ADDING REALISM TO SOURCE CHARACTERIZATION USING A GENETIC ALGORITHM

Luna M. Rodriguez*, Sue Ellen Haupt, and George S. Young
Department of Meteorology and Applied Research Laboratory
The Pennsylvania State University, University Park, Pennsylvania

1. INTRODUCTION

America's National Strategy for Homeland Security states that one of the nation's goals is to respond to and recover from harmful incidents that occur (Homeland Security Office 2007). Such incidents include an intentional release of hazardous chemical, biological, nuclear, or radioactive (CBNR) material into the atmosphere. It is important to be able to predict the transport and dispersion of these materials. However, sometimes there is inadequate source information to perform those predictions; therefore it becomes necessary to characterize the source of an airborne contaminant from remote measurements of the resulting concentration field. The characterization of a source involves back-calculating the source location and emission rate. There has been extensive work back-calculating source characteristics, for example, Thompson et al (2007) and Rao (2007). Some previous work that uses genetic algorithms (GA's) to optimize source characteristics include the work done by Allen (2006, 2007), Haupt (2005), Haupt et. al. (2006, 2007a, 2007b, 2007c), and Long et. al. (2007).

In addition to characterizing the source some of these efforts include back-calculating meteorological variables, such as wind speed, wind direction, and stability. Several of the previous papers include adding noise to the data to simulate errors in the sensor data, input parameters, and the inherent atmospheric turbulence.

The goal of the present study is to perform a sensitivity analysis to add an element of realism to the likely sensor constraints. It is done using an identical twin approach with a Gaussian Puff model, which then optimizes a solution by means of a GA, and finally finds the global minimum with the Nelder-Mead downhill simplex algorithm (NMDS). The sensitivity analysis is needed because some of the sensors are often limited in terms of saturation and detection levels. These are taken into account because they make the observations non-Gaussian, which means that the GA must be aware of the levels so that it can model the observed data instead of the ideal Gaussian.

*Corresponding author address: Luna M. Rodriguez, Department of Meteorology, The Pennsylvania State University, University Park, PA 16802; lmr257@psu.edu

2. PROCEDURES

The Gaussian puff model (1) is used to determine

$$C_r = \frac{Q\Delta t}{(2\pi)^{1.5}\sigma_x\sigma_y\sigma_z}\exp\left(\frac{-(x_r - Ut)^2}{2\sigma_x^{\,2}}\right)\exp\left(\frac{-y_r^{\,2}}{2\sigma_y^{\,2}}\right)\left[\exp\left(\frac{-(z_r - H_e)^2}{2\sigma_z^{\,2}}\right)+\exp\left(\frac{-(z_r + H_e)^2}{2\sigma_x^{\,2}}\right)\right] \quad (1)$$

concentration observations over five time steps and five grid sizes on a 16 km^2 domain (Table 1), where C_r is the concentration at receptor r, (x_r, y_r, z_r) are the Cartesian coordinates downwind of the puff, Q is the emission rate of the source, Δt is the length of time of the release itself, U is the wind speed, H_e is the effective height of the puff centerline, and (σ_x, σ_y, σ_z) are the dispersion coefficients that are computed from Beychok (1994).

Table 1. Characteristics of grids evaluated on a constant 16 km^2 domain

GRID SIZE	NUMBER OF GRID POINTS	SPACING BETWEEN GRID POINTS (KM)
2X2	4	16.00
4X4	16	5.33
6X6	36	3.20
8X8	64	2.29
16X16	256	1.07

Data is created by first applying (1) to generate concentration at grid points then clipping the data to simulate detection and saturation of the levels. A Pasquill stability class D is used in this study and Figure 1 shows how the concentration strength varies over 5 time steps for a 16X16 receptor grid. This concentration data is then clipped to simulate saturation and detection levels of the sensors.

The detection level is determined with respect to the maximum concentration strength value and any data under that level is set to zero. The detection levels are simulated by imposing a minimum level to our data, i.e., for a $1X10^{-16}$ cutoff, anything smaller than $1X10^{-16}$ of the maximum concentration is changed to 0 and similarly for the $1X10^{-12}$, $1X10^{-8}$, $1X10^{-4}$ kg m^{-3} cases.

Saturation level for this study means that any value over a determined percent of the maximum concentration strength is changed to that particular value. For a saturation level of 100% of the maximum

concentration strength, 1 kg m^{-3} is used as the cutoff value. For saturation of 50% of the maximum concentration strength 0.5 kg m^{-3} is the cutoff so anything above this value is set equal to 0.5 kg m^{-3}, likewise for 1% (0.1 kg m^{-3}) and 0.1% (0.01 kg m^{-3}). Examples of the detection and saturation cutoff levels are given in Figure 2 where the maximum concentration in this illustration is 1 kg m^{-3}. After the data is clipped it serves as our "true" observations.

The GA begins with a random population of "guesses" to the variables that fall within the criteria described in Table 2. These are then compared to our "true" observations by means of a cost function (2).

$$\text{cost function} = \frac{\sum_{t=1}^{5} \sqrt{\sum_{r=1}^{TR} \left(\log_{10}(aC_r + \varepsilon) - \log_{10}(aR_r + \varepsilon)\right)^2}}{\sum_{t=1}^{5} \sqrt{\sum_{r=1}^{TR} \left(\log_{10}(aR_r + \varepsilon)\right)^2}} \qquad (2)$$

Where, C_r is the concentration as predicted by the dispersion model given by (1), R_r is the observation data value at receptor r, TR is the total number of receptors, a and ε are constants added to avoid logarithms of zero.

Table 2. Variable Thresholds used to populate the GA

PARAMETER	MINIMUM VALUE	MAXIMUM VALUE
Location (x,y) (meters)	-8000	8000
Source Strength (kg m^{-3})	0	20
Wind Direction (°)	0	360

GAs work by evaluating an initial population via the cost function then selecting the best ranking individuals to reproduce, forming a new generation through the GA operators of crossover and mutation. These are then in turn evaluated and the process iteration. We use a population of 40 chromosomes, 640 iterations, and a mutation rate of 0.32.The mean and standard deviations of 10 Monte Carlo runs of the median of 10 runs for each saturation level were evaluated for each detection level and grid size. The true solutions for all of these cases is a solution strength of 1 kg/s, a location of (0,0), and a wind direction of 180°.

3. RESULTS

Figures 3-6 show the mean values of the 10 Monte Carlo runs, each figure with a different saturation level as a constant while varying the detection levels across the abscissa and the differencing grid sizes indicated by the colored lines. In the figures each parameter (wind direction, source strength, & x,y location) is plotted separately. For the wind direction we are seeking a value of 180°, for the concentration strength a value of 1 kg/s, and for the source release a location of 0,0 meters.

Figures 3 and 4 show that the GA retrieves the correct source characteristics for all the detection levels using the 8X8 and 16X16 grid. The other grid sizes did not perform as well and were inconsistent. When lowering the saturation level, Figures 5 and 6, every grid size smaller than a 16X16 becomes highly unreliable. Thresholding the data too severely eliminates so much information that retrieval quality goes down significantly, thus, more dynamic range in sensors lends to more accurate inversion for the variables. This dynamic range produces the most impact if it extends to the maximum concentration as is illustrated in Figure 7. In this figure the detection level is $1X10^{-16}$ and the saturation levels vary along the abscissa with the differing grid sizes indicated by the colored lines. In agreement with Figures 3-6, Figure 7 shows that the larger grid sizes, 8X8 and 16X16, are successful in retrieving the correct parameter values up to the 50% saturation level. The smaller grid sizes are less reliable after the 50% cutoff and none of the grid sizes are able to correctly identify all of the parameters for the lowest saturation levels.

4. CONCLUSION

The hybrid GA method used here (with NMDS) is successful in back-calculating source characteristics and wind direction with data that has been thresholded forming a clipped Gaussian. These thresholds simulate saturation and detection levels in sensors and if applied too severely they eliminate so much information that retrieval quality degrades significantly. The inversion is most successful if the sensor can detect the maximum concentrations, which means that the most effective sensors have this characteristic.

The next step in this project is to retrieve source and meteorological data using a time-dependent computational fluid dynamics large eddy simulation to create synthetic data. Such data inherently includes time-dependent behavior unique to each contaminant episode rather than the ensemble average predicted by the transport and dispersion model used in previous studies. Then a realistic sensor configuration will be considered as well as varying stability classes. Finally, we expect to use this model to back-calculate source characteristics and meteorological parameters with real field data observations

ACKNOWLEDGEMENTS

This work was supported by DTRA under grant number W911NF-06-C-0162.

REFERENCES

Allen, C.T., G.S. Young, and S.E. Haupt, 2006: Improving Pollutant Source Characterization by

Optimizing Meteorological Data with a Genetic Algorithm, *Atmospheric Environment,* **41**, 2283-2289.

Allen, C.T., S.E. Haupt, and G.S. Young, 2007: Source Characterization with a Receptor/Dispersion Model Coupled With A Genetic Algorithm, *Journal of Applied Meteorology and Climatology,* **46**, 273-287.

Beychok, M. R., 1994: Fundamentals of Gas Stack Dispersion, 3PrdP ed. Milton Beychok, pub., Irvine, CA, 193.

Haupt, S.E., 2005: A Demonstration of Coupled Receptor/Dispersion Modeling with a Genetic Algorithm, *Atmospheric Environment*, **39**, 7181-7189.

Haupt, S.E. G.S. Young, and C.T. Allen, 2006: Validation Of A Receptor/Dispersion Model Coupled With A Genetic Algorithm, *Journal of Applied Meteorology,* **45**, 476–490.

Haupt, S.E., R.L. Haupt, and G.S. Young, 2007: A Mixed Integer Genetic Algorithm used in Chem-Bio Defense Applications, submitted to *Journal of Soft Computing*.

Haupt, S.E., G.S. Young, and C.T. Allen, 2007: A Genetic Algorithm Method to Assimilate Sensor Data for a Toxic Contaminant Release, *Journal of Computers*, **2**, 85-93.

Homeland Security Office, cited 2007: National Strategy for Homeland Security. [Available online at http://www.whitehouse.gov/infocus/homeland/nshs/NSHS.pdf.]

Long, K.J., S.E. Haupt, and G.S. Young, 2007: Improving Meteorological Forcing and Contaminant Source Characterization Using a Genetic Algorithm. Submitted to *Journal of Environmental Management*.

Rao, K.S., 2007: Source estimation methods for atmospheric dispersion, *Atmos. Env.* **41**, 6963-6973.

Thomson, L.C., Hirst, B., G. Gibson, S. Gillespie, P. Jonathan, K.D. Skeldon, M.J. Padgett, 2007: An improved algorithm for locating a gas source using inverse methods, *Atmos. Env*., **41**, 1128-1134

FIGURE

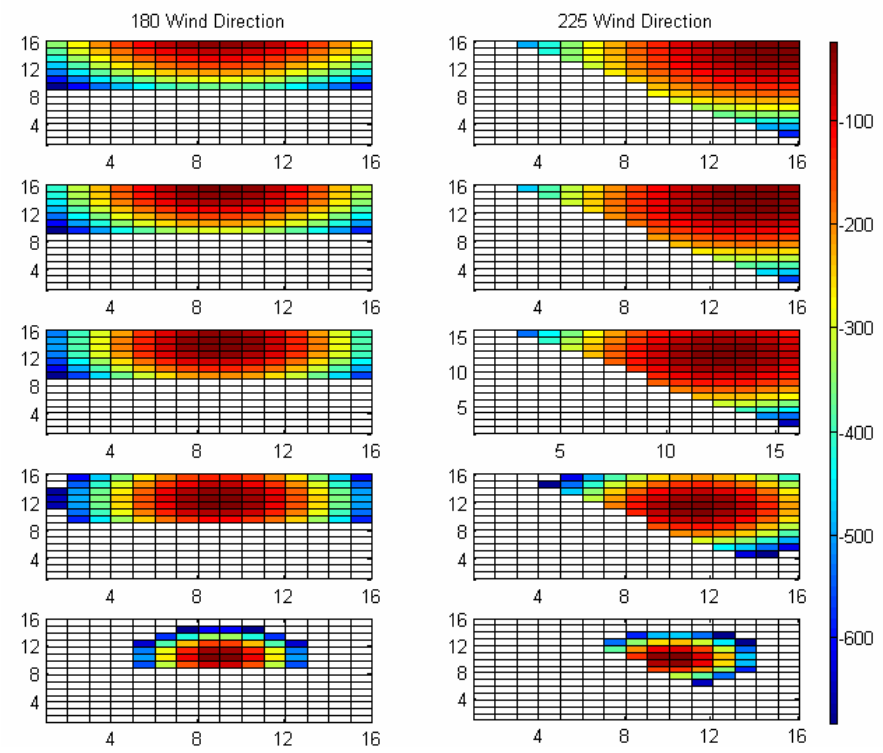

Figure 1. Concentration pattern over 5 time steps on a 16X16 receptor grid. The panel on the left shows the concentration for with a 180° wind direction and the panel on the right for a 225° wind direction.

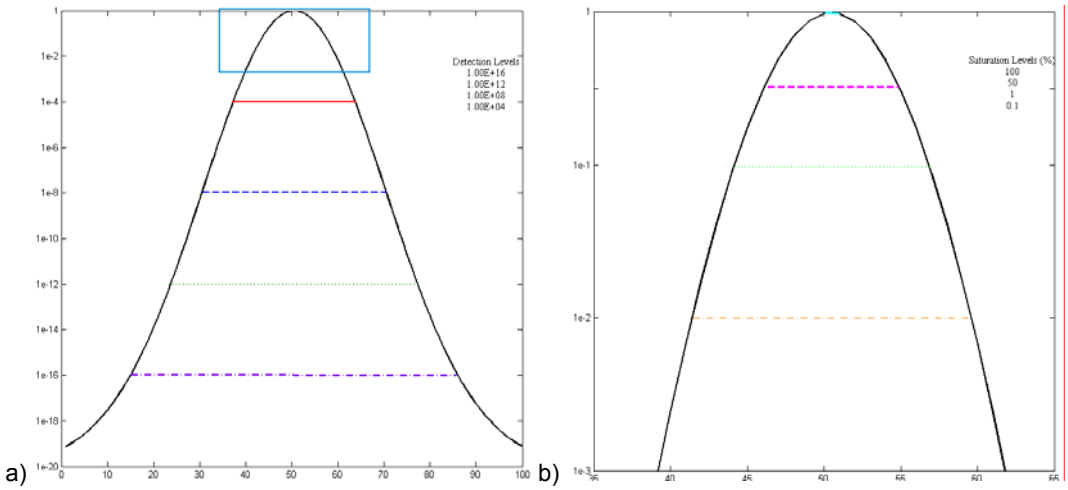

Figure 2. Data fit to a Gaussian. The maximum concentration normalized to 1 kg m^{-3}. Panel a indicates the threshold detection levels of $1X10^{-16}$, $1X10^{-12}$, $1X10^{-8}$, and $1X10^{-4}$. Panel b shows the saturation levels, set as a percentage of the full (100%, 50%, 1%, 0.1%).

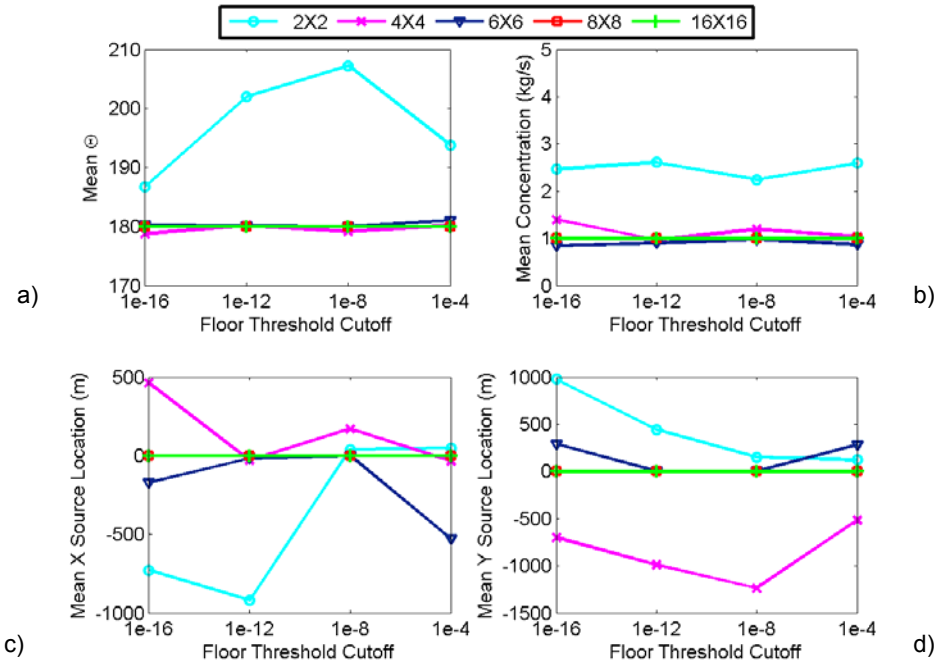

Figure 1. Results for all grid sizes and detection levels of the saturation level that at 100% of the maximum concentration strength value. Panel a shows the mean value of θ (wind direction, 180°), panel b indicates source strength (1 kg/s), and panels c and d show the location (0, 0) (in m) for x and y respectfully. All of these results are of 10 Monte Carlo runs of the median value of 10 individual runs.

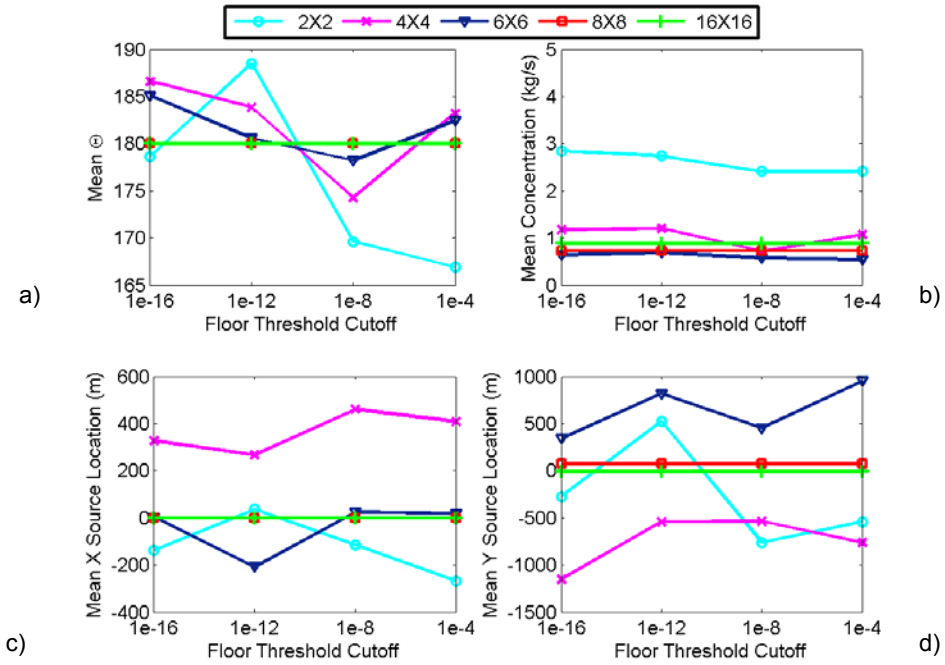

Figure 2. Results for all grid sizes and detection levels of the saturation level that at 50% of the maximum concentration strength value. Panel a shows the mean value of θ (wind direction, 180°), panel b indicates source strength (1 kg/s), and panels c and d show the location (0, 0) (in m) for x and y respectfully. All of these results are of 10 Monte Carlo runs of the median value of 10 individual runs.

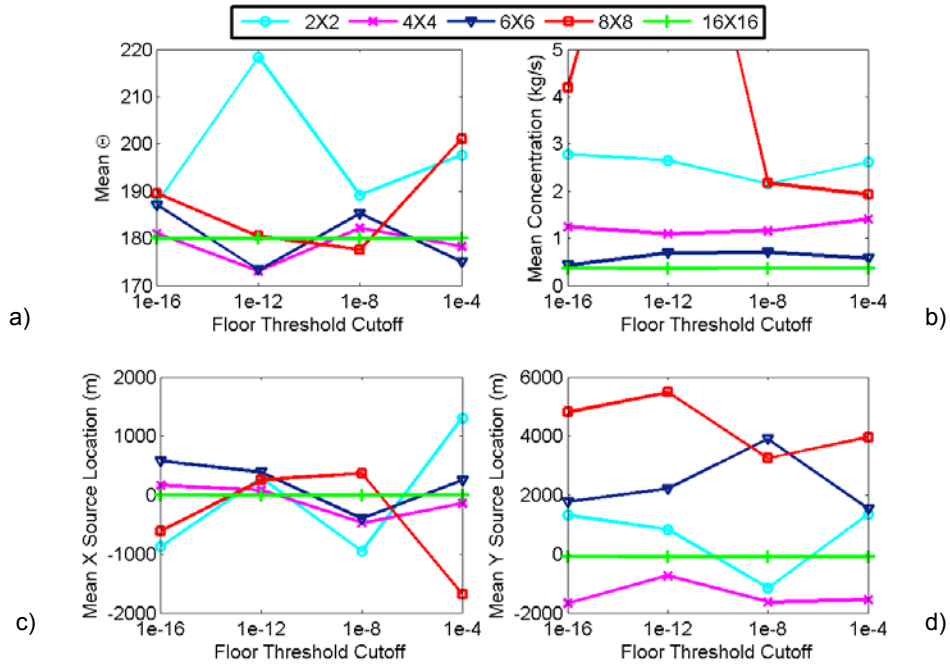

Figure 3. Results for all grid sizes and detection levels of the saturation level that at 1% of the maximum concentration strength value. Panel a shows the mean value of θ (wind direction, 180°), panel b indicates source strength (1 kg/s), and panels c and d show the location (0, 0) (in m) for x and y respectfully. All of these results are of 10 Monte Carlo runs of the median value of 10 individual runs.

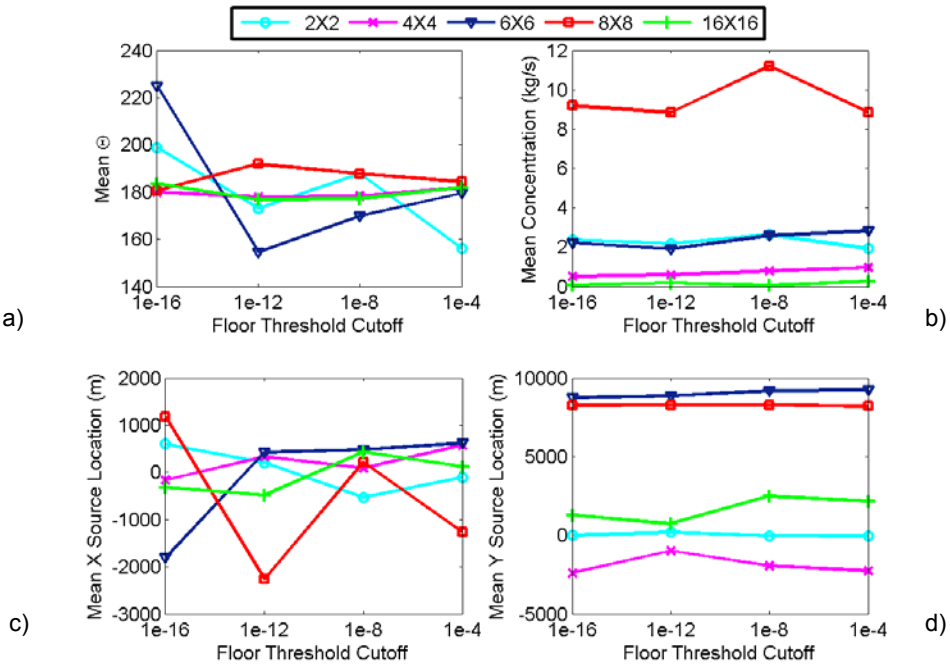

Figure 4. Results for all grid sizes and detection levels of the saturation level that at 0.1% of the maximum concentration strength value. Panel a shows the mean value of θ (wind direction, 180°), panel b indicates source strength (1 kg/s), and panels c and d show the location (0, 0) (in m) for x and y respectfully. All of these results are of 10 Monte Carlo runs of the median value of 10 individual runs.

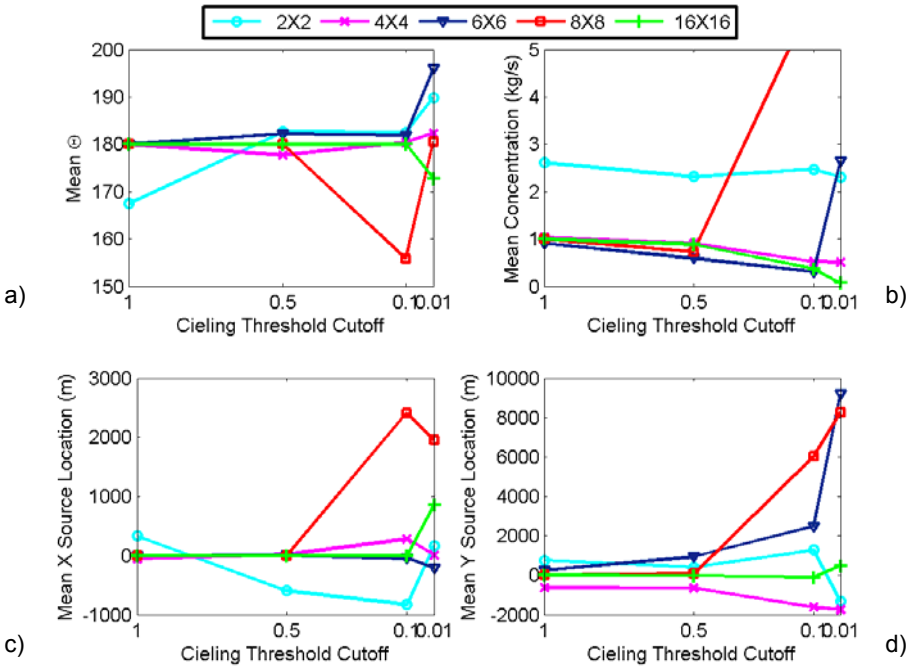

Figure 7. Results for all grid sizes and saturation levels of the detection level that at $1X10^{-16}$ of the maximum concentration strength value. Panel a shows the mean value of θ (wind direction, 180°), panel b indicates source strength (1 kg/s), and panels c and d show the location (0, 0) (in m) for x and y respectfully. All of these results are of 10 Monte Carlo runs of the median value of 10 individual runs.