

SPATIAL MODELS FOR CANADIAN TEMPERATURE INFERENCE

BY BRYAN ZANG^{1,a}

¹Department of Statistics, University of Waterloo, ^abszang@uwaterloo.ca

Climate change has been a hot topic over the past few decades and remains to be so; an important factor to measure then is temperature. We will show that under some naive assumptions, the global additive model provides the most interpretable results in understanding the spatial variance of Canadian temperature change over some defined regions.

1. Introduction.

With the progression in science and technology, it has been evident over the past decades that human activity has caused significant amounts of greenhouse gas (GHG) emissions. These GHG emissions trap radiation and heat from the sun, leading to global warming and climate change. To measure the rate and effect of these GHG emissions, we can use temperature as it is a key factor influencing various parts of the regional and global ecosystem ¹. We wish to model the temperature change in Canada hoping to identify possible variances in the regional data and if so, why? This will be done by fitting a selected model to the regional data and assessing the fit of the model.

2. Data.

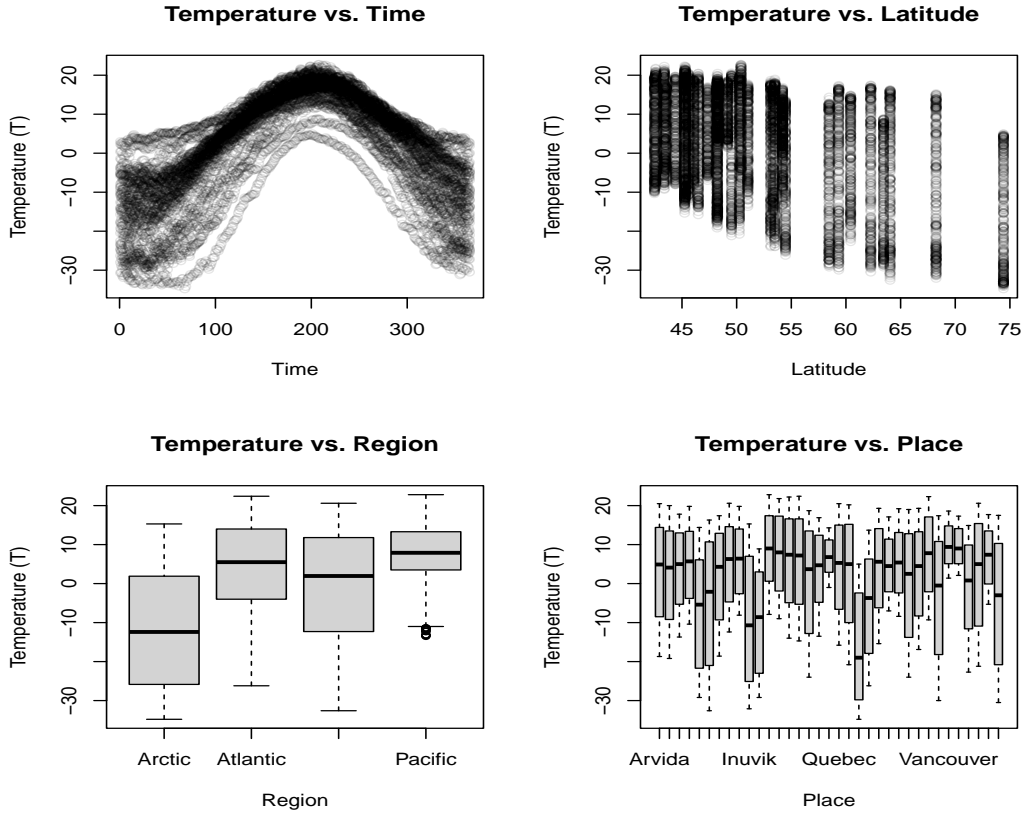
Our data comes from the R package `gamair` ², it is the `CanWeather` dataset which is a dataset describing Canadian temperature (averaged over the years 1960-1994) throughout a year at 35 different locations. There is a column of integers denoting the day in the year (1-365), a numerical column for the mean temperature for that day in Celsius, a column of categorical data denoting the general location (Arctic, Atlantic, Continental, or Pacific), a numerical column denoting the location latitude, and another categorical column of location names. We conducted some brief exploratory data analysis to better understand the data and temperature arises as the response variable of interest, and covariates such as time and region appear to be interesting in the sense that we are able to partition data into portions local to some variable (i.e., the temperature change in x days or the overall temperature model in the Atlantic region). Conducting a summary of the data we have the table below

time	T	region	latitude
Min. : 1	Min. :-34.800	Arctic :1095	Min. :42.48
1st Qu.: 92	1st Qu.: -6.700	Atlantic :5475	1st Qu.:45.58
Median :183	Median : 4.100	Continental:4380	Median :49.53
Mean :183	Mean : 1.878	Pacific :1825	Mean :51.84
3rd Qu.:274	3rd Qu.: 12.600	NA	3rd Qu.:54.47
Max. :365	Max. : 22.800	NA	Max. :74.41

¹[Environment and Climate Change Canada \(2023\)](#)

²[Ramsay and Silverman \(2006\)](#)

and the following plots



3. Methods.

To simplify the analysis we implemented two types of approach: the first being a localized model and the second being a global model. By inspection, it can be seen that the data follows a clear bell-shaped distribution and so a direct observation was made to see a polynomial regression may have been suitable. Initial models indicated that a global polynomial fit was not suitable and un-interpretable. To account for this issue, we fit a polynomial model to each region hence mitigating the fitting problem:

```
region$T ~ Poly(region$time, deg=6)
```

Polynomials of degree 6 were used to fit the models since they are the lowest degree polynomials that provided sufficiently well enough fits for the data — i.e., polynomials of lower degree did not fit the data well and polynomials of higher degree did not provide significant improvement in model fit for the corresponding computational costs.

With the additive models, we also fit in a similar manner of localizing a model to each region; but since the additive model can indeed incorporate regional indicators (specifically latitude) without losing significant goodness of fit, we also fit a global additive model with two covariates. This gives

```
region$T ~ Spline(region$time)
```

for the local spline regression model, and

```
T ~ Spline(time, deg=3) + Spline(latitude, deg=3)
```

for the global additive model. We use `mgcv`'s automatic selection feature to obtain our λ values and knot counts for the additive models and it is important to note in the global additive model we use B-splines for smoothing `time` and cubic regression splines to smooth `latitude` for the purposes of a better fit.

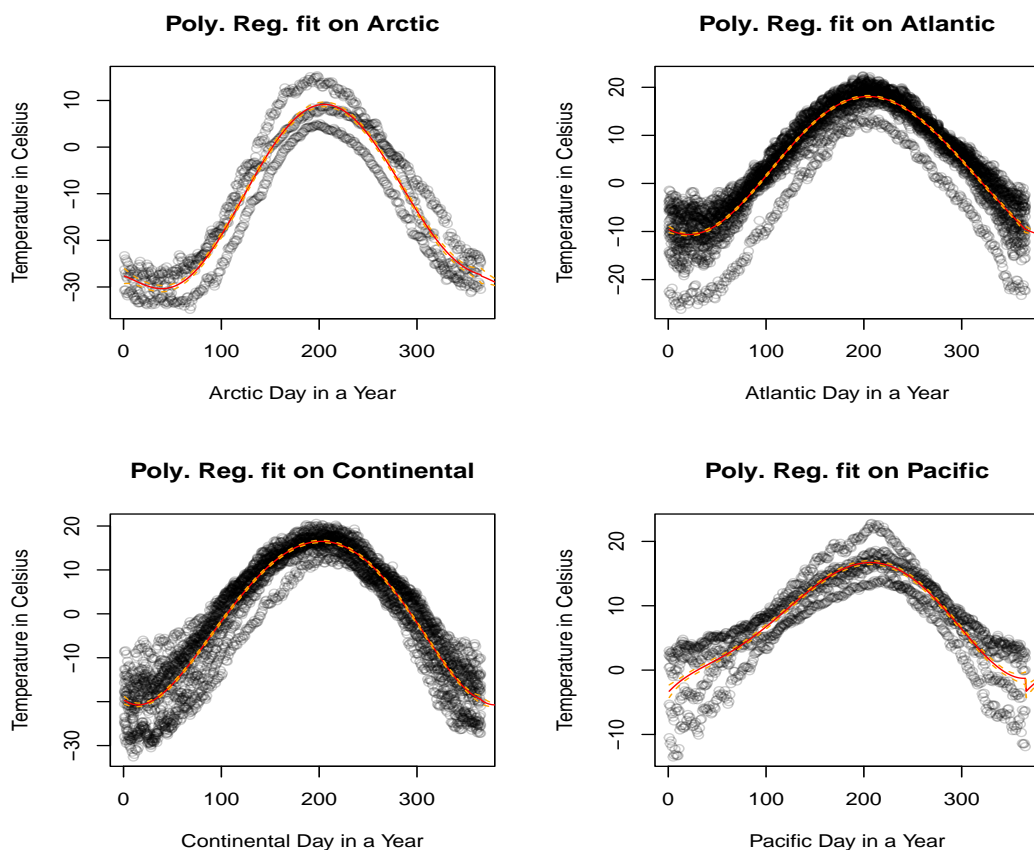
In determining the goodness of the model fits, we will use the GCV estimates (expected generalization errors) and the adjusted R^2 values as these two statistics both are indicators for how well the model explains the data. The mean adjusted R^2 is used for the global additive model and the mean non-adjusted R^2 is used for all other models, since there is only one covariate in each of the localized models (adjusted and non-adjusted R^2 values are approximately equal) and we wish to be able to interpret the local R^2 values relative to the global model. As for the GCV estimates, `mgcv` provides GCV estimates for all the additive models, so we only have to compute the GCV estimate for the polynomial regression models; to be able to interpret these GCV estimates with respect to the global additive model results, we compute the GCV estimate by assuming the true temperature value to be the average temperature over each place in a region, that is

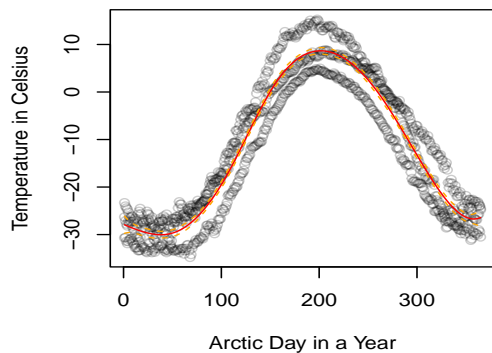
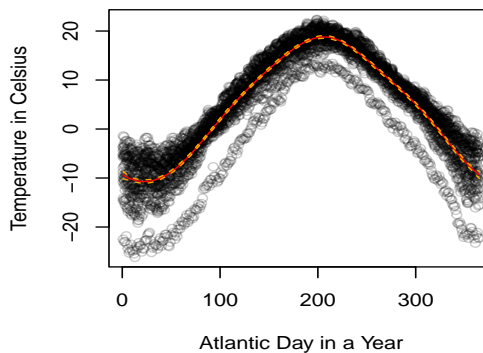
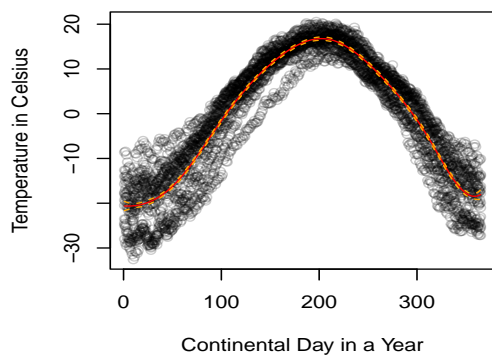
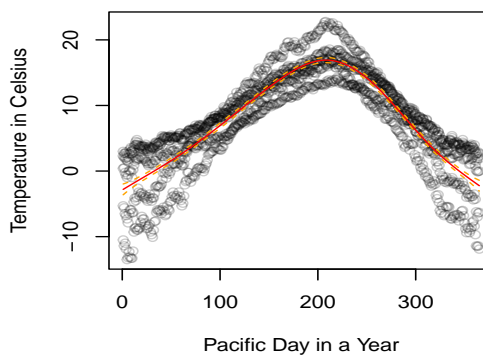
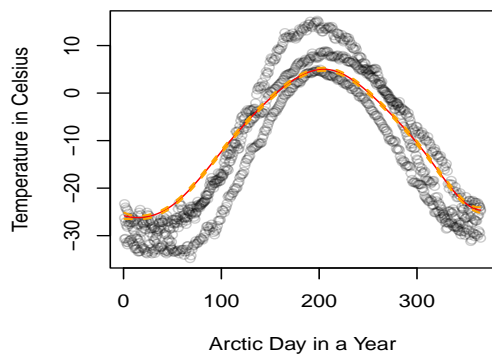
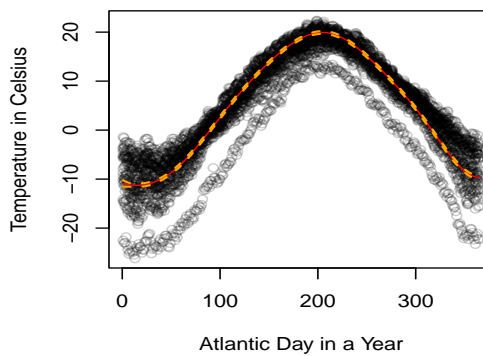
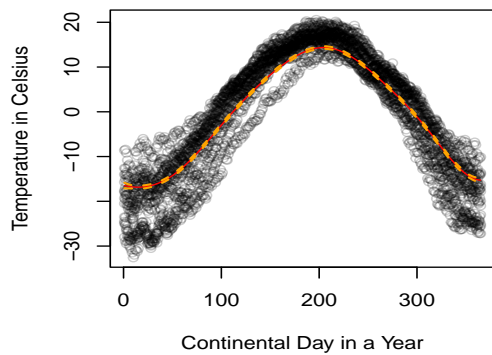
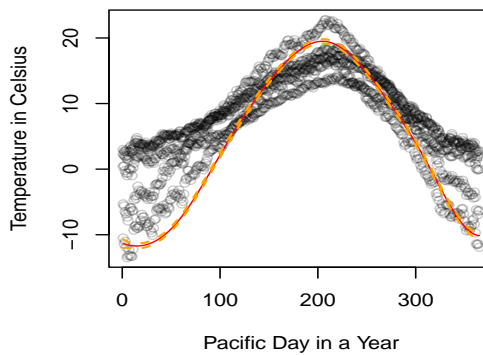
$$\text{true } T \text{ at region } r = \frac{1}{\# \text{ of places in } r} \sum_{\text{place in } r} (T \text{ at place})$$

hence we will compute the fitted values in a similar manner to match the dimension of the true temperature value vector.

4. Results.

Fitting our models to the data we get the following plots where the fitted line is in red and the confidence bands are in orange.



Spline Reg. fit on Arctic**Spline Reg. fit on Atlantic****Spline Reg. fit on Continental****Spline Reg. fit on Pacific****GAM: Temp vs. Arctic****GAM: Temp vs. Atlantic****GAM: Temp vs. Continental****GAM: Temp vs. Pacific**

Notice the polynomial models when plotted out have these "jumps" nearing the end of the year — the reason behind this requires a deeper understanding of the R functions used to fit and plot the models, however we suspect the reason behind this might be due to the repetition time (since it a sequence of numbers 1 to 365 repeated n times for n places in the region). Observe also that the plot of the global additive model for the Pacific region has a slightly higher amplitude than the two localized models. The cause of this is likely due to the fact that this model is fit globally, and since there is less data in the Pacific region than that of Continental and Atlantic, the influence of the trends in those two regions may have dominated that of the Pacific. Then to assess the goodness of fits, we compute the GCV estimates and R^2 values to get the results below.

	Polynomial	Spline	GAM
R squared	0.87900	0.8794453	0.8916382
GCV estimate	14.32027	14.2235375	17.8291651

Based on these values it can be concluded that while for the global additive model, the R^2 value is slightly better and the GCV estimate is slightly larger than that of the two localized models, in general there is minimal difference between the goodness of fit within each of the models.

5. Conclusions.

While each method and corresponding model(s) fits to the data relatively well, we recall that the point of the study is an attempt to understand the spatial variance of annual Canadian temperature with respect to region; with this in mind, it is sensible to select the global additive model as our prime choice in understanding the data. This is because the global additive model not only provides easily interpretable results — since we can plot out the partial effects and inspect the relationships between each covariate and T — it also provides insight into national temperature patterns and not just regional patterns. In restricting our models to be local to a region, we gain minimal improvements in the goodness of fits at the cost of difficult and even un-interpretable results in which we cannot use to compare the relative spatial patterns of the annual temperature. It also makes it very difficult to understand the relationship between each of the local models since each one is a separate model fit to one specific region, thus hindering the interpretability of the overall results. In addition, it is worth noting naive methods were implemented since the data used was relatively simple; if it were the case to see complicated and/or much larger amounts/types of data, other methods and models are recommended as they may be able to provide more accurate fits by taking into consideration more geological and meteorological phenomena.

REFERENCES

- ENVIRONMENT OF CANADA (2023). *Canadian Environmental Sustainability Indicators: Temperature change in Canada*. Available at [Canadian Environmental Sustainability Indicators: Temperature change in Canada](#). Government of Canada.
- RAMSAY J.O. AND B.W. SILVERMAN (2006). *Functional Data Analysis*, 2nd ed. Springer.