Institute for Defense Analyses

4850 Mark Center Drive • Alexandria, Virginia 22311-1882

12th International Conference on *Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*

Cavtat, Croatia

October 6-9, 2008

Evaluation Plan for Comparative Investigation of Source Term Estimation Algorithms Using FUSION Field Trial 2007 Data

Nathan Platt (nplatt@ida.org)

Steve Warner (swarner@ida.org)

Steve Nunes (snunes@ida.org)

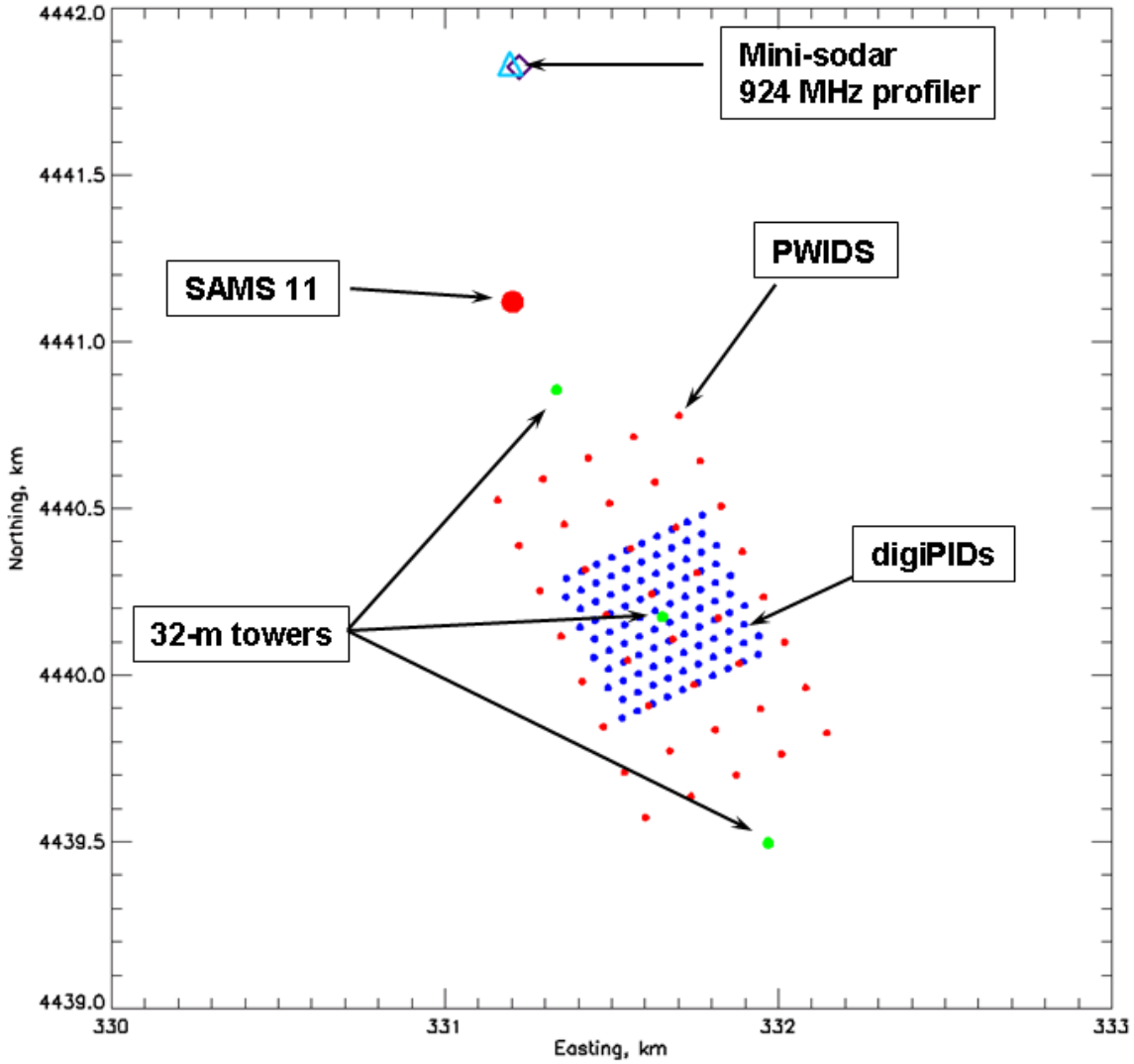
October 8, 2008

Outline:
Highlights of the STE Algorithm Evaluation Plan

- Motivation
- Evaluation Plan Structure
- Sample Creation of Simulated Chemical Sensor Output
- Outputs to be Provided by Modelers
- Comparative Evaluation Metrics

STE = Source Term Estimation

FFT 07 Field Trial Illustration



STE Algorithm Evaluation Plan

based in recently distributed IDA paper

We are expecting about 5-7 sets of predictions.

IDA Document D-3488, *Plan for Initial Comparative Investigation of Source Term Estimation Algorithms Using FUSION Field Trial 2007 (FFT 07),* March 2008

What is covered in the plan?

- **Idea of writing formal "Evaluation Protocol" document was suggested by Jim Bowers at Dugway Proving Ground meeting in late October 2007**
- **Draft STE Algorithm Evaluation Plan prepared by IDA with DPG help**
  - Purpose
  - Data Distribution
    - » Stages of evaluation
    - » Description of cases for each stage
    - » Creation of simulated chemical sensor output
    - » Meteorological input options
    - » Data "scrambling"
    - » Sequence of events / schedule
  - Outputs to be provided by modelers
  - Comparative evaluations metrics and plan

STE Algorithms Evaluation Protocol is Needed

- **To best allow for scientific insights from comparative analyses**
- **To provide for credible and fair comparisons among algorithms (in a *reasonably* realistic setting)**
  - To avoid perceived intentional, or more likely unintentional, model parameter tweaking to fit the unique data and observations of FFT 07
  - To give the most credible assessment of the state-of-the-art
- **To best allow information to be re-used for independent validation in the future (with newer algorithms)**
- **To clarify maturity of emerging STE algorithms for possible inclusion into Joint Effect Model**

Some uncertainty in source information should be maintained.

Basic SDF Parameters to Examine:
(covered by FFT07)

- **Release type**
  - instantaneous or continuous
- **Release time**
  - daytime or nighttime
- **Total number of sensors**
  - four or sixteen, includes nulls
    - » Compromise between “technical” and “operational” evaluation
- **Number of sources**
  - single, double, triple, and quadruple
- **Quality of sensor output**
  - binary sensors, “bar” sensors with 8 levels of discrimination, and future sensors capable of producing continuous concentrations
- **Available meteorological information (“MET”)**
  - close-in single tower profile at the center of the grid
  - MET measured some distance away
    - » Mini-SODAR + 924 MHz Profiler + SAMS 11 site

Need a careful balance between number of cases to run SDF versus adequate sampling of the parameters of interest affecting algorithm performance

Stages of the Evaluation

- **Run evaluation in up to three stages with each stage involving different level of simulated chemical sensor output**
  1. Continuous series of concentrations (JCAD)
  2. Bar-like threshold sensors (ACADA/CAM/ICAM)
  3. Binary threshold sensors (ACADA present day use)

  } ***Optional***

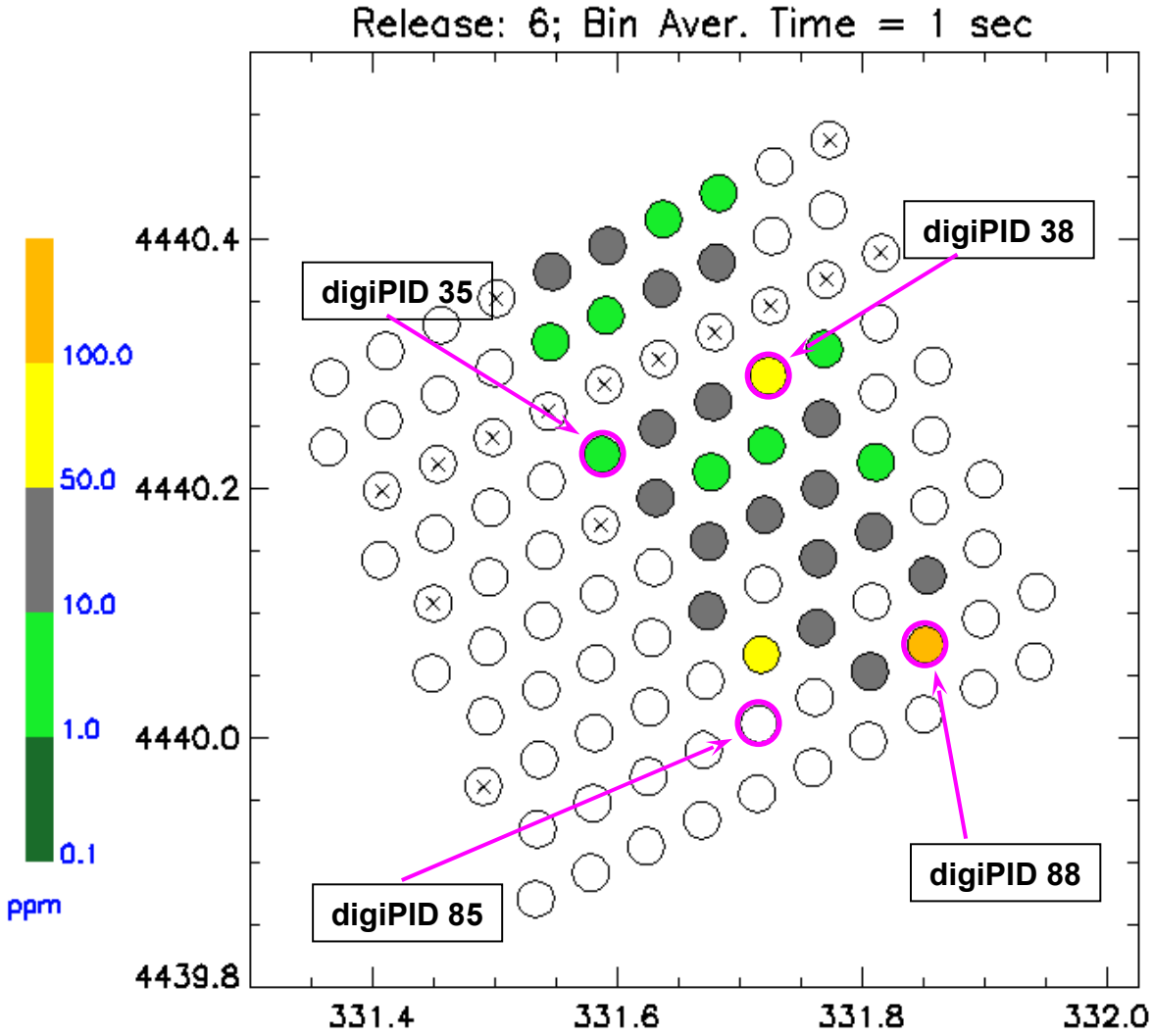
- **Each stage of the evaluation to consist of *104* individual cases for which predictions are sought**
  - Each stage is described in the document

If needed additional stages could be added at the end of this evaluation

Selection of Samplers and Creation of Simulated Sensor Output

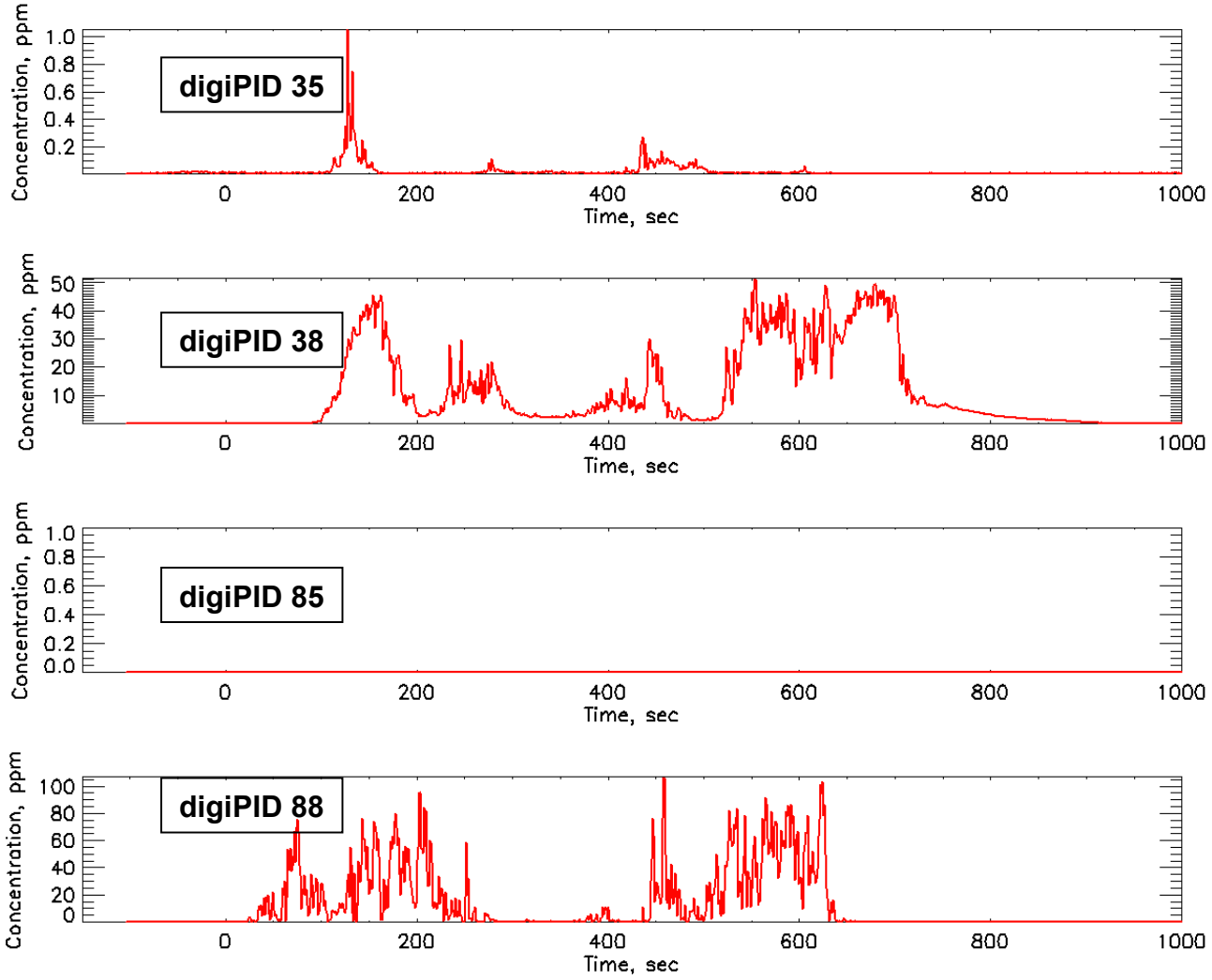
Selection of Samplers

(digiPIDs in this case)



Notional Demonstration on Creating Simulated Sensor Output for Stage 1 Evaluation

1-second bin averaged concentration data



Outputs Provided by Modelers

Outputs Provided by Modelers

Minimal Output

- **Minimal Output**
  1. Best estimate for the source location(s) (x and y) in UTM coordinates at the concluding time of each case.
  2. Source type, strength, and number of sources:
  3. Release start time (in seconds from the start)
  4. For continuous releases, release end time (or duration) should also be provided

It some of the data is not available (especially items 2 and 4), it is expected that modelers provide as much information as possible.

Outputs Provided by Modelers
Additional Desirable Output

- **Additional Desirable Output**
  1. Uncertainty in the source estimation, especially for location.
     - Need to "synchronize" uncertainty estimates among algorithm developers
  2. For algorithms that are based on a continuous estimate of the source term as a function of the simulated sensors time history, a time history of the source term estimates is requested (especially location and strength)
  3. For algorithms that use some form of T&D code to simulate backward and forward propagation of the tracer gas, a concentration time history based on the predicted sources(s) is requested
     - Based on "best source estimate"
     - need to provide a concentration time-history at every digiPID/UVIC location
     - evaluators will convert these to a "hazard region" based on suitably-defined critical thresholds.

Comparative Evaluation Metrics and Plan

Comparative Evaluation Metrics and Plan

Dealing directly with source term estimates

- Examine cross-, with-sampler array, and total distance between actual and predicted locations
- Examine fraction of source term mass predicted
- Calculate differences in the start time of the release
- If source term estimate uncertainty (location, mass, start and end time) is provided, and are comparable among different algorithms, then we will devise some analyses of this

Note that careful analyses and clear presentation of the results might be needed for the cases where there are mismatches in the types of the release (continuous versus instantaneous) and the number of sources.

Comparative Evaluation Metrics and Plan

Dealing with expected hazard regions

- **Main reason for using sensor fusion algorithms is**
  - to quickly estimate source terms characteristics to improve (forward) hazard prediction as compared to other means
    - » NBC messages
    - » ATP-45 template with “standard” amount of agent
  - to aid in forensic recovery of initial source term location and estimation of source term strength
    - » Usually follows by calculation of hazard area to determine effects (i.e. affected population, potential contamination, etc)
      - E.g. Gulf war syndrome

- **After obtaining source characteristics we need to estimate hazard area at all available samplers**
  - Ideally would prefer algorithm developer to do this with “native” software
  - Will substitute HPAC when needed
    - » Possible biasing of the results

- **Metrics to use in hazard area comparisons**
  - Standard statisticis
    - » NAD, FB, FAC_x
  - 2D-MOEs
    - » Comparison of the observed and predicted hazard region
      - samplers above critical threshold for a few threshold values
    - » Comparison of averaged concentration observed and predicted at each location.
      - note that average concentrations are closely related to dose (or dosages) that are predicted and observed at each location.

Fully released datasets

Some portion of the field trial data will be kept from STE modelers to insure that present/future STE algorithms have some "pristine" data to work with

"Full" data set release (in red)

- 53 total continuous disseminations
  - **36 nighttime releases**
    - » **15 one-source releases (9)**
    - » **11 two-source releases (7)**
    - » **8 three-source releases (5)**
    - » **2 four-source releases (2)**
  - **17 daytime releases**
    - » **7 one-source releases (3)**
    - » **5 two-source releases (3)**
    - » **3 three-source releases (2)**
    - » **2 four-source releases (2)**

- 29 total instantaneous disseminations
  - **20 nighttime releases**
    - » **7 one-source releases (3)**
    - » **7 two-source releases (3)**
    - » **4 three-source releases (3)**
    - » **2 four-source releases (2)**
  - **9 daytime releases**
    - » **4 one-source releases (2)**
    - » **2 two-source releases (1)**
    - » **1 three-source releases (1)**
    - » **2 four-source releases (2)**

- **Proposed total number of trials "fully" released: 50 (~61%)**
  - Puff trials include series of puffs resulting in multiple test cases for algorithm developer per each trial released
- **"Full" data release will not include processing done for stage 1 data release to STE algorithm developers**
  - Puff trials will not be broken into individual puffs
  - Sampler (digiPID) data will be at "native" 50Hz

Status

- **"Full" dataset containing *44* trials became available for download in early August.**
- **Stage 1 dataset consisting of *104* cases for which source term estimates are sought became available to algorithm developers on September 2, 2008**
- **STE predictions are expected to became available to us in late December 2008.**
- **Planning to distribute Stage 2 dataset sometimes in the second quarter of 2009.**

https://fft07-dpg.dpg.army.mil

10-sec Average Concentration Movies

Observations vs HPAC Predictions

Continuous Trial
06

Puff Trial
69

2-sec Average Concentration Movies

Observations

Puff Trial
37

Continuous Trial
64