J1.4 BACK TRAJECTORIES FOR HAZARD ORIGIN ESTIMATION: BACKHOE

George S. Young*, James A. Limbacher, Sue Ellen Haupt, Andrew J. Annunzio
The Pennsylvania State University, University Park, Pennsylvania

1. INTRODUCTION

Source characterization for an unknown contaminant puff release is usually achieved by inverting a transport and dispersion model, such as done by a genetic algorithm system (Allen et al. 2007, Haupt et al. 2007, Long et al. 2008) or Markov chain Monte Carlo Chow et al. 2008; Monarch et al. 2008, Keats et al. 2007). Such techniques test a statistically generated sequence of trial inputs to the forward model for how well they cause it to reproduce observed concentrations. There are, however, meteorological situations in which the deformation component of transport dominates the spread of an airborne contaminant, allowing successful back-calculation without recourse to a sequence of forward model runs. This situation arises when the puff size is approximately that of the dominant eddies, in which case the puff is deformed by those eddies faster than it is dispersed by the smaller eddies. Two common settings in which this occurs are the convective boundary layer (for puffs with a horizontal scale similar to the boundary layer depth) and in the mid-latitude troposphere (for puffs with a horizontal scale similar to that of baroclinic cyclones (Stohl 1996). In these settings back trajectories computed for contaminated air parcels tend to converge on the source location as time is rewound back to that of the release. Convergence is not perfect of course, both because of dispersion by smaller eddies and because the wind field in the dominant eddies will not be perfectly resolved by observations.

The use of back trajectories to locate the source of an airborne contaminant has been tested in a number of settings ranging from source-receptor studies (e.g., Occhipinti et al. 2008) to investigations of radioactive contamination (e.g., Lee et al. 2004). The method often works well on the synoptic scale for two reasons. First, for a continuous source, the time evolving flow ensures that parcels contaminated at different times follow different trajectories. Thus, parcels backtracked from multiple sensor-based contaminant reports will follow different paths, all of which will eventually cross the source location. Second, for an instantaneous source, the deformation in the synoptic scale flow often spreads contaminant trajectories (e.g., Ishikawa 1995). In this situation back tracking until the trajectories come together will reveal the source location.

* *Corresponding author address:* George S. Young, The Pennsylvania State University, Department of Meteorology, 503 Walker Building, University Park, PA, 18802; e-mail: young@meteo.psu.edu

Because boundary layer turbulence exhibits both rapid time evolution and large deformation this back trajectory approach can also be applied on scales similar to those of the largest boundary layer eddies, i.e. for transport distances of a few hundred meters to a few kilometers. This new application of back trajectory analysis is the focus of the current study.

2. PROCEDURES

For situations such as those described above we have developed a new algorithm: a back trajectory model for hazard origin estimation (BackHOE). The essence of the method is to backtrack those air parcels known from sensor observations to be contaminated. The method is tested here in the boundary layer setting, with transport times on the scale of minutes and distances of less than a kilometer. For this study we use observations obtained from the turbulence-resolving Eulerian-Semi-Lagrangian (EULAG) large eddy simulation (LES) model Smolarkiewicz and Margolin 1997) run of the Fusing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT07) domain (Platt and Warner 2008). Simulated concentration sensors are deployed at 20 m intervals in a 280 by 280 m area within the 2230 by 2230 m computational domain. Each sensor produces one measurement every 10 seconds.

A new back trajectory was initiated from every sensor measurement that exceeded a user specified concentration threshold of 0.2 percent of the maximum concentration. Thus, a single analysis of source location involves computation of hundreds of back trajectories, each extending a number of minutes back from the time of measurement. The trajectory calculations are limited to the horizontal plane of surface wind observations since vertical velocity data and detailed upper-level winds are rarely available in contaminant release situations. This, approximation limits the validity of BackHOE to those situations where the surface flow is divergent, i.e. where the back trajectories would be held against the surface by overlying downdrafts. This assumption imposes little additional restriction on the basic method because it requires a highly deforming flow in order to for the back trajectories to converge and, in boundary layer settings, such flows typically occur under convective downdrafts.

The winds necessary to compute the back trajectories are obtained from simulated surface observation stations collocated with the concentration sensors within the computational domain of the LES model. The concentration data are generated by driving a Lagrangian particle model (Biazar et al. 2005) with the horizontal wind field from the LES model. An instantaneous release of 350,000 particles at 1 m

altitude is used for his study. The source location was varied between runs to determine how BackHOE's performance depends on the turbulence structure into which the contaminant is released.

Each step of a parcel's back trajectory is calculated by interpolating the modeled surface winds in time and space to the current parcel location and then advecting the parcel upwind at that velocity for a time step of 10 seconds. An adjustment factor is available to the surface winds to obtain the advecting wind. This factor corrects for the increase in wind speed with altitude its effect on the advection of the turbulently deepened puff. Since our test data is derived from a purely two-dimensional Lagrangian particle model, this factor is set to 1.0 for the current study.

If a parcel advects off the observation grid the frozen wave approximation is applied to determine the advecting wind. In this situation each of the available wind grids is advected upwind at the mean wind speed for a number of seconds equal to the difference between its valid time and the current time along the back trajectory. Of those advected wind grids which contain the parcel, that with the closest valid time is used as a source for wind data to continue the parcel's advection along the back trajectory.

In settings where deformation dominates dispersion these tracked parcels should come together at the source location. Several metrics for diagnosing maximum convergence of the cloud of contaminated parcels, thus deducing the source location, were developed and tested. For a single source release the clumping of the parcels around their spatial centroid (i.e. mean position) indicates that the back trajectories have reached the source location. A number of clumping metrics were tested. The simplest is the mean-squared-distance (known here as position variance) between the parcels and their centroid. A similar metric based on the mean-squared-distance perpendicular to the mean wind direction allows the parcel clouds to deform along the wind direction while contracting in the crosswind direction. For each of these metrics the source location was diagnosed as the centroid location at the time when the metric reached its minimum value.

For diagnostic visual interpretation we tested two forms of spatial histogram. The first form depicts the number of parcels passing through each cell of a spatial grid. The histogram is first computed for each time step and then the maximum count for each grid cell is used in the final map. The grid size is adjusted to obtain a smooth contour analysis on this composite map. The second form weights each parcel's contribution to the count by the concentration measurement which triggered its creation. For both forms, the grid box with the highest count is the geographic location of the maximum convergence of the contaminated parcels.

For a multi-source release, application of the scalar metrics described above requires that we first apply cluster analysis to associate each parcel with a particular cloud and then quantify the evolution of each cloud's spread along its back trajectory. The K-means clustering algorithm (Wilks 2006) is used for this purpose. At each time step along the back trajectories each parcel is reassigned by the clustering algorithm. Parcels tend to remain with the same cluster however, since at each new time step the clustering algorithm is seeded with the cluster centroids from the previous step. The position variance metrics and histogram maps are then computed separately for each cluster. It is also possible to compute count-weighted or concentration-weighted position variance between cluster centroids instead of individual parcels. These metrics quantify the convergence of multiple clusters back to a single source, minimizing the need to match the number of clusters to the number of releases though *a priori* knowledge.

Since the number or sources is not known, a method was also developed for determining the source locations when the model is run with more clusters than sources. This is achieved by ranking the clusters in order of their minimum spread, i.e. how well each converges to a single source location. Those clusters which fail to converge along their back trajectory are attributed to air parcels that were deformed into the vicinity of the contaminant cloud and subsequently mixed into the cloud. Those clusters that do converge are examined and their convergence locations grouped in space and time. The number of groups indicates the number of sources.

BackHOE was tested for a number of source locations for both single and multiple sources in an effort to determine the conditions under which the method performed well and those for which it failed. Results are presented for three cases: single-source success, two-source success and a single-source failure. In all cases the surface wind is from the northwest. The wind field and sensor locations remain unchanged between case studies, only the source number and locations are changed.

3. RESULTS

Results of case study experimentation show that BackHOE succeeds in characterizing an unknown instantaneous source in those meteorological situations where puff deformation by the transporting wind field poses challenges for conventional forward-model-based artificial intelligence techniques. The histogram map in Figure 1 depicts the parcel cloud's back trajectory for a single-source release. The wind is from the northwest, so the back trajectories extend northwest from the square region gridded with sensors. The wind field deformation pulls the parcels together as they approach the source (going backward in time) and then apart again after they pass the source. The diagnosed source location, shaded deep red on the map, is outside the area sampled by the sensor grid. Figure 2 shows this same information from a quantitative statistical perspective, plotting a time history of the spatial spread (i.e. intra-cluster location variance) of each of two clusters of parcels. The cluster which tracks to the release point achieves a much lower minimum spread than the other, thereby identifying it as the correct choice for source location. The plot includes results from five BackHOE runs, each for a different start time, to demonstrate the repeatability of the results. For

these same five runs, Figure 3 shows the location of the cluster centroids at the time each cluster achieves minimum spread. The cluster with the smallest minimum was able to consistently locate the source to within a few decameters despite the source being located 15 decameters outside of the concentration sensor grid.

The second case study illustrates BackHOE's ability to distinguish and locate two sources whose evolving contaminant clouds overlap in the sensor grid. Figure 4 shows the histogram map of concentration-weighted parcel locations for the back trajectory to a pair of sources. The back trajectory clouds are somewhat more compact (i.e. narrow in the crosswind direction) than in the unweighted parcel histogram map shown in Figure 1. Nonetheless, both graphics accomplish the task of defining the location of the source or sources. For this second case, the maximum convergence of parcel locations occurs in two positions, one along each of the swaths of parcel back trajectories. One BackHOE source location estimate is directly under the right-hand sources and the other just to the southwest of the left-hand source.

In both of these successful cases, the releases took place under the areas of subsidence that fill the majority of the convective boundary layer (Young 1988). In contrast, BackHOE fails, as expected, when a release takes place under a convective updraft. In such a release the contaminant is lofted away from the surface, reducing the immediate risk to downwind surface locations. Although posing less of a threat to downstream areas, such releases violate the basic assumption of BackHOE that trajectories remain in the surface layer. This violation, coupled with the horizontal convergence of the surface layer flow into the convective updraft causes the back trajectories to diverge rather than converging on the true source location. The histogram map shown in Figure 5 illustrates such a failure in a single-release case study. Since the flow is converging, the back trajectories diverge. While the densest congregation of back trajectories does pass quite close to the true source location others fan out over a wide area. Moreover, the densest congregation of parcel locations occurs when the contaminant cloud passes over the sensor grid rather than at the time of release. Thus, BackHOE fails in those situations where convergent surface layer flow lofts the contaminant cloud away from surface.

4. CONCLUSIONS

Extension of the back trajectory of source-receptor analysis from synoptic scale problems in long-range transport to turbulence scale short-range contaminant source location problems is shown to be feasible. The BackHOE implementation introduced here was tested on synthetic wind and concentration data from a turbulence-resolving LES model that provides the flow field for a Lagrangian particle model. For the high-impact cases where the hazardous release occurs under a boundary layer downdraft and so remains in the surface layer, BackHOE is able to distinguish and locate multiple sources. It retains this capability even when the contaminant clouds overlap and when the source lies at moderate distances outside the sensor domain. In contrast, for the lower-impact cases where contaminants are released into a lofting updraft, BackHOE's basic assumptions are violated so it fails to locate the source.

BackHOE's success depends on sophisticated metrics of trajectory convergence and, in the case of multiple sources, on clustering of parcel trajectories. Multiple metrics are presented, the most successful of which are intra-cluster location variance and concentration-weighted parcel histogram maps.

The back trajectory source location problem is probably easier for continuous sources than for the instantaneous sources studied here, because eventually divergent flows will be sampled. As with prior, synoptic scale, back trajectory source-receptor studies, the ability to monitor a plume in an evolving flow field will eventually reveal the location where the majority of the back trajectories cross.

Acknowledgements: This research is supported by DTRA (contract numbers W911NF-06-C-0162 and DTRA01-03-D-0010-0012). We particularly thank John Hannan and Christopher Kiley for valuable feedback throughout this project. Luna Rodriguez provided code to read the data and Kerrie Long offered helpful suggestions.

References

Allen, C. T., G. S. Young, and S. E. Haupt, 2007: Improving Pollutant Source Characterization by Optimizing Meteorological Data with a Genetic Algorithm. *Atmos. Environ.*, **41**, 2283-2289.

Biazar, A.P., R.T. McNider, K. Doty, W.B. Norris, D. Moss, 2005: Modeling Activities at UaH in Support of TexAQSII: A final report to the Houston Advanced Research Center. 47 pp.

Chow, F.K., B. Kosovic, and S. Chan, 2008: Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations, submitted to *J. Appl. Meteorol. and Climatol.*

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.

Ishikawa, H., 1995: Evaluation of the Effect of Horizontal Diffusion on the Long-Range Atmospheric Transport Simulation with Chernobyl Data. *J. Appl. Meteor.* **34**, 1653–1665

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

Lee L. Y. L., R. C. W. Kwok, Y. P. Cheung, K. N. Yu, 2004: Analyses of airborne ^{7}Be concentrations in Hong Kong using back-trajectories. *Atmos. Environ.*, **38**, 7033-7040.

Long, K.J., S.E. Haupt, and G.S. Young, 2008: Improving Meteorological Forcing and Contaminant Source Characterization Using a Genetic Algorithm. To be submitted to *Optimization and Engineering.*

Monache, L.D., J.K. Lundquist, B. Kosovic, G. Johannesson, K.M. Dyer, R.D. Aines, F. K. Chow, R.D. Belles, W.G. Hanley, S.C. Larsen, G.A. Loosmore, J.J. Nitao, G.A. Sugiyama, P.J. Vogt, 2008: Bayesian inference and Markov Chain Monte Caro sampling to reconstruct a contaminant source at continental scale, submitted to *J. Appl. Meteorol. and Climatol.*

Occhipinti, C., V.P. Aneja, W. Showers, D. Niyogi, 2008: Back-trajectory analysis and source-receptor relationships: Particulate matter and nitrogen isotopic composition in rainwater. *J. Air & Waste Manage. Assoc.*, **58**, 1215-1222.

Platt, N., and S. Warner, 2008: Plan for Initial Comparative Investigation of Source Term Estimation Algorithms using FUSION Field Trial 2007 (FFT07), Institute for Defense Analyses Document D-3488, 34 pp.

Smolarkiewicz, P.K., and L.G. Margolin, 1997: On forward-in-time differencing for fluids: An Eulerian/semi-Lagrangian non-hydrostatic model for stratified flows. Atmos.–Ocean, XXXV, 127–152.

Stohl, A., 1996b: Trajectory statistics—A new method to establish source–receptor relationships of air pollutants and its application to the transport of particulate sulfate in Europe. *Atmos. Environ.,* **30,** 579–587.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*, Academic Press, Burlington, MA, 627 pp.

Young, G. S., 1988: Turbulence structure of the convective boundary layer II: PHOENIX 78 aircraft observations of thermals and their environment. *J. Atmos. Sci.*, **45**, 727-735.

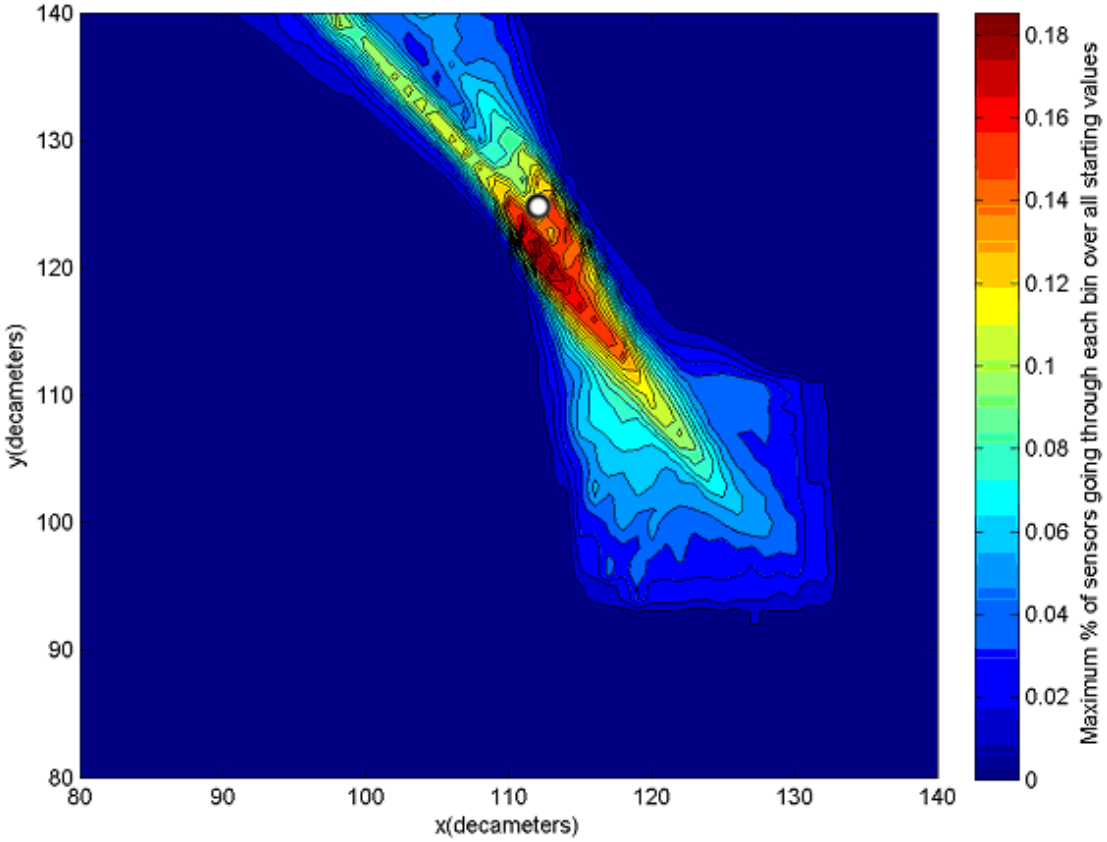

Figure 1. Histogram map of parcel locations for the back trajectory to a single source. The bullseye marks the true source location while the maximum convergence of parcel locations just to its south marks the BackHOE estimate of the source location.

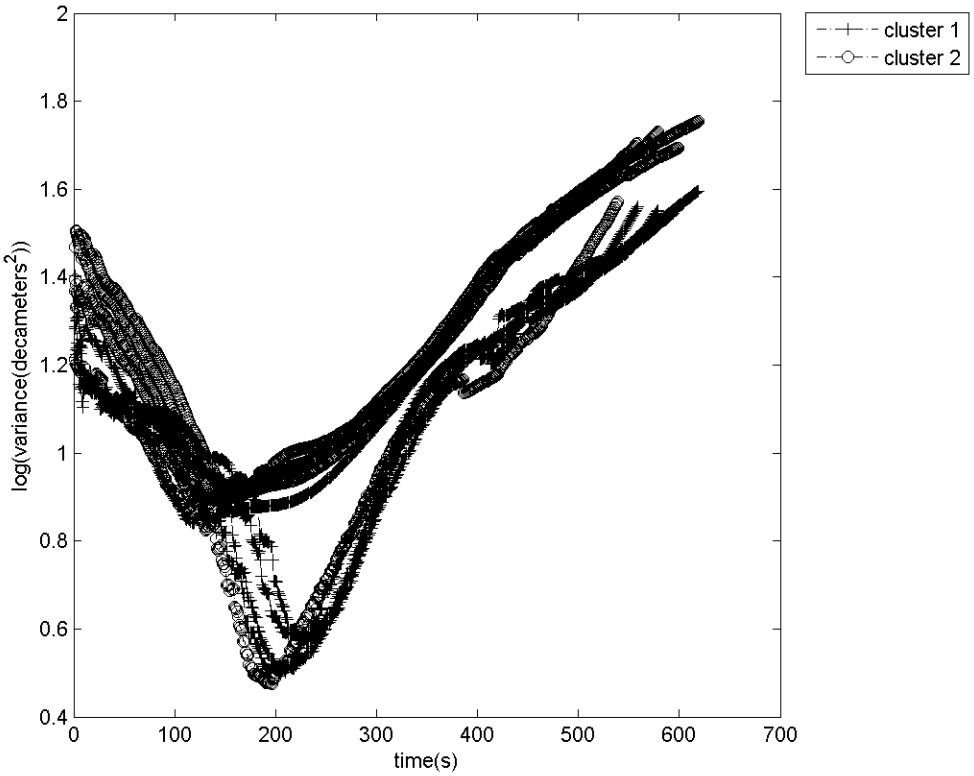

Figure 2. The time series of parcel cluster spreads, i.e. position variance around the cluster centroids, for a two cluster analysis of a single source location. BackHOE was run for 5 different start-times to demonstrate the repeatability of the results. In each case the cluster with the smallest minimum spread tracks close to the source, reaching a minimum spread near its point of closest approach.

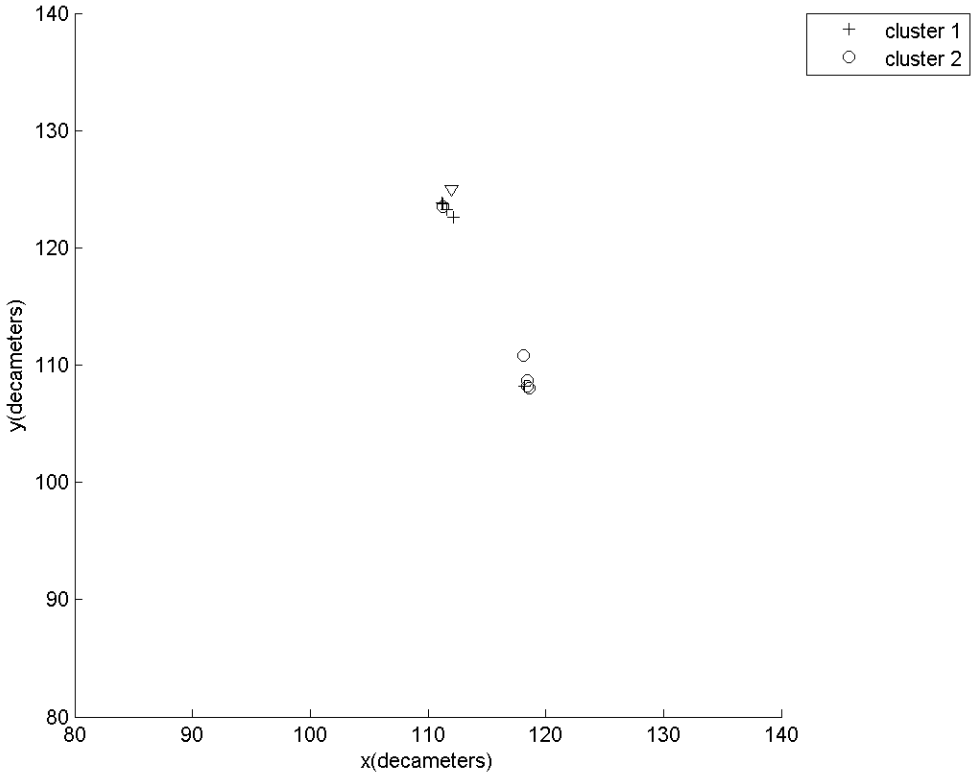

Figure 3. A map of the locations at which each cluster achieves its minimum spread. The true source location is shown as a triangle. The cluster with the smallest minimum spread achieved that value within a few decameters of the true source location. In contrast, the other (large-spread) cluster suggests a source position that is in error by almost 20 decameters. Thus, the magnitude of the intra-cluster spread correctly identifies both the cluster that tracks to the source and the position along that track at which the source is located.

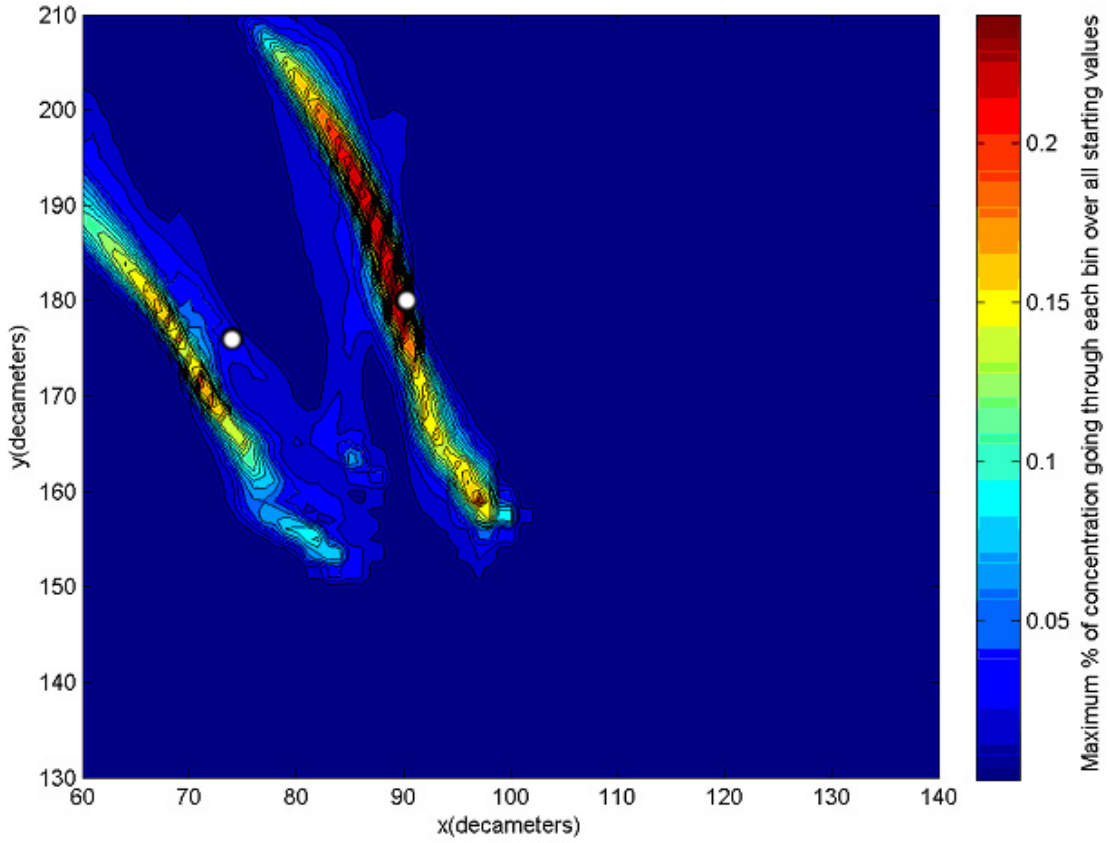

Figure 4. Histogram map of concentration-weighted parcel locations for the back trajectories to a pair of sources. The bull's-eyes mark the true source locations. The maximum convergence of parcel locations occurs in two positions, one directly under the right-hand source and the other just to the southwest of the left-hand source.

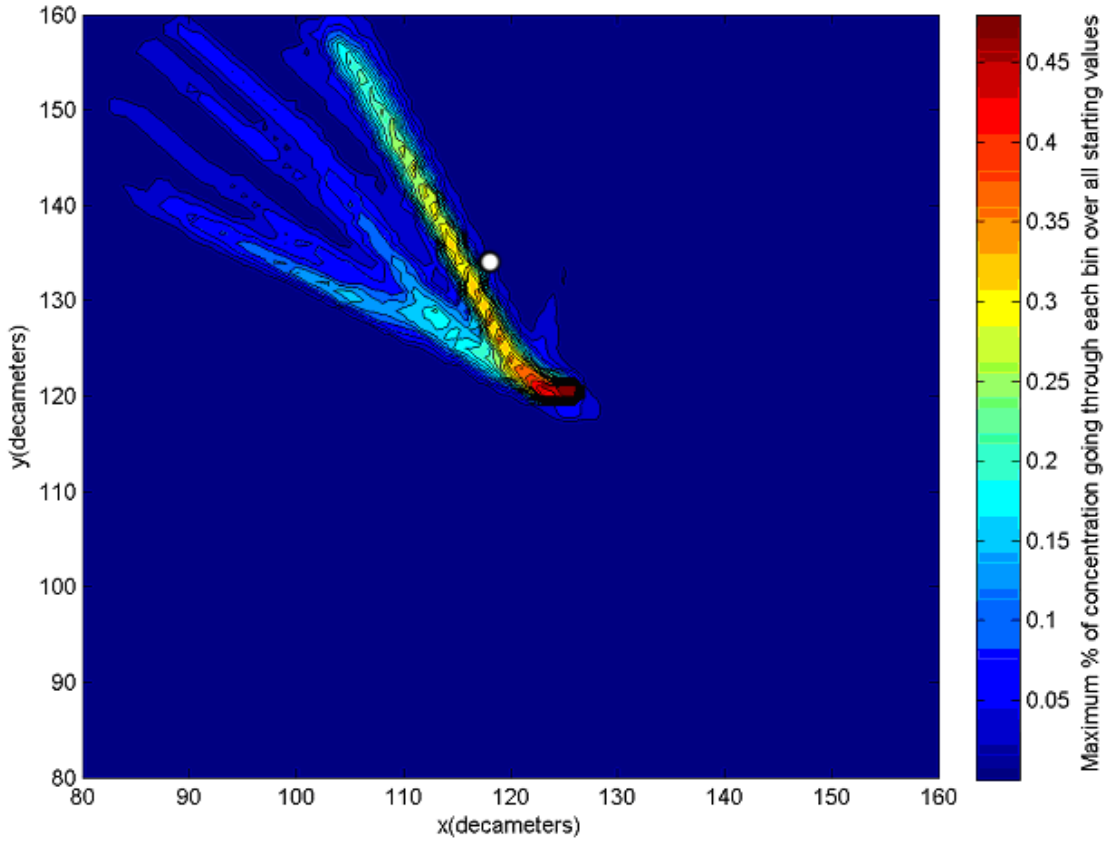

Figure 5. Histogram map of concentration-weighted parcel locations for the back trajectory to a single source. The bullseye marks the true source location while the maximum convergence of parcel locations near the start-point of the back trajectories marks the BackHOE estimate of the source location.