the Moderate Resolution Imaging Spectroradiometer (MODIS), which has a coarse spatial resolution (1 km) (Laaidi et al. 2012; Tomlinson et al. 2011). Only a few studies have used more finely resolved data to describe the spatial patterns of heat exposure (Hondula 2012; Johnson et al. 2012; Johnson et al. 2014; Smargiassi et al. 2009; Uejio et al. 2011), and none have compared LST with temperature metrics that might better describe human health risks.

In addition, the use of LST to quantify heat exposure in these studies seems primarily motivated by the relative ease with which this variable can be mapped, and not by considerations concerning its conceptual appropriateness as a proxy for human heat exposure. LST is relatively straight-forward to map from publicly available satellite data (Coll et al. 2010; Weng et al. 2004) because the longwave thermal radiation measured by the satellite sensor is strongly determined by the skin temperature under normal atmospheric conditions. However, it can only be mapped at the time of satellite overpass. For the Landsat sensors typically used to obtain neighbourhood-level spatial resolution, this means that the resulting maps describe the distribution of skin temperature at approximately 10 am local time. SEVIRI, AVHRR, MODIS and other sensors less frequently used in spatial epidemiology can provide information multiple times per day, but do so at much reduced spatial resolution.

Potential alternative measures of heat exposure that are practical for public health research include air temperature or apparent temperature, a measure of the ambient heat perceived by humans that is typically calculated using both air temperature and humidity (Basu 2009; Massetti et al. 2014; Steadman 1979). Either of these can be quantified as the daily minimum, mean, maximum, or as multi-day average values of these quantities. Both air temperature and apparent temperature are more frequently used than skin temperature to quantify heat exposure in the non-spatial epidemiology literature, but they are also more difficult to map at the urban/neighbourhood scale. Mapping air temperature requires spatial interpolation between a dense network of weather stations (Hattis et al. 2012), statistical approaches based on the presence of very dense vegetation in the local area (Prihodko & Goward 1997; Zakšek & Schroedter-Homscheidt 2009) or regression modeling with local calibration data (Ho et al. 2014; Xu et al. 2014). Uneven distribution of weather stations in most urban areas, and especially the lack of observations in densely built-up areas where the intra-urban difference of urban temperature is strongest, are serious shortcomings for the interpolation approach, and the presence of very dense vegetation cannot be assumed in all urban areas. The regression approach, however, has been shown effective for mapping intra-urban variation in air temperature, and application to apparent temperature is straight-forward.

While energy fluxes ensure a relationship between the skin temperature and the temperature and humidity of the air mass above it, the relationship between skin, air, and apparent temperature is complex as each temperature measure is also influenced by other factors such as wind speed and direction, the three-dimensional structure of the surface and its effect on both shading and turbulence, the land surface heat capacity, and near-surface water available for evapotranspiration. It is therefore not evident that the spatial pattern of one (e.g. skin temperature) is a good proxy for another (e.g. air temperature), nor that mapping mid-morning skin temperature, as is typically done using Landsat data, is the best approach to quantify human heat exposure in different parts of a city. Apparent temperature is not only more closely related to heat mortality, but has also been shown to be a better predictor of indoor temperature than is air temperature (Nguyen et al. 2014). As a proxy for heat exposure this is especially important given that people in most industrialized countries spend >90% of their time indoors (Hoppe &

Martinac 1998). It is therefore likely that maps of apparent temperature would be a good source of information on heat exposure in heat health studies.

Attempts to evaluate spatial variability due to the distribution of vulnerable populations have typically focused on social risk factors, such as socioeconomic status (Buscail et al. 2012; Morabito et al. 2015; Reid et al. 2009; Tomlinson et al. 2011). For example, Reid et al. (2009) developed a qualitative index for mapping regional vulnerability to heat across the US based on county-level indicators such as race, poverty, education, and social isolation. In further work, this vulnerability index was enhanced by including heat exposures as measured by LST, allowing isolation of areas assumed to be at most risk during extreme hot weather (Buscail et al. 2012; Tomlinson et al. 2011). However, such indices apply universal measures of social and material deprivation that might not capture the place-specific nuances of risk within individual cities, such as the importance of access to transportation during the 1995 event in Chicago (Semenza et al. 1996; Yardley 2011). Furthermore, the performance of such vulnerability indices has not been rigorously evaluated with historical mortality data, which has been flagged as a significant limitation of the work (Tomlinson et al. 2011). Even with excellent maps of exposure and vulnerability, it is important to consider that areas indicated to be at higher risk may not, in reality, experience excess deaths during extreme hot weather (O'Neill et al. 2009; Tomlinson et al. 2011).

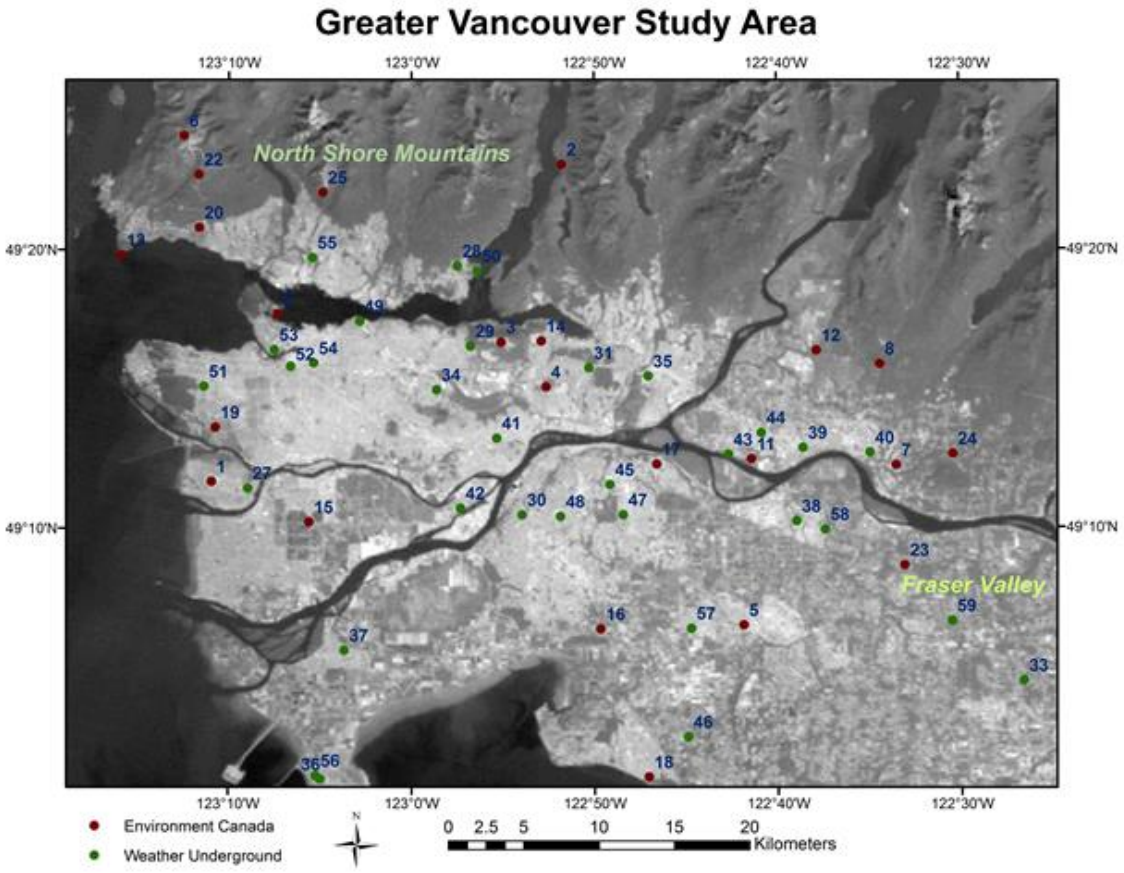
The only way to accurately delineate areas of higher and lower risk is with methods designed to assess spatial differences in the temperature-mortality relationships across a city (Greene et al. 2011; Hattis et al. 2012; Hondula 2012; Schaeffer et al. 2015; Smargiassi et al. 2009). Here I and my research group developed an approach that combines data from a single representative weather station with exposure and vulnerability maps in a temporally and spatially stratified case-crossover design. A case-crossover design is a biostatistical approach used to predict health risk, in which each case acts as its own control, i.e. a period in which death occurred is considered the 'case', and one or more other periods (in which death did not occur) act as the 'control(s)'. The temporal stratification uses control periods that are as comparable as possible to the case period, i.e. from the same year, same season, and same weekday, to minimize the influence of confounding factors. The spatial

stratification uses exposure and vulnerability data derived from maps to compare mortality risk across spatial units. This is an extension of previous work conducted by Smargiassi et al (2009) in Montreal, Canada. We applied these methods to the city of greater Vancouver, Canada where an extreme hot weather event during the summer of 2009 was associated with a 40% increase in mortality and more than 100 excess deaths over a 7-day period (Henderson and Kosatsky 2012; Kosatsky et al. 2012). The specific objectives of our study were to: 1) evaluate linear and non-linear statistical methods to map the intra-urban variability of air temperature during a typical hot summer day (Chapter 2); 2) compare the urban temperature distributions mapped using different temperature metrics (LST, air temperature, and apparent temperature) (Chapter 3); 3) assess spatial differences in the temperature-mortality relationship on extremely hot days using maps of LST, air temperature, and apparent temperature (Chapter 4); 4) assess spatial differences in the temperature-mortality relationship on extremely hot days using maps of social vulnerability indicators (Chapter 4); 5) evaluate whether areas at higher risk due to exposure and vulnerability have additive effects on the temperature-mortality relationship (Chapter 4); and 6) map higher risk area for the uses of emergency planning and public health protection during future extreme heat events (Chapter 4).

Therefore, we used greater Vancouver as a study site in our study. The metropolitan area of greater Vancouver (also known as Metro Vancouver) is on the west coast of British Columbia (BC), Canada (Figure 1), and has a population of more than 2 million people, according to Canadian Census at 2011. Vancouver is bordered by fold mountain ridges to the north, the Pacific Ocean to the west, and the semi-arid Fraser Valley to the east. This geographic context creates a complex microclimatic system that can lead to strong UHI effects on hot summer days (Runnalls and Oke 2000; Oke and Hay 1994; Oke 1976). During the summer, ocean breezes and winds from the mountain ridges can cool down the coastal regions, while the Fraser Valley can trap air masses and create a relatively hot zone (Oke and Hay, 1994). Temperature in the urban area is heavily influenced by cloud cover, while cooling from evapotranspiration is of little influence due to limited vegetation cover in the built-up areas. A hot summer day in greater Vancouver is typically cloudless with light winds. This can generate a strong UHI effect and substantially higher temperatures in the urban areas compared with their surroundings (Oke and Hay, 1994).

For summer days (July and August) from 1998-2014 the average high, mean, and low air temperatures measured at Vancouver International Airport (YVR) were 22°C, 18°C, and 14°C, respectively. However, during the unprecedented hot weather event in 2009, daytime high temperatures reached 34°C and were over 30°C for three consecutive days (Kosatsky et al. 2012). Although this is not hot by international standards, the 40% increase in mortality was a clear indication that greater Vancouver was adversely affected by ambient temperatures that were high relative to seasonal norms (Kosatsky et al. 2012).

Figure 1 – Map of Study Area. The red dots represent the weather stations from the Environmental Canada and the green dots represent the weather stations from the Weather Underground.



Chapter 2.

Evaluating the uses of linear and nonlinear statistical models to map intra-urban difference of temperature distributions

This chapter describes the development of a map of daily maximum air temperature for a typical hot summer day in the greater Vancouver area. Details have been published in the following peer-reviewed article: Hung C. Ho, Anders Knudby, Paul Sirovyak, Yongming Xu, Matus Hodul and Sarah B. Henderson. (2014). Mapping Maximum Urban Air Temperature on Hot Summer Days. Remote Sensing of Environment, 154, 38-45. My major contributions of this paper included: 1) to develop the idea of using support vector machine and linear regression model to compare with the random forest models, 2) to decide which variables should be used in the research, 3) data preparation, 4) data analysis, 5) to develop the idea of how to evaluate the map and weather stations by mean error, and 6) to write and revise the manuscript.

2.1. Data and methods

2.1.1. Satellite Data

The satellite data used in this study consist of all (n=6) cloud-free Landsat 5 TM and Landsat 7 ETM+ images available from 2001 to 2010 for hot summer days in the study area, here defined as days with daily maximum air temperature (Tmax) greater than 25°C at Vancouver International Airport (YVR) (Table 1). 2011 – 2014 data were not included because there were lack of cloud-free Landsat 7 ETM+ images at the day overpassing Vancouver with the maximum temperature higher than 25°C, and Landsat 8 images was launched at Feb 2013, after we started and completed this part of research. 25m Canadian Digital Elevation Data (CDED, http://www.geobase.ca) were used to

provide elevation information for the study area. Landsat 5 TM images and the DEM were resampled to 60m to match the spatial resolution of the ETM+ thermal band, and all data were projected to UTM zone 10N.

Table 1 – Landsat images used for this study. The images were obtained for cloud-free hot summer days in the study area.

Satellite Images	**Acquired Date**	**Daily Maximum Air Temperature at YVR**
Landsat 5 TM	13/08/2002	27.1
Landsat 5 TM	17/07/2004	27.1
Landsat 7 ETM+	02/08/2004	25.4
Landsat 5 TM	23/07/2006	26.7
Landsat 5 TM	12/07/2008	25.5
Landsat 7 ETM+	31/07/2009	28.7

2.1.2. Satellite-derived predictors

Several spatial data layers were derived from the Landsat and elevation data for use as predictors in regression models to map Tmax: LST, Normalized Difference Water Index (NDWI), elevation, skyview factor (SVF), and solar radiation. All layers except elevation were derived separately for each Landsat image.

LST was estimated from the thermal band in the Landsat images. NASA's Atmospheric Correction Parameter Calculator (NAALSED) from Barsi et al. (2003) was

used to estimate the upwelling atmospheric radiance ($L^{\uparrow}$), downwelling atmospheric radiance ($L^{\downarrow}$) and atmospheric transmittance τ, which were then used with the NAALSED emissivity values (ε) to obtain the blackbody radiance at temperature LST:

$$B(LST) = \frac{L_{sen} - L^{\uparrow}}{\varepsilon\tau} - \frac{1-\varepsilon}{\varepsilon} L^{\downarrow} \quad \text{(Coll et al, 2010)}$$

where B(LST) is the blackbody radiance at temperature LST and L_{sen} is the at-sensor radiance. An inversion of Planck's Law was then applied to derive the kinetic surface temperature using the following equation (Coll et al, 2010):

$$LST = \frac{k_2}{ln\left(\frac{k_1}{B(LST)} + 1\right)}$$

where k_1 and k_2 are the thermal band calibration constants found in the Landsat metadata files.

NDWI is an index designed to quantify vegetation water content (Gao, 1996), which strongly influences surface cooling through evapotranspiration. NDWI is defined as:

$$NDWI = (\rho_{NIR} - \rho_{MIR})/(\rho_{NIR} + \rho_{MIR})$$

where ρNIR and ρMIR are near-infrared band and mid-infrared band.

In this study, Landsat bands 4 and 5 were used to calculate NDWI for land areas. NDWI values are not meaningful over water and, as both ρNIR and ρMIR are very small over water, tend to be noisy. To allow the NDWI layer to function as a proxy for cooling by evapotranspiration, we found the maximum NDWI value on land and applied it to all water surfaces, replacing the results from the original NDWI calculation for water.

SVF can be defined as the portion of unobscured sky, which is related to the radiation received or emitted in an area (Chen et al., 2012; Su et al., 2008) and is influenced by both topography and building structure. SVF was mapped for each pixel using an empirically calibrated relationship with shadow proportion, which was derived

using partial spectral unmixing. Validation using independent lidar data for Vancouver shows the method to perform well (SVF Root Mean Square Error = 0.056).

Incoming solar radiation has also been shown to influence air temperature (Bristow & Campbell, 1984). Incoming solar radiation was estimated using an implementation of the direct solar radiation equation (Fu & Rich, 2002) in ArcGIS 10.1, which assumes that the total direct solar radiation is the sum of direct solar radiation from all sun map sectors. It is defined as:

$$Solar\ Radiation = S_{Const}\,\beta^{m(\theta)}\ \text{x}\ SunDur_{\theta,\alpha}\ \text{x}\ SunGap_{\theta,\alpha}\ \text{x}\ cos(AngIn_{\theta,\alpha})$$

where SConst = solar flux, β = atmospheric transmittance, m(θ) = optical path length, SunDurθ,α = time duration represented by sky sector, SunGapθ,α = gap fraction for the sun map sector, AngInθ,α = incident angle, θ = solar zenith angle.

Several additional factors known to correlate with local air temperature were not included in our model. Preliminary analyses showed that NDVI, which is designed to quantify photosynthetic activity, did not increase the predictive power of models that already included NDWI, which is designed to quantify vegetation water content and thus is more directly related to evapotranspiration and air temperature. A similar result was found by Benali et al. (2012). Land cover information was not directly incorporated in the model, but is indirectly quantified by LST and NDWI, which are both strongly influenced by land cover, and SVF, which is determined by three-dimensional surface structure and thus also influenced by land cover. Wind speed and direction, despite being known to strongly influence local air temperature, were not included because of their temporal variability as well as a lack of wind data with sufficient spatial resolution.

2.1.3. Calibration/validation data

In addition to the spatial data layers, Tmax observations from 59 weather stations in the study area were used for calibration and validation of the regression models for mapping air temperature (Figure 1). 26 of the weather stations are operated by Environment Canada (EC), and the other 33 stations are included in the Weather Underground (WU) network, a public participation weather project in which individuals

can deploy personal weather stations and have the resulting data uploaded to a publicly accessible geodatabase (Goodchild, 2007; Sieber, 2006). Not all weather stations were operational for all of the six days used in our study, including several of the WU stations for which data were only available for one or two of the latest dates.

2.1.4. Regression Models

Regression models, using predictors derived from satellite and elevation data, were first used to predict the absolute Tmax observed at weather stations for each of the six days for which Landsat data were available, and Tmax values were then normalized relative to the corresponding Tmax value at YVR. Ultimately the six maps of relative Tmax were averaged to describe Tmax, relative to YVR, for a typical hot summer day in the study area. Per-pixel predictions of absolute temperature were based on the pixel values of elevation, SVF and solar radiation, as well as spatially averaged values of LST and NDWI, calculated for circular areas with radii of 100, 200, 400, 600, 800, 1000m surrounding each pixel. The spatial averaging was used to account for the influence of temperature and moisture surface conditions in the areas surrounding each station (Liu and Moore, 1998; Zakšek, K., & Oštir, 2012), and the specific radii were selected to cover the range from the smallest meaningful radius, given the 60 m spatial resolution of the data, to the largest radius within which the predictors were expected to influence Tmax. Additional research on the spatial dependence of interaction between surface properties and air temperature could inform better selection of these radii in future studies.

Three types of statistical models were used to predict Tmax from the predictors and the sub-predictors above: 1) Ordinary least squares multiple linear regression, 2) Support Vector Machine (SVM), and 3) Random Forest. Ordinary least squares regression uses the observed data to estimate the parameters of a linear regression function (Kutner et al., 2004). SVM is a machine learning based regression model initially developed for classification problems. It determines the location of decision boundaries and uses the boundaries to predict optimal separation of classes (Mountrakis et al., 2011; Pal & Mather, 2005; Cortes & Vapnik, 1995). Previous remote sensing studies have primarily used this method for classification (Huang, 2002; Melgani, &

Bruzzone, 2004; Oommen et al., 2008; Pal & Mather, 2005; Ocak & Seker, 2013) but Bruzzone and Melgani (2005) developed an SVM-based conceptual multiple estimator system for biophysical parameter prediction. This system successfully develops a regression-based SVM and is widely applied to predict continuous output in environmental studies (Camps-Valls , 2006a; Camps-Valls , 2006b; Das et al, 2010; Durbha et al., 2007; Knudby et al., 2010a; Knudby et al., 2010b; Mountrakis et al., 2011; Sun et al., 2009). Random forest is a nonlinear ensemble machine learning model using multiple regression trees, where each tree is trained on a random subset of the training data and each split in each tree is determined using a random subsample of the available predictors (Breiman, 2001; Stumph & Kerle, 2011; Timm & McGraigal, 2012). R and its contributed packages (raster, e1071, and randomForest) were used to develop and apply the statistical models using default model parameter settings (Hijmans & van Etten, 2009; Liaw and Wiener, 2012; Meyer et al., 2012; R Core Development Team, 2008).

2.1.5. Accuracy Assessment

To quantify the prediction error expected for an unsampled area, each regression model was calibrated using observations from all weather stations except one, and predictions for that station were then compared to the observations to determine prediction error. This leave-one-out process was repeated using observations from each station for validation once. Using all data from a given weather station in either the calibration or validation data set ensured that this cross-validation procedure provided a good estimate of typical prediction errors for an unsampled location. Cross-validation is a widely recognized error assessment method frequently used to validate machine learning models such as Random Forest and SVM (Gislason et al, 2006). The mean absolute error (MAE) and root mean square error (RMSE) were used to quantify errors (Willmott & Matsuura, 2005), and the model with the lowest error values was ultimately used to create a relative Tmax map for Greater Vancouver.

2.1.6. Variance Importance Analysis

A model-independent permutation-based analysis was used to assess the influence of the individual predictors in the most accurate model (Strobl et al. 2009, Genuer et al. 2010, Knudby et al, 2010b). During this analysis, a model was calibrated using a subset (63.2%) of all observations and validated by quantifying the prediction error for the remaining observations. The values of one predictor were then randomly permuted in the validation data, new predictions produced, and the percentage increase in mean squared error (%IncMSE) caused by the permutation was calculated. This permutation was repeated for each predictor to quantify the relative increase in prediction error caused by permutation of the individual predictors.

2.2. Results and discussion

2.2.1. Model performance

The MAE and RMSE values were similar for Random Forest, support vector machine and linear regression (Table 2), ranging from 1.82 to 1.93°C and from 2.31 to 2.46°C respectively, which is comparable with previous remote sensing-based air temperature mapping studies (Benali et al., 2012; Czajkowski et al.; 2000, Prihodko & Goward, 1997; Vancutsem et al., 2010; Zakšek & Schroedter-Homscheidt, 2009, Zhu et al., 2013). The Random Forest model outperformed the other model types slightly, and was therefore retained for further analysis and mapping.

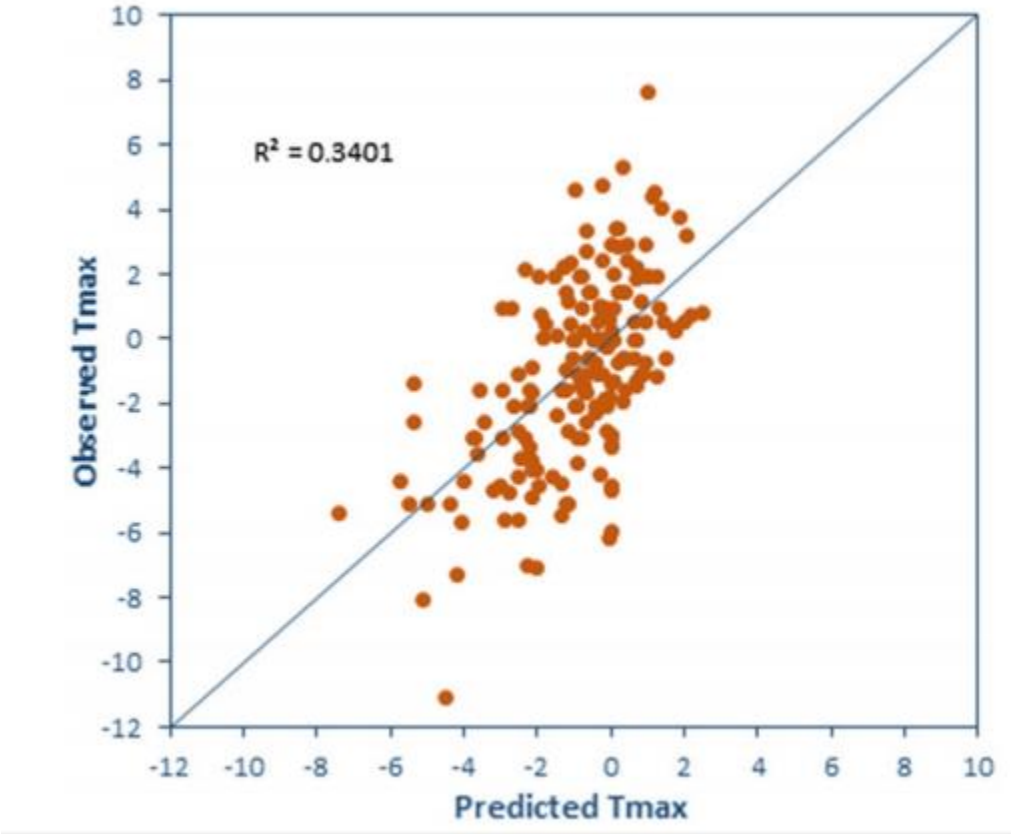
Table 2 – Maximum air temperature (Tmax) error estimates based on cross-validation, for each model type. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used to estimate the potential errors from predictions.

Model	MAE	RMSE
Random Forest	1.82 °C	2.31 °C

Support Vector Machine	1.91 °C	2.46 °C
Ordinary Least Squares Regression	1.93 °C	2.46 °C

Observations with very high Tmax values were consistently underestimated by the Random Forest model, and observations with very low Tmax values were consistently overestimated (Figure 2). This is partly an inevitable result of the model specification, where the air temperature predictions for a typical hot summer day are compared with observations for all hot summer days, some typical and some extreme. It may also result from the fact that several environmental variables with known influence on air temperature were not included as predictors in the model, such as wind speed and direction, turbulence, and anthropogenic heat, all of which may influence air temperature at very small spatial scales.

Figure 2. Scatter Plot of Validation Results from Air Temperature Mapping. The plot shows the comparison between the predicted and observed values from the best statistical model (Random Forest).

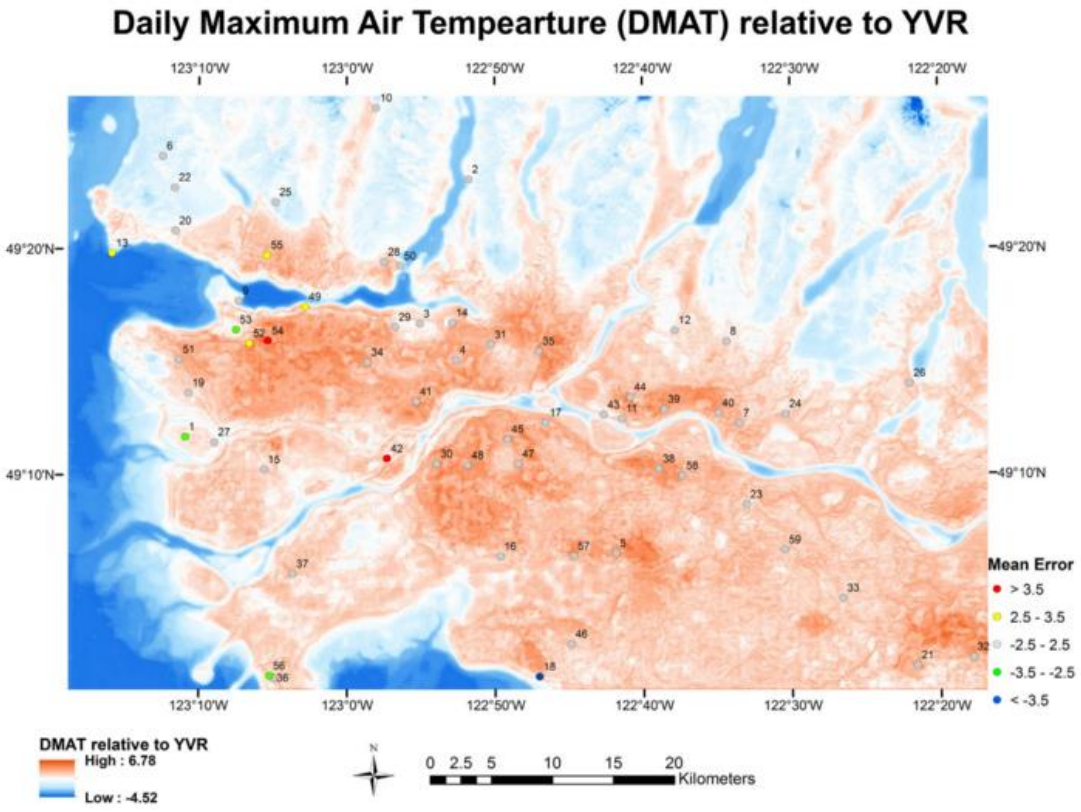


2.2.2. Spatial distribution of relative maximum air temperature

The map of Tmax distribution in the study area (Figure 3) was produced by application of the Random Forest model to each of the four Landsat 5 images and calculation of the average per-pixel Tmax prediction from each image. The two Landsat 7 images were excluded from this map production as their Scan-Line Corrector errors produced substantial local artifacts. The map presents a sensible description of the spatial distribution of air temperature in the study area. Low temperatures are predicted in the mountainous area north of Vancouver, particularly at high elevations, while high temperatures are predicted in the urban/suburban residential areas. Agricultural areas are predicted to be relatively cool, and forests are consistently predicted to be cool. These predictions correspond well with past studies of air temperature distribution in the area (Oke & Hay, 1994; Prihodko & Goward, 1997; Weng et al., 2004). Downtown Vancouver is predicted to be relatively cool compared with other parts of the metropolitan area. This can be the result of the proximity of downtown to the ocean,

which is known to be an important controlling factor of Tmax in Vancouver (Runnalls & Oke, 2000), especially when winds from the ocean cool down the coastal areas. Downtown also has low SVF values due to the high density of tall buildings. This reduces direct solar insolation and thus dampens the daily temperature cycle leading to lower Tmax. In addition, the tall buildings may also create an urban canyon with a “wall effect” (Wong et al., 2011) that controls the wind direction in the urban area and reduces airflow by trapping air masses within the downtown area. This isolating effect may create a relatively cold zone in the area dominated by high-rise buildings, where the cool air from the early morning and previous night is only slowly heated. Unfortunately, no field observations were available from downtown Vancouver for this study; future data collection will be used to update and validate the model predictions for this area.

Figure 3. Spatial Distribution of Air Temperature (Random Forest Model). Points indicate all weather stations used for calibration/validation. Point color indicates mean error values for each station.



2.2.3. Variable importance

All LST variables had high importance scores, as did solar radiation (Table 3). Positive importance scores for every variable is an indication that all variables have contributed to the model predictions. LST and solar radiation have previously been found to play a strong role in determining maximum air temperature for a typical hot summer day (Bristow & Campbell, 1984; Vancutsem et al, 2010), and receive the highest variable importance scores in our model. LST calculated at the largest radius (1000 m) had the highest score of any predictor, emphasizing the effect of LST in neighbouring areas on air temperature. The influence of NDWI could show the cooling effect of evapotranspiration in areas with relatively wetter vegetation and/or the shading effect of vegetation. The importance of SVF is relatively significant, and is more than half of the

most important variable. Given the concentration of low SVF areas around downtown Vancouver, where no weather stations are located, the actual influence of SVF on air temperature is likely to be underestimated by the variable importance results. A similar effect is likely the cause of elevation having the lowest variable importance, as only a few weather stations are located at elevations above 100 m. The permutation of both the SVF and elevation variables will therefore cause only a negligible change in the validation data.

Table 3 – Results from variable importance analyses (air temperature). The percentage increase in mean squared error (%IncMSE) quantifies the increase in prediction error caused by permutation of each individual predictor.

Variable	%IncMSE
LST (1000m)	17.49
LST (600m)	15.93
Solar	15.87
LST (800m)	14.85
LST (100m)	14.70
LST (400m)	13.24
SVF	13.17
NDWI (200m)	12.74

NDWI (600m)	12.71
LST (200m)	12.58
NDWI (400m)	12.30
NDWI (100m)	11.65
NDWI (800m)	11.42
NDWI (1000m)	9.39
elevation	7.44

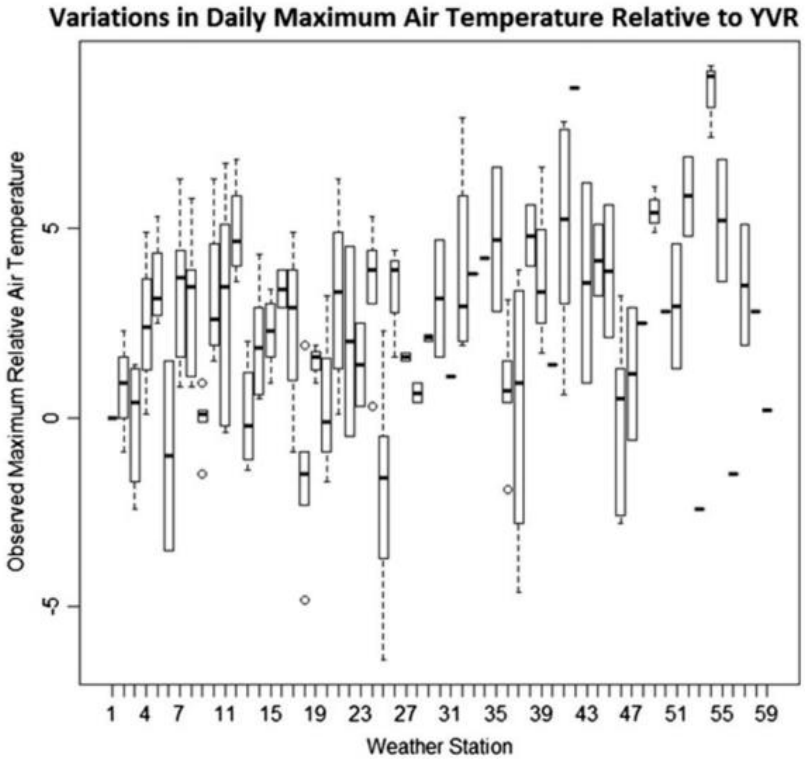
In addition, we note that it is difficult to assess the importance of any individual variable in a multivariate model, as the calculated importance of each variable is influenced by the presence of other correlated or interacting variables in a dataset, i.e. multicollinearity (Murray and Conner, 2009). Model structure can also affect variable importance, for example predictors related non-linearly to the response will have higher importance in nonlinear than linear models (Knudby et al, 2010b). More fundamentally, predictors that have themselves been mapped with low accuracy will, all other things being equal, have relatively low variable importance. Variable importance results, including ours, should thus only be interpreted with caution and in the light of other available evidence.

2.2.4. Temperature pattern variability

An implicit assumption behind mapping the distribution of Tmax relative to YVR for a typical hot summer day is that such a typical distribution exists, in other words that relative Tmax is reasonably stable for hot summer days in the area and that the notion of typical Tmax is meaningful. A typical weather pattern has indeed been shown to exist for

hot summer days in the area, with positive anomalies of the 500-hPa geopotential heights and 850-hPa temperatures centered over the British Columbia coast, and negative anomalies of sea level pressure centered off the coast of Washington State (Bumbaco et al. 2013), leading to cloudless skies and light winds from the Fraser Valley (Oke & Hay, 1994). Nevertheless, we observed considerable variability of Tmax relative to YVR (Figure 4). While some weather stations always recorded much higher temperatures than YVR (e.g. station 43) and some typically recorded lower temperatures (e.g. station 18), others (e.g. stations 4 and 20) show substantial variation in their temperatures relative to YVR. Further research should focus on identifying weather patterns or other factors that may control these variations.

Figure 4. Boxplot of Air Temperature Variation of Each Weather Station. The X-axis shows the ID of each weather station and the Y-axis shows the variation of air temperatures based on the weather records we used for predictions. The horizontal lines in each box represent the first, second and third quartiles, dashed lines extend to +/- 1.58 times the inter-quartile range divided by the square root of n, and outliers are denoted as circles.



To further investigate this issue, mean prediction errors (ME) for each station were calculated and used to separate stations into five groups: severely underestimated (ME > +3.5°C); underestimated (+3.5°C > ME > +2.5°C); unbiased (+2.5°C > ME > -2.5°C); overestimated (-2.5°C > ME > -3.5°C); and severely overestimated (-3.5°C > ME). A plot of these values on the map (Figure 3) suggests that stations close to the ocean are predicted more poorly than those further inland, as eight of the 21 weather stations located within 5 km of the ocean are either overestimated or underestimated, compared with two of the 38 stations located further inland. It is likely that variable wind patterns are responsible for the poorer prediction in areas near the ocean, as the air temperature here is strongly influenced by sea breezes bringing cold ocean air eastward, and land breezes bringing hot air westward from the Fraser Valley (Runnalls & Oke, 2000). In comparison, areas further inland are likely to be less influenced by wind direction. The lack of high spatial resolution data on wind speed and direction may thus be an important limitation for mapping of typical air temperature patterns in urban environments. EC weather stations are intended to monitor regional weather patterns, and are by design located outside densely populated areas. In the absence of appropriate data, modeling of local winds may be possible using simulation software (e.g. WindNinja, WindWizard) and high-resolution digital surface models. To our knowledge the accuracy of wind simulations for urban environments is unknown, as is their potential utility for air temperature mapping.

2.2.5. Weather underground data

Although WU provides guidelines for deploying weather stations, data quality is not controlled and was not accounted for in this study. Potential problems with the WU data are also illustrated in the mean error map (Figure 3), where it can be seen that air temperature is severely underestimated (ME > 2.5 °C) at six weather stations, of which five are from WU and only one is from EC. In other words, temperatures from 4 % of EC weather stations and 15 % of WU weather stations are severely underestimated. This problem is less acute with respect to overestimation of air temperature, which is seen for two EC and two WU weather stations. Greater prediction errors for WU weather stations could be a result of improper placement of some WU weather stations (e.g. next to heat sources or close to a highly reflective surface). The use of WU data in addition to EC

data for model calibration has both advantages and disadvantages. In our study area inclusion of WU data almost doubled the number of weather station locations for which data were available. In addition, the use of WU data complemented the spatial distribution of data from EC weather stations, as the former were primarily available from urbanized areas while the latter are deliberately located away from large buildings and other reflecting surfaces typical of the urban environment (World Meteorological Organization, 1996). However, the lack of quality control is likely to result in some stations being improperly placed or operated, introducing bias for model calibration. This issue will be of increasing importance in the future, as WU and other volunteered geographic information (VGI) projects produce increasing volumes of data with uncertain quality. One alternative to VGI for increasing the amount of calibration/validation data could be dedicated mobile temperature monitoring (Anderson et al. 2012), which lends itself better to quality control but requires additional resources.

2.2.6. Complementary Data Source

Finally, to incorporate spatial information on diurnal temperature variations, integration of MODIS-derived predictors may be feasible in future studies. The higher temporal resolution of MODIS images, and especially the early afternoon overpass of the Aqua satellite, may allow a description of energy balance dynamics that is better resolved in time and closer to the typical time of peak temperature (between 3 and 5 pm in our study area). The use of VGI, wind simulation software, and MODIS data may all be avenues to improve the modeling of air temperature in urban environments and achieve lower error values than those reported here.

Chapter 3.

Comparison of the intra-urban difference of temperature distributions mapped with different temperature metrics

This chapter expands on the air temperature mapping outlined in the previous chapter by producing additional and comparable maps of a) mid-morning land surface temperature and b) daily maximum apparent temperature (Humidex). The three maps are compared to identify similarities and differences and thus to determine the differences in spatial distribution between the three temperature measures. Details of this chapter have been published in the following peer-reviewed article: Hung C. Ho, Anders Knudby, Yongming Xu, Matus Hodul and Mehdi Amnipouri (2016). A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area. Science of the Total Environment, 544, 929-938. My major contributions of this paper included: 1) to develop the idea of mapping humidex, 2) to develop the idea of comparing humidex map with land surface temperature and air temperature maps, 3) to decide how to improve the statistical model to map Humidex, 4) to decide how to compare the temperature maps, 5) data preparation, 6) data analysis, and 7) to write and revise the manuscript.

3.1. Data and methods

3.1.1. Humidex field observations

Maximum air temperature and dew point data were obtained from 39 weather stations operated by Environment Canada or the Weather Underground; the smaller number of weather stations compared to those used to calibrate the air temperature model was due to not all stations collecting humidity data. Data were obtained for six hot

summer days in the period 2002-2009, for which cloud-free Landsat satellite images were also available (Table 1), all with Tmax >25°C at YVR. Using only data from hot summer days ensured that the results are calibrated to such days, when the risk of heat-caused human health problems is greatest. Not every weather station was operational for all six days used in this study; several Weather Underground and Environment Canada stations only had data available for one or two days of the six days. For each observation, maximum air temperature and co-incident dew point data were used to calculate apparent temperature using the Humidex, a well-known heat index used to quantify heat exposure in Canada (Burke et al., 2006; Gosling, 2014; Masterson & Richardson, 1979). The Humidex is defined as:

$$H = Ta + 0.5555 \times \left(6.11 \times e^{5417.753 \times \left(\frac{1}{273.16} - \frac{1}{dew}\right)} - 10\right)$$

where H is the Humidex, Ta is the air temperature, and dew is the dew point in degrees Kelvin.

3.1.2. Spatial data layers

Spatial data layers containing potential predictors of Humidex for use in regression modeling were derived from four different sources. 1) Cloud-free Landsat 5 TM and Landsat 7 ETM+ images were obtained for the six days for which field observations were available, and 2) a land surface emissivity data set was retrieved from the North American ASTER Land Surface Emissivity Database (NAALSED) (Hulley & Hook, 2009). In addition, 3) a Digital Elevation Model (DEM) was obtained from the 25m Canadian Digital Elevation Data (CDED, http://www.geobase.ca), and 4) the MODIS Precipitable Water product (MOD05) was used to provide information on atmospheric water vapor content for the six days with Landsat data.

The Landsat 5 TM images, the MODIS water vapor product and the DEM were all resampled to 60m cell sizes to match the spatial resolution of the ETM+ thermal band, and all data were re-projected to the Universal Transverse Mercator projection, Zone 10 N. Areas affected by Scan-Line Corrector (SLC) errors in each Landsat ETM+ image were masked out prior to further analysis.

3.1.3. Regression modelling

The following environmental data layers were derived from the satellite and digital elevation data and used as predictors of Humidex in a regression model: 1) Land Surface Temperature (LST), 2) Elevation, 3) the Normalized Difference Water Index (NDWI), 4) Sky View Factor (SVF), 5) Incident solar radiation, 6) Distance from the ocean, and 7) Atmospheric water vapor.

All data layers except a) distance from the ocean and b) atmospheric water vapour were produced using similar approaches to those outlined in section 2.1.2.

Distance from the ocean was estimated from the Vegetation Resources Inventory (VRI) land cover map by Ministry of B.C (Taylor, 1998). All terrestrial land cover classes were joined to create a binary classification with only “water” and “land” categories from which the distance to the ocean was calculated.

Atmospheric water vapor was derived from the MODIS Precipitable Water product (MOD05) near-infrared total-column precipitable water, which is known to be sensitive to near-surface water vapor in the canopy layer (King et al, 2003; Wong et al, 2015). The canopy layer is the layer within 2m above the land surface, which is a key component of intra-urban difference of temperature patterns and the atmospheric layer, and is most important to environmental health studies (Oke, 1976).

Spatial averaging was applied to the LST and NDWI layers to account for the influence of these surface variables in areas around each weather station (Liu and Moore, 1998; Zakšek, K., & Oštir, 2012). A range of circular kernels were tested for this spatial averaging, with radii ranging from 100m to 1000m in 100m intervals. To determine which LST kernel radius would lead to the best model performance, spatially averaged LST data layers were included individually in the model in combination with all other data layers, and cross-validation used to quantify model performance. The LST kernel radius that achieved the lowest RMSE value was retained. A similar process was performed to determine the best NDWI kernel radius. In subsequent model runs, only the best spatially averaged LST and NDWI data layers in combination with all other data layers were used as predictors of Humidex.

A Random Forest regression model was used to formalize the relationship between these spatial data layers and the Humidex data. We used the randomForest package (Liaw & Wiener, 2012) in the statistical software R, with default parameter settings (Hijmans & van Etten, 2009; Liaw and Wiener, 2012; Meyer et al., 2012; R Core Development Team, 2008), which had previously proven effective at similarly mapping air temperature (Ho et al. 2104). The model was first used to predict the maximum Humidex for each of the six days for which my research group had satellite data (Table 1), and the value predicted for YVR was then subtracted to produce a relative temperature measure. The mean of these predictions of maximum Humidex relative to YVR was then calculated to represent the relative Humidex on a typical hot summer day.

3.1.4. Accuracy assessment

The Random Forest model was validated using leave-one-station-out cross validation, as outlined in section 2.1.5.

3.1.5. Variance importance analysis

A permutation-based variance importance analysis similar to the one outlined in section 2.1.6 was used to determine the contributions of each predictor to the regression model (Strobl et al. 2009, Genuer et al. 2010, Knudby et al, 2010), however, because only the Random Forest model type was used in the Humidex model, the implementation of this analysis in the randomForest R package was used. For each regression tree, the data that were not used to 'grow' the tree were initially used to assess prediction error. For these data points, each predictor was then permuted in turn and the resulting increase in prediction error recorded. This was done for all trees, and the average reported as the percentage increase in mean squared error (%IncMSE).

3.1.6. Production of temperature maps

The Random Forest model was applied to the spatial data layers to produce maps of the distribution of maximum Humidex in the study area, relative to YVR, for the six days studied. The maps produced with the four Landsat 5 TM images were then

averaged to produce a single map representative of a typical hot summer day in the area; Landsat 7 ETM+ images were not used for this averaging due to the SLC errors. An identical approach to the one described above, but using maximum air temperature instead of maximum Humidex as the dependent variable in the Random Forest model, was used to produce a comparable map of maximum air temperature. This maximum air temperature map is similar in nature to the one described in section 2.1, but was produced with identical calibration data and spatial data layers to facilitate better comparison with the map of maximum Humidex. In addition, a land surface temperature map was created by calculating the mean of the LST values derived from each of the four Landsat 5 TM images.

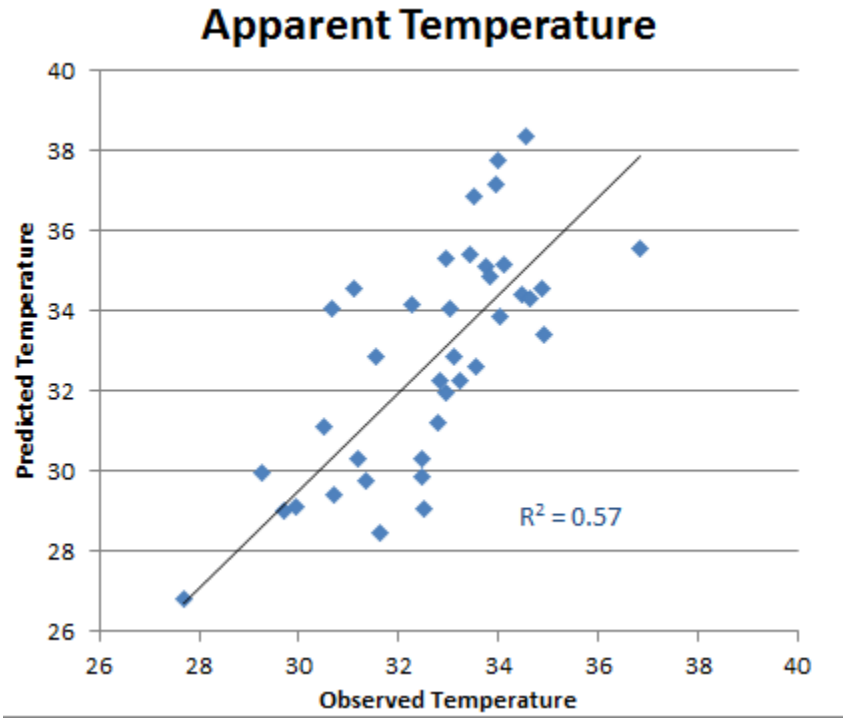
3.2. Results and discussion

3.2.1. Predictions of Humidex

In general, larger radii used for spatial averaging resulted in lower prediction errors for Humidex. LST averaged over the largest radii (1000m) produced the lowest prediction error, as did NDWI averaged over 500m. Using this spatially averaged LST and NDWI layers along with the other predictors, the MAE and RMSE values from the Random Forest model were 1.67 °C and 2.04°C respectively (R^2 = 0.57; Figure 5), which is considered fairly accurate compared to other air temperature mapping studies (Benali et al, 2012; Ho et al, 2012; Prihodko & Goward, 1997; Zakšek & Schroedter-Homscheidt, 2009), although Humidex and air temperature itself is not directly comparable. Very high values of Humidex were typically underestimated, and one low value was significantly overestimated. This is a known issue caused by the Random Forest model structure, in which predictions are not extrapolated beyond the values included in the training data. The issue may have been further affected by the presence of important predictors not included in the model, such as wind speed and anthropogenic heat generation (Oke, 1988). Similar results were seen for predictions of air temperature, with MAE and RMSE values of 1.50°C and 1.90°C respectively, and was also fairly accurate compared to

other air temperature mapping studies (Benali et al, 2012; Ho et al, 2012; Prihodko & Goward, 1997; Zakšek & Schroedter-Homscheidt, 2009).

Figure 5. Scatter Plot of Validation Results from Humidex Mapping. The plot shows the comparison between the predicted and observed values from the Random Forest model.



3.2.2. Variable importance (Humidex)

The variable importance results for all predictors are shown in Table 4. The most important predictor was LST, which is known to directly influence air temperature through radiative, conductive, and convective heat transfer (Bristow & Campbell, 1984; Vancutsem et al, 2010). Atmospheric water vapor and distance from the ocean also had high variable importance scores, which was expected as one directly quantifies humidity and the other influences the strength of the moderating influence of the ocean. SVF had only moderate variable importance, however, given that no weather stations were located in very low SVF areas such as downtown Vancouver, the actual influence of

SVF on Humidex in such areas is not captured in the model and the importance of SVF across the study area may thus be underestimated. A similar effect may have caused the low variable importance values of the NDWI and elevation predictors, as only a few weather stations are located in high NDWI (forest) and high elevation (mountain) locations.

Table 4 – Results from the variable importance analysis (Humidex). The percentage increase in mean squared error (%IncMSE) quantifies the increase in prediction error caused by the permutation of each individual predictor in the validation data.

Variable	%IncMSE
LST (1000m)	26.68
Atmospheric water vapor	17.65
Distance from ocean	14.81
SVF	9.01
Solar radiation	5.81
NDWI (500m)	2.84
Elevation	1.92

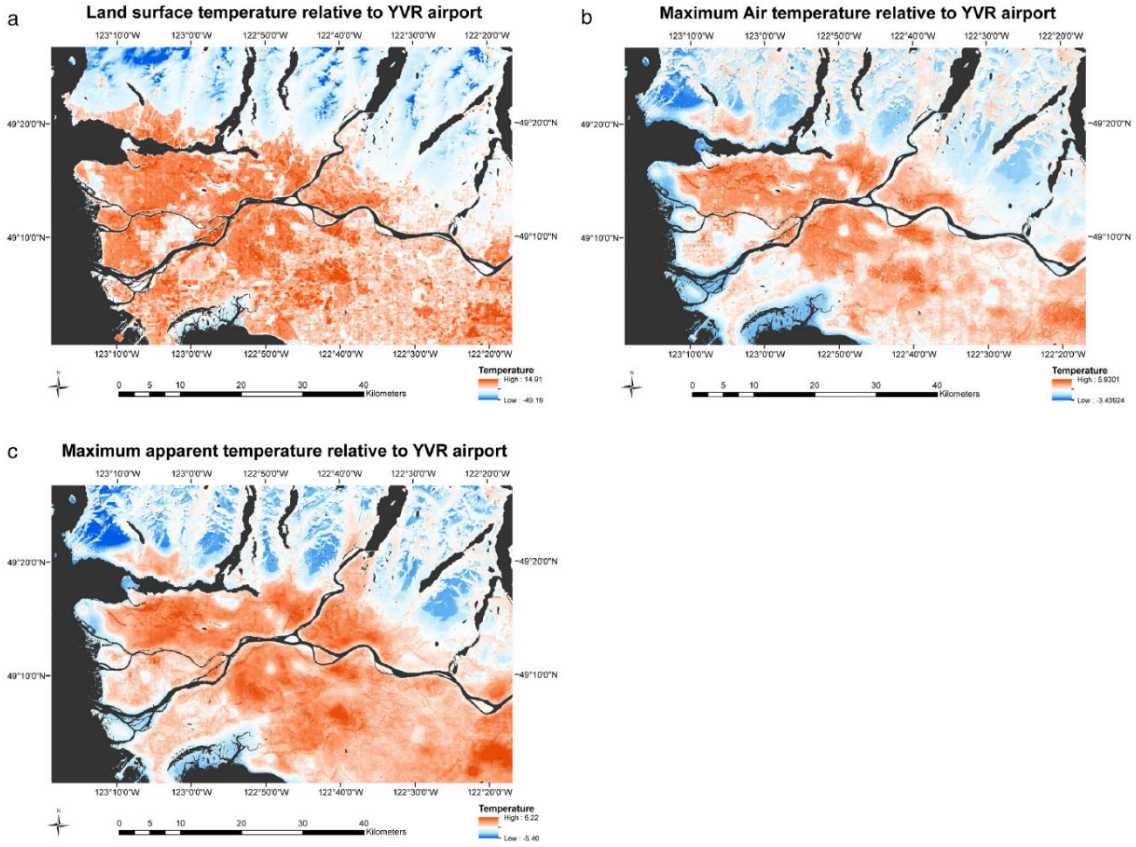
3.2.3. Spatial distribution of the three temperature measures

Maps showing the distribution of each temperature measure on a typical hot summer day in the study area are shown in Figure 6. The distribution of maximum Humidex (Figure 6c) corresponds well with past temperature research in the greater Vancouver area (Oke & Hay, 1994), and shows that relatively high Humidex are found in urban areas, while low Humidex are found in the mountainous areas of West and North Vancouver. Agricultural lands and forests were predicted as relatively cool, and downtown Vancouver was predicted to be relatively cooler than other major urban areas. This latter result could be due to a sea breeze cooling down the downtown core, which is

surrounded by water to the south, west, and northeast (Runnalls & Oke, 2000). A “wall effect” (Wong et al., 2011) in which cold air is trapped within the urban canyon at night and the early morning, reducing air flow and preventing direct solar radiation from increasing the temperature of the downtown air mass during the day, could also be a contributing factor.

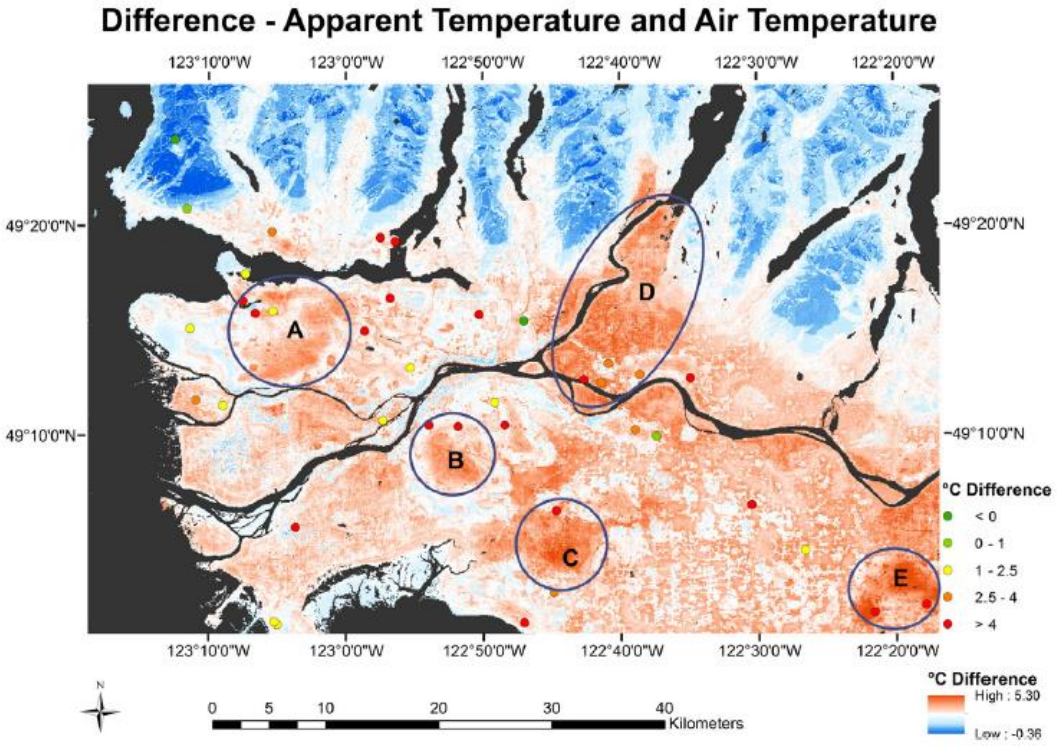
The spatial temperature patterns are broadly similar between the three temperature measures, with urban areas being relatively warm while forested and mountainous areas are relatively cool, however, notable differences also exist. LST values (Figure 6a) in the area range from almost 50 degrees cooler to almost 15 degrees warmer than YVR, while air temperature and Humidex values have a smaller value range. In addition, the LST map shows strong contrasts in temperature between adjacent areas, while values vary more gradually in the air temperature (Figure 6b) and Humidex (Figure 6c) maps. Considering how the maps have been produced, this difference in local contrast is a result of data layers with low local contrast having been used to produce the air temperature and Humidex maps, notably the spatially averaged LST and NDWI data layers as well as the ‘distance from the ocean’ layer. However, it is also a realistic description of LST being strongly dependent on the radiative and thermal properties of the surface material while air temperature and Humidex are relatively more influenced by advection and thus by surface conditions in a larger upwind area (Stewart and Oke 2012). As a result, the LST distribution is poorly correlated to both that of air temperature (R^2 = 0.38) and Humidex (R^2 = 0.39). This result is also confirmed by the observed data

Figure 6. Spatial Distributions of Temperature from Three Metrics. 6a (Top left) is the spatial distribution of land surface temperature, 6b (Top right) is the maximum air temperature from Random Forest model, predicted by both Landsat and MODIS product, and 6c (bottom) is the maximum Humidex map from Random Forest model.



from weather stations, for which the correlation between LST and the observed air temperature from all records of all weather stations was only 0.15. While the air temperature and Humidex distributions show much greater similarity (R^2 of model outputs = 0.87, R^2 of observed values from weather station = 0.74), several areas of interest are predicted to have relatively high Humidex, indicating high humidity and thus high Humidex values on hot summer days. Some of these areas are highlighted with blue circles in Figure 7, which shows the difference between the predicted Humidex and air temperatures. Hot spots of Humidex, relative to air temperature, include some of the densely built-up neighborhoods in East Vancouver and central Burnaby, downtown Surrey, the small community of Cloverdale, the confluence of the Pitt and Fraser rivers, and the city of Abbotsford. In these areas, predicted Humidex are up to 5 degrees higher than the corresponding air temperatures.

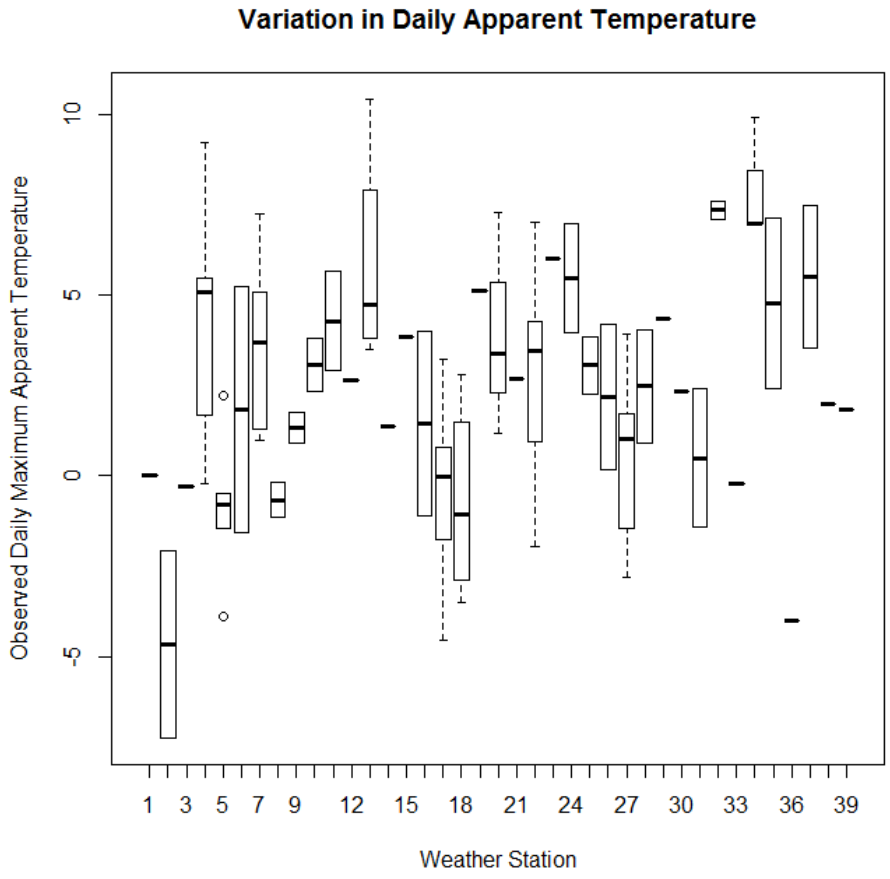
Figure 7. Comparison of Air Temperature and Humidex Maps. This map is calculated by Humidex minus air temperature. Circles indicate hot spots of Humidex. Top left (cir- cle A): Central Burnaby and East Vancouver. Bottom left: Central Surrey (circle B) and Cloverdale (circle C). Top right (circle D): Pitt Meadows and Port Coquitlam at the confluence of the Pitt and Fraser Rivers. Bottom right (circle E): Abbotsford. Points indicate average differences between Humidex and air temperatures observed at weather stations on hot summer days.



3.2.4. Humidex patterns and variability

Despite hot days in the area having a distinct weather pattern, Figure 8 shows that some variation exists in daily maximum Humidex, relative to YVR, observed at each weather station on the days used in our study. While several weather stations (e.g. stations 5, 16, 17, 18, 27) show variations in temperature that include values both higher and lower than YVR, other stations are either consistently warmer than YVR for the days included in the study (e.g. station 13), or consistently colder (e.g. station 2).

Figure 8. Boxplot of Humidex Variation of Each Weather Station. X-axis showed the ID of weather station and Y-axis showed the variation of air temperatures based on the weather records we used for predictions. The horizontal lines in each box represent the first, second and third quartiles, dashed lines extend to +/- 1.58 times the inter-quartile range divided by the square root of n, and outliers are denoted as circles.



Closer examination of each station's local landscape context showed that stations with high Humidex relative to YVR were typically found near sparsely built-up or agricultural areas, as were the case for Abbotsford and Pitt Meadows respectively. Stations with relatively low Humidex were found in the forested and mountainous areas in the northern part of the study area. Stations with high temperature variability were primarily found in suburban areas near the coast, a context that allows wind speed and direction great influence on the local temperature (Runnalls & Oke, 2000).

Apart from the local landscape context, imperfect data quality may also have contributed to the observed temperature variability. Many Environment Canada weather stations only provide records of temperature and dew point in the form of daily minima and maxima, which are not suitable for calculation of maximum Humidex because the maximum temperature and the maximum dew point may not occur at the same time. As a result, data from these stations could not be used in this study. Weather Underground stations typically provide hourly values of the same measurements, as a result 82% of the weather stations used in this study belonged to the Weather Underground. Environment Canada follows guidelines from the World Meteorological Organization (WMO) when deploying weather stations, which ensures that these are located away from large buildings, reflective surfaces, and other sources of heat or radiation (World Meteorological Organization, 1996). While this aims to ensure that measurements are representative of the air mass in the surrounding area, it also precludes deployment of weather stations in the urban areas which are of greatest interest from a human health perspective. Weather Underground data has potential data quality issues, which we had stated in Section 2.2.5.

3.2.5. Comparison and validation of temperature distributions

Despite the strong correlation between the predicted air temperature and Humidex, important differences exist in their spatial distribution, especially in the areas circled in Figure 7 where Humidex typically exceed air temperatures by up to 5 °C. Several urban centers show relatively large differences between Humidex and air temperature, including Abbotsford, Cloverdale, Surrey, and parts of Burnaby and Vancouver. While this is an intriguing result and weather station data confirm the existence of such differences in the study area, no weather stations whose observations could validate the result are located in the cores of these areas. However, coverage of weather stations in the Weather Underground network is rapidly expanding in the study area and it is hoped that this particular result can be validated with field observations in the near future. On the other hand, several weather stations are located near the confluence of the Pitt and Fraser rivers, confirming that this area experiences relatively high Humidex. Given that many large cities are located in river valleys, on deltas, or on historical floodplains (e.g. New York City, Hong Kong, London, and Shanghai), the

relatively high Humidex observed in both urbanized areas and at the confluence of the two rivers has important implications for mapping human heat exposure in urban areas and their surroundings. The differences in the spatial distributions of all three temperature measures are especially important because Humidex has been shown to be the most closely linked to urban heat health, indoor temperature (Nguyen et al. 2014), and mortality during extreme heat events (Almeida et al. 2010, Barnett et al. 2010, Zhang et al. 2014). It is likely that heat-related spatial epidemiology studies will benefit from using Humidex instead of LST to quantify human heat exposure. Furthermore, future studies can be developed based on this study, such as, comparing the temperature maps we developed in this study with the common temperature metrics (e.g. radiant temperature) that can be used to measure the human thermal comfort (Blazejczyk et al, 2012), in order to develop a better overview of how thermal environment influences the living conditions. However, since our paper is focused on evaluating the spatial distribution of temperature maps in order to enhance future public health planning and urban health assessment, research related to human thermal comfort is excluded from this study.

Chapter 4. Delineation of the spatial variability of the temperature-mortality relationship

This chapter uses the maps developed in the two previous chapters to spatially resolve relative heat exposure in greater Vancouver, and applies the resulting heat exposure variables to long-term mortality data from the area to assess the heat health risk associated with living in a relatively hot area. Details of this chapter are currently in revision with the following peer-reviewed journal: Environmental Health Perspectives, under a manuscript titled “Delineation of spatial variability in the temperature-mortality relationship on extremely hot days in greater Vancouver, Canada”, first-authored by Hung Chak Ho. My major contributions of this paper included: 1) to develop the idea of comparing heat mortality risk with different temperature metrics and social vulnerability data, 2) to apply the advices from graduate committees to apply a spatial delineation method of case-crossover study to evaluate the relationship between risks, exposures and vulnerability, instead of using the regression-based analysis proposed in the PhD proposal, 3) to decide which variables should be used in the case-crossover study, 4) data preparation, 5) data analysis, 6) to develop the idea of how to map the heat risk based on the final results from spatial delineation, and 7) to write and revise the manuscript.

4.1. Data and methods

4.1.1. Mortality data

The BC Centre for Disease Control (BCCDC) receives daily data from the BC Vital Statistics Agency for the purposes of routine surveillance and evaluation of public health threats. I used data from 1998 through 2014 for the analyses. The database includes the following relevant information about each decedent: date of death; age in

years; sex; underlying cause of death coded according to the 10th Revision of the International Classification of Diseases (ICD-10); and 6-digit residential postal code. An additional field describing the location of death (home, hospital, care facility, or other) was added in 2008. All deaths within greater Vancouver were geocoded by the residential 6-digit postal codes using the DMTI geodatabase (Statistics Canada, 2007), with 99.99% success. All deaths due to transport accidents (ICD-10 codes V01-V99) were omitted from the analyses.

4.1.2. Time-varying temperature data

Hourly data from the Environment Canada weather station at YVR were used as the time-varying measure of exposure in the time-stratified case-crossover analyses. Every decedent was assigned daily mean Humidex values for the day of death (case day) and for all occurrences of the same week day within the same calendar month (control days). The Humidex is a measure of apparent temperature routinely used to quantify heat exposure in Canada (equation noted at Section 3.1.1). Daily mean Humidex was chosen as the time-varying temperature metric because previous work has shown that it is more strongly associated with mortality than air temperature in the study area (Henderson et al. 2013).

4.1.3. Spatial heat exposure data

In order to evaluate how the temperature-mortality relationship was associated with different measures of spatial variability in temperature, we used the maps of LST, maximum daily air temperature outlined in chapter 2, and maximum daily Humidex outlined in section 3.1.6. Each geocoded death was assigned three heat exposure values, one for each of LST, maximum air temperature and maximum Humidex, to quantify heat exposure.

4.1.4. Spatial social vulnerability data

Deprivation indices are commonly used for describing social and health disparity across Canada (Bell et al. 2007; Bell and Hayes 2012; Pampalon and Raymond 2000;

Pampalon et al. 2009; Pampalon et al. 2012; Walker et al. 2014). The Vancouver Area Neighborhood Deprivation Index (VANDIX) is a combined measure of material and social deprivation that was developed specifically for health research in the study area. Researchers asked 27 local medical health officers to rank 21 census variables with respect to their perceived importance to public health in the region, and 10 officers returned the completed survey (Bell et al. 2007; Bell and Hayes 2012; Walker et al. 2014). Of the 21 variables, seven were used to construct the VANDIX in ranked order of importance: percent of the population that did not finish high school; unemployment rate; percent of the population with a university education; percent of single parent families; average income; percentage of homes owned; and labour participation rate. The latter refers to the percent of the adult population that is either employed or actively looking for work.

We used the VANDIX itself and all seven contributing variables to evaluate spatial differences in the temperature-mortality relationship related to social vulnerability. In addition, this research considered some census variables used for the US heat vulnerability index (Buscail et al 2012; Morabito et al. 2015; Reid et al. 2009; Tomlinson et al. 2011). These included: the density of the senior population (age ≥55); the density of persons living alone; and the density of older housing (built prior to 1970). The values for all variables and the resulting VANDIX were taken from the 2006 national census using the SimplyMap 3.0 database (http://geographicresearch.com/simplymap/). The data were mapped at the level of census dissemination area, where each spatial unit contains an approximate population of 400-700 persons (Statistics Canada 2007).

There were four variables for which increasing heat mortality risk was assumed to be associated with decreasing values: percent of the population with a university education; average income; percent of homes owned; and labour participation rate. To ensure consistency of interpretation between variables, we expressed these as: percent of the population without a university degree; average income ratio; percent of homes rented; and the labour nonparticipation rate. The average income ratio for each dissemination area was calculated as the minimum average income for all dissemination areas divided by the average income for that area.

4.1.5. Spatial delineation of the temperature-mortality relationship using case-crossover analyses

The residential location of every successfully geocoded record in the mortality dataset was assigned values from the heat exposure and social vulnerability maps. We considered these 14 variables to be potentially indicative of spatial differences in the temperature-mortality relationship assessed with time-varying mean Humidex measured at YVR. Given our interest in spatial variation during extreme hot weather, the mortality dataset was then reduced to all records with a case day or a control day in the 99.9th percentile (≥24°C) of daily mean air temperature values measured at YVR from 1998-2014. In other words, the analytic dataset included all deaths that occurred on extremely hot days, and all deaths that occurred on comparable non-extreme days. Conditional logistic regression was used to estimate the odds ratio (OR) for mortality associated with a 1°C increase in daily mean Humidex at YVR. All analyses were conducted in R with the 'survival' package, a package for survival analysis (Gail et al. 1980; Therneau 2014).

The spatial delineation was conducted for each of the 14 spatial variables as follows: (1) we took the values of the variable for all decedents and placed splits at the 1st and 99th percentiles; (2) we placed another 98 splits at equal intervals between these values; (3) for each of the 100 splits the complete dataset was subdivided into two groups, with decedents who lived in areas with below the split value in the first group (below-split) and those who lived in areas equal to or greater than the split value in the second group (above-split); (4) the conditional logistic regression model was run for both subsets to estimate the ORs and 95% confidence intervals for areas above and below the splits; and (5) we plotted the results for the below-split and above-split subsets as two smooth lines in one graph. Where there is increasing risk of mortality associated with increasing values of a spatial variable, we expect the above-split ORs to increase as the split value increases. For variables where there was a positive relationship with the above-split ORs, we identified a single "cut point" (a significant split value) to spatially delineate areas of higher risk based on the first separation of the below-split and above-split confidence intervals.

4.2. Results and discussion

4.2.1. Data Summary

The analytic mortality dataset included 1) cases who died on extremely hot days (mean air temperature ≥ 24°C at YVR) and 2) cases who died on cooler days (mean air temperature < 24°C at YVR) that had an extremely hot day from the same weekday of the same month as the control data. There were six extremely hot days during the 1998-2014 study period (two in 1998, one in 2006, and three in 2009), and 267 deaths occurred on those dates. There were 21 cooler days that had an extremely hot day as a control day, and 730 deaths occurred on those dates (Table 5). The mean Humidex at YVR on the extremely hot days was 25.6°C (standard deviation: 1.6°C), and the mean of the maximum values was 31.2°C (standard deviation: 2.1°C). The number of deaths on extremely hot days ranged from 26 to 65, with the highest number of deaths occurring on the hottest day. The number of deaths on the cooler days ranged from 15 to 47 (Table 5). Cases who died on extremely hot days had consistently higher heat exposures and indicators of social vulnerability, but differences between groups were only significant for three variables: VANDIX, average income ratio, and percent of homes rented (Table 5).

Table 5 - Summary statistics for the mortality analytic dataset. Summaries are presented for cases who (1) died on the extremely hot days (mean air temperature ≥ 24°C at YVR) and (2) who died on a cooler day that had one of the extremely hot days as a control day. Values in bold indicate a statistically significant difference between means (t-test with alpha = 0.05).

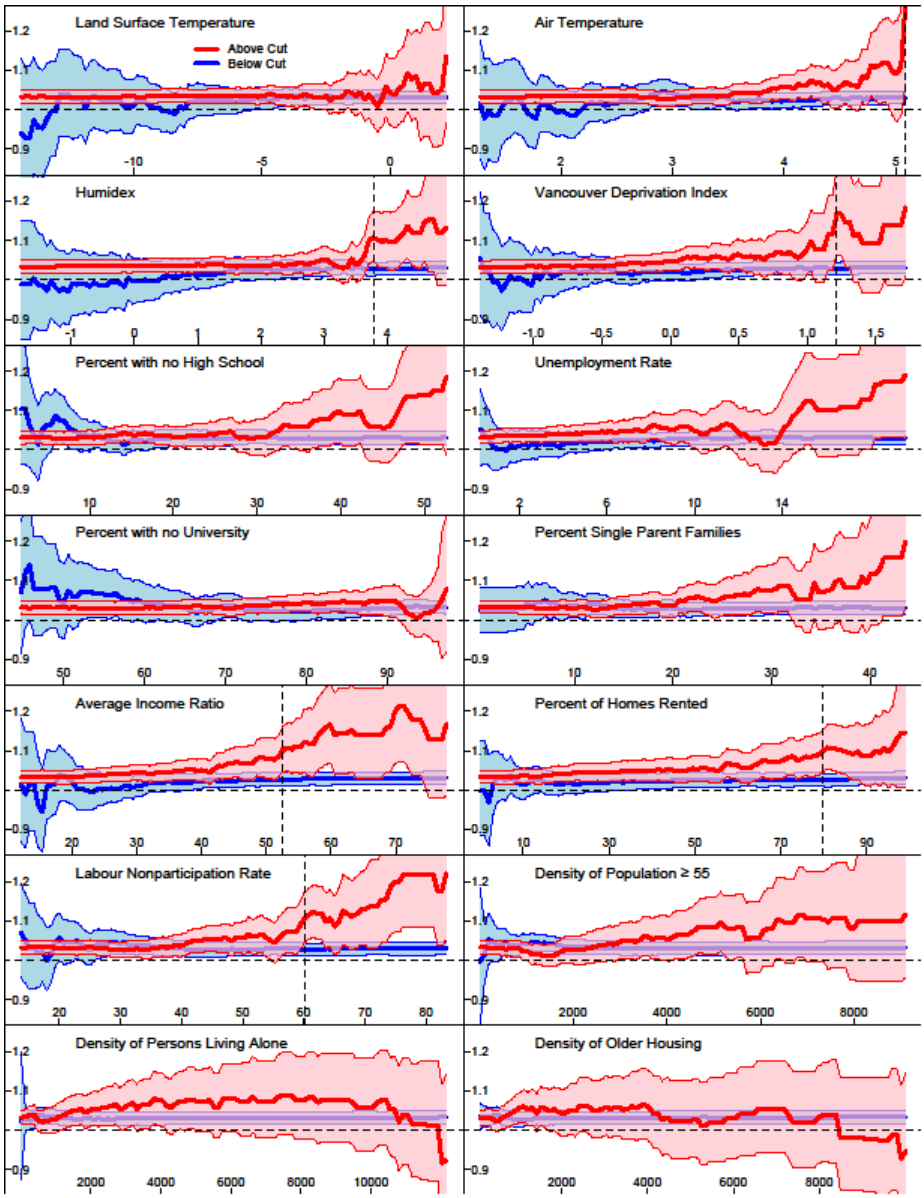
	Cases who died on six extremely hot days	Cases who died on 21 cooler days with an extremely hot day as a control day

Total	267	730
Average (range) deaths per day	44.5 (26-65)	34.1 (15-47)
Percent male	50.2	48.6
Mean (SD) age at death (years)	75.5 (15.7)	74.7 (18.6)
Mean (SD) land surface temperature relative to YVR (°C)	-4.7 (3.1)	-4.9 (3.2)
Mean (SD) air temperature relative to YVR (°C)	3.5 (0.87)	3.4 (0.86)
Mean (SD) Humidex relative to YVR (°C)	2.2 (1.4)	2.0 (1.4)
Mean (SD) VANDIX	**0.02 (0.67)**	**-0.09 (0.60)**
Mean (SD) percent with no high school graduation	19.6 (11.3)	18.8 (10.0)
Mean (SD) unemployment rate (%)	6.3 (4.2)	5.8 (3.9)
Mean (SD) percent with no university degree	77.9 (14.0)	77.3 (12.9)
Mean (SD) percent of single parent families	16.3 (10.6)	15.3 (8.6)
Mean (SD) average income ratio	**37.4 (13.5)**	**34.8 (11.4)**
Mean (SD) percent of homes rented	**37.7 (27.4)**	**33.9 (25.6)**
Mean (SD) labour nonparticipation rate	38.1 (15.7)	36.7 (13.0)
Mean (SD) density of population ≥ 55 years (km^{-2})	1627 (2056)	1551 (1661)
Mean (SD) density of persons living alone (km^{-2})	1145 (2269)	1070 (2311)
Mean (SD) density of housing built prior to 1970 (km^{-2})	831 (1630)	828 (1803)

4.2.2. Spatial variability in the temperature-mortality relationship by heat exposure

There were significant differences in risk between the LST, air temperature, and Humidex, based on the spatial values extracted from the geocoded mortality data (Figure 9). For LST there was near-complete overlap of the confidence intervals for odds ratios above and below all cut points (Details of cut point referred to Section 4.1.5). For air temperature there was some statistically insignificant separation of the lines starting at values approximately 4°C warmer than those at YVR, with the only significant difference observed in areas more than 5°C warmer. For Humidex there was a sharp increase in risk above the cut point for areas greater than 3.5°C warmer than YVR, with the first significant difference observed at 3.7°C. At this value the odds ratio (OR) for a 1°C increase in daily mean Humidex at YVR was 1.09 [1.04, 1.15] for data above the cut point (values within square brackets are the lower and upper confidential interval of the odds ratio), compared with 1.03 [1.01, 1.04] for data below the cut point. The area above the cut point included 12.0% of deaths on extremely hot days and 6.7% of deaths on the cooler days.

Figure 9. Delineation of Temperature-Mortality Relationship. Each figure indicates the odds of mortality associated with a 1°C increase in daily mean Humidex for decedents who lived in areas with the corresponding spatial values above (red lines) and below (blue lines) each cut point value on the x-axis. The first cut point at which there is separation between the lower confidence interval of above split and the upper confidence interval of the below split, and is marked with a vertical dashed line, where applicable.



4.2.3. Spatial variability in the temperature-mortality relationship by social vulnerability

We used 11 variables to assess spatial differences in the temperature-mortality relationship associated with social vulnerability. Of these, only four showed statistically significant separation of the odds ratios above and below any of the cut points (Figure 9). These were VANDIX (≥1.2), average income ratio (≥52%), percent of homes rented (≥80%), and the labour nonparticipation rate (≥60%). The strongest trend was for the labour nonparticipation rate. Above the cut point of 60% the OR for a 1°C increase in daily mean Humidex at YVR was 1.11 [1.04, 1.17] compared with 1.03 [1.01, 1.04] below the cut point. The area above the cut point included 9.4% of deaths on extremely hot days and 5.2% of deaths on the cooler days. Above a cut point of 75% the OR for a 1°C increase in daily mean Humidex at YVR was 1.20 [1.07, 1.34]. The area above the cut point included 3.7% of deaths on extremely hot days and 1.0% of deaths on the cooler days.

4.2.4. Sensitivity analyses

We conducted this study with administrative mortality data from the BC Vital Statistics Agency. One limitation is that all decedents were assigned values of the spatial variables using their residential postal codes, regardless of where they experienced excessive heat exposure or died. It is likely that some who died in hospital were not exposed at their residential locations on the day of death, meaning that they were spatially misclassified, particularly for the heat exposure variables. To assess these potential impacts of this misclassification we restricted the analytic dataset to the 27.7% subset of 276 decedents who were known to have died out-of-hospital. Those with values of 3.7°C or greater on the spatial Humidex map had an OR of 1.15 [1.06, 1.25], compared with 1.09 [1.04, 1.15] for the entire analytic dataset. Similarly, those living in areas where the labour nonparticipation rate was ≥ 35% had an OR of 1.23 [1.12, 1.35], compared with 1.11 [1.04, 1.17] for the entire analytic dataset.

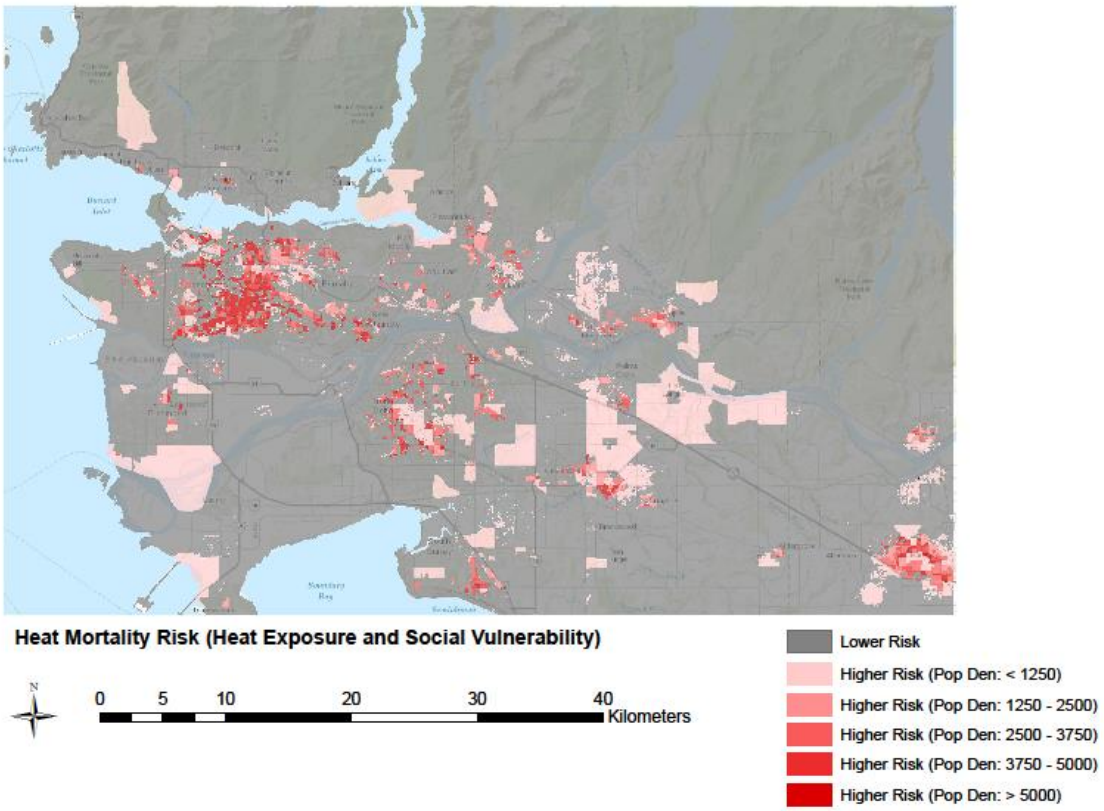
4.2.5. Combined heat exposure and social vulnerability index

One motivation for spatial delineation of the temperature-mortality relationship is to map areas of potential high risk for future environmental health assessment and emergency planning (Reid et al. 2009, Tomlinson et al. 2011). Based on results from the individual variables we wanted to evaluate whether the temperature-mortality relationship was further elevated in areas with combined risks from heat exposure and social vulnerability. However, maps of the areas above and below the Humidex and labour nonparticipation cut points (3.7°C and 60%, respectively) indicated little spatial overlap. The Humidex map showed that urban and suburban residential areas were at higher heat risk, while the labour nonparticipation map highlighted different areas of high deprivation, including the downtown eastside, one of the most socioeconomically deprived neighbourhoods in Canada (Kerr et al. 2003). Although extreme values of these two variables are spatially disparate for greater Vancouver, it could be an interactive effect between these two variables at lower spatial values to develop an elevated heat risk.

To better quantify this, we integrated the Humidex and labour nonparticipation maps, which normalized the spatial bias from exposure and vulnerability, and enhanced the visualization of higher risk areas for use in emergency planning (Figure 10). To build the composite index both variables were first reclassified to percentile maps using their high risk cut points and their minimum/maximum values from the analytic dataset (Brewer and Pickle 2002; Erhardt et al. 2012). Locations lower than the cut points were reclassified from 0% to 50% by taking percentiles from the minimum value to the cut point value such that each bin contained 0.5% of the data. The same method was applied to the locations above the cut points, where the variables were reclassified from 50% - 100%. This approach normalized the Humidex and labour nonparticipation variables using the cut points as the standard value at the 50th percentile. By taking the average of the reclassified maps, a composite index map was produced, and then evaluated using the same methods applied to all other variables. The results showed no statistical separation of ORs above and below any cut point, but the point of maximum separation of the confidence intervals was at 46%. However, when the dataset was restricted to out-of-hospital deaths only, there was clear separation of the confidence

intervals at 35% (Figure 10). At this value the OR for a 1°C increase in daily mean Humidex at YVR was 1.12 [1.07, 1.17] for data above the cut point, compared with 1.03 [1.00, 1.07] for data below the cut point.

Figure 10. Heat Risk Map from Heat Exposure and Social Vulnerability. Values of the index greater than 35% are classified as higher risk, and are shown here by increasing population density (per km^2), in order to display the relationship between higher risk and population distribution in the greater Vancouver study area.



4.2.6. Discussion and limitations

We have used data from extremely hot days to assess spatial variability in the temperature-mortality relationship by differences in heat exposure and social vulnerability in greater Vancouver, Canada. We found that maps of the Humidex and labour nonparticipation rates showed the strongest spatial differences in this study area, and we developed a combined index of these variables to further delineate areas of increased risk. The analytic approach used a time-stratified case-crossover design to examine differences between the ORs for cases with residential locations above and below a complete range of cut points for each spatial variable. Where complete separation of the ORs above and below the cut points was observed, the odds of mortality for those above the cut point was increased by 10-11% for each 1°C increase in daily mean Humidex at YVR, compared with 2-3% for those below the cut point.

These results are somewhat consistent with other, comparable studies (Gronland et al. 2015; Hondula et al. 2015; Smargiassi et al. 2009). For example, Hondula et al. (2015) used baseline mortality at 10°C and observed mortality on hot days to report a 0 – 10% increase in risk for areas with temperatures greater that 30°C in multiple US cities including Seattle, WA, which is located 200 km south of greater Vancouver on the Pacific coast. In comparison, locations between 20-30°C had increases from -2.5 to 2.5% (Hondula et al. 2015). In another spatially-stratified case-crossover analysis of social vulnerability, Gronlund et al. (2015) reported no significant differences in ORs for cardiovascular or respiratory mortality between all cases and those cases that occurred in areas of Michigan presumed to be vulnerable. The one exception was for respiratory mortality in areas with older housing (Gronlund et al. 2015). Finally, Smargiassi et al. (2009) reported a 6-7% increase in odds of non-accidental mortality for areas of Montreal, Canada where LST was used to identify areas more than two degrees hotter than the fixed reference site.

One strength of our study was its comparison between LST, air temperature, and Humidex for mapping differences in the temperature-mortality relationship, where others have used LST without further evaluating its relative suitability for this purpose. Although we observed a small increase in the OR when LST values were higher than those at YVR, it dropped rapidly and no separation of ORs was observed at any cut point. This

could be because the LST map over-emphasized the temperatures of impervious surfaces, especially when expressed relative to YVR where there is extensive impervious surface without any vegetative cover. Most urban areas had spatial LST values that were lower than those at YVR, yet we expect the time-varying temperatures to be lower at YVR than throughout most of the study region because of the strong coastal influence. This combination could suggest that LST was less appropriate for this particular study than for previous work, but further evaluation of LST for meaningful delineation of the temperature-mortality relationship is needed. In comparison, we found that our previously-developed maps of maximum air temperature and Humidex on hot days (Ho et al. 2014; Ho et al., 2016) showed clear increases in the temperature-mortality relationship as differences from YVR increased, with the stronger relationship observed for the Humidex map. Our methods indicated a statistically significant difference in risk for those living in areas ≥3.7°C warmer than YVR, which includes approximately 7% of the greater Vancouver population.

Another strength of this study was its data-driven comparison of several variables reflecting different social vulnerabilities with the potential to be associated with increased risk of mortality during extreme heat. Of the 11 variables evaluated, only four showed significantly increased risk in some areas: VANDIX; average income ratio; percent of homes rented; and the labour nonparticipation rate. The three latter variables all contribute to the modified composite VANDIX, along with four variables for which no significantly increased risks were found: percent with no high school; unemployment rate; percent with no university; and percent single parent families. While it is valuable to know that the VANDIX does delineate risk in the greater Vancouver area for which it was designed, it is also valuable to know that not all of the contributing variables can adequately capture the heat health risks associated with deprivation. Furthermore, no areas of significantly increased risk were observed for variables reflecting the density of the senior population, persons living alone, and older housing in greater Vancouver. These are important observations because some spatial indices of hot weather vulnerability use such variables based on a general association that may not, in fact, exist in specific areas to which they are applied (Reid 2009; Yardley 2011). Here we report that labour nonparticipation was the strongest indicator of social vulnerability for

greater Vancouver, but this finding may not hold in other cities with different geographic and demographic contexts.

The methods we have presented here use spatially stratified case-crossover analyses to statistically delineate areas of higher and lower mortality risk on extremely hot days. This approach is limited by its strict separation of areas that might be data contiguous by the values extracted by the spatial locations and, therefore, at similar risk. It would be valuable for future work to develop methods that account for such data clustering while maintaining adequate power for the analyses. We did experiment with binning data by temperature ranges that included 20%, 10%, 5%, and 2% of the analytic dataset, but found that the larger ranges did not produce clear spatially-stratified results and that the smaller ranges did not achieve statistical significance. Other approaches may include moving windows, or fuzzy boundaries that include all deaths within some tolerance of spatial values of the strictly delineated areas. Such advances may also help to improve maps of composite risk from combined heat exposure and social vulnerability, thereby improving environmental health planning for future events (Tomlinson et al. 2011).

Another limitation of our study was the definition of the analytic dataset using daily mean air temperature ≥24°C at YVR. Given our objective to assess the mortality risks on extremely hot days, we hypothesized that the inclusion of data from lower temperatures would serve to attenuate the spatially-stratified results we were hoping to highlight. To test the sensitivity to this decision we repeated all analyses using values of 20-26°C to create the analytic dataset. When results for the Humidex map and the nonparticipation rate were compared with those reported here, we found that similar spatially stratified results were evident at the ≥20°C restriction, but that statistical separation was not clear until the ≥24°C restriction. This separation persisted for the Humidex map at the ≥25°C and ≥26°C restrictions, despite the reduced power of the smaller datasets. However, for the purposes of heat risk mapping it might be informative to work with a larger analytic dataset restricted by some other variable. For example, our sensitivity analyses with the 27.7% subset of cases who died out-of-hospital appeared more sensitive to spatial differences than the entire subset defined by temperature. Given that extreme hot weather is associated with increased incidence of out-of-hospital

deaths (Kosatsky et al. 2012; O'Neill et al. 2003), we recommend that future work could focus on this subset of decedents to most effectively delineate areas of risk.

Chapter 5. Conclusion

Few studies have attempted to map temperature distributions in urban areas using remote sensing, and none have applied the resulting temperature map in a heat health study. My research first demonstrated the ability of a regression model, calibrated with multiple Landsat scenes and weather station data for hot summer days, to produce a sensible map of the daily maximum air temperature distribution in Greater Vancouver. Prediction errors (RMSE = 2.31 °C for air temperature) are comparable with those reported in previous studies, despite the complex landscape of the area. In addition, the mapped predictions of Tmax relative to a reference station (YVR) correspond well with previous descriptions of air temperature distribution in the area, and are explained well by local differences in land cover, surface temperature, elevation, and distance to the ocean. Specifically for downtown Vancouver, the low skyview factor and proximity to the ocean results in relatively low Tmax values compared with the surrounding suburban areas. The ability to map the spatial distribution of Tmax in a large metropolitan area contributes a spatial dimension to urban heat studies, which in the past have relied on a single reference station to approximate air temperature for large and often heterogeneous areas.

Secondly, my research demonstrated that weather station observations and spatial data layers derived from satellite imagery and a digital elevation model could also be combined in a Random Forest regression model to produce predictions of the daily maximum Humidex for unsampled areas (cross-validated RMSE = 2.04 °C). Predictions were normalized to values at Vancouver International Airport, and the model was applied to the spatial data layers to produce a map of relative daily maximum Humidex for a typical hot summer day in the greater Vancouver area, Canada. Comparable maps were produced for daily maximum air temperature and mid-morning land surface temperature. Land surface temperature is poorly correlated to the two other temperature measures (R^2 = 0.38 and 0.39 respectively). While air temperature and Humidex show greater

similarity (R^2 = 0.87), the Humidex exceeds the air temperature by more than 5 °C in several urban areas as well as around the confluence of the Pitt and Fraser rivers. Given that air temperature and Humidex have been shown to be more closely related to the human health impacts of extreme hot weather, satellite-derived land surface temperature is not an ideal proxy for human heat exposure, and spatial epidemiology studies can benefit from mapping Humidex using an approach similar to the one reported above, to better quantify differences in heat exposure that exist across an urban landscape.

Finally, my research has presented a novel application of the time-stratified case-crossover study design for spatial delineation of differences in the temperature-mortality relationship on extremely hot days. We have also evaluated the utility and appropriateness of different heat exposure maps and measures of social vulnerability. The results showed Humidex and labour nonparticipation were most strongly associated with spatial differences in the temperature-mortality relationship for greater Vancouver, while there was no association for most variables. A composite index showed some areas where combined heat exposure and social vulnerability led to increased risk. This highlights the need for data-driven mapping of the temperature-mortality relationship to best prepare individual cities for public health protection during extreme hot weather events.

References

Almeida, S. P., Casimiro, E., & Calheiros, J. (2010). Effects of apparent temperature on daily mortality in Lisbon and Oporto, Portugal. Environmental Health, 9(1), 12.

Anderson, A., Chapman, M., Drobot, S., Tadesse, A., Lambi, B., &Wiener, G. (2012). Quality of mobile air temperature and atmospheric pressure observations from the 2010 development test environment experiment. Journal of Applied Meteorology and Climatology, 51(4), 691–701.

Barnett, A. G., Tong, S., & Clements, A. C. A. (2010). What measure of temperature is the best predictor of mortality? Environmental research, 110(6), 604-611.

Barsi, J. A., Barker, J. L. & Schott, J. R. (2003). An Atmospheric Correction Parameter Calculator for a single thermal band earth-sensing instrument. Toulouse, France: IGARSS IEEE International Geoscience and Remote Sensing Symposium.

Basu, R. (2009). High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. Environ Health, 8(1), 40.

Bell, N., Schuurman, N., Oliver, L., Hayes, M. V. (2007). Towards the construction of place-specific measures of deprivation: a case study from the Vancouver metropolitan area. The Canadian Geographer/Le Géographe canadien 51(4): 444-461.

Benali, A., Carvalho, A., Nunes, J., Carvalhais, N. & Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. Remote Sensing of Environment, 124(2012), 108-121.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

Brewer, C. A., Pickle, L. (2002). Evaluation of methods for classifying epidemiological data on choropleth maps in series. Annals of the Association of American Geographers 92(4): 662-681.

Bristow, K. & Campbell, G. (1984). On the Relationship Between Incoming Solar Radiation and Daily Maximum and Minimum Temperature. Agricultural And Forest Meteorology, 31, 159-166.

Bruzzone, L., & Melgani, F. (2005). Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data. IEEE Transactions on Geoscience and Remote Sensing, 43(1), 159–174.

Bumbaco, K.A., Dello, K.D., Bond, N.A. (2013). History of Pacific Northwest heat waves: Synoptic pattern and trends. Journal of applied meteorology and climatology 52, 1618:1631.

Burke, M., Sipe, N., Evans, R. & Mellifont, D. (2006). Climate, geography and the propensity to walk: environmental factors and walking trip rates in Brisbane. Papers of the 29th Australasian Transport Research Forum, Gold Coast, 27-29 September 2006.

Buscail C., Upegui, E. & Viel, J-F. (2012). Mapping heatwave health risk at the community level for public health action, International Journal of Health Geographics, 11:38.

Camps-Valls, G., Bruzzone, L., Rojo-Alvarez, J. L., & Melgani, F. (2006a). Robust support vector regression for biophysical variable estimation from remotely sensed images. IEEE Geoscience and Remote Sensing Letters, 3(3), 339–343.

Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Vila-Francés, J., Amorós-López, J., & Calpe-Maravilla J. (2006b). Retrieval of oceanic chlorophyll concentration with relevance vector machines. Remote Sensing of Environment, 105(1), 23-33.

Chen, L, Ng, E., An, X., Ren, C., Lee, M., Wang, U. & He, Z. (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.

Coll, C., Galve, J.M., Sánchez, J.M., Caselles V. (2010). Validation of Landsat-7/ETM+ Thermal-Band Calibration and Atmospheric Corrections with Ground Based Measurements. IEEE Transactions on Geoscience and Remote Sensing, 48(1), 547-555.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Curriero, F. C., Heiner, K. S., Samet, J. M., Zeger, S. L., Strug, L., Patz, J. A. (2002). Temperature and mortality in 11 cities of the eastern United States. American journal of epidemiology 155(1): 80-87.

Czajkowski, K. P., Goward, S. N., Stadler, S., &Walz, A. (2000). Thermal remote sensing of near surface environmental variables: Application over the Oklahoma Mesonet. The Professional Geographer, 52(2),345–357.

Das, S., Samui, P., Sabat, A., & Sitharam, T. (2010). Prediction of swelling pressure of soil using artificial intelligence techniques. Environmental Earth Sciences, 61, 393–403.

Durbha, S., King, R., & Younan, N. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. Remote Sensing of Environment, 107(1–2), 348–361.

Erhardt, E. B., Nandram, B., Choi, J. W. (2012). Bayesian simultaneous intervals for small areas: an application to variation in maps. International Journal of Statistics and Probability 1(2): p229.

Fenner, D., Meier, F., Scheret, D., Polze. (2014). Spatial and temporal air temperature variability in Berlin, Germany, during the years 2001–2010. Urban Climate. DOI: 10.1016/j.uclim.2014.02.004

Filleul, L., Cassadou, S., Médina, S., Fabres, P., Lefranc, A., Eilstein, D., et al. (2006). The relation between temperature, ozone, and mortality in nine French cities during the heat wave of 2003. Environmental Health Perspectives 1344-1347.

Fu, P. & Rich, P. (2002). A Geometric Solar Radiation Model with Applications in Agriculture and Forestry. Computers and Electronics in Agriculture, 37, 25–35.

Gail, M. H., Santner, T. J., & Brown, C. C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. Biometrics 255-266.

Gao, B.-C. (1996). NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58, 257 – 266.

Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. (2010). Variable Selection using Random Forests. Pattern Recognition Letters, 31(14), 2225-2236.

Gislason, P., Benediktsson, J. & Sveinsson J. (2006). Random Forests for Land Cover Classification. Pattern Recognition Letters 27, 294-300.

Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211–221.

Gosling, S. N., Bryce, E. K., Dixon, P. G., Gabriel, K. M., Gosling, E. Y., Hanes, J. M., Hondula, D.M., Liang, L., Mac Lean, P. A., Muthers, S., Nascimento, S.T., Petralli, Martina, Vanos, J.K. & Wanka, E. R. (2014). A glossary for biometeorology. International journal of biometeorology, 58(2), 277-308.

Greene, S., Kalkstein, L. S., Mills, D. M., Samenow, J. (2011). An examination of climate change on extreme heat events and climate-mortality relationships in large US cities. Weather, Climate, and Society 3(4): 281-292.

Gronlund, C. J., Berrocal, V. J., White-Newsome, J. L., Conlon, K. C., O'Neill, M. S. (2015). Vulnerability to extreme heat by socio-demographic characteristics and area green space among the elderly in Michigan, 1990–2007. Environmental research 136: 449-461.

Hart, M. A., Sailor, D. J. (2009). Quantifying the influence of land-use and surface characteristics on spatial variability in the urban heat island. Theoretical and Applied Climatology 95(3-4): 397-406.

Hajat, S., Sheridan, S. C., Allen, M. J., Pascal, M., Laaidi, K., Yagouti, A., et al. (2010). Heat-health warning systems: a comparison of the predictive capacity of different approaches to identifying dangerously hot days. American journal of public health 100(6): 1137.

Hattis, D., Ogneva-Himmelberger, Y., & Ratick, S. (2012). The spatial variability of heat-related mortality in Massachusetts. Applied Geography, 33, 45-52.

Hawkins, T. W., Brazel, A. J., Stefanov, W. L., Bigler, W., Saffell, E. M. (2004). The role of rural variability in urban heat island determination for Phoenix, Arizona. Journal of Applied Meteorology 43(3): 476-486.

Henderson, S. B., Kosatsky, T. (2012). A data-driven approach to setting trigger temperatures for heat health emergencies. Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique 227-230.

Henderson, S. B., Wan, V., Kosatsky, T. (2013). Differences in heat-related mortality across four ecological regions with diverse urban, rural, and remote populations in British Columbia, Canada. Health & place 23: 48-53.

Hijmans, R. J. & van Etten, J. (2009). Raster. Geographic analysis and modeling with raster data: R package version 0.9.8-40/r681. See http://www.R-Forge.R-project.org/projects/raster/.fic essay, and social science essay)

Ho, H. C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., Henderson, S. B. (2014). Mapping maximum urban air temperature on hot summer days. Remote Sensing of Environment 154: 38-45.

Ho H.C., Knudby, A., Xu Y., Hodul, M., Aminipouri, M. (2016). A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area. Science of the Total Environment, 544, 929–938.

Hondula, D., Davis, R., Leisten, M., Saha, M., Veazey, L., Wegner C. (2012). Fine-scale spatial variability of heat-related mortality in Philadelphia County, USA, from 1983-2008: a case-series analysis. Environmental Health 11:16

Hondula, D. M., Davis, R. E., Saha, M. V., Wegner, C. R., Veazey, L. M. (2015). Geographic dimensions of heat-related mortality in seven US cities. Environmental research, 138: 439-452.

Höppe, P., Martinac, I. (1998). Indoor climate and air quality. International journal of biometeorology, 42(1): 1-7.

Huang, C., Davis, L., & Townshend, J. (2002). An assessment of support vector machines for land cover classification. International Journal of Remote Sensing, 23(4), 725–749.

Huang, W., Kan, H., & Kovats, S. (2010). The impact of the 2003 heat wave on mortality in Shanghai, China. Science of the total environment, 408(11), 2418-2420.

Hulley G.C. & Hook, S.J. (2009). The North American ASTER Land Surface Emissivity Database (NAALSED) version 2.0. Remote Sensing of Environment, 113(9), 1967-1975.

INVS (2004). Étude des facteurs de risqué de décès des personnes âgées résidant à domicile durant la vague de chaleur d'août 2003. Institut de Veille Sanitaire. Available from http://www.invs.sante.fr/publications/2004/chaleur2003_170904/rapport_canicule.pdf

Johnson, D. P., Stanforth, A., Lulla, V., Luber, G. (2012). Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. Applied Geography 35(1): 23-31.

Johnson, D. P., Webber, J. J., Urs Beerval Ravichandra, K., Lulla, V., & Stanforth, A. C. (2014). Spatiotemporal variations in heat-related health risk in three Midwestern US cities between 1990 and 2010. Geocarto International 29(1): 65-84.

Kaiser, R., Rubin, C. H., Henderson, A. K., Wolfe, M. I., Kieszak, S., Parrott, C. L., & Adcock, M. (2001). Heat-related death and mental illness during the 1999 Cincinnati heat wave. The American journal of forensic medicine and pathology, 22(3), 303-307.

Kerr, T., Wood, E., Small, D., Palepu, A., Tyndall, M. W. (2003). Potential use of safer injecting facilities among injection drug users in Vancouver's Downtown Eastside. Canadian Medical Association Journal 169(8): 759-763.

King, M. D., Menzel, W. P., Kaufman, Y. J., Tanré, D., Gao, B. C., Platnick, S., ... & Hubanks, P. A. (2003). Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS. Geoscience and Remote Sensing, IEEE Transactions on, 41(2), 442-458.

Knudby, A., Brenning, A., & LeDrew, E. (2010a). New approaches to modelling fish-habitat relationships. Ecological Modelling, 221, 503–511.

Knudby, A., LeDrew, E. & Brenning, A. (2010b). Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. Remote Sensing of Environment, 114(6), 1230–1241.

Kosatsky, T., Henderson, S. B., & Pollock, S. L. (2012). Shifts in Mortality During a Hot Weather Event in Vancouver, British Columbia: Rapid Assessment With Case-Only Analysis. American journal of public health, 102(12), 2367-2371.

Kutner, M., Nachtsheim, C., & Neter, J. (2004). Applied linear regression models (4th ed.). McGraw Hill.

Laaidi, K, Zeghnoun, A., Dousset, B., Bretin, P., Vandentorren, S., Giraudet, E. & Beaudeau P. (2012). The Impact of Heat Islands on Mortality in Paris during the August 2003 Heat Wave. Environmental Health Perspectives, 120(2), 254-259.

Liaw, A., & Wiener, M. (2012). Breiman and Cutler's Random Forests for classification and regression, R package.

Liu, J. G., & Moore, J. M. (1998). Pixel block intensity modulation: adding spatial detail to TM band 6 thermal imagery. International Journal of Remote Sensing, 19, 2477–2491.

Massetti, L., Petralli, M., Brandani, G., & Orlandini, S. (2014). An approach to evaluate the intra-urban thermal variability in summer using an urban indicator. Environmental Pollution, 192, 259-265.

Masterson, J & Richardson, F.A. (1979). Humidex. A method of quantifying human discomfort due to excessive heat and humidity. Environment Canada, Downsview

Meehi, G.A., Tebaldi C. (2004). More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. Science 305: 994-997.

Meyer D., Dimitriadou, E., Hornik K., Weingessel, A. & Leisch, F. (2012). Misc Functions of the Department of Statistics (e1071), TU Wien, R package.

Morabito, M., Crisci, A., Gioli, B., Gualtieri, G., Toscano, P., Di Stefano, V. et al. (2015). Urban-Hazard Risk Analysis: Mapping of Heat-Related Risks in the Elderly in Major Italian Cities. PLOS one e0127277.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(2011), 247–259.

Murray, K., & Conner, M. M. (2009). Methods to quantify variable importance: Implications for the analysis of noisy ecological data. Ecology, 90, 348–355.

Nguyen, J. L., Schwartz, J., & Dockery, D. W. (2014). The relationship between indoor and outdoor temperature, apparent temperature, relative humidity, and absolute humidity. Indoor air, 24(1), 103-112.

Ocak, I., & Seker, S. (2013). Calculation of surface settlements caused by EPBM tunneling using artificial neural network, SVM, and Gaussian processes. Environmental Earth Sciences, http://dx.doi.org/10.1007/s12665-012-2214-x.

Oke, T. (1976). The distinction between canopy and boundary -layer urban heat islands. Atmosphere, 14(4), 268-277.

Oke, T. (1988). The urban energy balance. Progress in Physical Geography, 12, 471-508.

Oke, T. & Hay, J. (1994). The Climate of Vancouver (2nd Edition). B.C. Geographical Series, 50. Vancouver, British Columbia.

O'Neill, M. S., Carter, R., Kish, J. K., Gronlund, C. J., White-Newsome, J. L., Manarolla, X., et al. (2009). Preventing heat-related morbidity and mortality: new approaches in a changing climate. Maturitas 64(2): 98-103.

O'Neill, M.S., Zanobetti, A., Schwartz, J. (2003). Modifiers of the temperature and mortality associations in seven US cities. Am J Epidemiol. 157: 1074–1082.

Oommen, T., Misra, D., Twarakavi, N., Prakash, A., Sahoo, B., & Bandopadhyay, S. (2008). An objective analysis of support vector machine based classification for remote sensing. Mathematical Geosciences, 40(4), 409–424.

Pal, M., & Mather, P.M. (2005). Support vector machines for classification in remote sensing. International Journal of Remote Sensing, 26(5), 1007–1011.

Pampalon, R., Hamel, D., Gamache, P., Philibert, M. D., Raymond, G., Simpson, A. (2012). An area-based material and social deprivation index for public health in Québec and Canada. Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique S17-S22.

Pampalon, R., Hamel, D., Gamache, P., Raymond, G. (2009). A deprivation index for health planning in Canada. Chronic Dis Can 29(4): 178-91.

Pampalon, R., Raymond, G. (2000). A deprivation index for health and welfare planning in Quebec. Chronic Dis Can 21(3): 104-113.

Pengelly, I.D., Campbell, M.E., Cheng, C.S., Fu, C., Gingrich, S.E. & Macfarlane, R. (2007). Anatomy of heat waves and mortality in Toronto. Revue Canadienne de santé publique, 98(5), 364-368.

Prihodko, L. & Goward, S. N. (1997). Estimation of air temperature from remotely sensed surface observations. Remote Sensing of Environment, 60(3), 335–346.

R Core Development Team. (2008). R: A language and environment for statistical computing. http://www.R-project.org/

Reid, C.E., O'Neill, M.S., Gronlund, C.J., Brines, S.J., Brown, D.G., Diez-Roux, A.V., et al. (2009). Mapping Community Determinants of Heat Vulnerability. Environmental Health Perspectives 117(11): 1730-1736.

Roth, M., Oke, T. R., Emery, W. J. (1989). Satellite-derived urban heat islands from three coastal cities and the utilization of such data in urban climatology. International Journal of Remote Sensing 10(11): 1699-1720.

Runnalls, K & Oke, T. (2000). Dynamics and Controls of the Near-surface heat island of Vancouver, British Columbia. Physical Geography, 21(4), 283-304.

Schaeffer, L., de Crouy-Chanel, P., Wagner, V., Desplat, J., Pascal, M. (2015). How to estimate exposure when studying the temperature-mortality relationship? A case study of the Paris area. International journal of biometeorology, 1-11.

Schifano, P., Cappai, G., De Sario, M., Michelozzi, P., Marino, C., Bargagli, A. M., et al. (2009). Susceptibility to heat wave-related mortality: a follow-up study of a cohort of elderly in Rome. Environ Health 8(50): 1-14.

Semenza, J.C., Rubin, C.H., Falter, K.H., Selanikio, J.D., Flanders, W.D., Howe, H.L. &Wilhelm, J.L. (1996). Heat-related deaths during the July 1995 heat wave in Chicago. New England Journal of Medicine, 335, 84–90.

Sieber, R. (2006). Public Participation Geographic Information Systems: A Literature Review and Framework. Annals of the Association of American Geographers, 96(3), 491-507.

Smargiassi, A., Goldberg, M.S., Plante, C., Fournier, M., Baudouin, Y. & Kosatsky, T. (2009). Variation of daily warm season mortality as a function of micro-urban heat islands. Journal of epidemiology and community health, 63(8), 659-664.

Statistics Canada. (2007). Greater Vancouver, British Columbia (Code5915) (table). 2006 Community Profiles. 2006 Census. Statistics Canada Catalogue no. 92-591-XWE. Ottawa. Released March 13, 2007.

Steadman, R.G. (1979). The assessment of sultriness. Part I. A temperature-humidity index based on human physiology and clothing science. Journal of Applied Meteorology, 18, 861-873.

Stewart, I.D., Oke, T.R. (2012). Local climate zones for urban temperature studies.. Bulletin of the American Meteorological Society, 93, 1879-1900.

Strobl, C., Malley, J. & Tutz, G. (2009). An Introduction to Recursive Partitioning: Rational, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. Psychological Methods, 14(4), 323-348.

Stumpf, A. & Kerle, N. (2011). Object-oriented mapping of landslides using Random Forests. Remote Sensing of Environment. 115, 2564-2577.

Su, J., Brauer, M. & Buzzelli, M. (2008). Estimating urban morphometry at the neighborhood scale for improvement in modeling long-term average air pollution concentrations. Atmospheric Environment, 42, 7884-7893.

Sun, D., Li, Y., & Wang, Q. (2009). A unified model for remotely estimating chlorophyll a in Lake Taihu, China, based on SVM and in situ hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 47(8), 2957–2965.

Taylor, C. G. (1998). A REVIEW OF THE BC VEGETATION RESOURCES INVENTORY SAMPLING DESIGN AND THE PROPOSED ESTIMATORS (Doctoral dissertation, Simon Fraser University).

Timm, B. & McGraigal, K. (2012). Fine-scale remotely-sensed cover mapping of coastal dune and salt marsh ecosystems at Cape Cod National Seashore using Random Forests. Remote Sensing of Environment, 127, 106-117.

Therneau TM (2014). survival: A Package for Survival Analysis in S. R package version 2.37-7, URL http://CRAN.R-project.org/package=survival.

Tomlinson, C., Chapman, L., Thornes, J. & Baker C. (2011). Including the urban heat island in spatial heat health risk assessment strategies: a case study for Birmingham, UK. International Journal of Health Geographic, 10:42.

Trenberth, K. E., & Fasullo, J. T. (2012). Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010. Journal of Geophysical Research: Atmospheres (1984–2012), 117(D17).

Uejio, C. K., Wilhelmi, O. V., Golden, J. S., Mills, D. M., Gulino, S. P., Samenow, J. P. (2011). Intra-urban societal vulnerability to extreme heat: the role of heat exposure and the built environment, socioeconomics, and neighborhood stability. Health & Place 17(2): 498-507.

Vancutsem, C., Ceccato, P., Dinku, T. & Connor, S. J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sensing of Environment, 114(2), 449–465.

Walker, B. B., Schuurman, N., Hameed, S. M. (2014). A GIS-based spatiotemporal analysis of violent trauma hotspots in Vancouver, Canada: identification, contextualisation and intervention. BMJ open 4(2): e003642.

Weng, Q., Lu, D. & Schubring, J. (2004). Estimation of land surface temperature - vegetation abundance relationship for urban heat island studies. Remote Sensing of Environment, 89, 467-483.

Willmott, C. J. & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30, 79-82.

Wong, M. S., Jin, X., Liu, Z., Nichol, J., & Chan, P. W. (2015). Multi-sensors study of precipitable water vapour over mainland China. International Journal of Climatology, 35(10), 3146-3159.

Wong, M., Nichol, J. & Ng, E. (2011). A study ofthe “wall effect” caused by proliferation of high-rise buildings using GIS techniques. Landscape and Urban Planning, 102, 245-253.

World Meteorological Organization (1996). Guide to meteorological instruments and methods of observation WMO-No. 8 (6th ed.) (Geneva).

Xu, Y., Knudby, A., & Ho, H. C. (2014). Estimating daily maximum air temperature from MODIS in British Columbia, Canada. International Journal of Remote Sensing, 35(24), 8108-8121.

Yardley, J., Sigal, R. J., Kenny, G. P. (2011). Heat health planning: The importance of social and community factors. Global Environmental Change 21(2): 670-679.

Zakšek, K. & Oštir, K. (2012). Downscaling land surface temperature for urban heat island diurnal cycle analysis. Remote Sensing of Environment, 117, 114–124.

Zakšek, K., & Schroedter-Homscheidt, M. (2009). Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. ISPRS Journal of Photogrammetry and Remote Sensing, 64(4), 414–421.

Zhang, K., Li, Y., Schwartz, J. and O'Neill, M. (2014). What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods. Environmental Research, 132, 350-359.

Zhang, P. (1993). Model selection via multifold cross validation. The Annals of Statistics, 21(1), 299-313.

Zhu, W., Lu, A., & Jia, S. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sensing of Environment, 130(2013), 62–73.