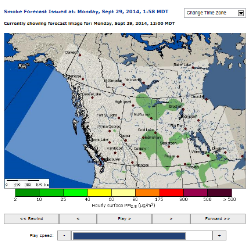
Improvements to the 2012 Evaluation of the Western Canadian BlueSky Smoke Forecasting System

Amy Thi[1], David Lyder[2]
1. Engineering Co-op Program, University of Alberta; 2. Alberta Environment and Sustainable Resource Development, Government of Alberta
(Corresponding author: David.Lyder@gov.ab.ca)

Background

The BlueSky Forecasting System aims to predict $PM_{2.5}$ concentrations in the air due to forest fire activity. Currently, the Western Canadian section is being evaluated against observed $PM_{2.5}$ ambient concentrations in several sites across Alberta. Past methods of evaluation include qualitative analyses of the BlueSky time series against the observed ambient time series and visual comparisons of Google Earth smoke plume animations and MODIS satellite images. To combat the obstacle of other sources contributing to the observed $PM_{2.5}$ levels, the ambient data was manipulated to remove non-smoke components by subtracting a polynomial fit of the average $PM_{2.5}$ levels from previous years; another approach included adding the average to BlueSky instead. Conclusions drawn from these past methods illustrated the need for ground information as well as the inclusion of carry over smoke in the model processing.



http://www.bcairquality.ca/bluesky/west/index.html

Ambient data:
Hourly $PM_{2.5}$ concentration was collected from the Casa Data Warehouse:
http://www.casadata.org/

BlueSky Data:
http://www.bcairquality.ca/bluesky/data/

Objective

The objective of this project was to investigate post processing techniques to further improve the Western Canadian BlueSky Forecasting System during the 2012 fire season (from April to October). Some of the attempted methods include using the Kalman Filter then applying smoothing techniques such as windowed spline, sliding spline, moving average, and least squares fitting.

Method

Learning from past evaluations, hourly averages were calculated by aligning observed ambient data from previous years on the Canadian Victoria Day as a reference point to account for variability in anthropogenic activity during different days of the week. The site-specific average background concentrations were then added to BlueSky data for a new series of improved forecast values.

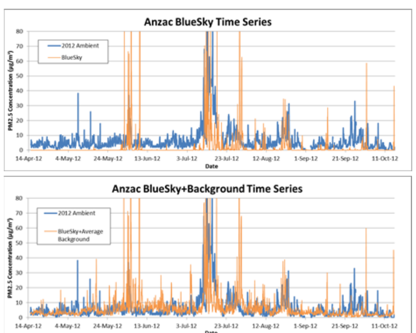


Figure 1. Top: Overlay of BlueSky data and 2012 observed ambient. Bottom: Overlay of BlueSky+Average Background concentrations and 2012 observed Ambient.

	BS	BS+Bckgrnd
Pearson	0.208471	0.21157016
MB	-4.064421	0.7503635
NMB	-65.17435	12.0323287
RMSE	12.628558	12.5691327
NME	105.18735	91.544703

Figure 2. Statistics comparing BlueSky with the observed ambient and BlueSky+Average Background with the observed ambient.

Method Continued

Kalman Filter

Although the Pearson Correlation remains poor after adding the average background values, there are super hourly trends in the ambient data e.g. increase of $PM_{2.5}$ from car emissions during the day, that should match the BlueSky+Average Background values. The Kalman filter utilizes this information to produce a more statistically optimal estimate. Applying the Kalman filter for the 1 hour forecast greatly increased the Pearson correlation in many sites (for the Edmonton Central site: from 0.11 to 0.89!).

Kalman Filter Algorithm (easily extended to determine forecasts further into the future)

$$\text{Corrected PM2.5 Forecast} = b \times \frac{A}{a}$$

where b = current hour (BlueSky+Background) value
A = previous hour ambient value
a = previous hour (BlueSky+Background) value

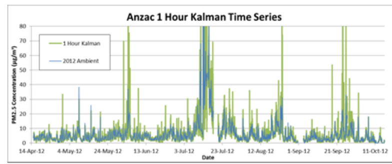


Figure 3. Overlay of the 1 hour Kalman predicted value for Anzac. Compared to the raw BlueSky or the BlueSky+Average Background time series, this is visually much more similar to the observed ambient

Problems with the Kalman Filter:

- Increased amplitude
- Degradation of correlation with increasing forecast hours; from the operational perspective, a reasonable forecast of at least 6 hours is desired

				Amb vs. Kalman								
Site	Stat	Amb vs BS	Amb vs BS+bckgrnd	1hr	3hr	6hr	9hr	12hr	15hr	18hr	21hr	24hr
Anzac	Pearson	0.208471	0.211570156	0.56958	0.40666	0.37833	0.32107	0.29386	0.25474	0.22398	0.21033	0.22343
	MB	-4.064421	0.750363504	0.74108	1.71485	2.38397	3.32437	3.87422	3.9282	4.06778	3.6988	3.39472
	NMB	-65.17435	12.03232865	12.3846	28.6528	39.8268	55.5147	64.6873	65.5641	67.8689	61.6937	56.604
	RMSE	12.628558	12.5691327	13.3594	18.7288	19.6382	27.5135	35.8917	37.3706	38.7731	28.5193	26.9103
	NME	105.18735	91.54470298	43.1844	78.0898	102.448	125.985	137.698	140.126	143.029	136.252	130.914

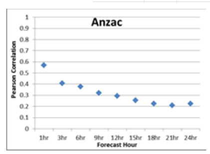


Figure 4. Top: Table of the statistical comparison between the observed ambient with the Kalman predicted value. Left: Scatter plot showing the degradation of correlation as the prediction hour increases. Bottom: Table of the average statistical comparison between the observed ambient with the Kalman predicted values between 15 urban and 9 rural sites, based on location, population, and baseline trends.

				Amb vs. Kalman								
Average	Stat	Amb vs BS	Amb vs BS+bckgrnd	1hr	3hr	6hr	9hr	12hr	15hr	18hr	21hr	24hr
Urban Sites (15)	Pearson	0.1472809	0.130417349	0.73973	0.46515	0.35116	0.29211	0.28989	0.25855	0.228	0.21721	0.21178
	MB	-7.655552	-0.276781762	0.45246	1.18472	1.82898	2.09393	2.0932	2.06909	2.10796	2.06815	1.88576
	NMB	-91.66541	0.318151628	5.90606	15.3763	24.0718	27.189	26.8141	26.4457	26.9435	25.9969	23.891
	RMSE	14.408974	13.01657197	12.7087	25.9092	33.8428	32.6005	31.7913	27.4457	28.7177	31.8753	26.5673
	NME	98.120156	78.23265797	41.2294	67.054	86.4034	94.243	97.2162	99.5016	101.517	100.012	98.284
Rural Sites (9)	Pearson	0.189526	0.156362982	0.72807	0.39322	0.35129	0.28679	0.28619	0.24447	0.20626	0.21953	0.22484
	MB	-6.94651	-0.061442785	0.46626	1.25967	1.89704	2.20749	2.16597	2.03877	2.07106	1.95113	1.74484
	NMB	-80.47976	1.057794286	6.2826	16.5014	25.3861	28.5885	27.5314	26.2155	26.7282	24.9558	22.7342
	RMSE	12.23188	12.01042115	14.1927	30.8255	38.2829	36.9293	35.8556	27.9743	29.9481	29.5218	24.3723
	NME	76.627949	73.93751365	41.875	70.5103	90.517	98.0843	100.231	101.989	104.042	100.767	98.0228

Techniques used in attempt to smooth Kalman filtered values:

1. Windowed Spline on Kalman – fitting a cubic equation to 4 hour periods of the Kalman filtered values
2. Sliding Spline on Kalman – fitting a cubic equation to sliding 4 hour periods of the Kalman filtered values
3. Moving Average of Kalman –using the average of Kalman filtered values over previous 4 hours as the desired hour's forecast value (also experimented with other hour averages as well)

Method Continued

4. Least Squares Fitting of Kalman (with different sized model sets)
   - The most promising method so far
   - Monte Carlo simulation with 100 and 1000 subsets to determine most representative term coefficients
   - The equation was then applied on the remaining validation set and compared to the observed ambient values

Ideal Result of Kalman filter:

$$B = b \times \frac{A}{a}$$

Actual equation:

$$\ln B = \alpha_1 \ln b + \alpha_2 \ln A - \alpha_3 \ln a$$

where b = current hour (BlueSky+Background) value
A = previous hour ambient value
a = previous hour (BlueSky+Background) value
B = current hour ambient value

Results

This is still a work in progress and all the methods have only been applied to a few select sites up to date. Naturally, variations between different sites due to location and subset size are expected.

- Least squares fitted Kalman has highest correlation and least degradation with time
- Visually, the least squares manipulation successfully decreased the amplitude
- General trends in time series captured
- Depending on the application of BlueSky, the certainty of concentration magnitude may or may not be an issue. E.g. Alberta Health has a threshold concentration of 30 g/m^3 $PM_{2.5}$ representing the level of concern.
- Complications with the least squares fitting: possibility of over-smoothing and the forecast trends lagging behind by a few hours

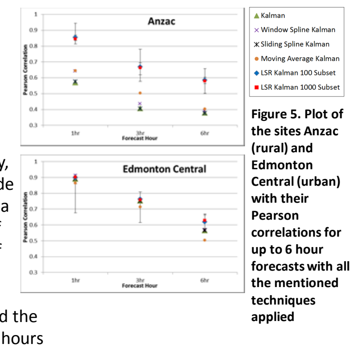


Figure 5. Plot of the sites Anzac (rural) and Edmonton Central (urban) with their Pearson correlations for up to 6 hour forecasts with all the mentioned techniques applied

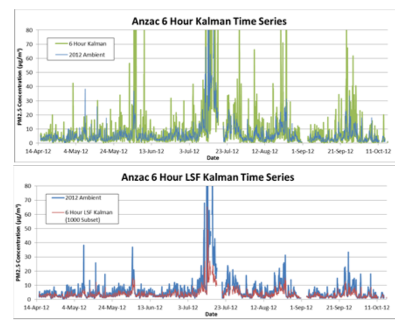


Figure 6. Overlay of the 6 hour Kalman prediction with the observed ambient (top) compared to the overlay of the 6 hour least squares fitted Kalman prediction based on a 1000 value subset with the observed ambient (bottom).

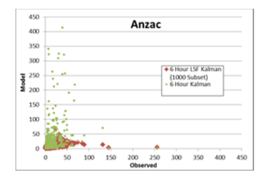


Figure 7. Scatter plot of the 6 hour least squares fitted and the raw 6 hour Kalman forecast against the observed ambient.

Conclusion

An investigation of the possible post processing techniques to improve the 2012 Western Canadian BlueSky Forecasting System revealed many advantages and disadvantages in Kalman Filter and smoothing using spline, moving averages and least squares fitting. Further analysis is required determine whether these improvements hold for additional sites as well for other years of evaluation.