Merging Modelled and Observed Data in Spatial Analysis of Ozone Distribution over Ontario, Canada, using a Kriging Technique

A. Chtcherbakov, R. Bloxam, L. Huang, S. Wong, Y. Hall (Ontario Ministry of Environment and Climate Change)

Introduction

Analysis of spatial distribution of ozone and $PM_{2.5}$ based only on observed data or modelled results can be challenging. O_3 and $PM_{2.5}$ observed data is available from the Ontario's National Air Pollution Surveillance Program (NAPS) network of 39 sites and Environment Canada's network of 33 Canadian Air and Precipitation Monitoring Network (CAPMoN) sites (4 of which are in Ontario). The majority of the stations are located in the more populated cities of southern Ontario resulting in gaps in monitoring data and thus a lack of reliable data for interpolation, especially in the northern part of the province. Modelled concentrations are available for the entire province, but modelling results are often biased due to a number of model uncertainties and assumptions.

This study describes an approach for merging modelled and observed data for spatial analysis using a kriging technique.

Kriging Technique

Kriging is an advanced geo-statistical procedure that generates an estimated surface from a scattered set of points with observed values. Kriging is based on statistical models that include autocorrelation (the statistical relationships among the measured points) and has the capability of producing both a prediction surface and providing a measure of the certainty or accuracy of the predictions. Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. Kriging fits a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location. Kriging is a multistep process; it includes exploratory statistical analysis of the data, variogram modelling, and creating predicted surface. The general formula for kriging is formed as a weighted sum of the data:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

where:

$Z(s_i)$ = the measured value at the i^{th} location,

λ_i = an unknown weight for the measured value at the i^{th} location,

s_0 = the prediction location, and

N = the number of measured values.

In kriging, the weight, λ_i, depends on a model fitted to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location. To make a prediction with the kriging interpolation method, two tasks are necessary: uncover the dependency rules and make the predictions. The first task involves creating the variograms and covariance functions to estimate the statistical dependence (so called spatial autocorrelation) values that depend on the model of autocorrelation (fitting a model).

Analysis Set-up

This kriging approach has been applied to modelled and observed concentrations over Ontario, Canada for the year 2010. The analysis set up is as follows.

The modelled concentrations were adjusted by observed data:

$$[M_{adj}] = [M_{mod}] \times [R],$$

where $[M_{adj}]$ = adjusted with observations modelled concentrations,
$[M_{mod}]$ = modelled concentrations, and
$[R]$ = ratio (or difference) of modelled to observed concentrations pairs calculated in the monitoring station locations and then extrapolated (using kriging) to the entire province.

Results

The described method has been used for the spatial analysis of several air quality metrics in Ontario such as annual concentrations of $PM_{2.5}$, 4^{th} highest 8 hour maximum O_3 concentration, etc. Modelled concentrations were extracted from outputs from Community Multi-scale Air Quality (CMAQ) modelling system run for the entire year of 2010. Observed data for the corresponding year was obtained from NAPS and CAPMoN networks. Fig.1 shows air quality monitoring stations location across the province. Most of the stations are concentrated in southern Ontario, leaving central and especially northern Ontario with limited monitored data.

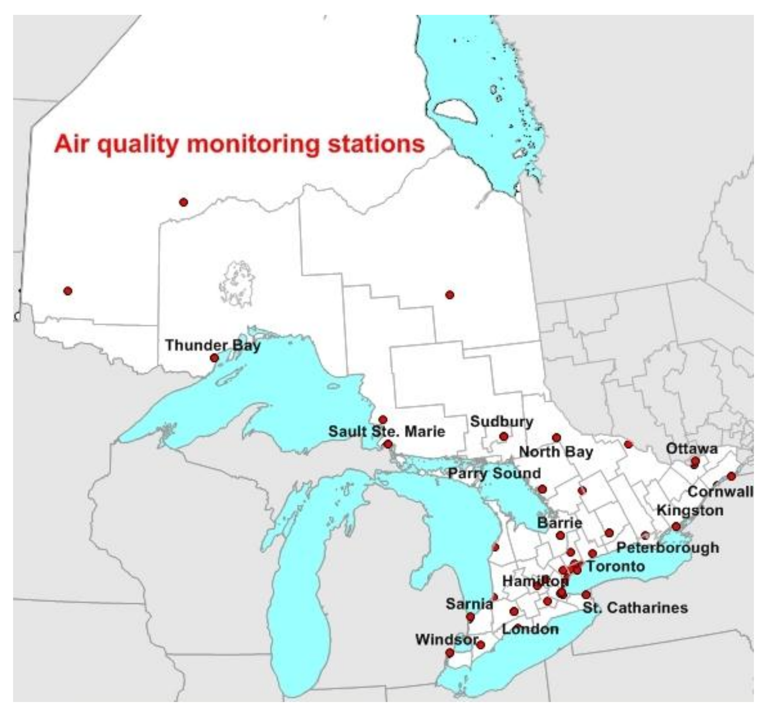

Fig 1

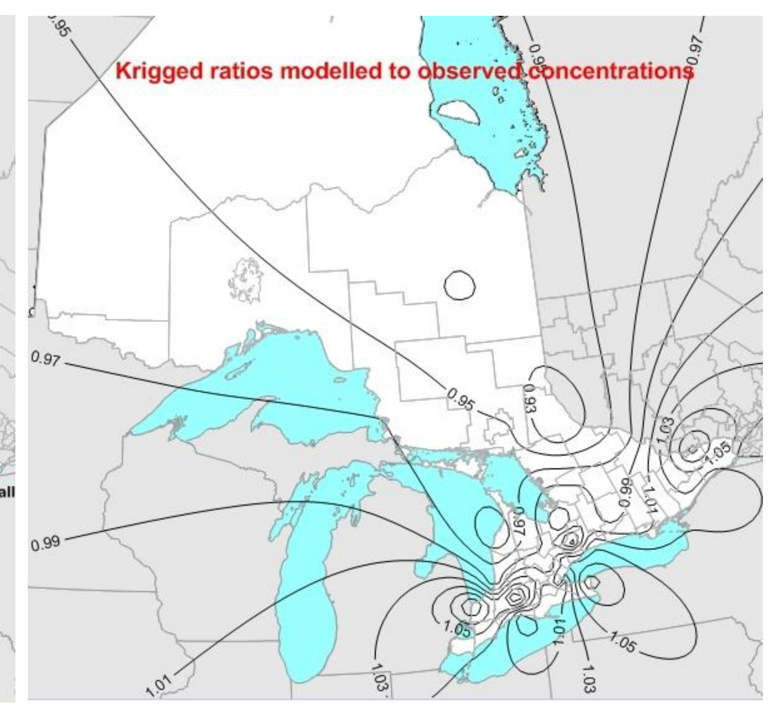

Fig 2

Preliminary analyse showed that applying kriged ratios rather than kriged differences of modelled-observed data pairs for adjusting modelled concentrations, produced more reasonable results. An example of kriged ratios of modelled/observed concentrations is shown on fig.2. The sharpest gradients are located in southern Ontario, in the Greater Toronto and Hamilton area, suggesting high level of variability due to the influences of large cities combined with the Great Lakes effects on concentrations. In northern Ontario kriged ratio does not show such a variance with close to linear or even distribution.

Fig. 3 shows distribution of 4^{th} highest O_3 concentration based only on interpolation of observed concentrations. The technique captures the general distribution of the metric, but has limits due to mostly inland stations locations. Particularly since O_3 formation over water bodies such as the Great Lakes and transport it onshore should result in higher O_3 concentrations.

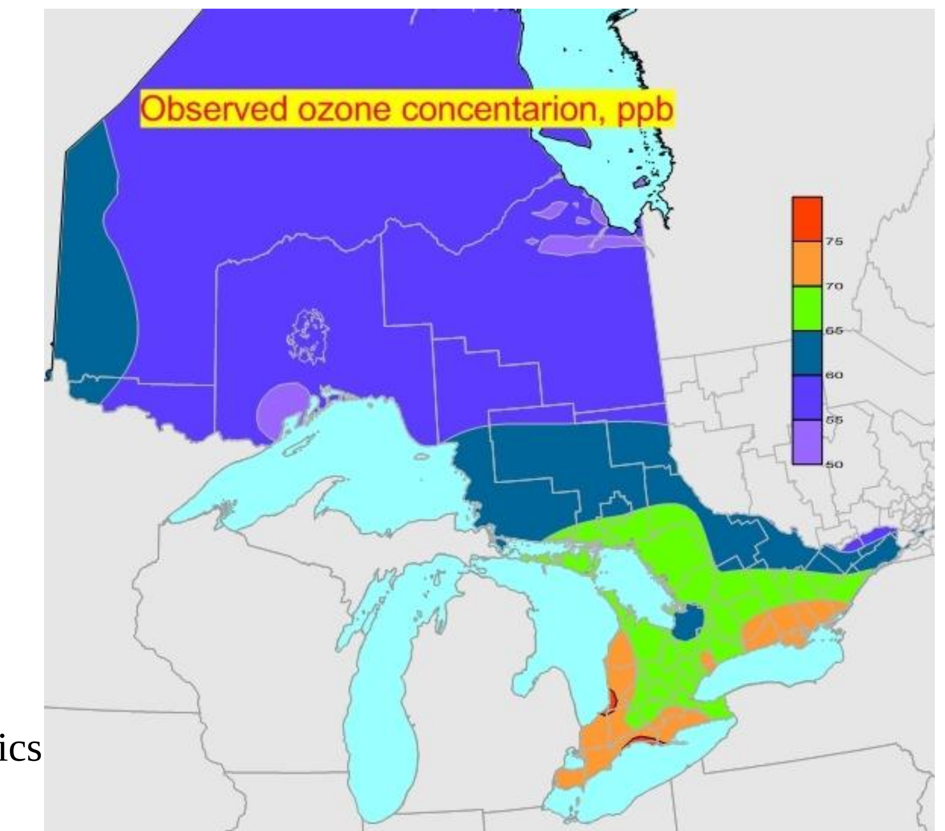

Fig 3

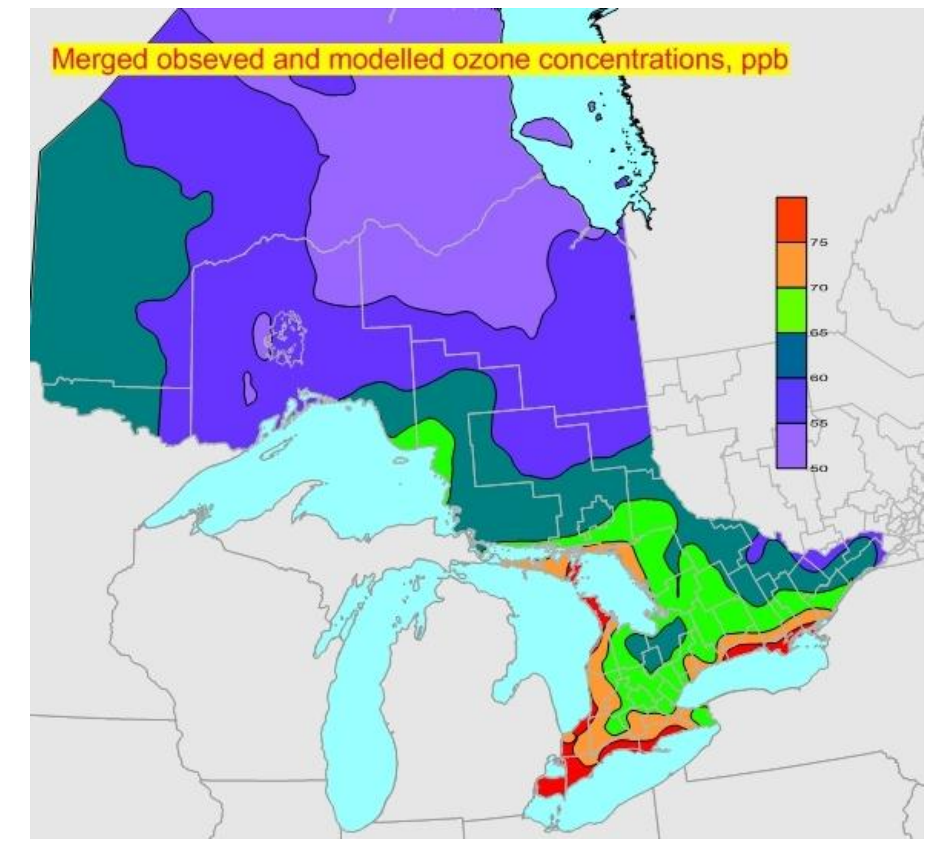

Fig 4

Fig. 4 shows the distribution of the same metric, but it is adjusted according to the above described technique. It can be seen that zones of high O_3 concentrations cover larger areas along the shorelines of the Great Lakes with O_3 concentrations increasing towards the water bodies. This is a more realistic O_3 distribution and more scientifically sound as these area are filled with adjusted modelled outputs which takes in account sophisticated chemistry and physics of O_3 formation and transport instead of simple interpolation of concentrations in areas between stations.

The accuracy and the certainty of the fused observed-modelled results depends on spatial density of observed-modelled pairs and if the observed values are below certain threshold values. One of approaches to address this problem is to set ellipses of influence, defining the number of analysis pairs and the distance to the given location, blanking all other areas. An example of this additional analysis is shown on fig.5.

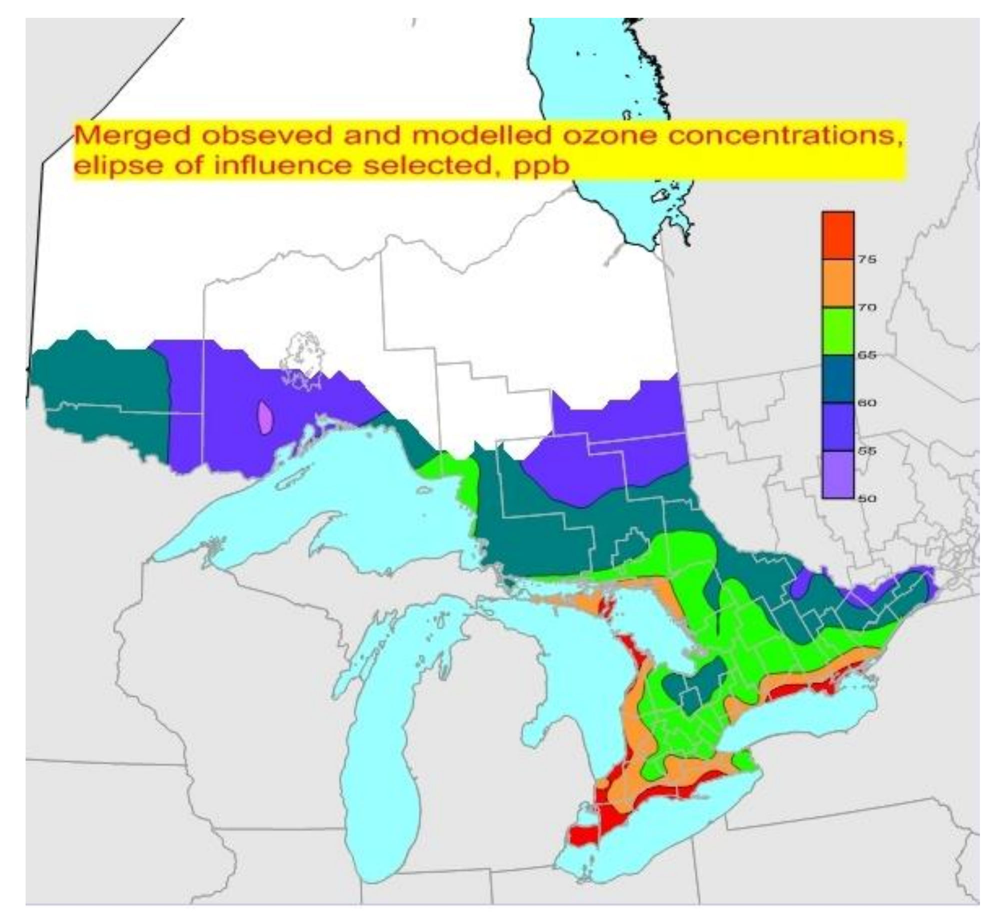

Fig. 5

Conclusions

Analysis of air quality in Ontario, based only on observed data gives good results, but they are representative only for the locations, in closer proximity to the monitoring stations. Merging together modelled and observed data using a kriging technique allows gaps between stations to be filled in with scientifically sound modelling outputs. The examples provided show the improved distribution of a selected metric (e.g. 4^{th} highest O_3 concentration) over Ontario. Similar analysis performed on other species ($PM_{2.5}$, NO_2) with different metrics has demonstrated similar results.

Protecting our environment. Ontario