8-1-2013

Stochastic Reconstruction of Multiple Source Atmospheric Contaminant Dispersion Events

Derek Wade
Boise State University

Inanc Senocak
Boise State University

Stochastic Reconstruction of Multiple Source Atmospheric Contaminant Dispersion Events

Derek Wade and Inanc Senocak

Department of Mechanical and Biomedical Engineering
Boise State University
1910 University Dr. Boise, ID 83725-2085

Abstract

Reconstruction of intentional or accidental release of contaminants into the atmosphere using concentration measurements from a sensor network constitutes an inverse problem. An added complexity arises when the contaminant is released from multiple sources. Determining the correct number of sources is critical because an incorrect estimation could mislead and delay response efforts. We present a Bayesian inference method coupled with a composite ranking system to reconstruct multiple source contaminant release events. Our approach uses a multi-source data-driven Gaussian plume model as the forward model to predict the concentrations at sensor locations. Bayesian inference with Markov chain Monte Carlo (MCMC) sampling is then used to infer model parameters within minutes on a conventional processor. The composite ranking system enables the estimation of the number of sources involved in a release event. The ranking formula allows plume model results to be evaluated based on a combination of error (scatter), bias, and correlation components. We use the 2007 FUSION Field Trial concentration data resulting from near-ground-level sources to test the multi-source event reconstruction tool (MERT). We demonstrate successful reconstructions of source parameters, as well as the number of sources involved in a release event with as many as three sources.

Keywords:
Event Reconstruction, Bayesian Inference, Source Term Estimation, Gaussian Plume Model

1. Introduction

Environmental awareness plays an important role in public safety, health, and threat mitigation. The release of harmful contaminants into the atmosphere could come by intentional or accidental means, and a quick response is key to limiting possible hazard to the population. Researchers have proposed event reconstruction (ER), also called source-term estimation (STE), methods (Annunzio et al., 2012a; Chow et al., 2008; Keats et al., 2007; Senocak et al., 2008; Stohl et al., 1998) that use contaminant concentration data from a network of well-placed sensors to characterize a dispersion event in terms of its source location

and emission rate. STE methods have been studied for many applications including defense and air quality management (Watson and Chow, 2004).

Most ER models adopt an inverse problem methodology along with a forward model to predict the plume dispersion. In cases where contaminant dispersion takes place over flat terrain on a scale of several kilometers or less, Gaussian plume models have been an effective forward model in ER methods (Senocak et al., 2008; Allen et al., 2007). At the continental scale with variable meteorological conditions, Monache et al. (2008) used the Lagrangian Operational Dispersion Integrator (LODI) as the forward model in a stochastic reconstruction method to determine the location of a radioactive release in Algeciras, Spain. At the urban neighborhood scale, Keats et al. (2007) adopted a computational fluid dynamics (CFD) model to better capture the effects of complex buildings on contaminant dispersion.

Researchers have adopted different methodologies to formulate a STE problem. Both deterministic and probabilistic algorithms have been proposed. By and large Bayesian inference methods form the basis for most of the probabilistic approaches. Johannesson et al. (2004) presented dynamic Bayesian models using both the well-established Markov chain Monte Carlo (MCMC) method and the sequential Monte Carlo for target tracking and atmospheric dispersion event reconstruction problems. Chow et al. (2008) extended the work presented in Johannesson et al. (2004) to neighborhood scale (building-resolved) atmospheric dispersion events using CFD models. Keats et al. (2007) combined a Bayesian inference method with an adjoint approach to reduce the computational time to reconstruct a release event in an urban environment using CFD based models. Senocak et al. (2008) developed a data-driven approach within a Bayesian inference framework whereby empirical turbulence diffusion parameters of the Gaussian plume model are estimated as part of the inverse problem in addition to characterizing the dispersion event. The practice led to substantial improvements over the empirically tuned Gaussian plume model.

Some researchers have favored a deterministic approach in which an optimization method is used to solve the inverse problem. Henze et al. (2009) discusses the use of adjoint models to inversely model $PM_{2.5}$ (particles with diameter less than 2.5μm) emissions. Akcelik et al. (2005) describes an optimization method which uses a conjugate gradient method to solve systems of partial differential equations. This method takes advantage of parallel computing to improve speed and efficiency of the otherwise lengthy optimizations for single-source event reconstructions. Another optimization method, proposed by Annunzio et al. (2012b), uses a Genetic Algorithm (GA) to carry out the optimizations in order to determine the source location of a single source release.

A contaminant dispersion event can involve releases from multiple sources. The source type may vary (e.g., point, line, area, volume) as well as the source elevation (e.g., ground level, stack, elevated line from aircraft). The release may also be categorized by the manner in which it is released, such as instantaneous (puff), continuous, or time-varying. Based on the methods described in Annunzio et al. (2012b), Annunzio et al. (2012a) introduced the Multi-Entity Field Approximation (MEFA) method for cases involving one or more ground-level point sources. With regards to continuous release scenarios, MEFA uses available wind data, and constrains any multiple releases to fall within a hazard area predicted by calculating the

spread far downwind for a single-source plume approximation. MEFA then searches within this hazard area for the optimal source locations while incrementing the number of possible sources. A cost function is used in part to determine the number of sources involved in the dispersion event. Field data is used to show that the method is capable of providing good approximations for multi-source events.

Platt and DeRiggi (2012, 2010) analyzed the blind predictions from STE models provided by eight different research groups, as applied to the FUSION Field Trials of 2007 (FFT-07) dataset (Storwold Jr., 2007). The comparative investigation provided useful information as to how well existing STE models perform relative to other STE models under different release scenarios. Platt and Deriggi applied a linear regression analysis to determine the significant factors that affected the reconstruction results obtained from various models. The present Bayesian inference method (Senocak et al., 2008) with a single-source, continuous release capability was also a part of the investigation. A subset of the results has revealed the advantages of a Bayesian inference method over other inverse methods that used the same forward model (i.e. Gaussian plume model).

Reconstruction of a multi-source contaminant release event is more challenging than reconstruction of a single source event. Yee has shown remarkable success using Bayesian inference techniques to reconstruct multi-source events with the number of sources unknown a priori (Yee, 2008, 2012a,b). Yee incorporates the unknown number of sources into the Bayesian inference framework in a principled fashion, which results in a posterior probability density for the number of sources. In our approach, we propose an alternative method to source number quantification by extending the Bayesian inference method presented in Senocak et al. (2008) to reconstruct contaminant dispersion events from multiple sources and couple it with a model ranking system. We adopt a data-driven multi-source Gaussian plume model as the forward model in the Bayesian inference method, and suggest a separate ranking system to estimate the number of sources involved in a release event. We apply the combined method to FFT-07 trial cases with up to three sources.

2. Forward Model

We adopt a data-driven Gaussian plume model as the forward model, because it is a suitable model for short range releases, over flat terrain under steady wind conditions, such as the FFT-07 trials considered in the present study. It is also computationally inexpensive. Therefore it can be used rapidly in the sampling process within the Bayesian approach. We are able to achieve accurate reconstructions in under two minutes on a conventional workstation with an Intel E8400 3.0 GHZ processor. Speed is an important aspect of STE when the intended use is first-response. Sophisticated forward models should be preferred for contaminant dispersion problems where a Gaussian plume model might not be suitable.

Stockie (2011) presents a derivation of the Gaussian plume model with single and multiple contaminant sources. For a single source release, the Gaussian plume model can be written as follows:

$$C_m(x,y,z) = \frac{Q}{2\pi U \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \times \left\{\exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right)\right\}, \quad (1)$$

where C_m is the concentration at location (x, y, z), Q is the rate of emission for the point source, U is the average wind speed, and H is the height of the release. We set z to 2m, the same height as the samplers used in the FFT-07 field experiments. In the FFT-07 trials, the contaminant was released from a near-ground-level source, therefore H is also set to 2m. H can also be set as an unknown and estimated using the Bayesian inference method as was shown in Senocak et al. (2008). Additionally, we combine $\frac{Q}{U}$ into a single parameter. The release rate, Q can then be estimated by calculating an average wind speed from local wind measurements at sensor height over the duration of the experiment.

We use an open-country Pasquill D type stability (Hanna et al., 1982) to define turbulent diffusion parameters σ_y and σ_z as follows:

$$\sigma_y = \zeta_y x(1 + 0.0001x)^{-0.5}, \sigma_z = \zeta_z x(1 + 0.0015x)^{-0.5} \quad (2)$$

where σ_y and σ_z are the standard deviations used in Equation 1 for the horizontal and vertical plume directions normal to the streamwise plume direction. Here, x refers to the distance along the streamwise plume direction. The parameters ζ_y and ζ_z are left as unknown parameters to be estimated by the Bayesian method, making the forward model a data-driven one. The practice results in significantly better estimates for the concentration field (Senocak et al., 2008; Senocak, 2010). Data-driven forward modeling gives better predictions than the baseline forward model when there are sufficient and reliable sensor data.

3. Bayesian Inference Method for Multi-Source Release Events

The Stochastic Event Reconstruction Tool (SERT) (Senocak et al., 2008) uses Bayesian inference to estimate information (i.e., source location, release height, emission rate, wind direction and speed) about the dispersion event. In this section, we present the Bayesian inference framework in SERT and extend it to multiple source releases. The number of sources involved in an event is then estimated separately using a ranking formula.

Generally speaking, the inverse problem can be formulated as follows:

$$\mathbf{m} \approx F^{-1}(\mathbf{d}), \quad (3)$$

where $\mathbf{d}$ is a vector of observed concentration values and $\mathbf{m}$ is a vector of forward model parameters to be estimated. F is the forward model, which is the Gaussian plume model in our case. Given the observed data, $\mathbf{d}$, our goal is to estimate forward model parameters, $\mathbf{m}$. In most Bayesian inference methods, Bayes' rule is simplified into the following proportionality:

$$P(\mathbf{m}|\mathbf{d}) \propto L(\mathbf{d}|\mathbf{m})P(\mathbf{m}), \quad (4)$$

where $P(\mathbf{m}|\mathbf{d})$ refers to the posterior probability density of the forward model parameters, $L(\mathbf{d}|\mathbf{m})$ is the likelihood function which calculates the likelihood of the observations given the model parameters, and $P(\mathbf{m})$ is the prior probability for the model parameters (Congdon, 2010).

Prior probabilities for model parameters are set based on certain expectations about each of the model parameters. All model parameters except Q/U and σ^2 are assigned proper uniform prior distributions. The normalized emission rate Q/U is given a Jeffrey's prior as follows:

$$p(Q/U) \propto \frac{1}{Q/U}. \tag{5}$$

To avoid division by zero we set a small minimum value for Q/U. Q/U is scaled using this minimum value to ensure that the maximum prior value is unity.

There are sensors capable of detecting trace amounts of a material in the atmosphere. But they have their limitations. Sensors can register a nominally zero value when, in fact, local concentration level, d_i, can be non-zero and below the detection threshold of the sensor. In such cases, we assign a probability to detecting a zero concentration level as follows:

$$d_i = \begin{cases} 0, & \text{with probability } \exp(-\alpha \cdot \hat{C}_i) \\ \xi_i, & \text{with probability } 1 - \exp(-\alpha \cdot \hat{C}_i) \end{cases} \tag{6}$$

where ξ_i is a concentration measured by a theoretically ideal sensor, d_i is the concentration measured by an actual sensor, and $\hat{C}_i$ is the concentration predicted by the model at the sensor location. Given the model parameters, ξ_i has a lognormal distribution with the following density:

$$p(\xi_i|\mathbf{m}) = \frac{1}{\sqrt{2\pi}\sigma\xi_i} \exp\left(-\frac{1}{2\sigma^2}(\ln \xi_i - \ln \hat{C}_i)^2\right), \tag{7}$$

When a sensor makes an observation at the sensor's detection threshold, C_{th}, we assume that it does so with a probability of $1/2$. Based on this assumption and Equation 6, α can be computed in the following manner:

$$1 - \exp(-\alpha \cdot C_{th}) = \frac{1}{2} \rightarrow \alpha = \frac{1}{C_{th}} \ln(2). \tag{8}$$

Given Equation 6, the conditional likelihood function is written as follows:

$$L(d_i|\mathbf{m}) = \begin{cases} \exp(-\alpha \cdot \hat{C}_i), & \text{if } d_i = 0 \\ \dfrac{1 - \exp(-\alpha \cdot \hat{C}_i)}{\sqrt{2\pi}\sigma d_i} \exp\left(\dfrac{1}{2\sigma^2}(\ln d_i - \ln \hat{C}_i)^2\right), & \text{if } d_i > 0 \end{cases}, \tag{9}$$

where σ^2, is the variance, which takes into account modeling and measurement errors cumulatively. We assume that the variance has an inverse gamma prior distribution with hyper

parameters $\alpha = 1.0$ and $\beta = 1000.0$.

SERT's previous design focused on single source continuous releases. In this study, we first modify the forward model to extend SERT to multiple source events. For a multi-source plume of a non-reactive and non-buoyant contaminant, the concentration at any point (x, y, z) is the sum of the contributions from each source (Stockie, 2011).

$$C_{total}(x, y, z) = \sum_{s=1}^{n} C(x'_s, y'_s; Q_s), \tag{10}$$

where n is the number of sources, and Q_s is the source emission rate. As in Stockie (2011), the shifted coordinates, x'_s and y'_s, are defined as follows:

$$x'_s = x - X_s, y'_s = y - Y_s, \tag{11}$$

where, x and y are the Cartesian coordinates, X_s and Y_s are the coordinates of source s. The origin is shifted to the source location, (X_s, Y_s), and the positive x-direction extends in the downwind direction.

Next, we introduce additional parameters required by the multi-source model into the Bayesian inference framework. Hereinafter we will refer to the multi-source event reconstruction tool as MERT. For multiple source releases, we define a reference source, and all other sources are defined relative to the reference source based on the distance to the source, d, and an angle, φ, measured from the global x-axis, as shown in Figure 1. Each source has its own emission rate normalized by the mean wind speed, $\frac{Q}{U}$. For example, the complete set of forward model parameters for a dual source model can then be written as follows:

$$\mathbf{m} = \left[x_{s1}, y_{s1}, \left(\frac{Q}{U}\right), \theta, \zeta_y, \zeta_z, \sigma^2, d_2, \varphi_2, \left(\frac{Q}{U}\right)_2 \right], \tag{12}$$

where (x_{s1}, y_{s1}) is the primary source location, and θ is the wind direction. We use Markov chain Monte Carlo (MCMC) sampling with the Metropolis Algorithm (Metropolis et al., 1953) to estimate the posterior distribution of the model parameters. In our approach, the candidate state is sampled from a Gaussian distribution centered on the current state.

Figure 1 shows a dual source plume with sufficient distance between two sources, such that overlap of the plumes does occur downstream and yet the sources are not too close together to consider them as a single entity. We assume that the distance between the two sources, d, is relatively small compared to the size of the search region. Therefore, for the current study with a sensor grid that covers an area of approximately 500m by 500m with 50m spacing between sensors, we set an upper limit of 5 times the spacing between sensors as the maximum cross-wind distance allowable between sources. If the sources are farther apart than this upper limit, they can be treated as individual single-source events in the present study. This reasoning also extends to sources that are extremely close to each other in the cross-wind direction, such that plumes overlap heavily to behave as a single source release. Therefore, a lower limit of one fifth of the spacing between sensors is used, below which we assume that plumes overlap and can be considered a single source release.

The additional parameters, d and φ, are used to calculate the location of the second source, (x_{s2}, y_{s2}), relative to the reference source. Equations 13 and 14 show the conversion from polar to rectangular coordinates with respect to the location of the primary source. It is not necessary to specify a primary source prior to the sampling process, because a source location, (x_{s1}, y_{s1}), is estimated from the MCMC sampling process, which will then serve as the reference source for other sources. Note that the other source locations are calculated using the estimated parameters d and φ. The polar configuration allows for additional sources to branch off of the primary source.

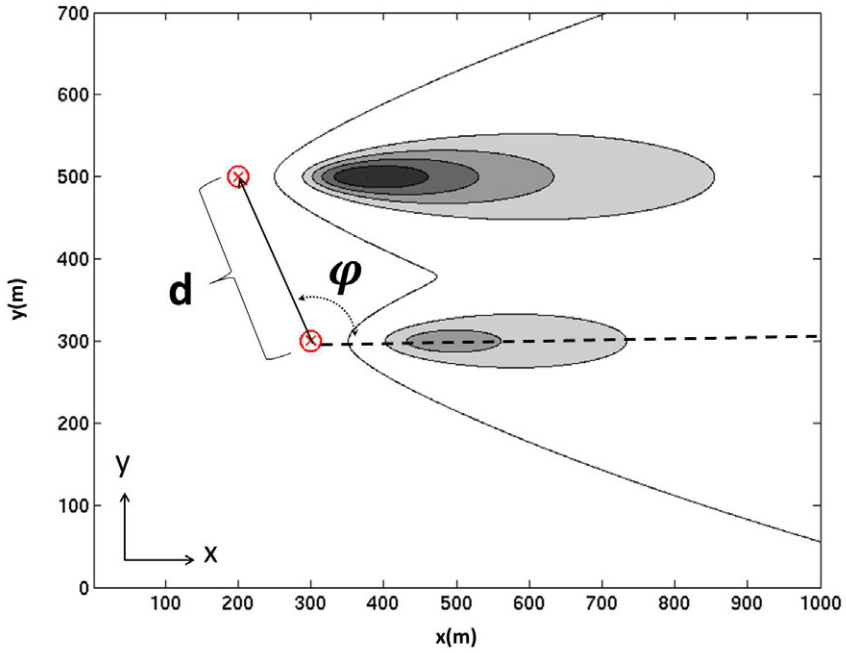

Figure 1. Sample dual source Gaussian plume colored by contaminant concentration at 2m above ground level. Source locations are shown as circles, d is the distance between sources, and φ is the angle between sources with respect to the positive x-axis.

$$x_{si} = x_{s1} + d_i \cos(\varphi_i) \tag{13}$$

$$y_{si} = y_{s1} + d_i \sin(\varphi_i) \tag{14}$$

where $i = 2, 3, ..., N$ and N is the maximum number of possible sources.

4. Composite Ranking to Determine the Number of Sources

Concentration or dosage of contaminant measured at the sensors can be an outcome of releases from single or multiple sources. However, in the ER problem we do not know the number of sources involved in a dispersion event, even for a single source release. A concentration field resulting from a multiple source release can come close to matching a concentration field from a single source with a different emission rate and source location.

The Bayesian framework that we presented in the previous section does not provide any inference on the number of sources involved in a dispersion event. Therefore, we propose a composite ranking approach to estimate the number of sources involved in an event. The ranking system is independent of the Bayesian inference to locate the source and emission rate. We consider single, dual, and three-source releases, but the overall method is applicable to more than three sources. In our approach, we execute MERT for each release possibility independently. Once the runs are completed, we extract the most probable value for each of the model parameters from the corresponding posterior probability distributions. We then run the forward model using the most probable parameters to calculate the concentrations at each of the sensor locations. We designate these model concentrations as $\hat{C}$ and compare them to the measured concentrations, C, for single, dual, and three-source assumptions separately using a ranking method.

In atmospheric dispersion applications, it is typical to use multiple performance metrics to effectively evaluate the predictive capability of a dispersion model. Researchers suggested a variety of metrics (Stohl et al., 1998; Pullen et al., 2005; Svensson, 1998; Chang et al., 2003; Hanna et al., 1993). We propose a composite ranking model that is inspired by the recent Environmental Protection Agency protocol to determine the best performing air quality model (EPA, 2012). The literature is mostly in agreement that error (scatter), bias, and correlation are important metrics in the evaluation process, all of which are included in some form in the global statistics portion of Mosca et al. (1998). Each of these metrics is weighted equally in our ranking model to decide whether a specific model achieves better results using a single or multiple source setting. We then identify the setting with the higher ranking as the release event containing the correct number of sources.

Our ranking model has three parts. The first component of the model's rank is the FAC2, which is a quantity measuring the fraction of predictions that fall within a factor of two of the corresponding observations (Chang et al., 2003), as shown in Equation 15 . This operation is performed to obtain a measure of error, or scatter, when comparing the observed and predicted values.

$$\text{FAC2} = \text{fraction of data for which } 0.5 \leq \frac{\hat{C}}{C} \leq 2.0, \tag{15}$$

where C is the observed concentration at the sensor and $\hat{C}$ is the estimated concentration calculated by using the most probable parameters, obtained from posterior distributions, in the forward model.

The next performance metric used in the ranking model is the Fractional Bias (FB). The FB is used to indicate a bias towards underprediction or overprediction of concentration data by the model. It has been used as a validation parameter for other dispersion models and is a robust indicator of model performance (Stohl et al., 1998). The FB ranges from -2 (extreme underprediction) to +2 (extreme overprediction), and 0.0 is a perfect score for this component. As part of the United States Environmental Protection Agency's (EPA) performance evaluation protocol (EPA, 1992), the FB is defined as follows:

$$FB = 2\left(\frac{\bar{C} - \bar{\hat{C}}}{\bar{C} + \bar{\hat{C}}}\right), \tag{16}$$

where $\bar{C}$ is the average measured concentration across all sensors, and $\bar{\hat{C}}$ is the average of the predicted concentrations computed by the model at all sensor locations.

The final component to our ranking model is the Pearson's Correlation Coefficient (R), which contributes a measure of correlation to the ranking model. R ranges from -1.0 to +1.0 with +1.0 corresponding to "perfect positive correlation"(EPA, 2012). An R value close to 0.0 would indicate that the predicted data and the measured data are not related. R is defined as follows:

$$R = \frac{\sum_i (C_i - \bar{C}) \cdot (\hat{C}_i - \bar{\hat{C}})}{\left[\sqrt{\sum_i (C_i - \bar{C})^2}\right]\left[\sqrt{\sum_i (\hat{C}_i - \bar{\hat{C}})^2}\right]} \tag{17}$$

The three components described above are combined to form the following ranking model

$$RANK = FAC2 + \left(1 - \frac{|FB|}{2}\right) + R^2 \tag{18}$$

The ranking model contains a measure of error (scatter), bias, and correlation in a composite fashion. These metrics provide a concise and quantitative description of how well the model performs with a varying number of sources. The composite rank ranges from 0 to 3, with 3 corresponding to a perfect score. The higher the $RANK$, the better the model did at matching the concentration predictions with sensor observations. We use the highest ranking model to make a decision on the correct number of sources involved in the dispersion event.

5. Results

In 2007, the Defense Threat Reduction Agency (DTRA) proceeded to address some of the unmet requirements in the current Joint Effects Model (JEM), which is to be used as the standard hazard prediction model at the Department of Defense (Storwold Jr., 2007). One of these requirements was to evaluate source term estimation models used to detect chemical and biological (CB) activity and estimate the characteristics of the source(s) in question. A large data set, FFT-07, was created for the evaluation and improvement of STE algorithms. The FFT-07 database provides detailed meteorological information and trace gas concentration measurements for short range (500m) dispersion experiments. These experiments were performed with single and multiple sources for continuous and puff (instantaneous) releases.

5.1. Evaluation with FFT-07 Trials

We use data from Trials 7, 27, 28, and 40 of FFT-07. In trials 27 and 40, there are two sources with different tracer emission rates. Trial 7 is a single source trial that we include

in our study to demonstrate that the ranking model will identify the correct number of source terms, even in a single source case. Trial 28 is a three source case. The true source locations and emission rates are known from the field data and used to assess the accuracy of the reconstructed model parameters. In working with the FFT-07 concentration data, we ignored sensor data that reports an error message for more than 50% of the sampling time.

In FFT-07, a grid of 100 digital photoionization detectors (digiPID) were spaced evenly on a square grid at 50m apart and 2m above the ground. A tracer of propylene gas was released from multiple locations at approximately 2m above ground and at constant flow rates for approximately 15 minutes per trial. We time-averaged the concentration data from sensors for the continuous release trials, which are the focus of the present study.

FFT-07 Trial 40 had very few poor readings (sensors reporting an error more than 50% of the time). This abundance of reliable sensor data and fairly uniform wind conditions resulted in reconstructions of the source locations that are approximately 8 and 6 meters from the true source locations, as seen in Figure 2. For this case, 48 of the 100 sensors are used in the estimation. Note that we include all positive reading sensors. Figure 3 shows the tight posterior distributions for the two source locations in which the true values fall within, or very close to, the 50% contour line. This inner contour line encompasses 50% of the posterior samples and the outer contour line includes 90% of the posterior samples. The range of each cell is normalized in both the horizontal and vertical direction with the limits corresponding to the minimum and maximum values for each parameter in the posterior samples. The normalization enables us to assess accuracy in percentage form in a global fashion over the parameter space. The plots along the diagonal show the marginal distributions of each parameter. Trial 40 is a good example of successful reconstruction from reliable sensor data.

Figure 4 shows a comparison of results from FFT Trial 40, where the left image uses a single source setting, the middle uses a dual source setting, and the third image uses a three source setting. The predicted values for the dual source setting (middle) match more closely to the data measured by the sensors. A perfect match would lie directly on the solid diagonal line running through the origin. This view of the data allows us to see the difference between an estimate with a single, dual, an three source setting. It also shows the points which fall within a factor of two of the observed values (FAC2) as well as the over or underestimation (Bias). From this figure, we can visually deduce that the dual source setting is most likely the correct answer, but we need a quantitative measure. Therefore, we proposed a composite ranking model as described in Section 4.

The more reliable the sensor data, the more accurate the reconstruction will be. However, operational data may be less reliable than desired. Hence we use Trial 27 from the FFT-07 data set, which has much less reliable data than the Trial 40, to test how less reliable or sparse data affects the reconstruction.

Figure 5 shows the layout of the 57 sensors used in Trial 27, as well as the true and estimated source locations. We observe that the most probable source locations are approximately 15 and 25 meters from the true source locations. The true source locations are illustrated with squares and the source location estimates with $\times$'s. Ideally, the estimates would fall directly on the true locations (e.g. $\times$'s on top of squares). The distances may not

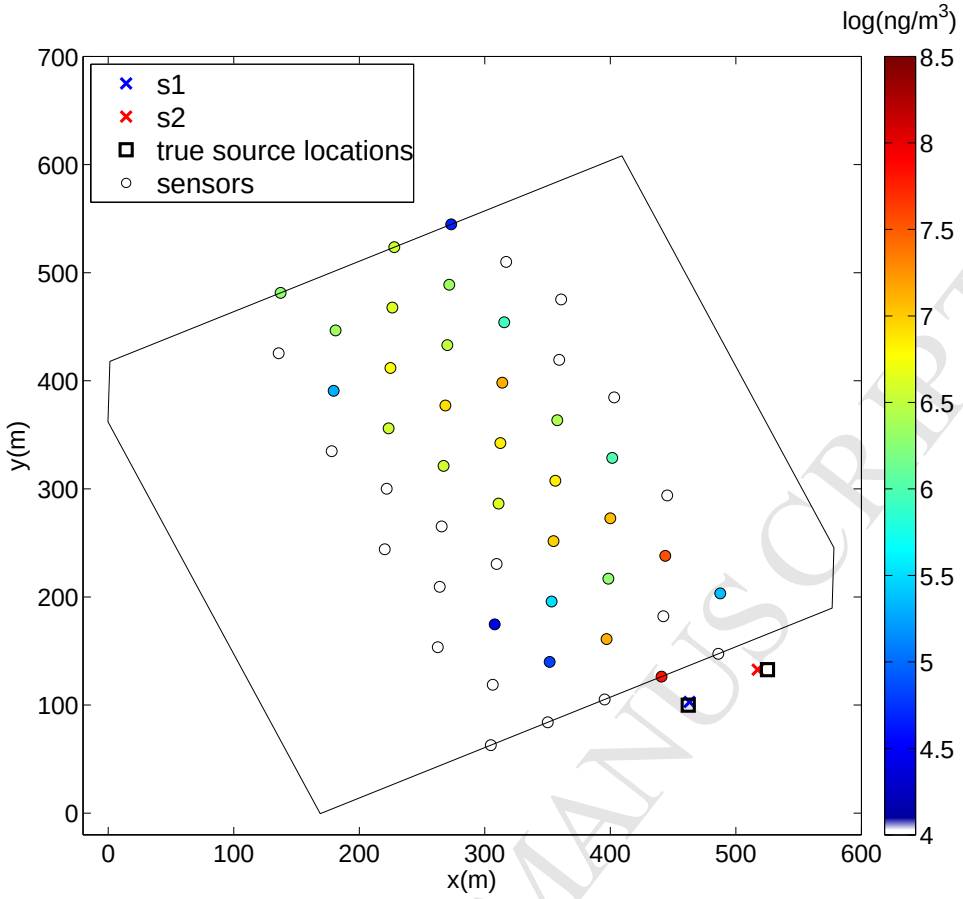

Figure 2. FFT-07 Trial 40 source location estimates (s1 and s2) with approximate errors of 8m and 6m. Sensors reporting nominally zero concentration are colored white. Large rotated box is the FFT-07 sensor network boundary.

be ideal (and not as small as Trial 40), but can still be very useful from an operational point of view since they would at least put any rapid response personnel in close proximity to the true locations.

FFT-07 Trial 28 is a three-source release event and a similar layout plot can be seen in Fig. 6. In this figure we can see that estimated source locations for sources 1 and 3 are fairly close to the true values, but the estimated source location for source 2 is approximately 48m from the true source location. We do note, however, the estimated sources are in a somewhat linear arrangement, as are the true source locations, and they are of approximately correct spacing with respect to one another.

5.2. Composite Ranking Model Results

Thus far, we have presented reconstruction of source locations and emission rates for dual and three-source releases. We have not made an attempt to estimate the number of sources involved in a dispersion event. The composite ranking model that we proposed in Section 4

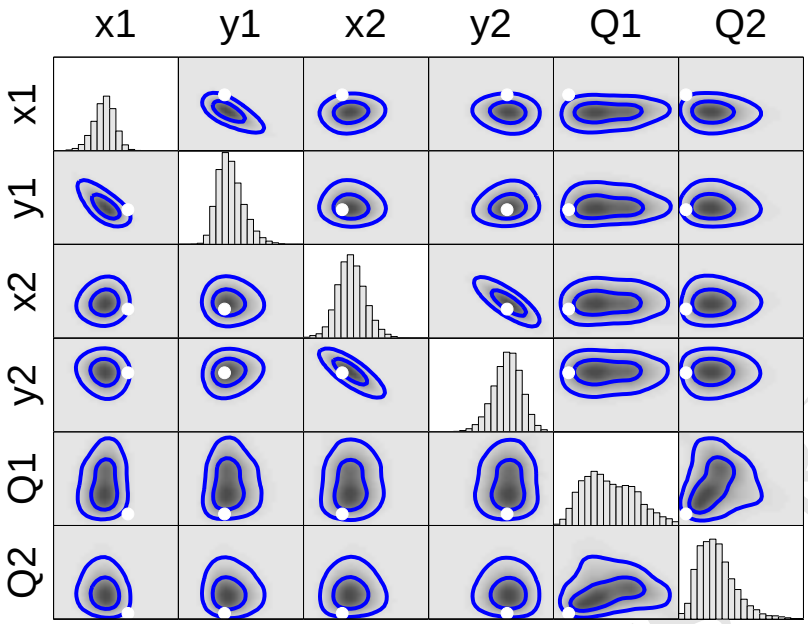

Figure 3. FFT-07 Trial 40 bivariate posterior distributions for location and emission strength. The range of each cell is normalized with respect to the minimum and maximum values for each parameter and distances can be viewed as percent error. The plot is colored by probability density and the darkest regions are the most probable. The outer contour encompasses 90% of the posterior samples and the inner contour includes 50%. White markers represent true values.

enables us to estimate the number of sources. We calculate $RANK$ using Equation 18. The components that make up the $RANK$ are: FAC2, FB, and R.

Figure 7 shows the composite ranking for all cases in this study, and is colored by contribution from each component in the rank. A rank of 3.0 corresponds to a perfect score. For all the cases tested, the model with the correct number of sources ranked higher. For instance, Trial 7 is a single source release, and our ranking model gives the highest score to the single source assumption correctly. In all the other cases the correct source-number assumption received the highest score, as expected.

6. Conclusions

We have extended a Bayesian inference method to reconstruct single-source contaminant release event, SERT, to reconstruct near-ground-level multiple-source release events, MERT. We proposed a composite ranking system to identify the number of sources involved in an event. The ranking formula is independent of the Bayesian method and can potentially be adopted in other event reconstruction methods.

We have applied the combined approach to releases from up to three sources, but the method can be extended to more than three sources. In the Bayesian framework we used a data-driven Gaussian plume model where turbulent diffusion parameters are inferred given the concentration data. The practice significantly improves the performance of the standard

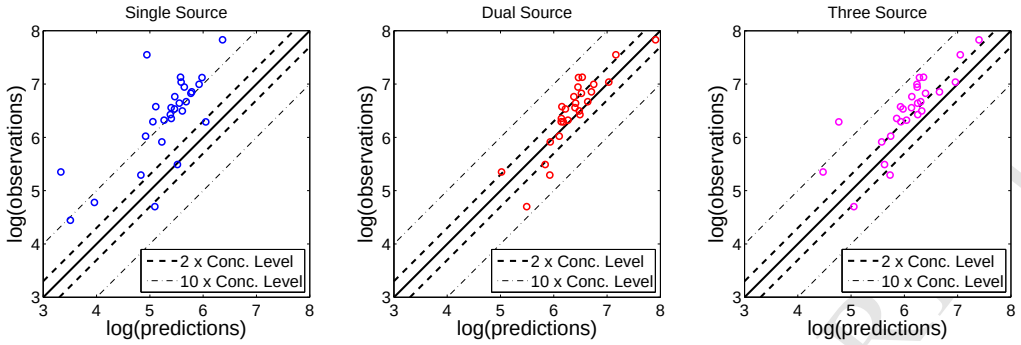

Figure 4. Observed sensor concentrations for FFT-07 Trial 40 vs. computed sensor concentrations using the most probable model parameters.

Gaussian plume model. However, for complicated dispersion events, sophisticated dispersion models should be preferred as the forward model.

Acknowledgments

This work is supported by a grant from the U.S. National Science Foundation (Award #1043107). Authors would like to thank the anonymous reviewer for his contribution to the final form of the composite ranking formula.

References

Akcelik, V., Biros, G., Draganescu, A., Hill, J., Ghattas, O., Waanders, B., 2005. Dynamic data-driven inversion for terascale simulations: Real-time identification of airborne contaminants, in: SC '05 Proceedings of the 2005 ACM/IEEE conference on Supercomputing.

Allen, C.T., Young, G.S., Haupt, S.E., 2007. Improving pollutant source characterization by better estimating wind direction with a genetic algorithm. Atmospheric Environment 41, 2283–2289.

Annunzio, A., Young, G., Haupt, S., 2012a. A multi-entity field approximation to determine the source location of multiple atmospheric contaminant releases. Atmospheric Environment 62, 593–604.

Annunzio, A., Young, G., Haupt, S., 2012b. Utilizing state estimation to determine the source location for a contaminant. Atmospheric Environment 46, 580–589.

Chang, J., Franzese, P., Chayantrakom, K., Hanna, S., 2003. Evaluations of CALPUFF, HPAC, and VLSTRACK with two mesoscale field datasets. Journal of Applied Meteorology 42, 453–466.

Chow, F., Kosovic, B., Chan, S., 2008. Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. Journal of Applied Meteorology and Climatology 47.

Congdon, P., 2010. Applied Bayesian Hierarchical Methods. Chapman and Hall/CRC.

EPA, 1992. Protocol for Determining the Best Performing Model. Technical Report EPA-454/R-92-025. United States Environmental Protection Agency.

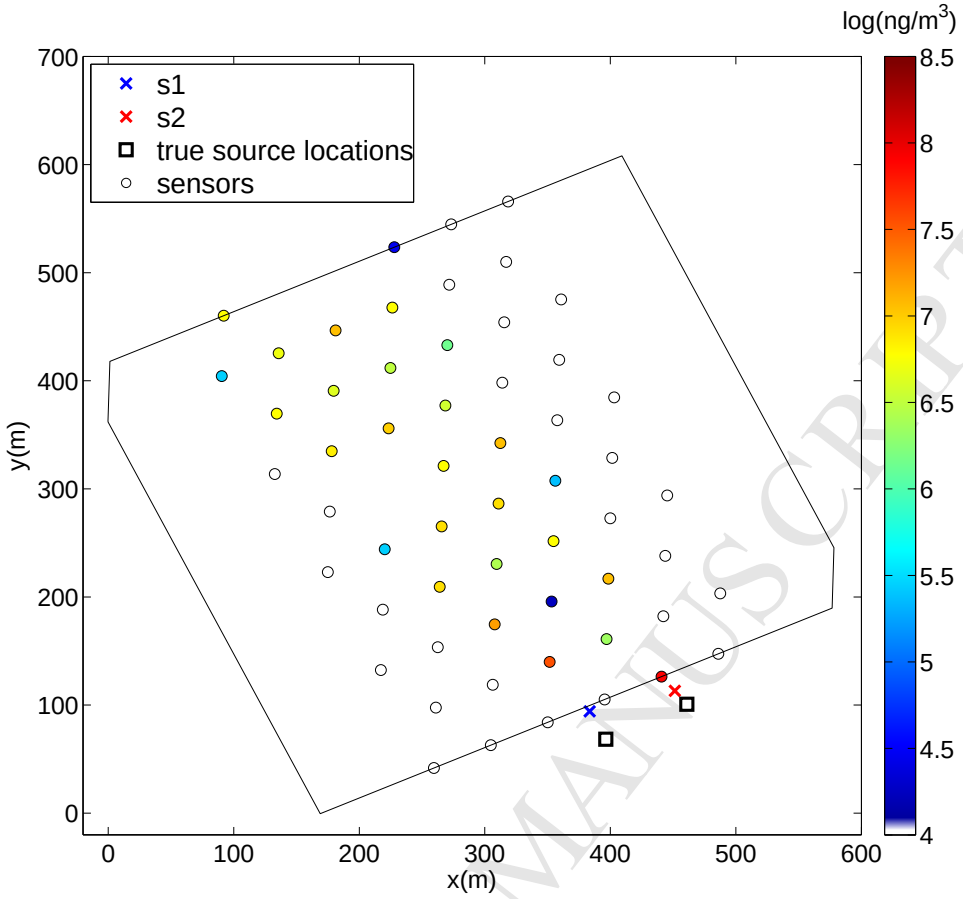

Figure 5. FFT-07 Trial 27 source location estimates (s1 and s2) with approximate errors of 15m and 25m. Sensors reporting nominally zero concentration are colored white. Large rotated box is the FFT-07 sensor network boundary.

EPA, 2012. Documentation of the Evaluation of CALPUFF and Other Long Range Transport Models Using Tracer Field Experiment Data. Technical Report EPA-454/R-12-003. United States Environmental Protection Agency.

Hanna, S., Briggs, G., Hosker Jr., R., 1982. Handbook on Atmospheric Diffusion. U.S. Department of Energy: Office of Energy Research. doe/tic-11223 edition.

Hanna, S., Chang, J., Strimaitis, D., 1993. Hazardous gas model evaluation with field observations. Atmospheric Environment 27A, 2265–2285.

Henze, D., Seinfeld, J., Shindell, D., 2009. Inverse modeling and mapping US air quality influences of inorganic $PM_{2.5}$ precursor emissions using the adjoint of GEOS-Chem. Atmospheric Chemistry and Physics 9, 5877–5903.

Johannesson, G., Hanley, B., Nitao, J., 2004. Dynamic Bayesian Models via Monte Carlo - An Introduction with Examples. Technical Report UCRL-TR-207173. Lawrence Livermore National Laboratory.

Keats, A., Yee, E., Lien, F., 2007. Bayesian inference for source determination with applications to a complex urban environment. Atmospheric Environment 41, 465–479.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, E., 1953. Equations of state calculations by fast

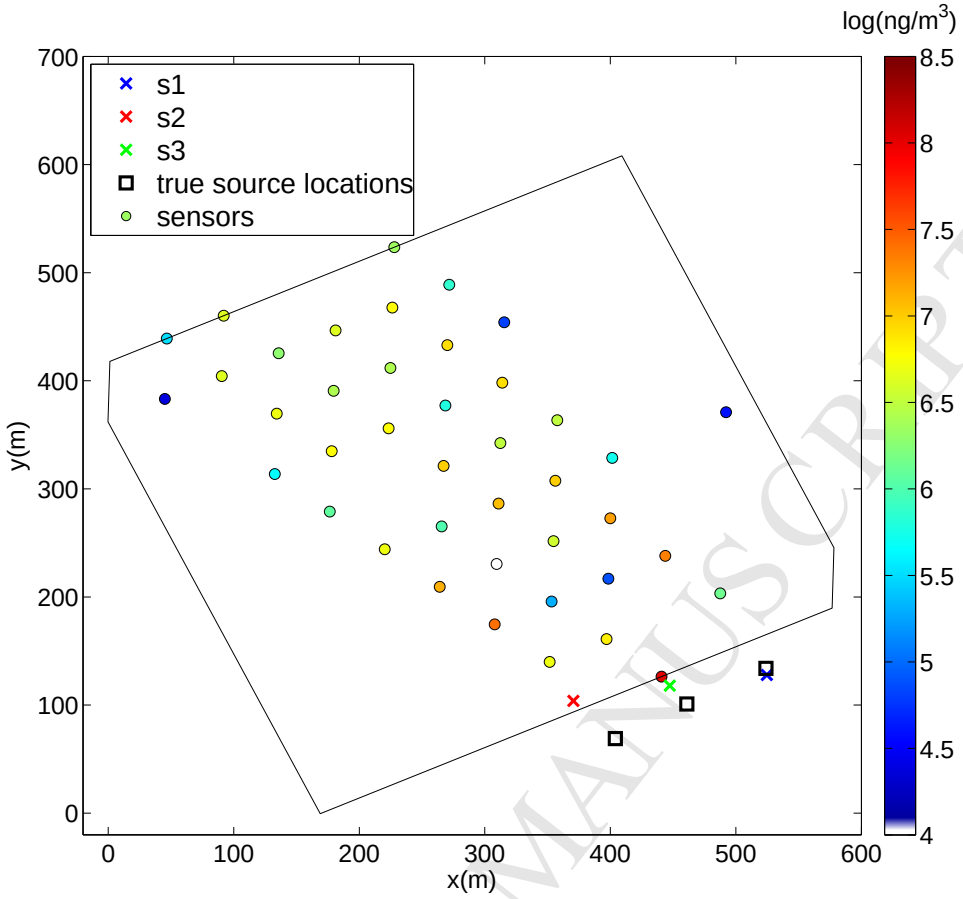

Figure 6. FFT-07 Trial 28 source location estimates (s1, s2, and s3) with approximate errors of 7.5m, 21.7m, and 48.3m, respectively. Sensors reporting nominally zero concentration are colored white. Large rotated box is the FFT-07 sensor network boundary.

computing machines. Journal of Chemical Physics 224, 560–586.

Monache, L.D., Lundquist, J., Kosovic, B., Johannesson, G., Dyer, K., Aines, R., Chow, F., Belles, R., Hanley, W., Larsen, S., Loosmore, G., Nitao, J., Sugiyama, G., Vogt, P., 2008. Bayesian inference and markov chain monte carlo sampling to reconstruct a contaminant source on a continental scale. Journal of Applied Meteorology and Climatology 47, 2600–2613.

Mosca, S., Graziani, G., Klug, W., Bellasio, R., Bianconi, R., 1998. A statistical methodology for the evaluation of long-range dispersion models: An application to the ETEX exercise. Atmospheric Environment 32, 4307–4324.

Platt, N., DeRiggi, D., 2010. Comparative investigation of source term estimation algorithms using FUSION field trial 2007 data. AMS 16th Conference on Air Pollution Meteorology .

Platt, N., DeRiggi, D., 2012. Comparative investigation of source term estimation algorithms using FUSION field trial 2007 data: linear regression analysis. International Journal of Environment and Pollution .

Pullen, J., Boris, J., Young, T., Patnaik, G., Iselin, J., 2005. A comparison of contaminant plume statistics from a gaussian puff and urban CFD model for two large cities. Atmospheric Environment 39, 1049–1068.

Senocak, I., 2010. Application of a bayesian inference method to reconstruct short-range atmospheric disper-

sion events, in: Bayesian Inference and Maximum Entropy Methods in Science and Engineering, American Institute of Physics, Chamonix, France.

Senocak, I., Hengartner, N., Short, M., Daniel, W., 2008. Stochastic event reconstruction of atmospheric contaminant dispersion using bayesian inference. Atmospheric Environment 42, 7718–7727.

Stockie, J., 2011. The mathematics of atmospheric dispersion modeling. SIAM Review .

Stohl, A., Hittenberger, M., Wotawa, G., 1998. Validation of the lagrangian particle dispersion model flexpart against large-scale tracer experiment data. Atmospheric Environment 32, 4245–4264.

Storwold Jr., D., 2007. Detailed Test Plan for the FUsing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT-07). Technical Report WDTC-TP-07-078. United States Army.

Svensson, G., 1998. Model simulations of the air quality in athens, greece, during the MEDCAPHOT-TRACE campaign. Atmospheric Environment 32, 2239–2268.

Watson, J., Chow, J., 2004. Receptor models for air quality managment. Environmental Management , 27–36.

Yee, E., 2008. Theory for reconstruction of an unknown number of contaminant sources using probabilistic inference. Boundary-Layer Meteorology 127, 359–394.

Yee, E., 2012a. Inverse dispersion for an unknown number of sources: Model selection and uncertainty analysis. ISRN Applied Mathematics .

Yee, E., 2012b. Probability theory as logic: Data assimilation for multiple source reconstruction. Pure and Applied Geophysics 169, 499–517.

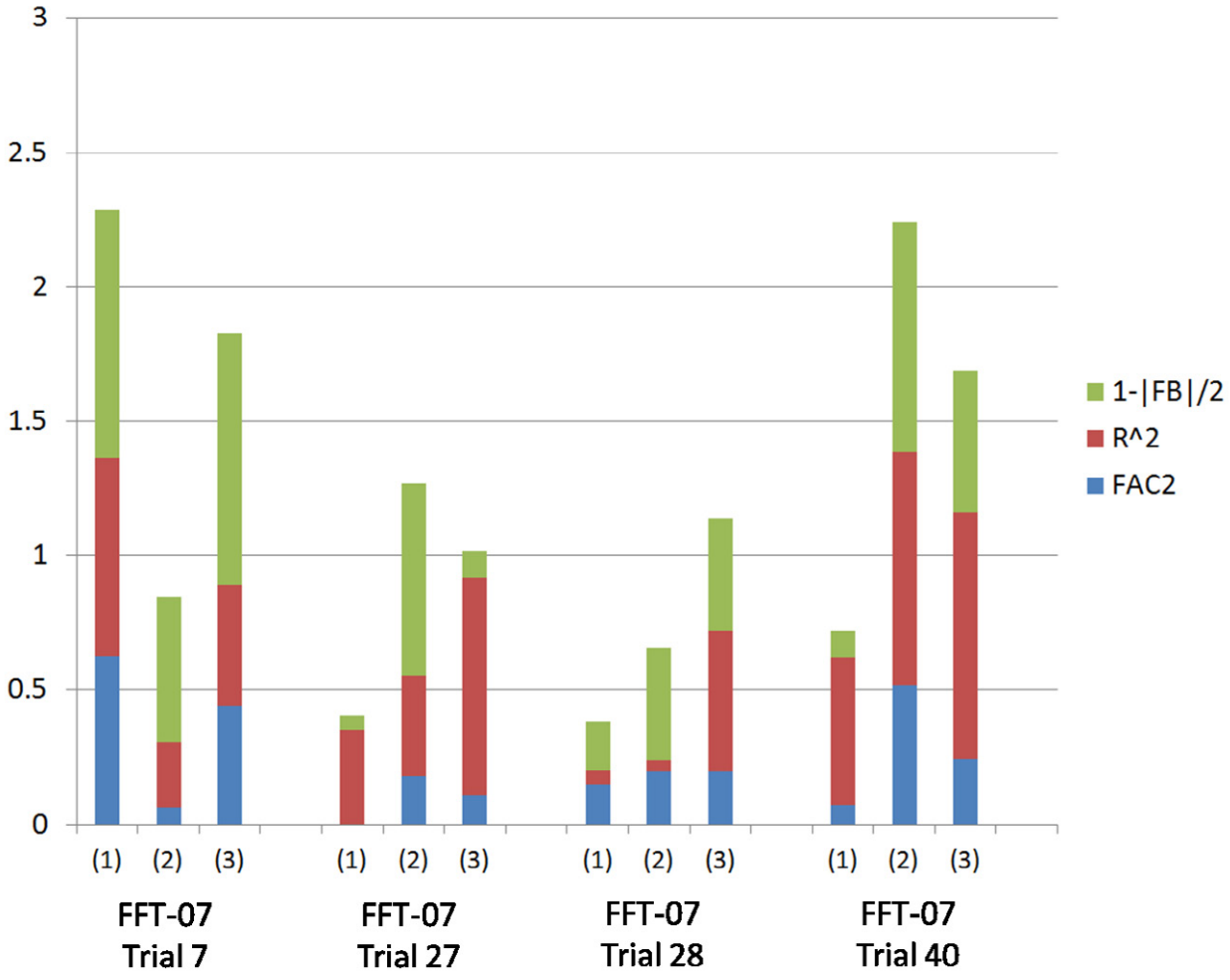

Figure 7. Ranking for each case tested. Colors correspond to individual rank components (e.g. R^2, FAC2, $1 - |FB|/2$) as shown in legend. (1),(2),(3) refer to single, dual, and three source settings, respectively. FFT-07 Trials 27 and 40 are truly dual source releases. Trial 7 is a single source release and Trial 28 is a three source release.