VAH: A Compartmental Modeling Approach for Forecasting Urban Heat Vulnerability

1. Research Question

Can we create a predictive tool that identifies not just *when* a heatwave is dangerous, but *who* is most at risk at the neighborhood level?

2. Background

Extreme heat is a growing public health threat in Canada. While meteorological forecasts can predict severity of heatwaves, they do not capture heterogeneity of risk within a city.

- Senior citizens in poorly ventilated and high-rise apartments face fundamentally different level of danger than a young family in a leafy suburb

Gap: the need for a dynamic system to guide targeted public health interventions in Canada. This thesis proposes the development of a kind of early warning system for *vulnerability*, giving officials intelligence to allocate resources appropriately (like cooling centers, public messaging, etc) to the specific populations and neighborhoods that need them most.

3. Hypothesis

A compartmental model, adapted from the Susceptible-Infected-Recovered (SIR) models used in epidemiology, can plausibly simulate the flow of an urban population through states of heat-related distress. **Vulnerable-Affected-Hospitalized (VAH) model**.

4. Materials & Methods

4.1. Framework:

The core of the thesis will be the development of the VAH model, a system of ordinary differential equations (ODEs). This framework will model the daily movement of people between three states

- Vulnerable (at risk population)
- Affected (suffering acute heat stress),
- Hospitalized (requiring emergency care or died).

4.2. Data Sources:

The model will be developed using publicly available data for major Canadian cities (e.g., Toronto, Vancouver).

- **For Geospatial Data:** Census Tract (CT) boundary files from Statistics Canada will define our neighborhoods.
- **For Demographic Data:** CT-level census profiles from Statistics Canada, including median income, age distribution, housing characteristics (e.g., % in high-rise apartments), education levels, and ethnicity.

- **Environmental Data:** Historical daily temperature data from Environment Canada weather stations.
- **Health Outcome Data:** Publicly available (or requested) daily data on all-cause mortality or heat-related hospitalizations (primary outcome for model)

4.3. Plan:

1. **First,** establish a local relationship between daily temperature and health outcomes.
2. **Second,** pool the neighborhood-specific parameters derived from Stage 1.
    a. Use a regression model to explain the variation in these parameters based on the socio-demographic characteristics of each neighborhood.

4.4. Implementation:

The model and all associated analyses will be implemented in Python, using standard scientific computing libraries such as

- pandas (data manipulation)
- geopandas (spatial analysis)
- scipy.integrate.solve_ivp (ODE solver)
- matplotlib (visualization).

The final project is expected to be an open source tool repository on Github, available and made public for accessibility and usability (app or otherwise)

VAH: A Heat Vulnerability Warning System for City Neighborhoods

Research Question: My project will build a tool to predict which neighborhoods are most vulnerable to heatwaves.

Background: Current heat warnings are city-wide and miss neighborhood-level differences in risk.

Stage 1 (VAH Model): I will build a VAH compartmental model where the heat exposure rate is driven by daily temperature.

Stage 2 (Stats): "I will use statistical methods to discover how neighborhood features like income and housing type determine the model's risk parameters."

Data: This model will use public census, weather, and hospitalization data.

Implementation: The tool will be built in Python and delivered as an open-source repository.

Citations:

- Eric's 4 papers
- Chapter 12 and Chapter 4 of Stacey's Book
- EQUALSS Paper

Heat is killing more Canadians. Two-thirds of all heat deaths in the last 40 years happened in the last 7. Hospital trips for heat illness jumped 367% in 2021 alone (Government of Canada, 2025). This problem is worse for seniors - people over 80 have the highest reported death rates. (Mention the current method, then talk about how it fails, with a specific example - We use city-wide interventions like so so , very specific, but it fails to so so so, then we follow up with the next sentence as illustration which is this → A city-wide heat alert for Ottawa does not help an 85 year-old widow on the 15th floor of a brick apartment in Centretown. Today's science only studies these deaths after they happen (We need to mention the current standard, being very specific). We need a tool that predicts risk before people die. (be more specific)

My thesis is to adapt a proven tool in mathematical modelling of disease, the SIR model. (we need to be specific about what the SIR model is and explain it with an example, so explain it in detail first, then give a concrete, canonical, non hypothetical example to showcase it) and adapt it to predict heat risk. My model is called VAH: Vulnerable, Affected, Hospitalized. Since SIR was a gold standard, trusted idea from disease modelling (how is it trusted? Is there tangible evidence? , we hypothesize it will work for solving a new climate problem. (how is the adaptation made? Explain that, so we adapted S → V, I → A, and R → H (and saying how we changed recovered to hospitalizwed to make it fit better for example) .

Mention that there were efforts to actually adapt the model before by other people, which validates our approach (We are not the firs tto do this). we argue that the Susceptible-Infected-Quarantined-Recovered (SIQR) model is more appropriate than the commonly employed SIR model given the data collected (number of active cases). → for COVID-19 disease modelling. (paper from title: Revisiting classical SIR modelling in light of the COVID-19 pandemic)

Both parts of the thesis use free, public data. (did we talk about the data? Specifically? First, I will build the VAH math model for Ottawa. It will predict how many people in each neighbourhood get sick or go to the hospital each day of a heatwave. Daily temperature data from Environment Canada and local hospital capacity numbers (what does this mean? Make it layman, the hospital capacity thing) will feed the model (add the complexity here, so beta? Gamma? Things like that, but also explain them. Like how will it feed the model say). A lot of complexity can go here, because we want to show concrete method already established which is the SIR, so elaborate).

Second, I use a computer program (Python. I want that level of specificity. Do not use abstract language like computer program. Say Python and specifically what in python, what libraries?) to find reasons why some neighbourhoods (such as?) face higher risk of extreme heat deaths than others (is it deaths? Not sure. Specific. Can we a statistic here). Looking at census data - income (specific, mention age brackets), age (specific), housing type (choices between houses, so is it apartment? Single house?, language (be specific, so french, english),

and find the mix? (i need a sophisticated term here, mix is not good) for high risk (of what?). Is it the combination of low income and old brick buildings, or is it (etc etc).

The final result is a heat risk map of Ottawa. Before a heatwave, Ottawa public health emergency preparedness can look at the map, and see that Vanier is colored red, while Kanata is colored green, which means they can stop just telling everyone to drink water (too informal? I feel like we cna balance the seriousness of this project with also the inefficiency of city-wide warnings), instead, we can send a cooling bus to one specific apartment building in Vanier (Vanier North, Vanier South, etc). (Can we also mention another specific, tangible intervention for extreme heat? What are they). They can check ont he seniors living there by name. My work gives Ottawa public health the tool to do this. (Can we estimate

- More statistics overall (specific statistics such as the ones already talked about).
- We need to also estimate implications - so given this many deaths or this many affected, we estimate this would save etc etc etc (at the very end). And knowing how to give good estimate sort of thing. It is very good to at least estimate

Canada is warming at twice the global rate, and extreme heat events are becoming deadlier each year (Government of Canada, 2025).
BIAS ON THE GAP

OGS draft

motivation, creativity, significance, communication skills)

Academic Record (weight 40%):
- Transcripts
- Other academic accomplishments
- Prior scholarships and awards
- Publications and presentations

Don't make reviewer think

When talking to Stacey, we need the letter to highlight leadership

—-

References

https://www.canada.ca/en/health-canada/services/publications/healthy-living/heat-related-morbidity-mortality.html

https://link.springer.com/article/10.17269/s41997-023-00782-5 ↠ # Future temperature-related excess mortality under climate change and population aging scenarios in Canada

Kalachev, L., Landguth, E. L., & Graham, J. (2023). Revisiting classical SIR modelling in light of the COVID-19 pandemic. ***Infectious Disease Modelling*****, *8*(1), 72–83.**
https://doi.org/10.1016/j.idm.2022.12.002

—

Final draft

New Thesis Direction

Canada is warming at twice the global rate and extreme heat events are becoming deadlier [1]. Current research seems to focus on why certain neighbourhoods are more vulnerable to heatwaves, as opposed to asking which neighbourhoods will have the most hospitalizations. This thesis aims to address this gap by developing a machine learning model that forecasts which Ottawa neighbourhoods are at highest risk of heat-related hospitalizations during a heatwave. The model will be trained on census tract data between 2010-2020, and tested for predictive accuracy against real-world outcomes from past heatwaves that the model has never seen. The model's output will then be compared to validated data (actual hospitalizations) to assess predictive power. The final result is envisioned as a tool to provide researchers ability to forecast which Ottawa neighbourhoods are facing the highest risk of heat-related hospitalizations during a given heat wave, and therefore ability to deploy targeted interventions.

The how is a grandiose waste of time. We care about the what (using statistics to see final answer, the how will always be of endless debate).
]

(Data will be split into training data and test data)

Rationale: Current research is focused on the

Answer: use data from 2010-2022 to forecast which neighborhoods are going to be most affected by heat using combination of factors to predict.

- The heat health association → peop le getting sick from heat

- Area level factors, science is looking only at singular factors, not multiple (effect measure modification
- Neighborhoorhood → census tract
- Handing in the end: a model that predicts → streamlit. Like basically it actually happened, and I forecasted correctly. How can i make my forecast super easy for myself.
- Goal: predict heat-related hospital admissions per 10,000 residents at the census tract level during a declared heatwave, using massive old data to forecast. RMSE. Training model from 2010-2020.

State problem, 1-2 sentences max.

State gap in research, 102 sentences.

State the thesis -2 sentences (this thesis will aim) - The What

How the thesis - The how

The final output of the thesis in concrete language as if it is here now - The final result + implications