# Source-term estimation for rapid hazard assessment

13th International conference on Harmonisation within Atmospheric Dispersion Modelling - 4th June 2010

Dr. Gareth Brown & Dr. Peter Robins

Defence Science Technology Laboratory (Dstl)

United Kingdom

# Rapid hazard assessment

- The problem:
  - *Successful defence against harmful atmospheric releases requires timely situational awareness*
- The aim:
  - *Optimally assimilate available data to obtain a probabilistic description of the release parameters and the resulting hazard*
- The solution:
  - *Implement a real-time Monte Carlo Bayesian Data Fusion (MCBDF) algorithm for release inference*


© Dstl 2010

# Bayesian data fusion

- MCBDF algorithm uses Bayesian inference over a sample set of hypothesised source-terms and Met. variables
  - it allows the set to be updated when new information is received

$$\theta = (\underbrace{x, y, t, m, a,}_{Source\text{–}term} \underbrace{u, v, L, z_0}_{Met})$$

- The posterior distribution is calculated using Bayes' rule:

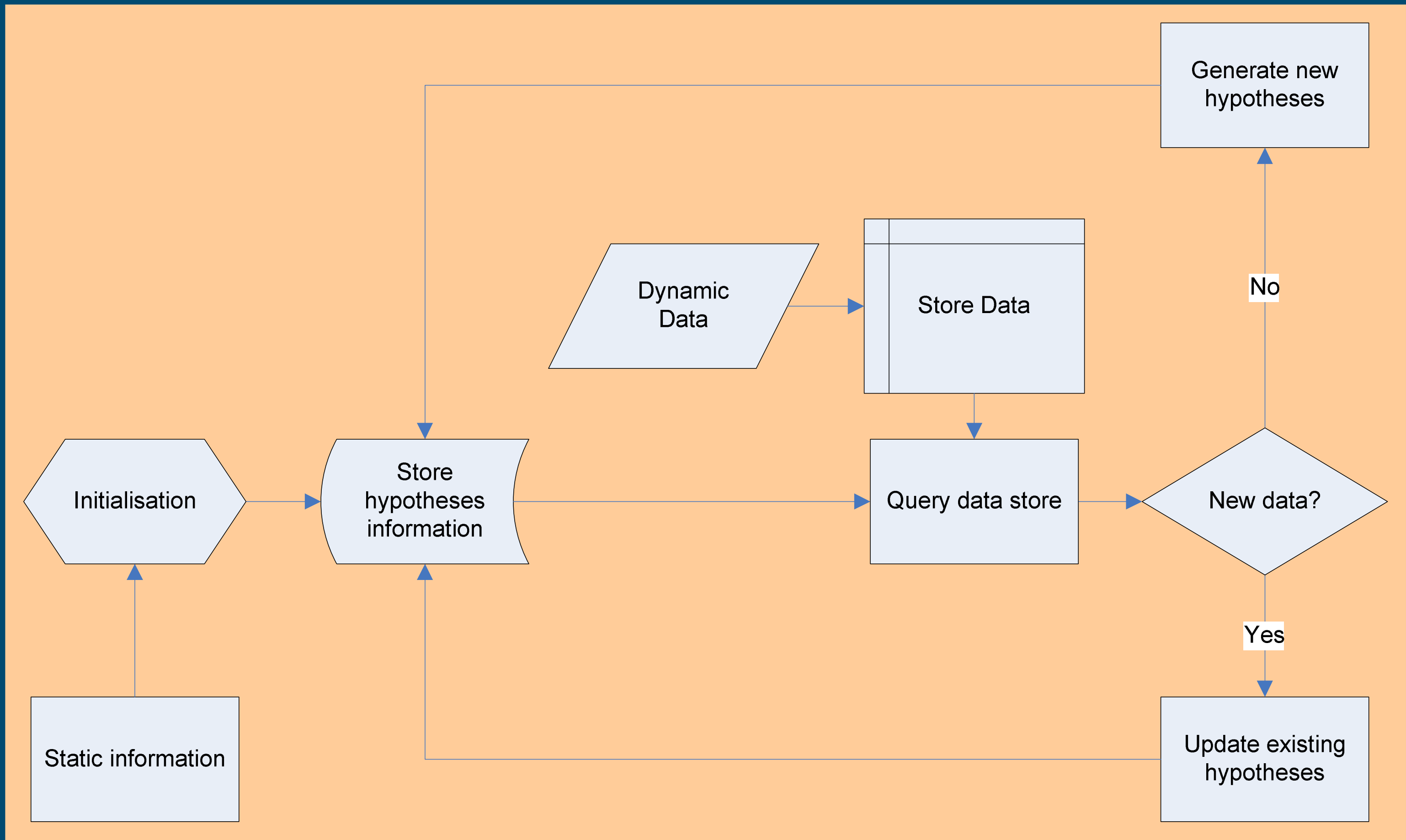
$$\underbrace{p(\theta|\mathbf{D})}_{Posterior} \propto \underbrace{p(\theta)}_{Prior} \underbrace{p(\mathbf{D}|\theta)}_{Likelihood}$$

# Simplistic overview of Monte-Carlo approach to source-term estimation

- Data is constantly arriving – cannot use standard Markov Chain Monte Carlo (MCMC)
- Release is a fixed point in space-time – cannot use standard Sequential Monte Carlo (SMC)
- Use a hybrid solution:
  - Fixed length, sliding data window to keep computational complexity constant
  - Hypotheses sampled and dispersion models run using MCMC in idle time
  - Hypothesis sample weights updated using stored dispersion model data when new data is received
- The combination of hypothesis weight and clustering in hypothesis space describes the *posterior* and permits probabilistic hazard assessment


© Dstl 2010

# Structure of MCBDF algorithm



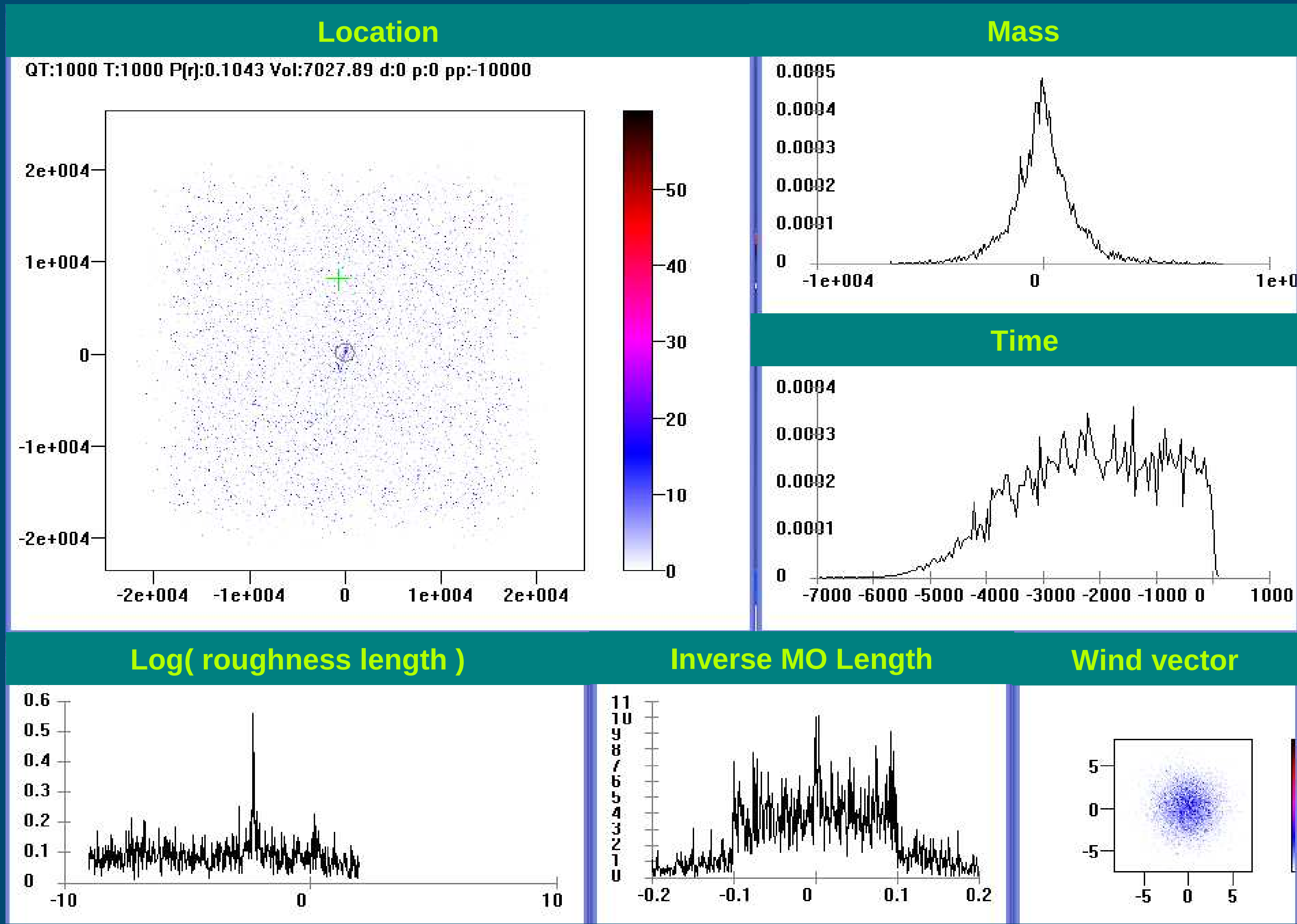
# The prior distribution

- Incorporates expert knowledge and previous experience about the shape of the hypothesis space
- Includes defining
  - Likely release locations
  - Realistic wind speeds, stabilities and surface roughness
  - Probable release times and release masses
- The mass prior is given as a double exponential

$$p(m^*) = \frac{1}{2} e^{-\left| m^* \mu_m \right|} \qquad \mu_m \equiv \text{ scale for release mass}$$


© Dstl 2010

# MCBDF: example prior distribution






© Dstl 2010

# Met Likelihood calculations

- When Met. data from a sensor is passed to MCBDF, MCBDF calculates how likely it is that the data could be generated from each hypothesis

$$\theta_k^{(i)} \equiv \text{Hypothesis } i \text{ at time slice } k$$

- The weight of each hypothesis is updated as

$$w_{k+1}^{(i)} = w_k^{(i)} p\left(d \middle| \theta_k^{(i)}\right)$$


© Dstl 2010

# The wind-vector likelihood model

- When wind-vector measurements are passed to MCBDF, the likelihood is calculated as

$$p\left(d \middle| \theta_k^{(i)}\right) = p\left(\mathbf{u} \middle| \boldsymbol{\mu}_u, \boldsymbol{\Sigma}\right) = \phi\left(\mathbf{u} \middle| \boldsymbol{\mu}_u, \boldsymbol{\Sigma}\right)$$

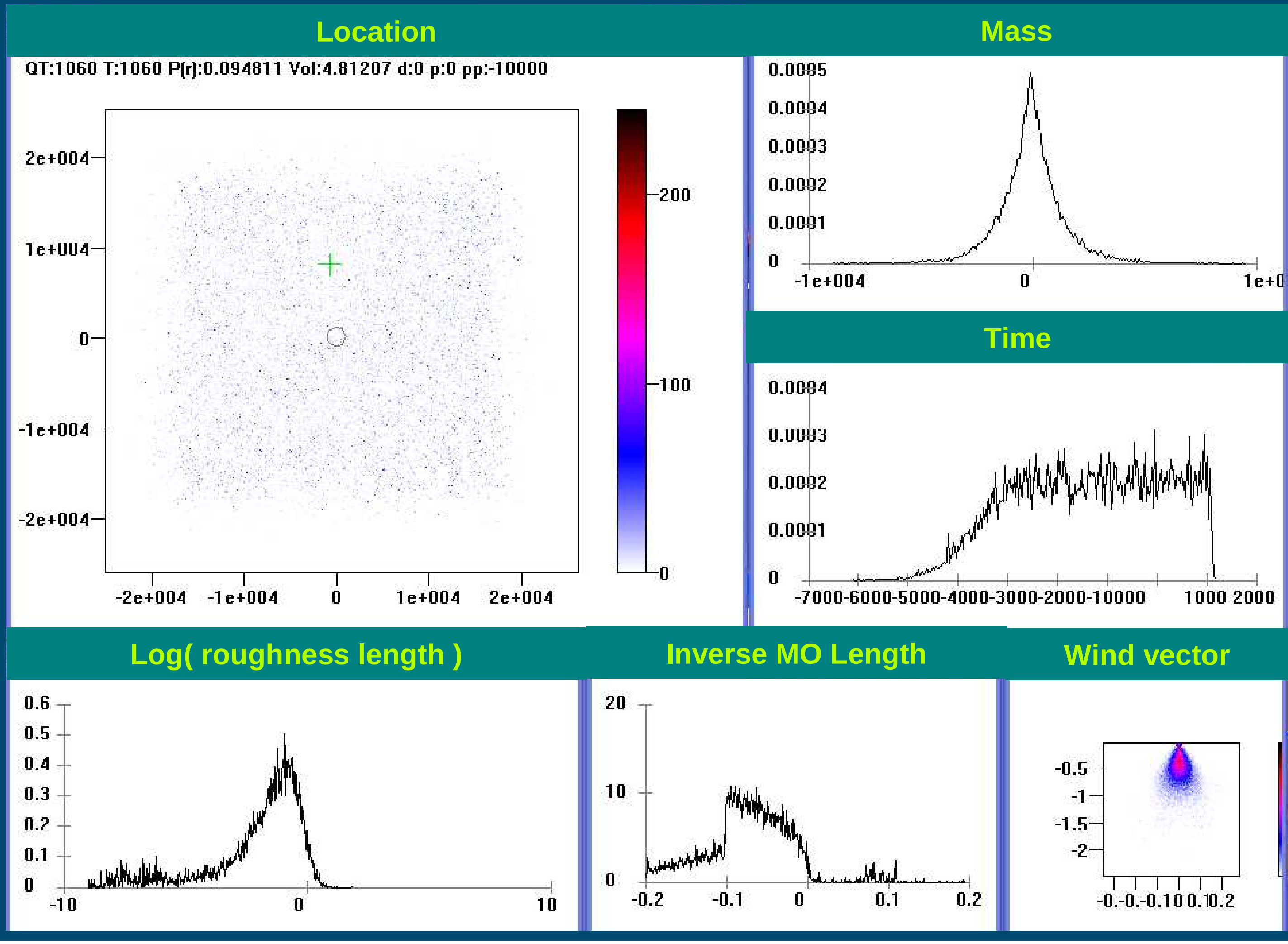
bi-variate normal distribution

$u \equiv$ Measured wind-vector

$\boldsymbol{\mu}_u \equiv$ Dispersion model's simulated mean wind vector at same time and location as measurement

$\boldsymbol{\Sigma} \equiv$ Measurement uncertainty covariance matrix

# MCBDF: upon receipt of met data





© Dstl 2010

# Detector likelihood calculations

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

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

$d$ $\equiv$ Detector measurement

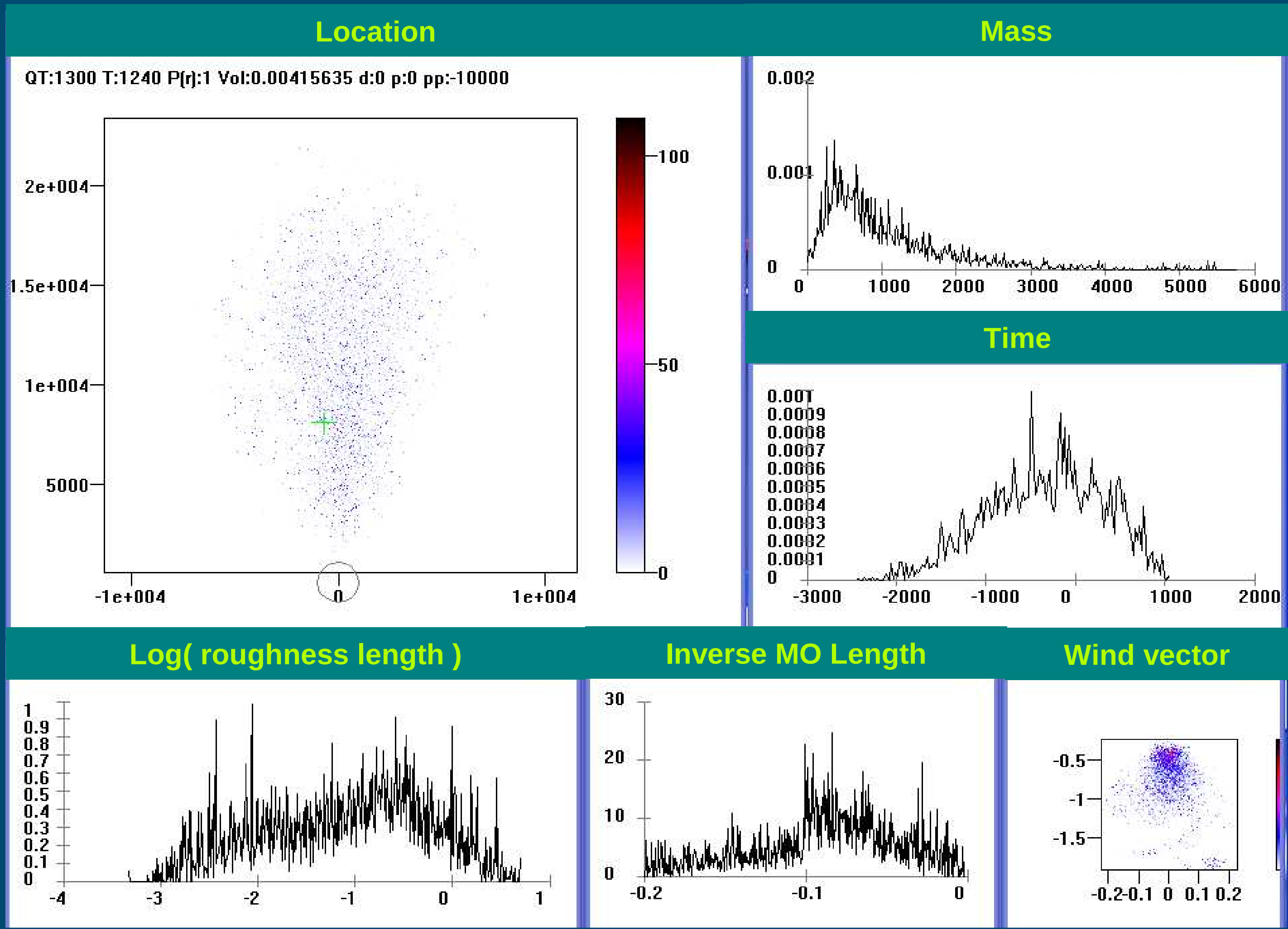
$\mu$ $\equiv$ Mean mass-concentration from dispersion simulation

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

$c$ $\equiv$ Unobserved ground-truth concentration


© Dstl 2010

# MCBDF: upon receipt of a detection



# Hypothesis generation

- MCBDF estimate of the posterior improves as it generates more and more hypotheses compatible with the data and rejects those that are incompatible
- New hypotheses are generated using Differential Evolution Markov Chain (DE-MC) Monte Carlo
- A candidate hypothesis, $\theta_i^*$, is generated by "mixing" up three current distinct hypotheses *i, j* and *k*

$$\theta_i^* = \theta_i + \gamma\left(\theta_j - \theta_k\right) + \varepsilon$$


© Dstl 2010

# Hypothesis generation

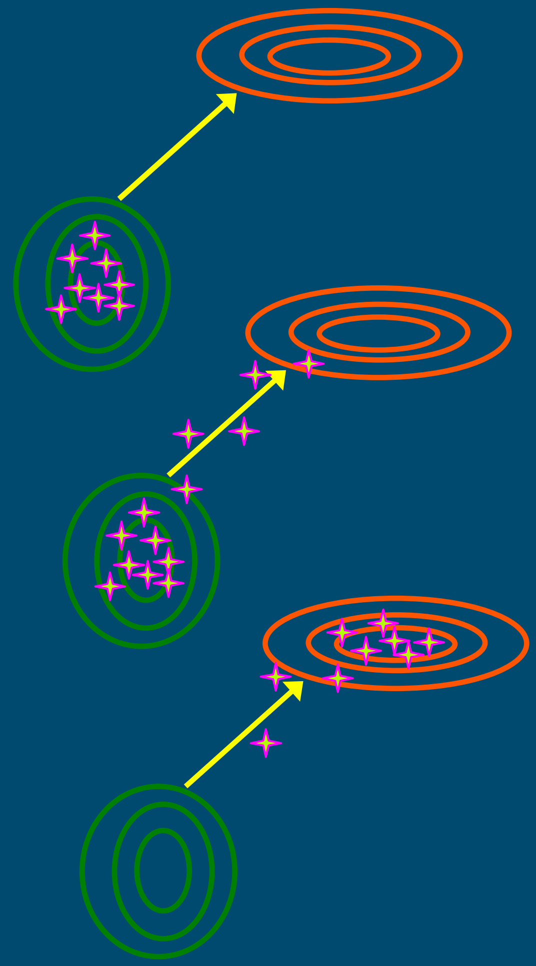
- The candidate hypothesis, $\theta_i^*$, is probabilistically added to the sample set based on the standard Metropolis accept-reject step:

$$U(0,1) < \frac{p(\theta_i^*) \prod_{j=1}^{N_d} p(d_j \mid \theta_i^*)}{p(\theta_i) \prod_{j=1}^{N_d} p(d_j \mid \theta_i)}$$

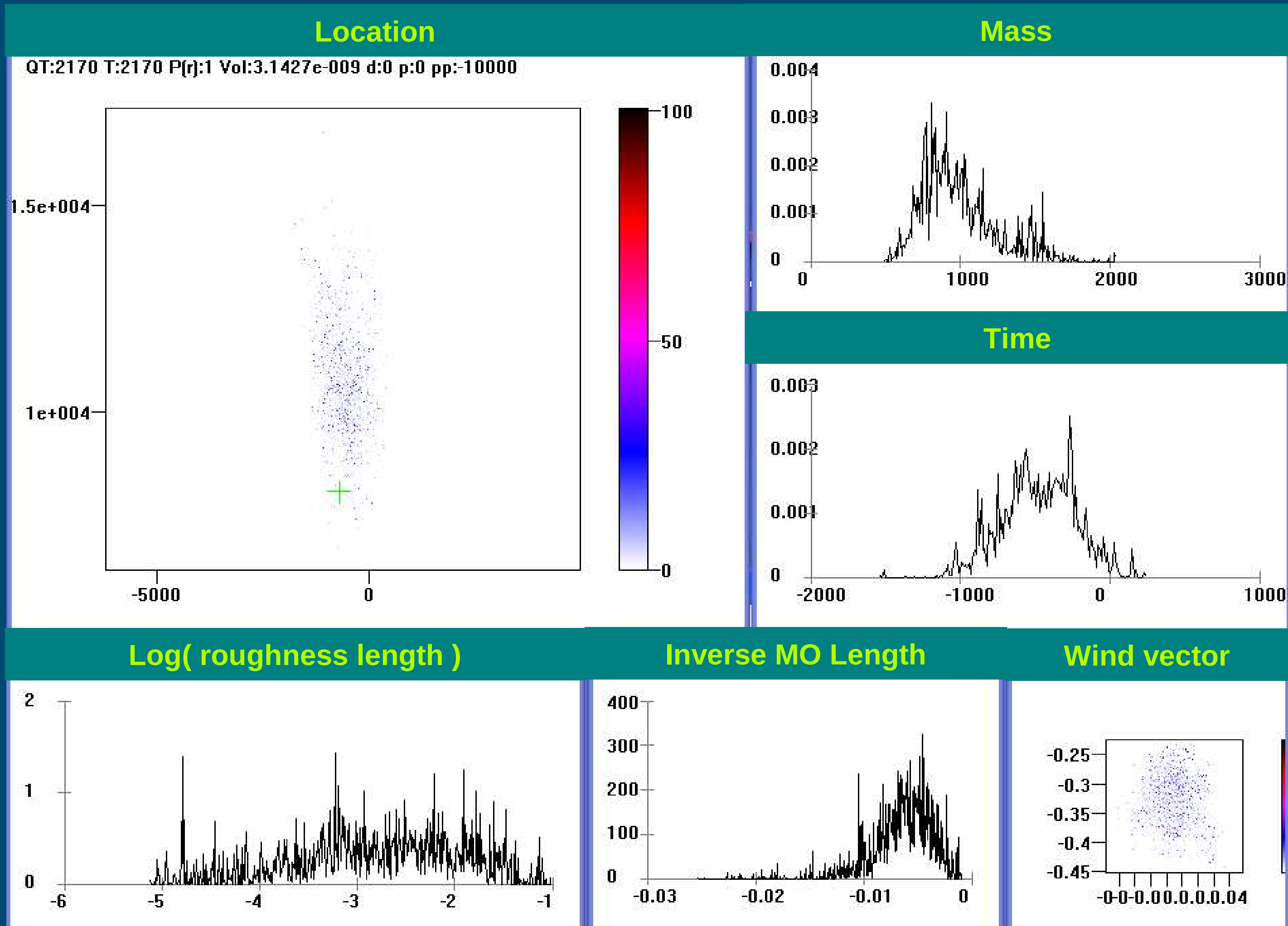

© Dstl 2010

# Burn In and Convergence

- Initial hypotheses may be far from the peak of the posterior
- But rapid answers are required
  - Incoming data continually changing posterior so population struggles to get to target
  - Once there, limited time for new sample weights to make the old ones insignificant



# MCBDF: estimate converged




17 May 2010

© Dstl 2010

DSTL/CP45472

HARMO13

Dstl is part of the Ministry of Defence

# Hazard calculation

- An operational system must first determine if a hazard is present, the probability of release is calculated as

$$P\left(m^* > 0 \middle| \mathbf{D}\right) = \frac{P\left(m^* > 0\right)\sum_i w_k^{(i)} \mathrm{I}\left(m_i^* > 0\right)}{P\left(m^* > 0\right)\sum_i w_k^{(i)} \mathrm{I}\left(m_i^* > 0\right) + \left(1 - P\left(m^* > 0\right)\right)\sum_i w_k^{(i)} \mathrm{I}\left(m_i^* \leq 0\right)}$$

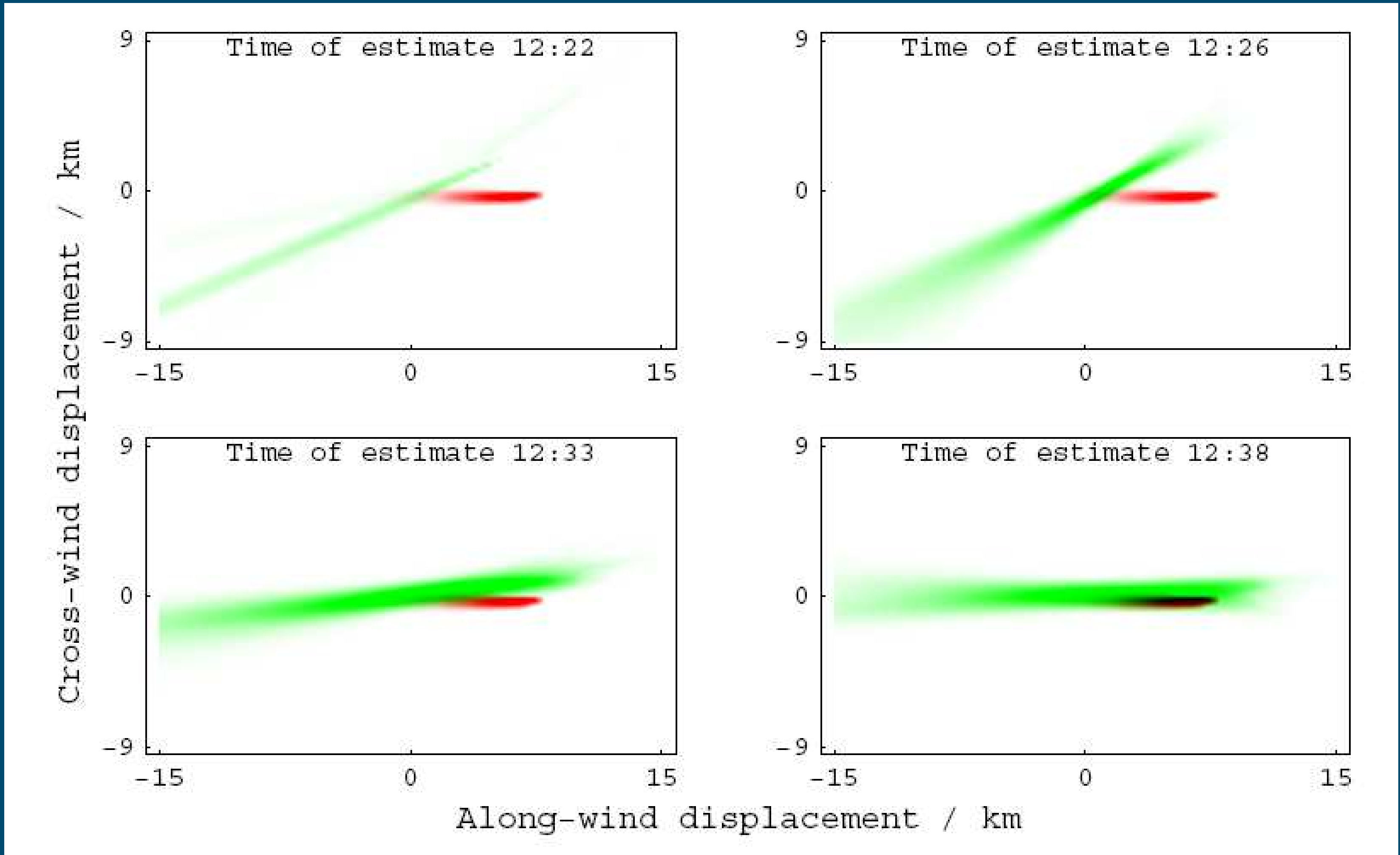
$w_k^{(i)} \equiv$ weight of the *i*th hypthesis

$P(m^* > 0) \equiv$ prior on a release occuring

$\mathrm{I}(m_i^* > 0) \equiv$ indicator function


© Dstl 2010

# Example hazard calculation






© Dstl 2010

# Summary

- Dstl have developed a real-time Bayesian inference engine capable of providing probabilistic hazard assessments
- The MCBDF algorithm can infer the uncertainty associated with the dispersion models' Met. parameters as well as release parameters
- The algorithm can therefore handle erroneous Met. input as demonstrated by its hazard calculator


© Dstl 2010

# Future work

- Inference on release duration in real-time on a desktop machine
- Extension to inference on multiple releases
- Extension to inference on the dispersion model and the form of the concentration probability distribution function
- More validation with trials data
  - Demonstration with FFT07 data


© Dstl 2010