Predicting long-term missing ground-level ozone using precursors, solar and climate data.

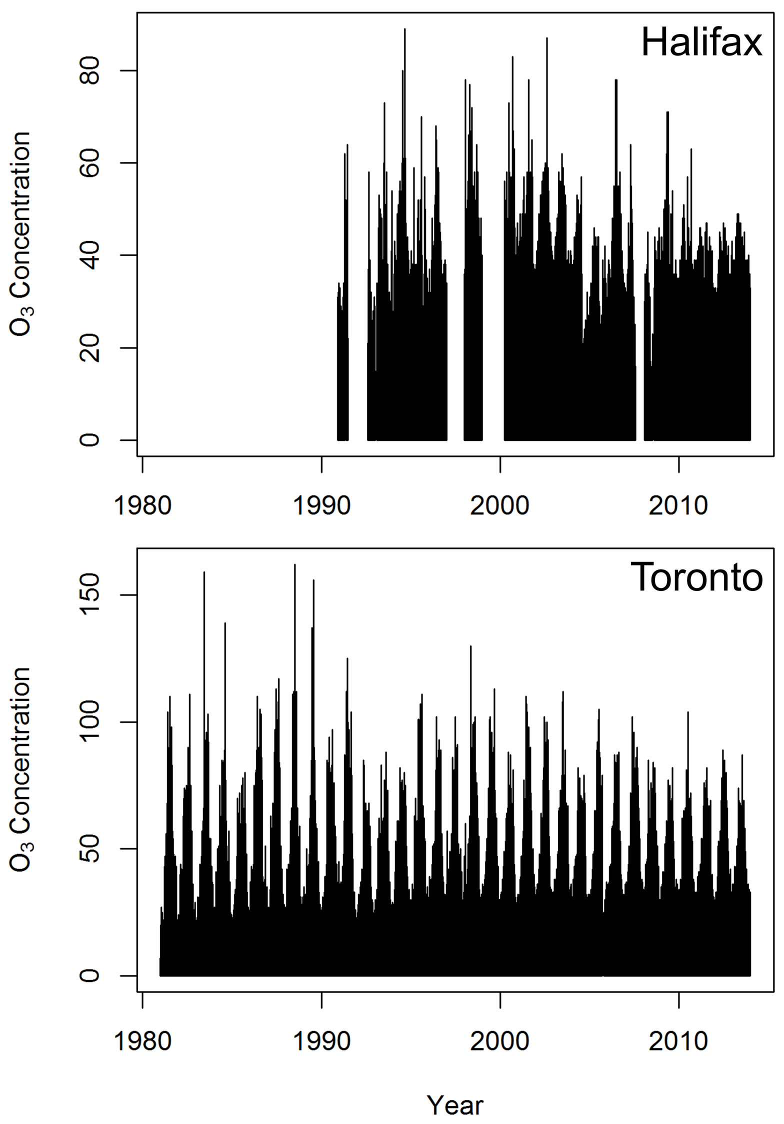
Aaron Springford & David J Thomson

aaron.springford@queensu.ca · djt@mast.queensu.ca

Queen's University

Health Canada Santé Canada

Recent years have seen a general decline in ambient ground-level ozone concentrations in developed nations. However, Health Canada researchers are investigating what causes variation in ozone-related mortality risk in larger Canadian cities. As part of this investigation, we are examining the ground-level ozone records carefully, and generating predicted ozone concentrations when none are available. We present two case studies – Halifax and Toronto – in which local ozone concentrations appear to be primarily driven by different mechanisms, and develop methods of prediction for each. Ozone concentrations in Toronto appear to be driven by local processes such as NO_x concentration, insolation, and temperature, whereas ozone concentrations in Halifax appear to be driven by regional processes such as wind speed and direction. We decided to pursue a regression tree approach for prediction for three reasons: (1) the possibility of complicated relationships and interactions between covariates and ozone measurements; (2) algorithms exist to handle missing covariate values, which are abundant; and (3) the method is tuned to optimize prediction performance.



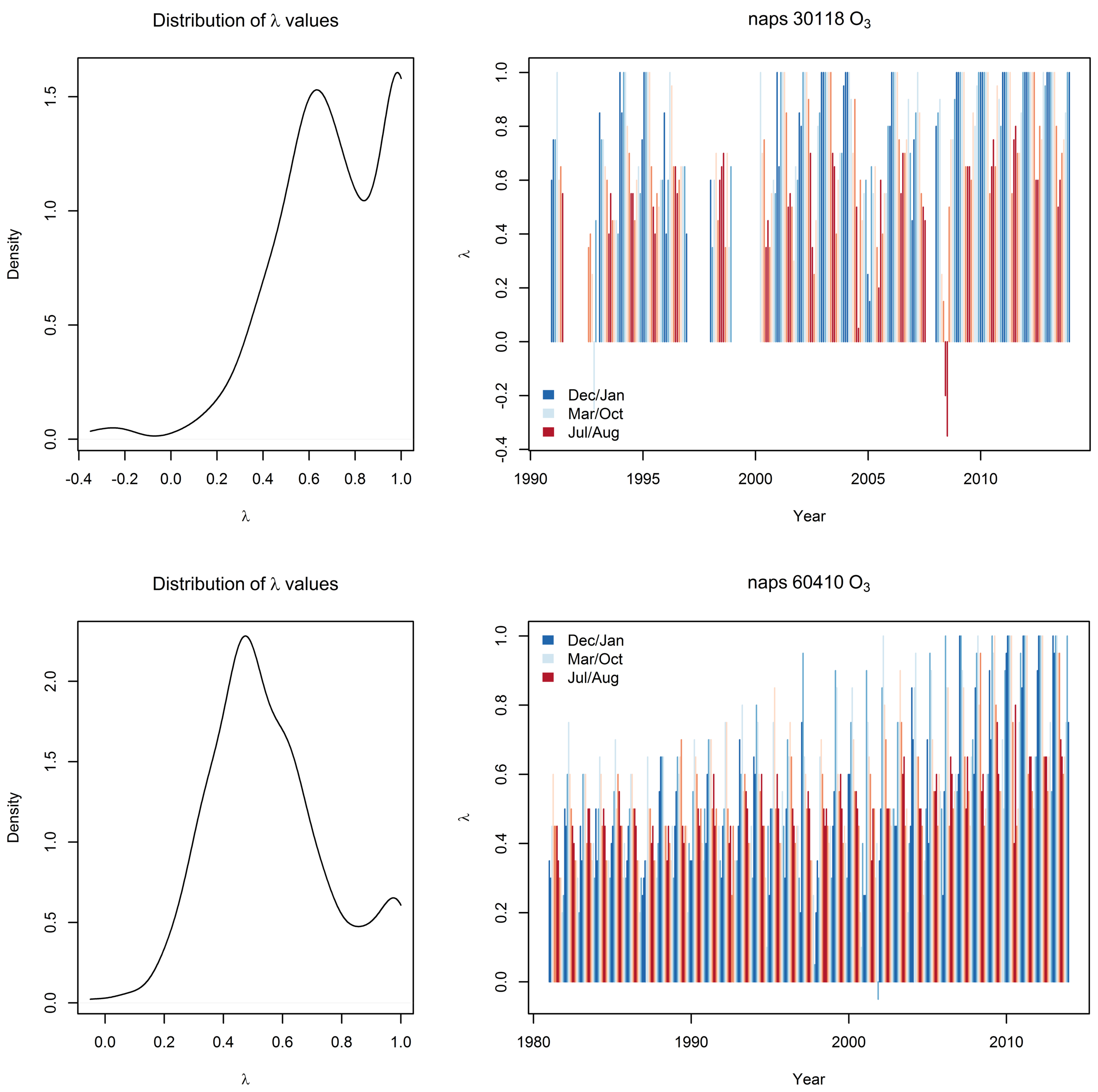
At what concentration does ground-level ozone become dangerous to health? Health Canada researchers are investigating this question using past records of ozone and cardiovascular death, but ozone records for many cities are incomplete. Halifax NAPS station 30118 (top) and Toronto NAPS station 60410 (bottom) have the most complete O_3 series in their cities for the years 1981—2013 (inclusive).

We examined ground-level ozone concentration data[1] collected in Halifax and Toronto with the ultimate goal of developing a method to predict ozone concentrations in cases where ozone measurements are not available. Precursor data included:

- NO_x concentration[1] (hourly)
- Volatile organic compounds[2] (VOCs) (~ 6-day)
- Ultraviolet radiation[3] (UV) (daily)
- Insolation[4] (hourly)

Weather data[2] were obtained from nearby weather stations. Temperature is a well-known driver of ozone production, but barometric pressure, wind speed, and wind direction were also examined.

The ozone data are non-Gaussian and non-stationary (right, above). Ozone records at Halifax contain much unexplained variation, but appear to depend to a large extent on transport from areas to the southwest. Ozone records at Toronto, by comparison, appear strongly related to local variables such as temperature and insolation. Both locations show an obvious relationship to NO_x that are consistent with VOC-limited scenarios. We found no apparent relationship between ozone and VOC reactivity, and excluded VOC data from prediction models. In the case of Halifax, we included other regional ozone measurements since missing ozone data coincides with missing NO_x data. We also transformed wind speed and direction data into North and East components. We fit regression trees to the ozone data, using surrogates for splitting when covariates were missing. Tree sizes were determined using cross-validation. For Halifax, the method had a relative prediction error of 55%, whereas for Toronto the method had a much lower relative prediction error of 28%.



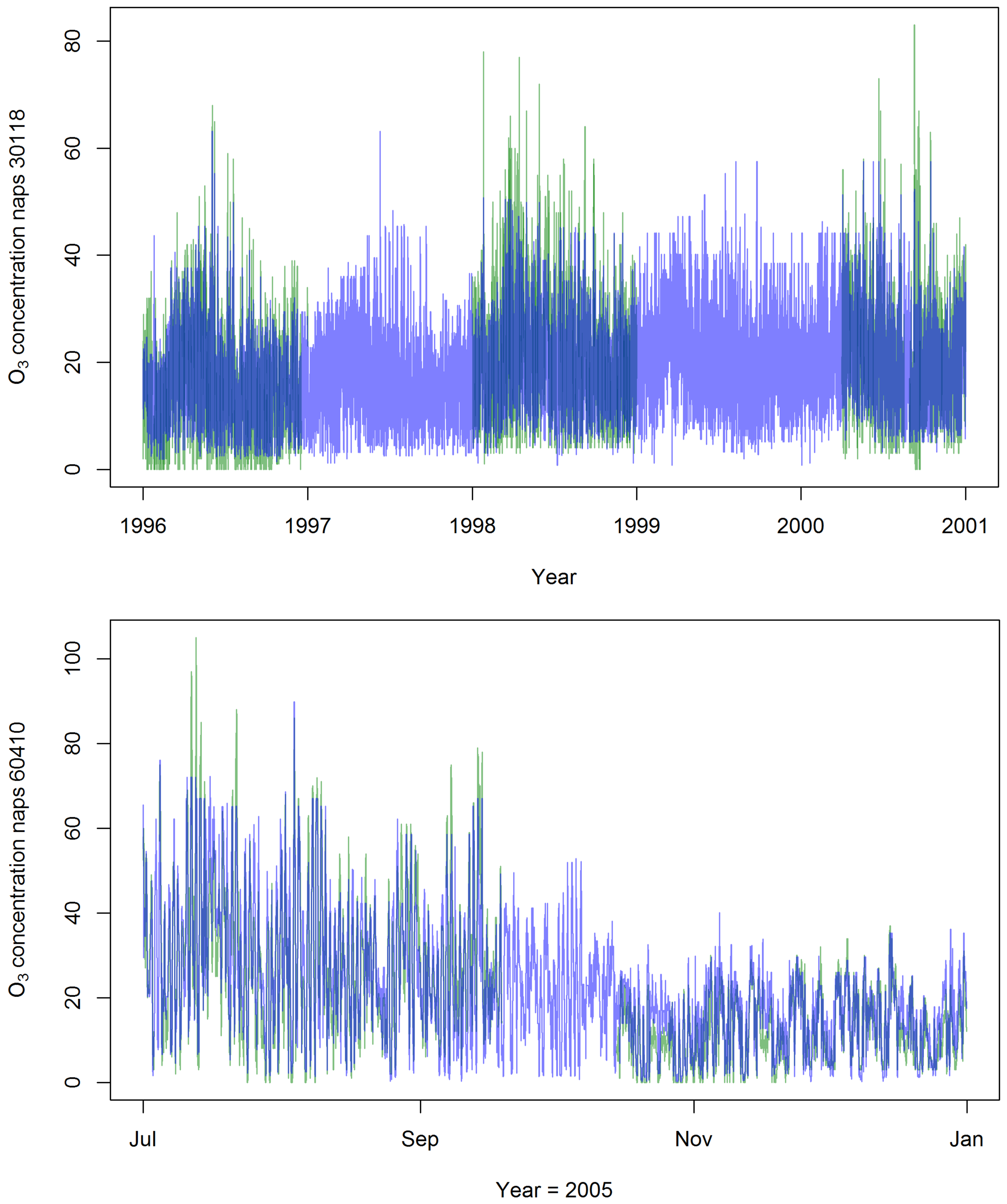
Distribution of monthly Box-Cox maximum likelihood power transformation estimates for Halifax NAPS station 30118 (top) and Toronto NAPS station 60410 (bottom). The left panels show the distribution of estimates. The right panels show estimates as a function of time. Seasonality is indicated by colour, with summer months in red and winter months in blue. A seasonal pattern is evident, as is a year-over-year trend in Toronto. Clearly, methods that rely on stationarity of time series or even cyclo-stationarity will have difficulty describing these data.

Halifax Variable	Relative Importance
O_3 (Kejimkujik)	0.373
O_3 (St John, +1hr)	0.179
UV index	0.112
O_3 (Yarmouth)	0.092
Wind (East) (10945)	0.060
Insolation	0.056
Wind (North) (10945)	0.042
Temperature (10945)	0.034
Temperature (27223)	0.022
Wind (East) (27223)	0.017
Wind (North) (27223)	0.013

Toronto Variable	Relative Importance
NO_x	0.591
Insolation	0.161
Temperature (4841)	0.161
Wind (North) (4841)	0.035
UV index	0.030
Wind (East) (4841)	0.023

Left: The relative importance of each regression tree input (station ID in brackets) for Halifax and Toronto.

Below: The tree-predicted ozone concentration (light blue) and actual data (green). Overlapping values appear dark blue. Top panel is Halifax from 1996—2001, showing the predicted missing years ~1997 and ~1999. Bottom panel is Toronto showing the longest gap in the series, October 2005.



Data sources

1. Hourly ground-level ozone and NO_x concentration data were collected across a national monitoring network and were obtained from the NAPS website at http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx?lang=en.
2. VOC data and local weather station data were obtained from Health Canada.
3. We obtained an index of solar UV activity based on the emission of the Mg(II) doublet from the institute of Environmental Physics at the University of Bremen http://www.iup.uni-bremen.de/gome/gomemgii.html.
4. Local insolation depends on the location on the Earth, the Earth's rotation, its orbit around the Sun, and its tilt relative to its orbit. We obtained hourly insolation to a horizontal surface using the calculator at https://www.nrel.gov/midc/solpos/solpos.html separately for Halifax and Toronto.