Source Term Estimation for Hazardous Releases

Fukushima Workshop

NCAR, 22nd – 23rd February 2012

Dstl/DOC61790

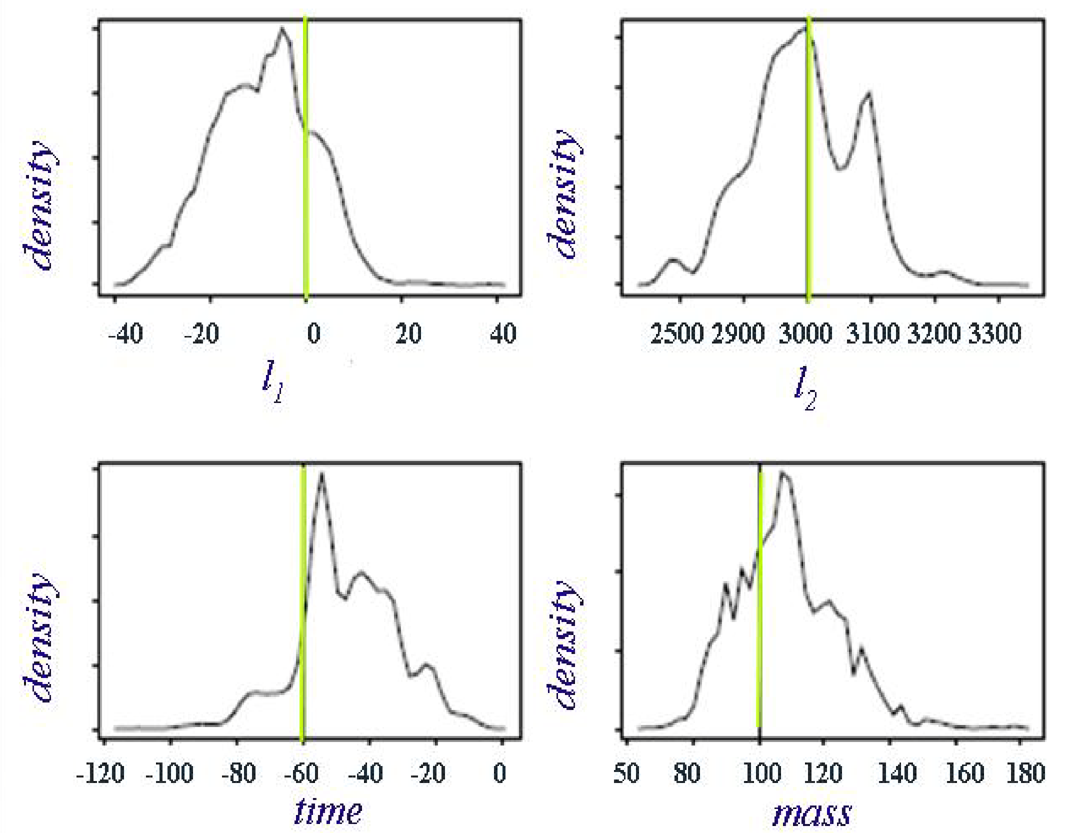
Dr. Gareth Brown

Presentation Outline

- Dstl's approach to Source Term Estimation
- Recent developments
  - Spatially and temporally gridded urban met object
  - Processing of deposition measurements
- Toy example
- Application to Fukushima

Source Term Estimation Monte Carlo Bayesian Data Fusion (MCBDF)

- MCBDF provides a set of likely source terms for calculating probabilistic hazards
- Accounts for uncertainty in source, meteorology, sensor performance & human reporting



P. Robins, V. Rapley, N. Green, *Realtime Sequential Inference of Static Parameters with Expensive Likelihood Calculations* (2009)

Bayes' Theorem

$$p(\theta|y) \propto p(y|\theta)\,p(\theta)$$

Posterior
probability density of hypothesis θ given the data y

Likelihood
probability density of the data y conditional on the hypotheses θ

Prior
probability density of θ based on prior knowledge

- Iterative update of belief in a hypothesis
- Likelihood calculation particularly costly
  - Necessary to run dispersion model for sensor data
  - Sensor likelihood function complex

MCBDF: Bayesian data fusion

- MCBDF uses Bayesian inference over a sample set of hypothesized source-terms and Met. variables

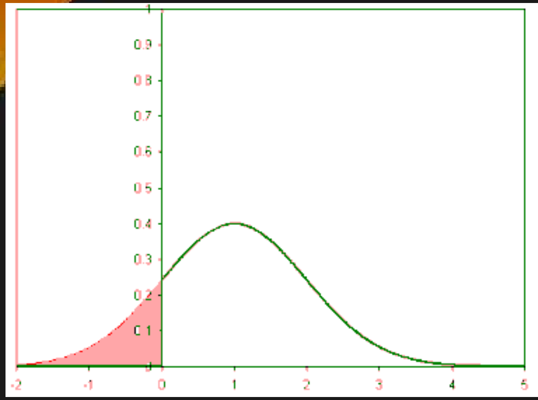
$$\theta = (\underbrace{x, y, t, m, d, a}_{Source\text{–}term}, \underbrace{u_{*x}, u_{*y}, \tfrac{1}{L}, \ln\left(z_0\right)}_{Met}, \underbrace{D, P}_{Model})$$

- The dimensionality of problem depends on its complexity
  – Complex source terms
  – Gridded meteorology in time and space
  – Urban dispersion

Physics Modelling

- Dispersion is a turbulent process
  - **Impossible to model accurately**
  - **Use ensemble puff dispersion models**
  - **SPRINT model optimised for STE**
  - **Implements SCIPUFF closure relations**

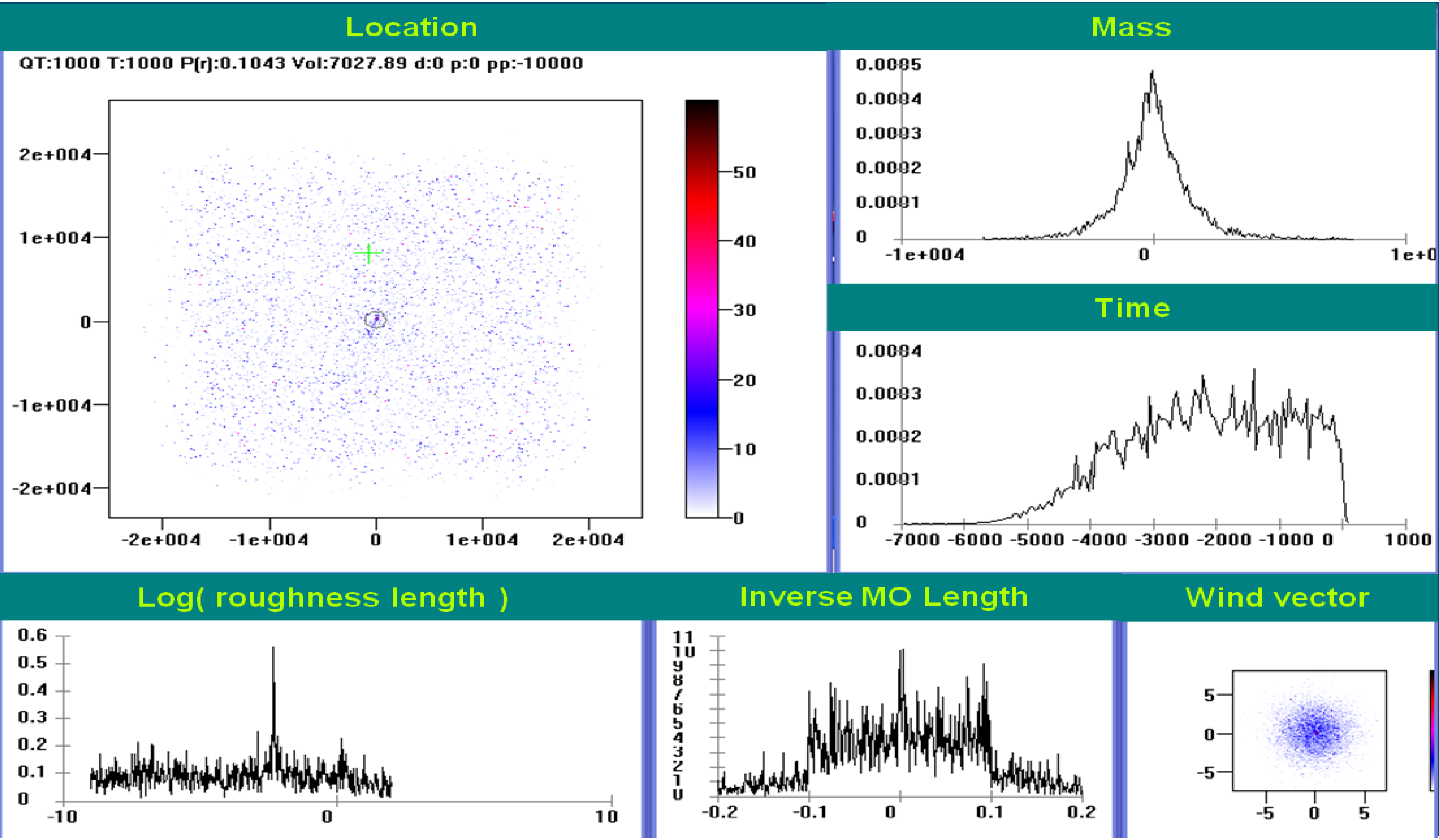
$$\left.\begin{array}{c}\theta \\ (\mathbf{x}, t_1, t_2)\end{array}\right\} \xrightarrow{\text{dispersion}} p\begin{pmatrix} c \,\middle|\, \overline{c}, \overline{c'^2} \\ \chi \,\middle|\, \overline{\chi}, \overline{\chi'^2} \\ \chi_d \,\middle|\, \overline{\chi_d}, \overline{\chi_d'^{\,2}} \end{pmatrix}$$



UK UNCLASSIFIED

Dstl is part of the Ministry of Defence

MCBDF: Example Prior



dstl

UK UNCLASSIFIED

Dstl is part of the Ministry of Defence

Sensor Likelihood Calculations

- When CB sensor measurements are passed to MCBDF the likelihood is calculated as

$$p\left(y \mid \mu, \sigma^2\right) = \int_0^\infty \underbrace{p\left(y \mid c\right)}_{\substack{\text{measurement}\\ \text{density}}} \underbrace{p\left(c \mid \mu, \sigma^2\right)}_{\substack{\text{concentration}\\ \text{density}}} \mathrm{d}c$$

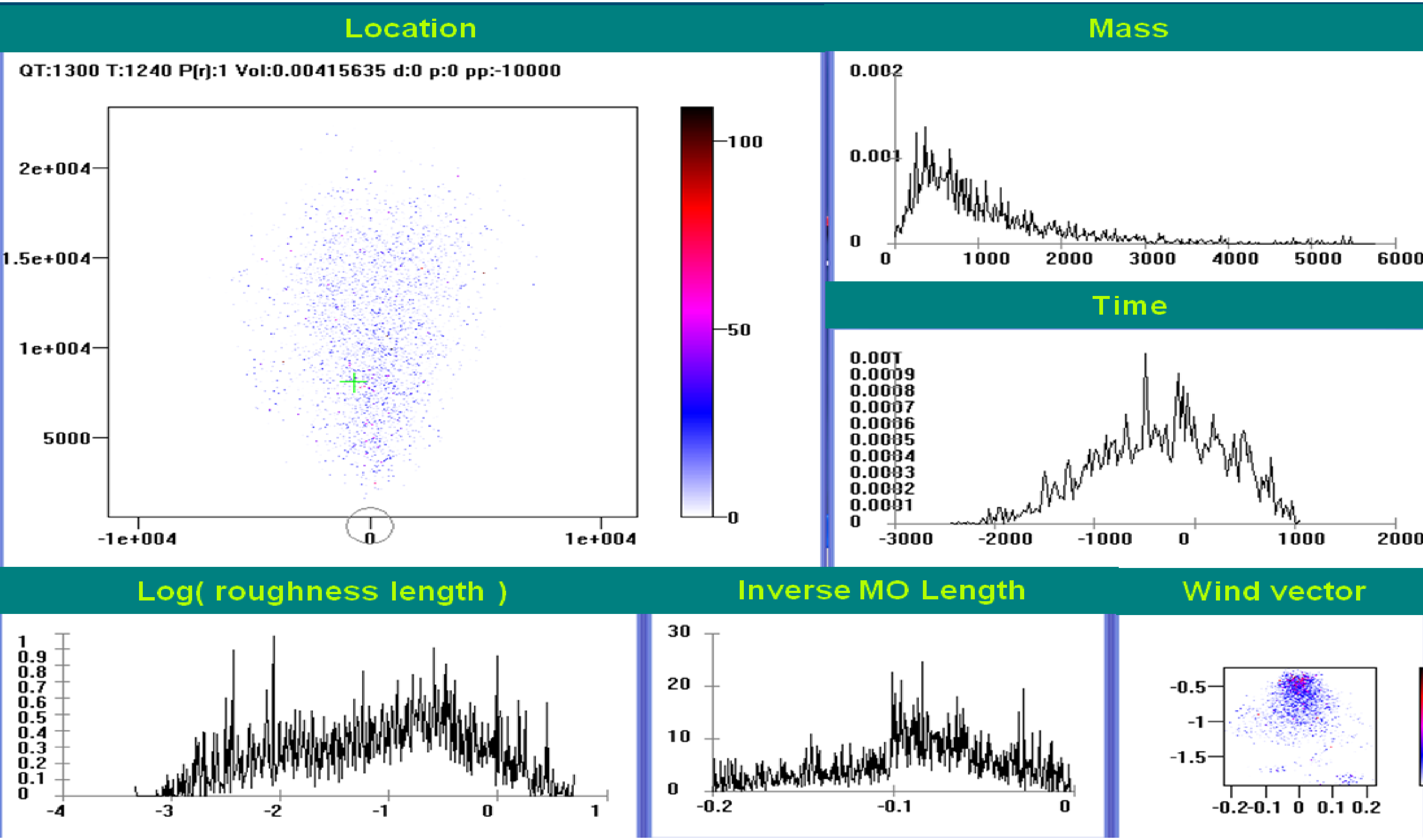
$y \equiv$ Sensor measurement

$\mu \equiv$ Mean mass-concentration from dispersion simulation

$\sigma^2 \equiv$ Mass concentration variance from dispersion simulation

$c \equiv$ Unobserved ground-truth concentration

MCBDF: Upon Receipt of a Detection



UK UNCLASSIFIED

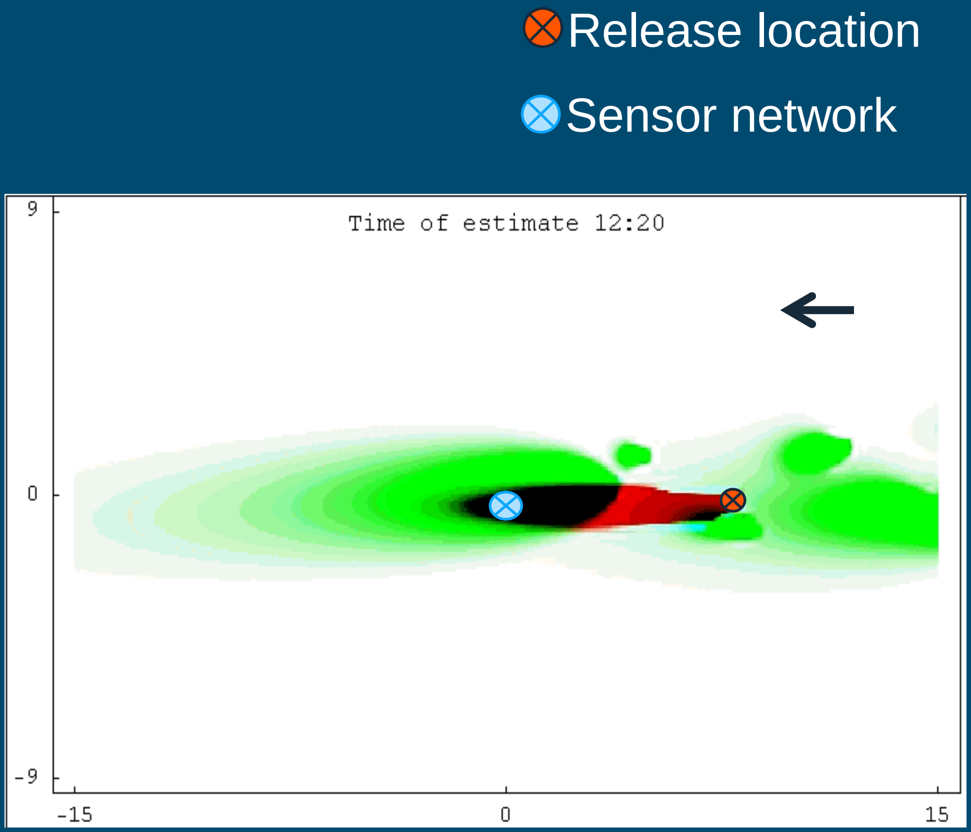
Estimating the Posterior Distribution

$$p\left(\theta \middle| \mathbf{y}\right) \propto \prod_i p\left(y_i \middle| \theta\right) p\left(\theta\right)$$

- Calculated as the product of the individual likelihoods and the prior
- Posterior estimated by sampling lots of hypotheses
- MCBDF uses Differential Evolution Markov Chain (DEMC) Monte Carlo to generate new hypotheses

Calculating Hazard Areas

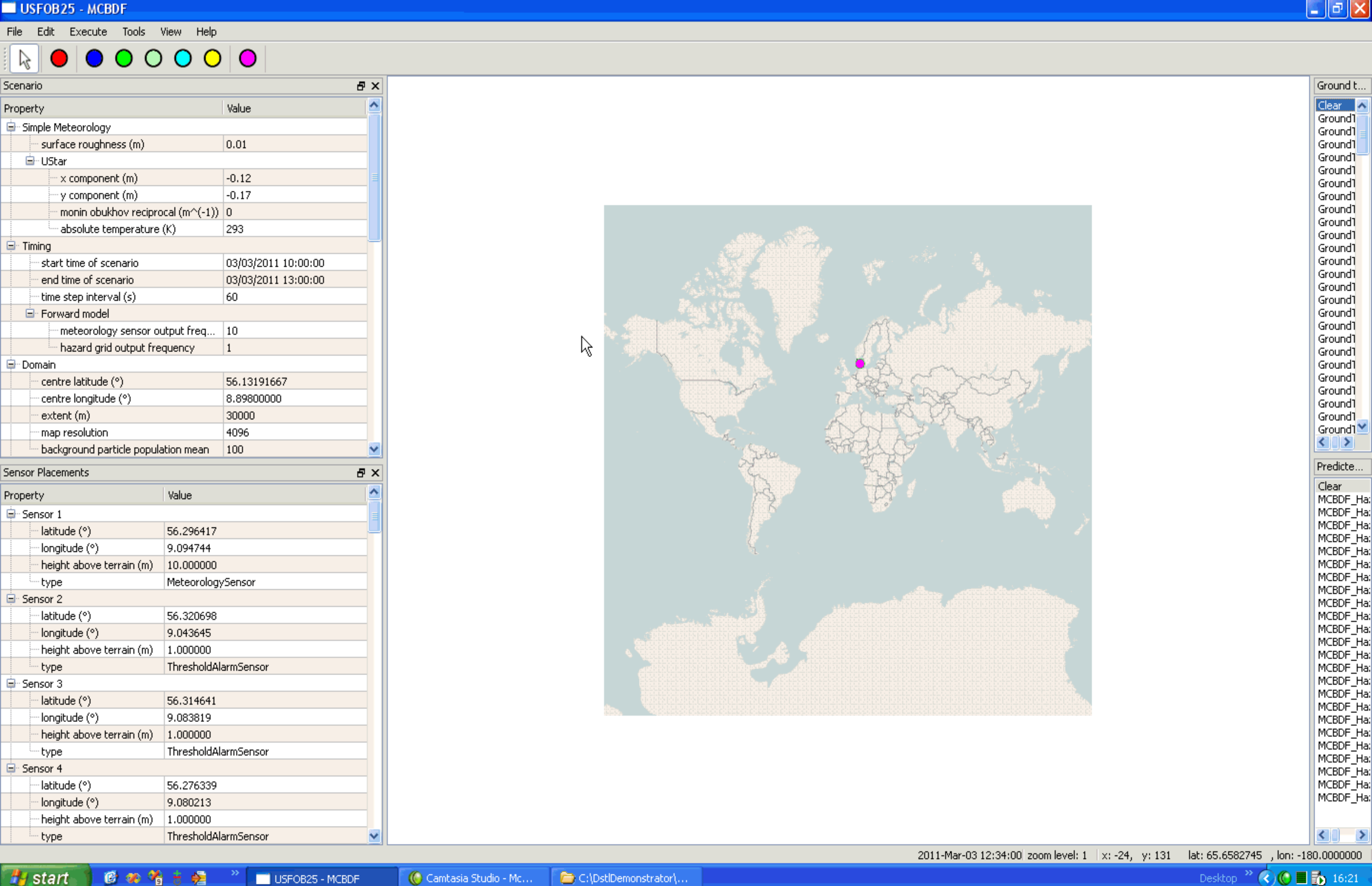
- The Hazard Calculator obtains the probability of exceeding a specified threshold dosage
- A weighted sum of these gridded probabilities is used to define the probable hazard area



Red: ground truth; green: predicted

UK UNCLASSIFIED

Dstl is part of the Ministry of Defence



[dstl]

UK UNCLASSIFIED

Dstl is part of the Ministry of Defence

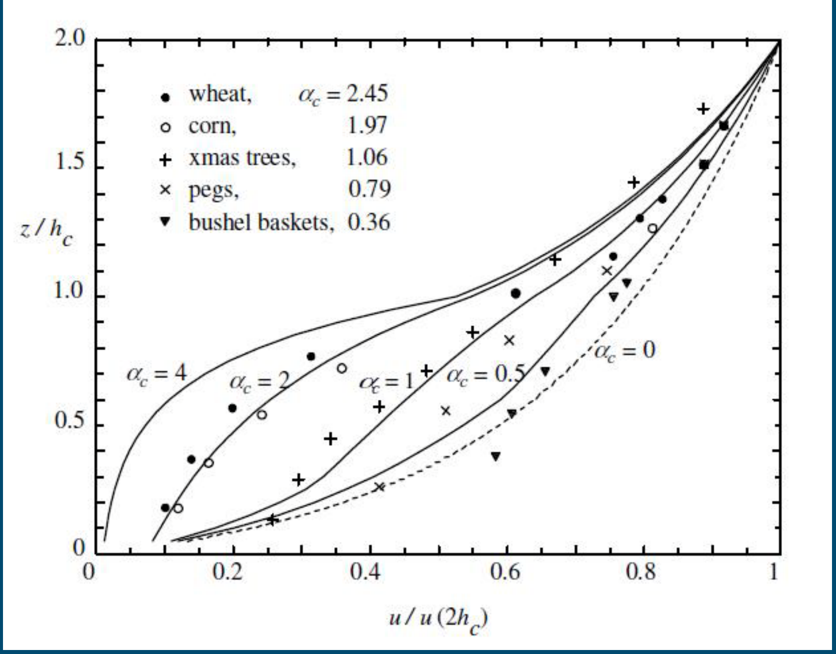
Recent Developments

- MCBDF initially developed for rapid warning of chemical attacks based on input from sensor networks
  - Short effective duration approximates to spatially and temporally invariant meteorology
- However STE for Bio and Rad releases in major cities requires
  - Fast dispersion models accounting for urban areas
  - Much longer temporal windows
  - Efficient methods to compute long duration dosage / deposition
  - Spatially varying meteorology for large scale dispersion

Meteorological Inference

Sykes et al, SCIPUFF Tech Doc

- Urban capability
- Gridded urban meteorological model defined
  - Wind flow components gridded in 2D space and time
  - Surface properties gridded in 2D space
  - Atmospheric stability gridded in time
- 60 – 400 extra dimensions in parameter space
  - MCMC techniques still efficient if converged
  - Much more difficult to achieve convergence



$$f_c(z, h_c, \alpha_c) = \exp\left[-\alpha_c\left(1 - \frac{z}{h_c}\right)\right]$$

Canopy height

Canopy flow parameter

Toy Problem Example

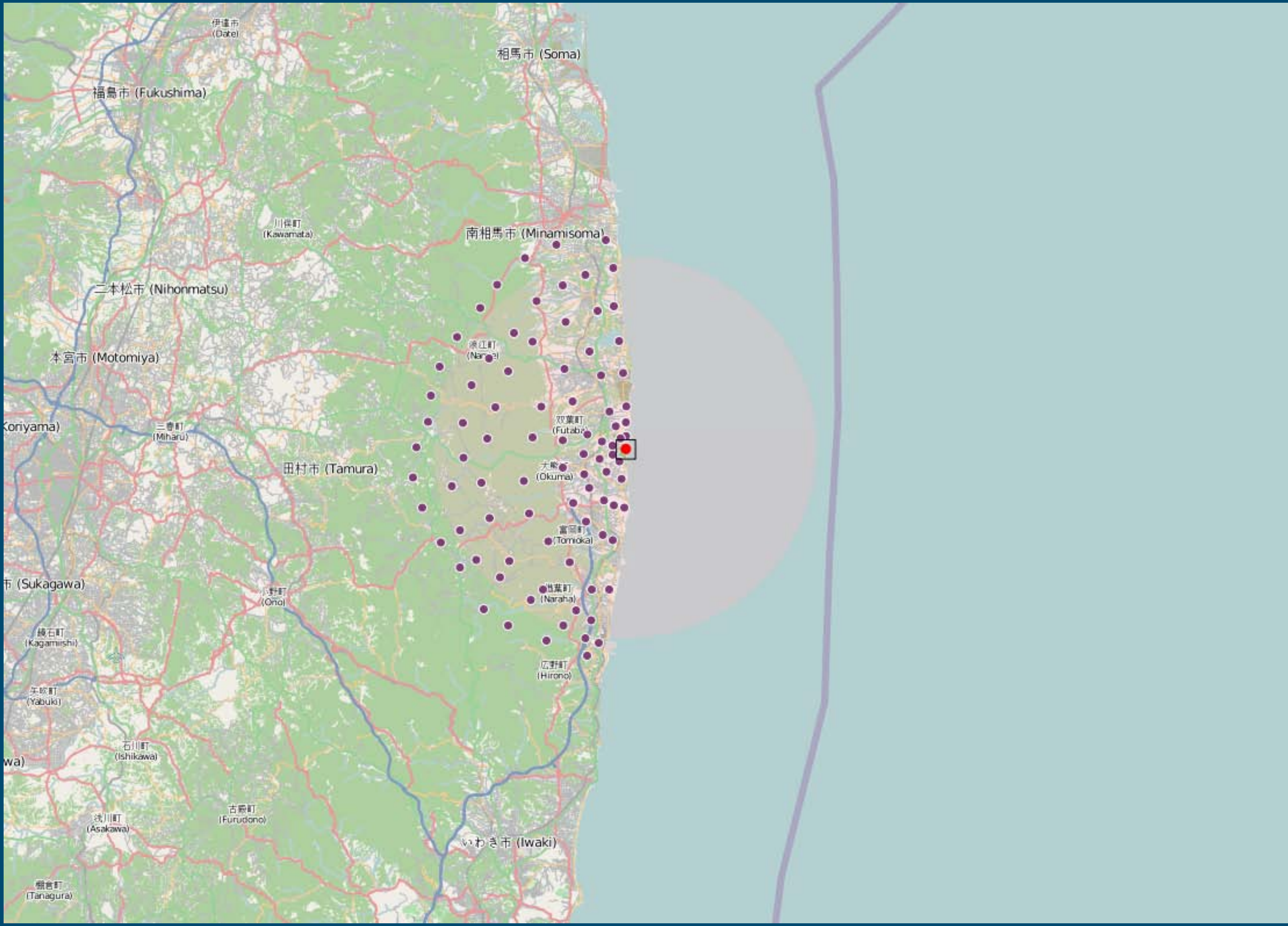
- Single, short duration release
- Single isotope
- Quantitative deposition measurements
- Spatially gridded, but temporally invariant meteorology
- No transport due to rain run-off before measurements taken

Deposition data

- Likelihood model for quantitative measurement of deposited material
  - Assume local detection only, alpha, beta, not long-range gamma
  - Uncertainty per measurement, not fixed
    - Time taken to collect Poisson count statistics
    - Energy spectrum uncertainty on which isotope is being measured
    - Naturally occurring background (with uncertainty) subtracted
    - Simple, normally distributed measurement error model
    - Lower limit and saturation can be input

Sample Points

- 88 points
- 20km exclusion zone shown
- Release point known
- No met data

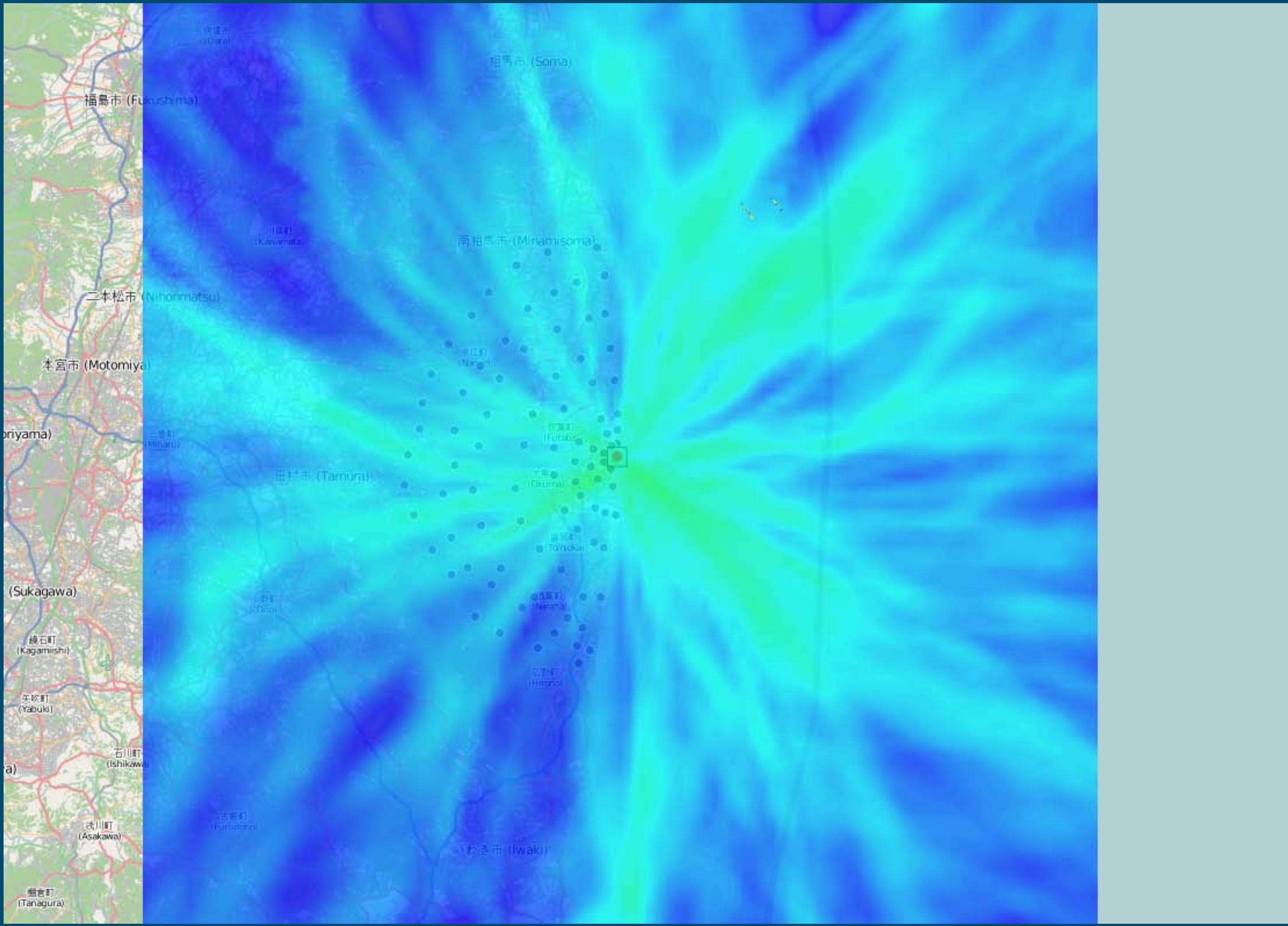


UK UNCLASSIFIED

Dstl is part of the Ministry of Defence

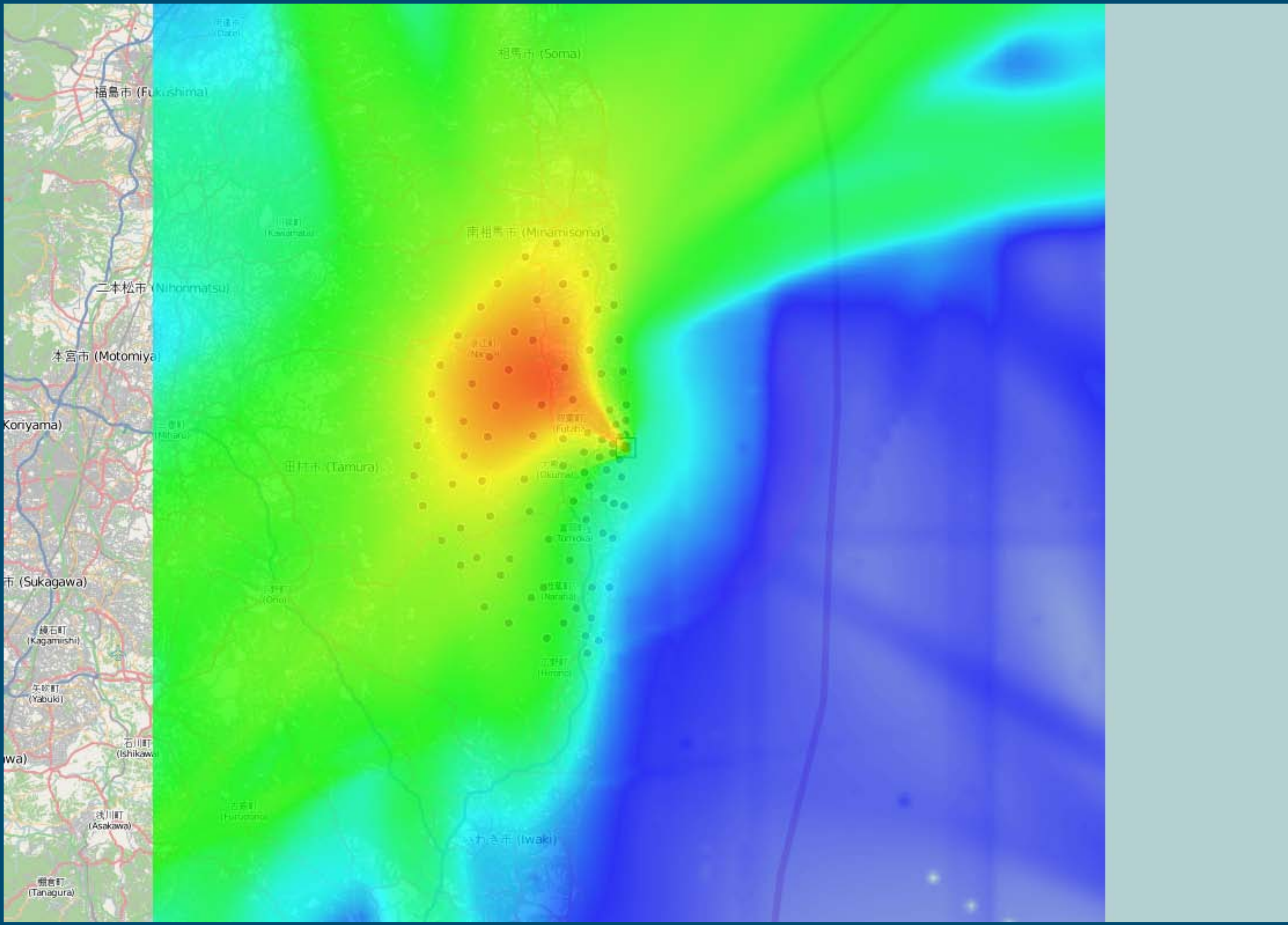
Hazard Plot

- First measurement above background
- Close to source
- Huge met uncertainty



Hazard Plot

- First 3 rings of data
- Less uncertainty close to source

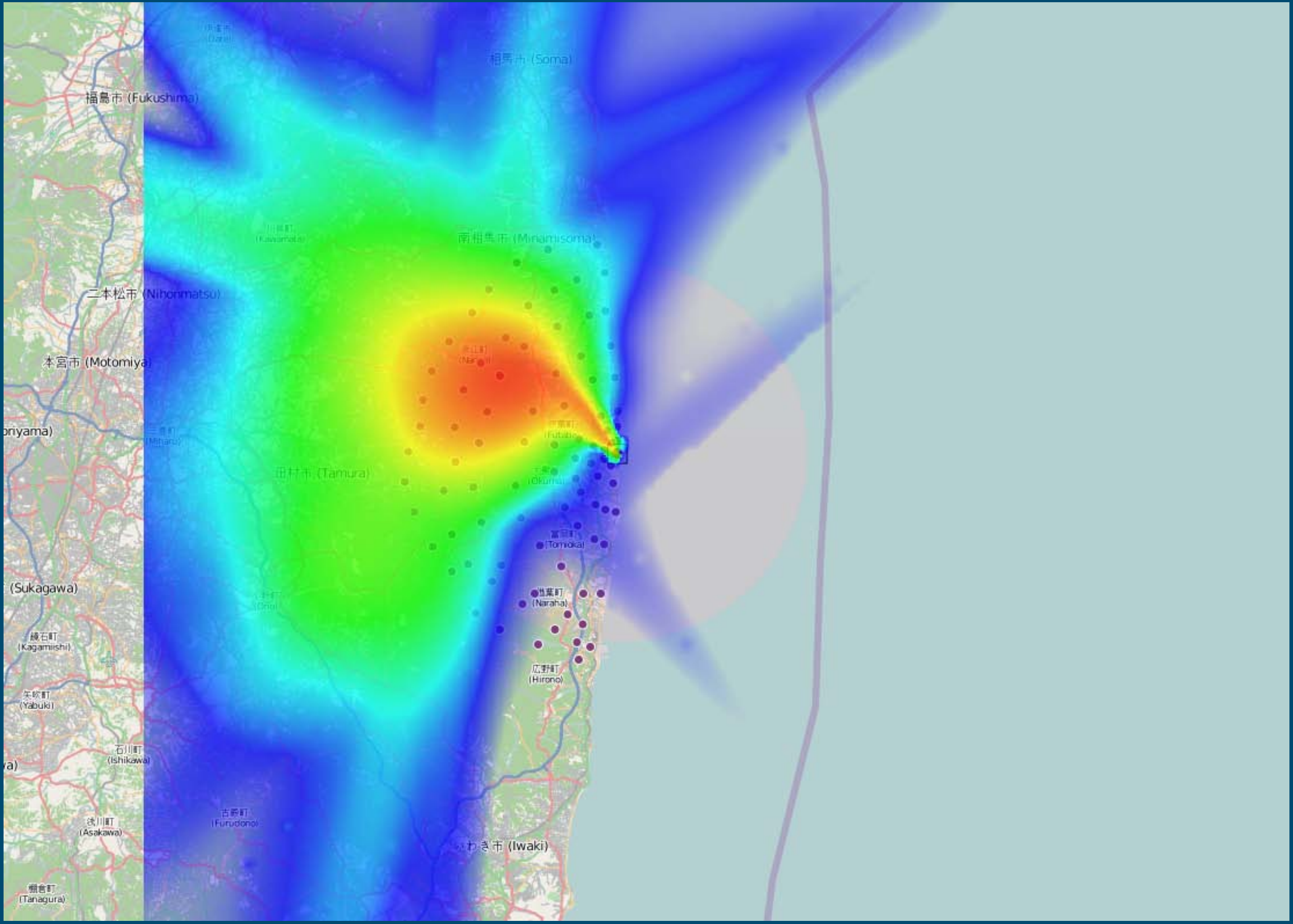


UK UNCLASSIFIED

Dstl is part of the Ministry of Defence

Hazard Plot

- 4 rings of data

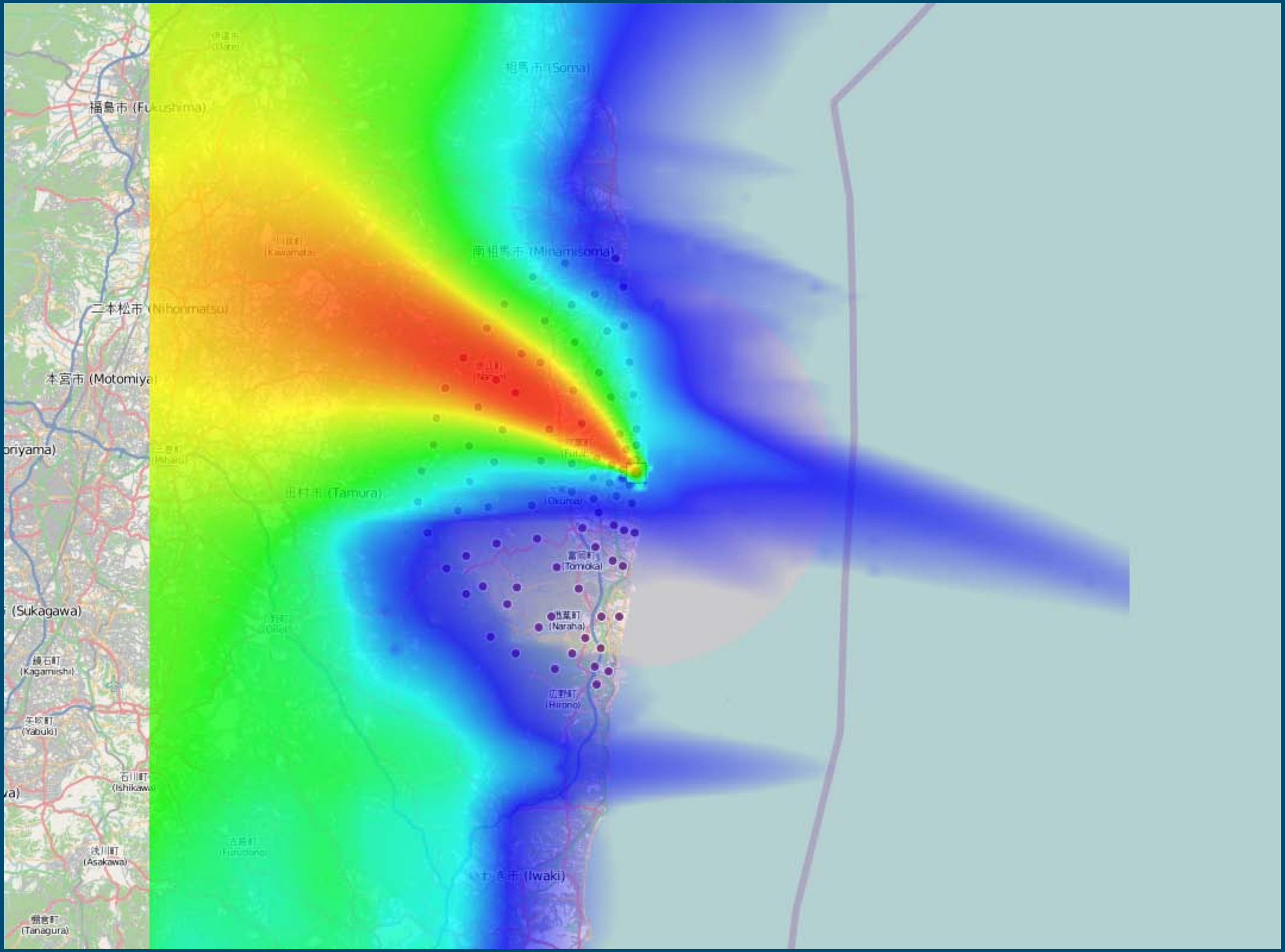


UK UNCLASSIFIED

Dstl is part of the Ministry of Defence

Hazard Plot

- 6 rings of data

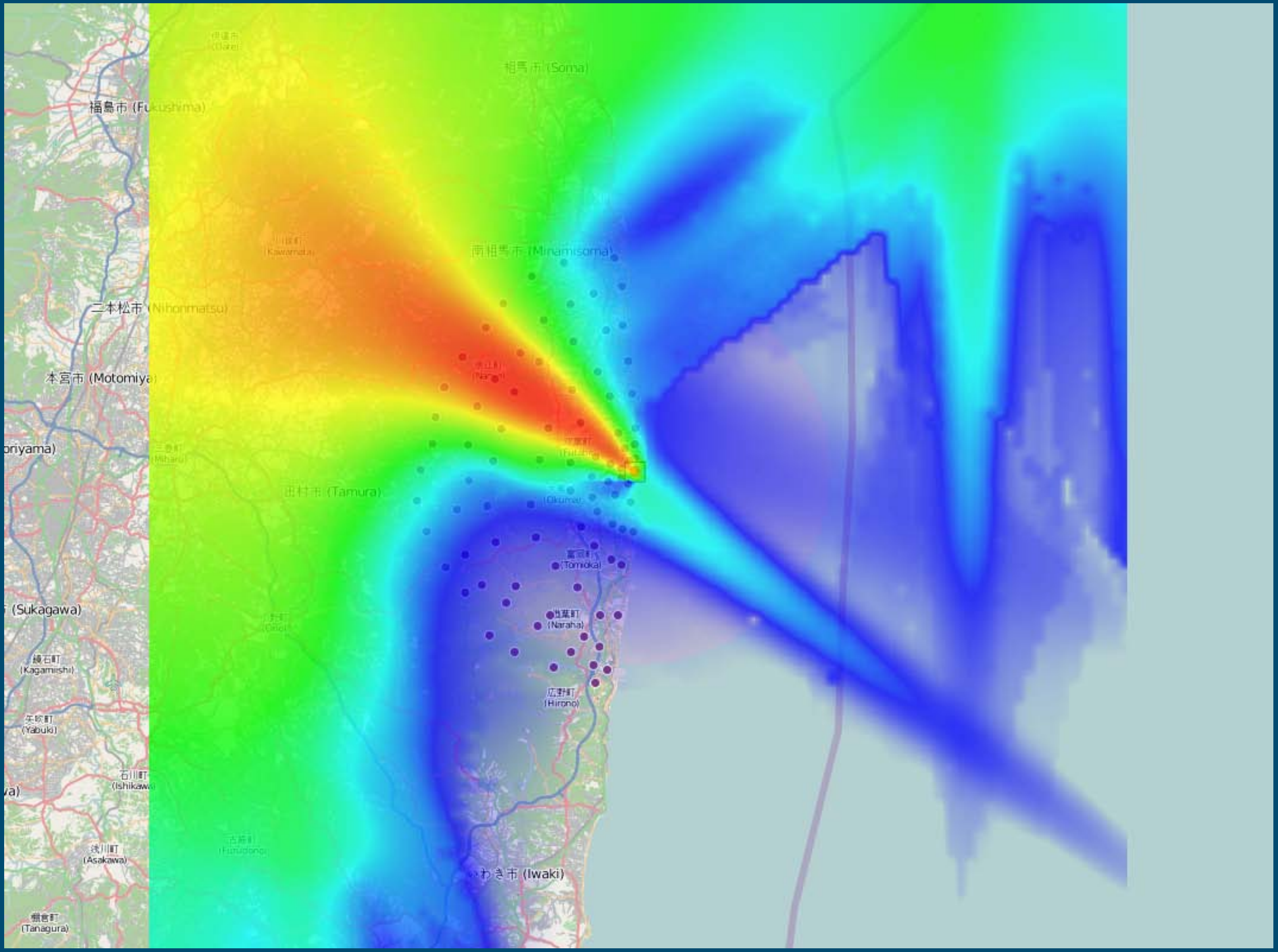


UK UNCLASSIFIED

Dstl is part of the Ministry of Defence

Hazard Plot

- All data
- More time for convergence



UK UNCLASSIFIED

Problems to solve for Fukushima

- Parameters to infer:
  - Mass released in each hour over the last year
  - Gridded meteorology for each hour over the last year
    - 10km spatial resolution
- Modelling:
  - Gamma sensors pick up spatially averaged dose
    - Function of range needs integrating
      - $1/r^2$, attenuation through air, scatter build up factors
  - Longer range second order closure dispersion model needed
    - SCIPUFF
    - Other?
  - Further transport of deposited material due to water run-off

Backup

The Concentration Sensor Likelihood Model

- Critical to the performance of MCBDF is an accurate probabilistic description of the detector's response

$$p(y|c) = \begin{cases} \Phi\left(\underline{L} \middle| c, \sigma_e^2\right) & y = \underline{L} \\ \phi\left(y \middle| c, \sigma_e^2\right) & \underline{L} < y < \overline{L} \\ 1 - \Phi\left(\overline{L} \middle| c, \sigma_e^2\right) & y = \overline{L} \end{cases}$$

normal distribution CDF

$\sigma_e^2 \equiv$ Measurement error variance

$\overline{L} \equiv$ Sensor saturation point

$\underline{L} \equiv$ Sensor limit of detection

Deposition modelling

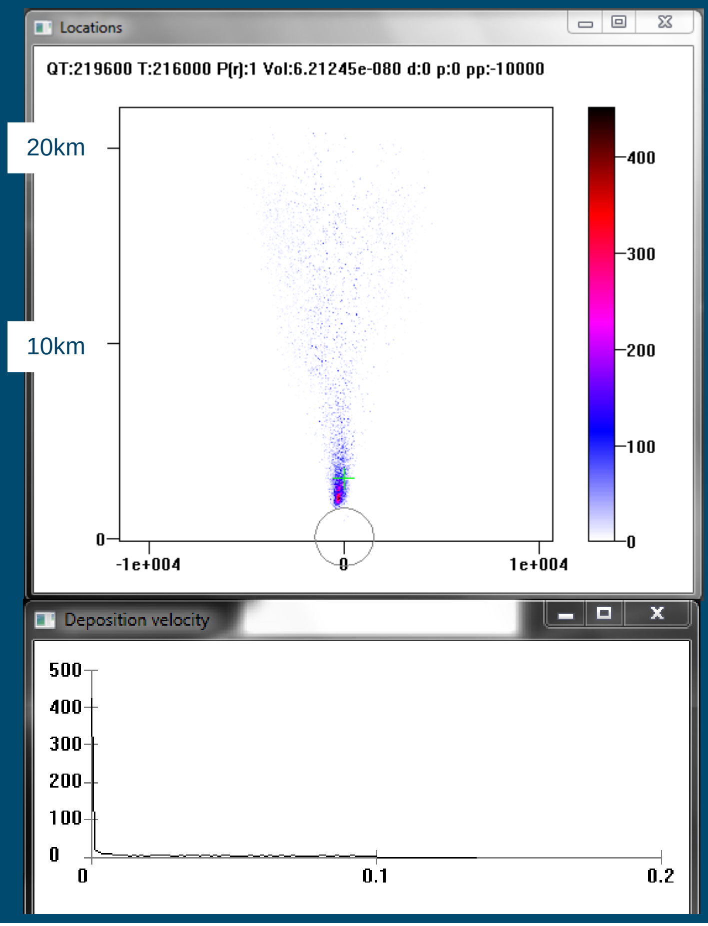
- Improved physics improves consistency of sensor data likelihood modelling
  - More accurate source term estimation in presence of deposition
- Amount of deposition inferred from data and/or uncertainty correctly passed to hazard calculations

Deposition data

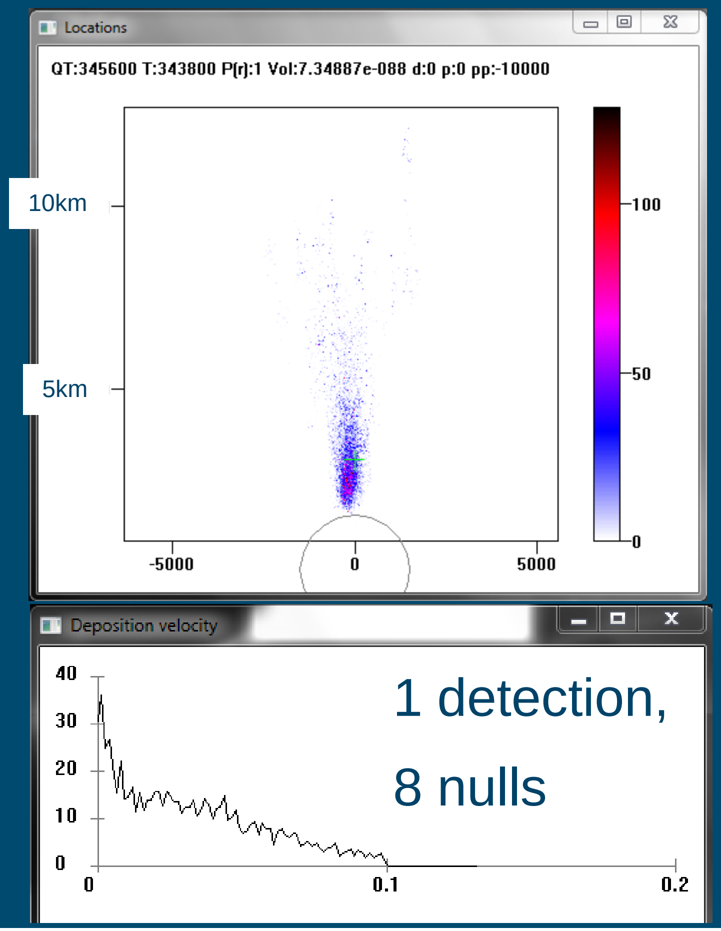
- Likelihood model for detection of deposited material already in place
  - Biological hazards
  - Similar to a probit model
  - Extended to include:
    - Probability of false alarm
    - Probability of false negative
  - Integrate out unobserved true amount of deposited material

Prelim. deposition results (simulated met. constant)

Collectors + Identifiers (dosage)

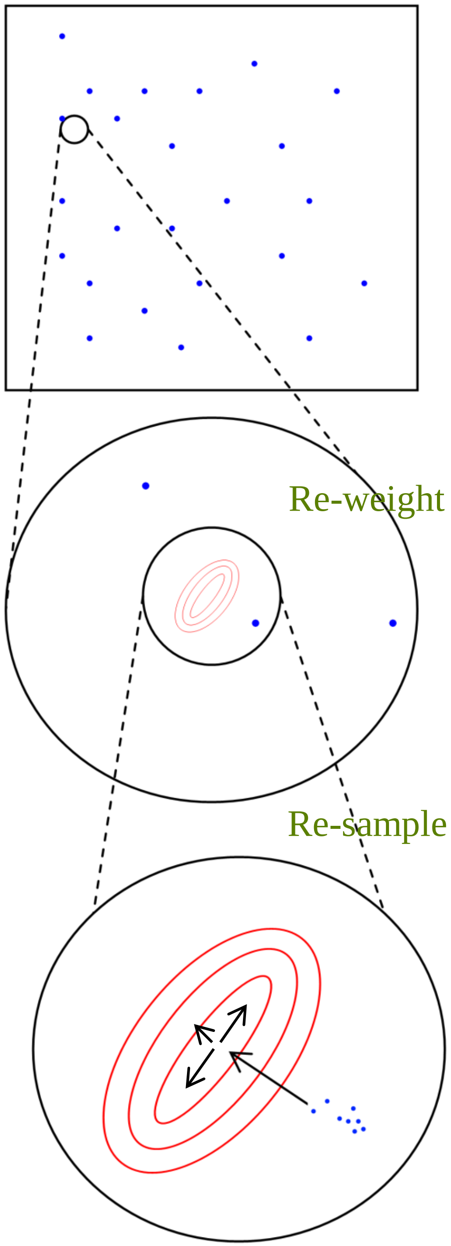


+ Survey ID (deposition)



Burn In and Convergence

- Initial hypotheses may be far from the peak of the posterior
- But rapid answers are required
  - Data continually changing the posterior
  - limited time for new sample weights to make the old ones insignificant
  - Limited time for samples to spread out and capture true uncertainty
- Detection of non-convergence delays message processing



Hazard Calculation

- Probit slope model
  - S used to indicate uncertainty in appropriate value for χ_{d50}

$$p\left(Hazard \middle| \chi_d\right) = \frac{1}{2}\left(1 + erf\left(\frac{S}{\sqrt{2}}\log_{10}\left(\frac{\chi_d}{\chi_{d50}}\right)\right)\right)$$

$$p\left(Hazard \middle| \overline{\chi_d}, \overline{\chi_d'^2}\right) = \int_0^\infty p\left(\chi_d \middle| \overline{\chi_d}, \overline{\chi_d'^2}\right) p\left(Hazard \middle| \chi_d\right) d\chi_d$$

- Average over 1000 release/met samples from posterior.

Deposition data likelihood (clipped normal)

$$p(y|\chi_d)=\begin{cases}\Phi(\underline{L}|\chi_d,\sigma_e^2) & y=\underline{L}\\ \phi(y|\chi_d,\sigma_e^2) & \underline{L}<y<\overline{L}\\ 1-\Phi(\overline{L}|\chi_d,\sigma_e^2) & y=\overline{L}\end{cases}$$

$$\left(\overline{\chi_d},\overline{\chi_d'^2}\right)\xrightarrow{unclipped}\left(\mu_N,\sigma_N^2\right)$$

$$p(y|\mu_N,\sigma_N^2)=\begin{cases}\int_0^\infty \Phi(\underline{L}|\chi_d,\sigma_e^2)\left[\Phi(0|\mu_N,\sigma_N^2)\delta(\chi_d)+\phi(\chi_d|\mu_N,\sigma_N^2)\right]d\chi_d & y=\underline{L}\\ \int_0^\infty \phi(y|\chi_d,\sigma_e^2)\left[\Phi(0|\mu_N,\sigma_N^2)\delta(\chi_d)+\phi(\chi_d|\mu_N,\sigma_N^2)\right]d\chi_d & \underline{L}<y<\overline{L}\\ \int_0^\infty \left(1-\Phi(\overline{L}|\chi_d,\sigma_e^2)\right)\left[\Phi(0|\mu_N,\sigma_N^2)\delta(\chi_d)+\phi(\chi_d|\mu_N,\sigma_N^2)\right]d\chi_d & y=\overline{L}\end{cases}$$

$$\phi(x|\mu,\sigma^2)\equiv\phi(\mu|x,\sigma^2)$$

$$\phi(x|\mu_1,\sigma_1^2)\phi(x|\mu_2,\sigma_2^2)\equiv k\phi(x|\mu,\sigma^2)$$

$$\sigma^2=\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

$$\mu=\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)$$

$$k=\frac{\sigma}{\sigma_1\sigma_2\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}-\frac{\mu^2}{\sigma^2}\right)}$$

$$p(y|\mu_N,\sigma_N^2)=\begin{cases}\Phi(\underline{L}|0,\sigma_e^2)\Phi(0|\mu_N,\sigma_N^2)+\int_{-\infty}^{\underline{L}} k(y)\left[1-\Phi(0|\mu(y),\sigma^2(y))\right]dy & y=\underline{L}\\ \phi(y|0,\sigma_e^2)\Phi(0|\mu_N,\sigma_N^2)+k(y)\left[1-\Phi(0|\mu(y),\sigma^2(y))\right] & \underline{L}<y<\overline{L}\\ \left[1-\Phi(\overline{L}|0,\sigma_e^2)\right]\Phi(0|\mu_N,\sigma_N^2)+\int_{\overline{L}}^{\infty} k(y)\left[1-\Phi(0|\mu(y),\sigma^2(y))\right]dy & y=\overline{L}\end{cases}$$

Romberg (closed and semi-infinite) numerical integration

Deposition data likelihood (clipped gamma)

$$p\left(y\middle|\chi_d\right)=\begin{cases}\Phi\left(\underline{L}\middle|\chi_d,\sigma_e^2\right) & y=\underline{L}\\ \phi\left(y\middle|\chi_d,\sigma_e^2\right) & \underline{L}<y<\overline{L}\\ 1-\Phi\left(\overline{L}\middle|\chi_d,\sigma_e^2\right) & y=\overline{L}\end{cases}$$

$$\left(\overline{\chi_d},\overline{\chi_d'^2}\right)\longrightarrow\left(s,k^*,\lambda\right)$$

$$p\left(y\middle|\mu_N,\sigma_N^2\right)=\begin{cases}\int_0^\infty\Phi\left(\underline{L}\middle|\chi_d,\sigma_e^2\right)\left[\left(\frac{\chi_d+\lambda}{s}\right)^{k^*-1}\frac{\exp\left(-\left(\frac{\chi_d+\lambda}{s}\right)\right)}{s\Gamma(k^*)}+(1-\gamma)\delta(\chi_d)\right]d\chi_d & y=\underline{L}\\ \int_0^\infty\phi\left(y\middle|\chi_d,\sigma_e^2\right)\left[\left(\frac{\chi_d+\lambda}{s}\right)^{k^*-1}\frac{\exp\left(-\left(\frac{\chi_d+\lambda}{s}\right)\right)}{s\Gamma(k^*)}+(1-\gamma)\delta(\chi_d)\right]d\chi_d & \underline{L}<y<\overline{L}\\ \int_0^\infty\left(1-\Phi\left(\overline{L}\middle|\chi_d,\sigma_e^2\right)\right)\left[\left(\frac{\chi_d+\lambda}{s}\right)^{k^*-1}\frac{\exp\left(-\left(\frac{\chi_d+\lambda}{s}\right)\right)}{s\Gamma(k^*)}+(1-\gamma)\delta(\chi_d)\right]d\chi_d & y=\overline{L}\end{cases}$$

$$p\left(y\middle|\mu_N,\sigma_N^2\right)=\begin{cases}\Phi\left(\underline{L}\middle|0,\sigma_e^2\right)(1-\gamma)+\int_0^\infty\Phi\left(\underline{L}-\chi_d\middle|0,\sigma_e^2\right)\left[\left(\frac{\chi_d+\lambda}{s}\right)^{k^*-1}\frac{\exp\left(-\left(\frac{\chi_d+\lambda}{s}\right)\right)}{s\Gamma(k^*)}\right]dc & y=\underline{L}\\ \phi\left(y\middle|0,\sigma_e^2\right)(1-\gamma)+\int_0^\infty\phi\left(y\middle|\chi_d,\sigma_e^2\right)\left[\left(\frac{\chi_d+\lambda}{s}\right)^{k^*-1}\frac{\exp\left(-\left(\frac{\chi_d+\lambda}{s}\right)\right)}{s\Gamma(k^*)}\right]dc & \underline{L}<y<\overline{L}\\ \left(1-\Phi\left(\overline{L}\middle|0,\sigma_e^2\right)\right)(1-\gamma)+\int_0^\infty\left(1-\Phi\left(\overline{L}-\chi_d\middle|0,\sigma_e^2\right)\right)\left[\left(\frac{\chi_d+\lambda}{s}\right)^{k^*-1}\frac{\exp\left(-\left(\frac{\chi_d+\lambda}{s}\right)\right)}{s\Gamma(k^*)}\right]dc & x=\overline{L}\end{cases}$$

Romberg (closed and semi-infinite) numerical integration

Urban meteorology

- Displacement height $z_d(h_c,\alpha_c) = 0.7 h_c F_c(\alpha_c)$
- Canopy blending $F_c(\alpha_c) = 1 - \exp\left(-\frac{(\alpha_c)^2}{0.25 + 0.5\alpha_c}\right)$
- Mean wind vector $\bar{\mathbf{u}}(x,y,z,t) = g(x,y,z)\bar{\mathbf{u}}'(x,y,z,t)$
- Shifted wind vector

$$\bar{u}_x'(x,y,z,t) = \frac{f_{sl}(z',z_0,L)}{k} u_{*x}(x,y,t)$$

$$\bar{u}_y'(x,y,z,t) = \frac{f_{sl}(z',z_0,L)}{k} u_{*y}(x,y,t)$$

$$\bar{u}_z'(x,y,z,t) = 0$$

$$z'(x,y,z) = \begin{cases} h_c - z_d & \text{if } \begin{cases}(h_c < z_s + z_d) \text{ and } (z < h_c) \\ \text{or } (h_c \geq z_s + z_d)\end{cases} \\ z - z_d & \text{if } (h_c < z_s + z_d) \text{ and } (h_c \leq z < z_s + z_d) \\ z_s & \text{if } (h_c < z_s + z_d) \text{ and } (z \geq z_s + z_d) \end{cases}$$

- Surface layer profile $f_{sl}(z,z_0,L) = \ln\left(\frac{z}{z_0} + 1\right) - \Psi_m\left(\frac{z}{L}\right)$
- Canopy layer profile $f_c(z,h_c,\alpha_c) = \exp\left[-\alpha_c\left(1 - \frac{z}{h_c}\right)\right]$
- Blending function $g(x,y,z) = F_c(\alpha_c) f_c(z,h_c,\alpha_c) + (1 - F_c(\alpha_c))\frac{f_{sl}(z,z_0,L)}{f_{sl}(h_c,z_0,L)}$

Wind vector spatial derivatives

- Calculated by chain rule, e.g.:

$$h_c < z_s + z_d$$

$$\frac{\partial g(x,y,z)}{\partial x_i} = \begin{cases} \frac{\partial g(x,y,z)}{\partial \alpha_c}\frac{\partial \alpha_c}{\partial x_i} + \frac{\partial g(x,y,z)}{\partial h_c}\frac{\partial h_c}{\partial x_i} \\ + \frac{\partial g(x,y,z)}{\partial z_0}\frac{\partial z_0}{\partial x_i} + \frac{\partial g(x,y,z)}{\partial z}\frac{\partial z}{\partial x_i} & if \quad z < h_c \\ 0 & if \quad z \geq h_c \end{cases}$$

$$h_c \geq z_s + z_d$$

$$\frac{\partial g(x,y,z)}{\partial x_i} = \begin{cases} \frac{\partial g(x,y,z)}{\partial \alpha_c}\frac{\partial \alpha_c}{\partial x_i} + \frac{\partial g(x,y,z)}{\partial h_c}\frac{\partial h_c}{\partial x_i} \\ + \frac{\partial g(x,y,z)}{\partial z_0}\frac{\partial z_0}{\partial x_i} + \frac{\partial g(x,y,z)}{\partial z}\frac{\partial z}{\partial x_i} & if \quad z < z_s + z_d \\ \frac{\partial g(x,y,z_s+z_d)}{\partial \alpha_c}\frac{\partial \alpha_c}{\partial x_i} + \frac{\partial g(x,y,z_s+z_d)}{\partial h_c}\frac{\partial h_c}{\partial x_i} \\ + \frac{\partial g(x,y,z_s+z_d)}{\partial z_0}\frac{\partial z_0}{\partial x_i} & if \quad z \geq z_s + z_d \\ + \frac{\partial g(x,y,z_s+z_d)}{\partial z}\left(\frac{\partial z_d}{\partial \alpha_c}\frac{\partial \alpha_c}{\partial x_i} + \frac{\partial z_d}{\partial h_c}\frac{\partial h_c}{\partial x_i}\right) \end{cases}$$

Hypotheses

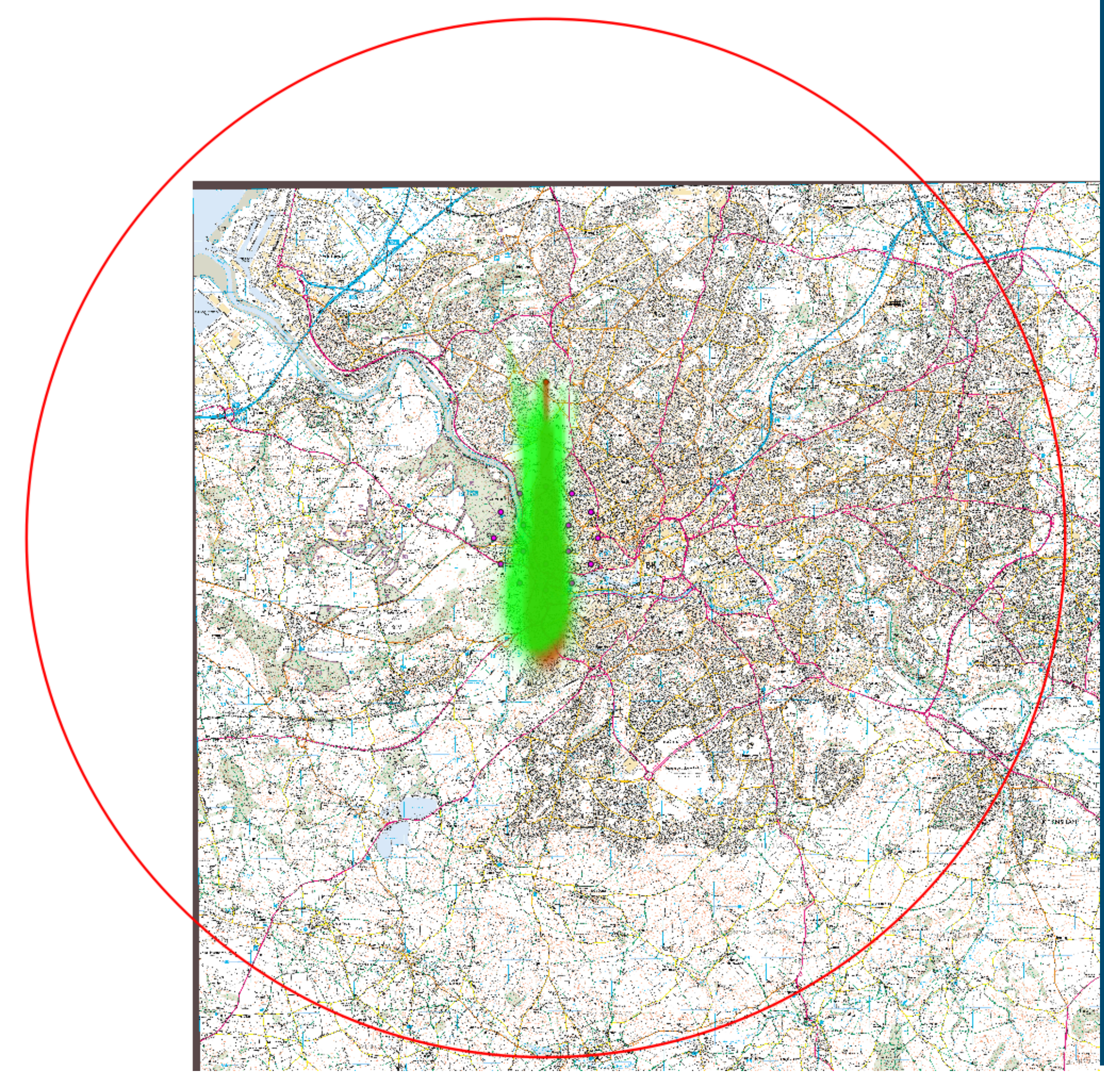
- Temporal gridding
- Spatial gridding

$$\theta = (\underbrace{l_1, l_2, t, m, \ln(d), \ln(v_d), a}_{\text{Source-term}}, \underbrace{u_{*x}, u_{*y}, \tfrac{1}{L}, \ln(z_0), \ln(h_c)}_{\text{Meteorology}}, \ln(\alpha_c), \underbrace{DP}_{\text{Model}})$$

- Location
- Time
- Mass
- Duration
- Deposition velocity
- Agent
- Friction velocity components
- Reciprocal Monin Obukhov length
- Surface roughness
- Canopy Height
- Canopy Flow
- Dispersion model
- Dispersion model output PDF

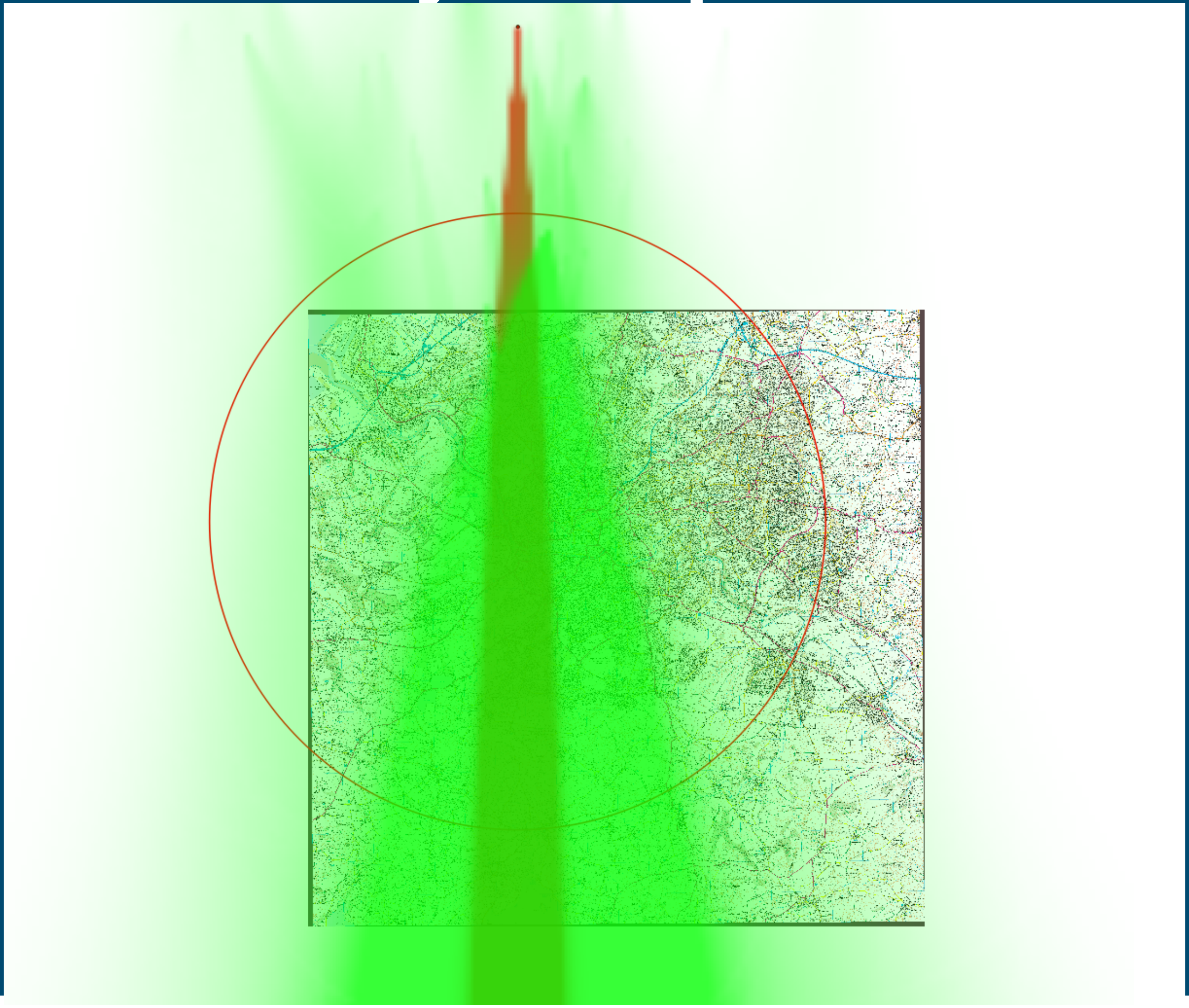
- Floating point values used to index discrete values

Accuracy Compared to ATP45



- Concentration sensor network
- Meteorology sensor
  - Shear LIDAR
  - SODAR
  - Multiple anenometers
- Probability of hazard effect
  - Ground Truth
  - Inferred

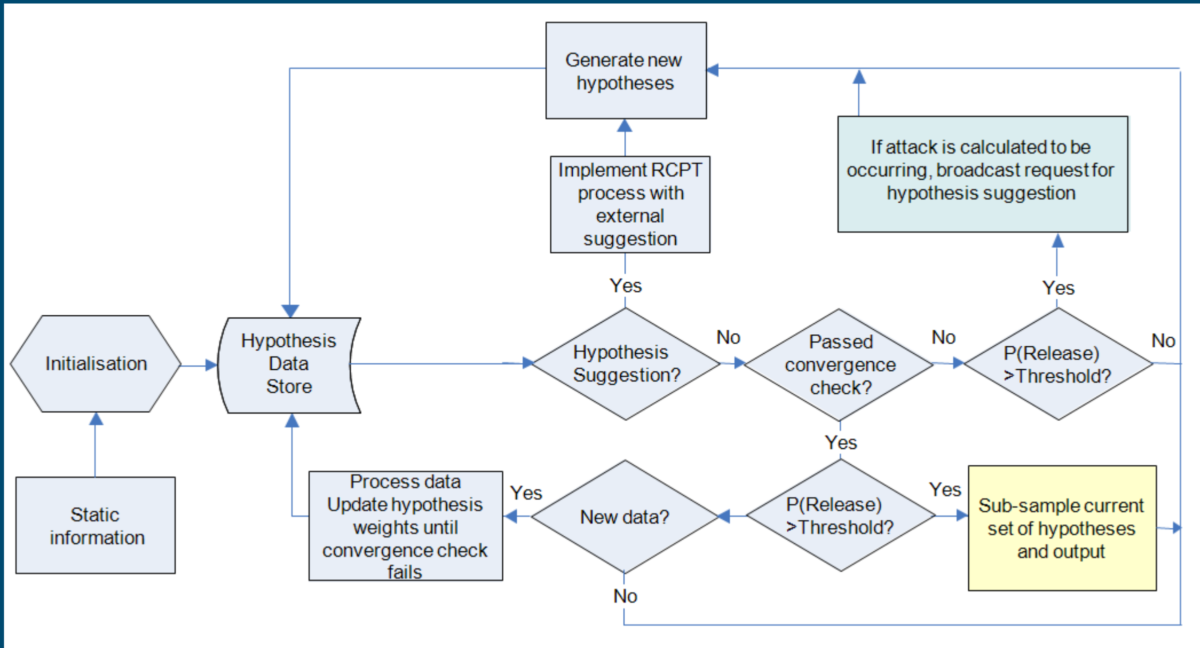
Uncertainty Compared to ATP45



- Single prompt alarm
- Forecast meteorology
- Probability of hazard effect
  - Ground Truth
  - Inferred

Processing dynamic sensor data

- MCBDF performs source-term estimation in real-time
  - A time window is maintained - typically 30 minutes into the past for chem, 2 days for bio
- On receipt of new data, old hypotheses and old data will become obsolete and are removed as they exit the current time window
  - Total likelihood housekeeping
- Remaining hypothesis weights are modified by the likelihood of new data



Exam Question

- Source term – means to an end
- How much material is in the soil
  - Can the land be used for habitation or farming.

Summary

- Deposition survey data as a data source appears effective
- Modelling/inference extensions required