

## Introduction

In an effort to explore the efficiency of mass receptor models in distinguishing the source contributions of gas and diesel vehicles in airshed samples, we have created artificial airshed data from the Los Angeles data with which we have produced various simulations for 3 proposed solutions to the mass receptor modeling problem. The following sections discuss how the data was created followed by summaries of the implementation and results of mass receptor modeling applied to the simulated data with the Effective Variation Solution, a bayesian approach, and Positive Matrix Factorization.

## The Data

The airshed data is made up of profiles for gas, smoker, and diesel vehicles multiplied by normally distributed source contributions for each profile. The gas profile consists of the No Smoke columns 0b NREL and 1 EMFAC for both cold and warm starts(4 columns) from the Averaging SI worksheet. These 4 profiles were averaged to create the gas profile. The smoker profile is the average profile for the 13 smoker vehicles listed under the headings Good Smokers Only, Bad Smokers Only, and New-Defined Smokers. The diesel profile is the average profile for the 22 diesel profiles from the Averaging CI worksheet. The 3 resulting profiles were then multiplied by their respective source contributions to form a proposed fleet of vehicles made up of 80% gas vehicles, 5% smoker vehicles, and 15% diesel vehicles. The source contributions are normally distributed, centered at 15, .937, and 2.8215 respectively(These contributions were chosen to have the correct proportions, thier values do not have further significance). Once the 3 profiles were multiplied by their respective source contributions they were summed together and the simulated data sets result from adding lognormal error to the created fleet for n days. The uncertainties from the L.A. data were used to calculate the variances of the simulated data.

## Effective Variance Solution

In order to test the EV solution as to its effectiveness in predicting source contributions for gas, smoker, and diesel vehicles, 10,000 days were simulated at each of 3 levels of CV. The EV solution does very well with clean data as seen in the MSE for CV=.01, but MSE increases substantially with the introduction of more noise as seen in the results for CV=.2 and CV=.5.

**Estimated Source Contributions for EV Solution and MSE**

CV on Y	Gas Vehicles	Smoker Vehicles	Diesel Vehicles	Gas MSE	Smoker MSE	Diesel MSE
Actual	15	.937	2.8215			
CV=.01	14.9134	.9665	2.8577	.0075	.0009	.0013
CV=.2	14.6691	.8720	2.6209	6.2685	1.1334	.3590
CV=.5	12.5991	.8081	2.0735	31.4499	4.2154	1.8240

## Bayesian Solution in Jags/WinBUGS

For our bayesian approach we consider the receptor model  $Y_{p \times n} = \Lambda_{p \times k} \times F_{k \times n} + E_{p \times n}$ . We assume the data to be lognormally distributed leading to a lognormal likelihood for Y. For this summary we used a specific gamma prior for every  $\lambda_{pk}$ , solving for  $\alpha_{pk}$  and  $\beta_{pk}$  with our data and uncertainties as our means and variances. We used a flat gamma prior with a mean around 6 for F, but more specific priors for the 3 factors can also be used. Below is the likelihood of the data along with its priors. The results of our simulations in JAGS(Just Anothoer Gibbs Sampler) are displayed at the top of the next page, these results seemed less affected by noise in the data as compared to the results of the EV solution.

$$\text{Log}(Y)_{[p,n]} \sim \text{Normal}\left(\text{Log}(\lambda \times F)_{[p,n]} - .5\sigma^2, \sigma^2\right)$$

$$\lambda_{[p,k]} \sim \text{Gamma}\left(\alpha_{[p,k]}, \beta_{[p,k]}\right)$$

$$F_{[k,n]} \sim \text{Gamma}\left(\alpha, \beta\right)$$

$$\sigma_{[p,n]}^2 = \text{Log}(cv^2 + 1)$$

**Estimated Source Contributions for Bayesian Solution and MSE**

CV on Y	Gas Vehicles	Smoker Vehicles	Diesel Vehicles	Gas MSE	Smoker MSE	Diesel MSE
Actual	15	.937	2.8215			
CV=.01	15.1515	1.0920	2.9553	.0244	.0239	.0181
CV=.2	15.3114	.9222	2.5603	.0991	.0016	.0687
CV=.5	12.7541	.7917	3.1794	5.0486	.0237	.1286

## Positive Matrix Factorization (PMF)

For our experiments with PMF, 500 days were simulated and the data was G-keyed in order to pull the  $\lambda_{pk}$ 's closer to the profile data. As seen with the EV solution, the MSE for the simulated f's dramatically increase when we turn up the CV and introduce noise into the data.

**Estimated Source Contributions for PMF and MSE**

CV on Y	Gas Vehicles	Smoker Vehicles	Diesel Vehicles	Gas MSE	Smoker MSE	Diesel MSE
Actual	15	.937	2.8215			
CV=.01	15.1009	.9288	2.8142	.0107	.000099	.000095
CV=.2	15.2551	.7499	2.7302	.2833	.0448	.0254
CV=.5	19.3268	.4574	.6909	29.4083	.2559	4.5455

## Conclusion

When the data is clean and the cv is very low(.01), EV and PMF do very well with low MSE's, but as the CV is turned up, the MSE's for EV and PMF are greatly influenced by the noise in the data while the MSE for the bayesian solution stays relatively low. The bayesian solution appears to be more resistant to noisy data which is very appealing knowing that measurements in real world situations will never be as clean as having CV=.01, therefore we are optimistic that this type of model using prior knowledge and updating could be an effective method in determining source contributions and profiles for vehicle pollution data. Through further research and simulations we hope to further develop an understanding of these 3 solutions to the mass receptor modeling problem as related to data dealing with vehicle pollution.