Assessment of environmental compliance of waterbodies through integration of monitoring and modelling

Joseph V. McGovern B.E. (Envir. Eng) NUIG
Supervisor Dr. Michael Hartnett
Environmental Informatics
Environmental Change Institute

ECI
Environmental Change Institute
National University of Ireland, Galway, Ireland

epa
Environmental Protection Agency
An Ghníomhaireacht um Chaomhnú Comhshaoil

OÉ Gaillimh
NUI Galway

NDP
National Development Plan 2007 - 2013

Overview

- Primary and secondary work objectives
- Bayesian Maximum Entropy overview
- BME process in detail
- Uncertainty assessment
- Interpretive maps for monitoring and assessment
- Conclusions and questions

Project objectives

- Primarily:
  - Develop methodology - accurate & cost effective monitoring programmes for mandatory compliance assessment
  - Geostatistical tool for assessment of historic dataset uncertainty
  - Assessment of the spatial representation of water quality based on historical monitoring data – spatial limit for one monitoring point
- Additionally:
  - Generic format
  - Regular updating of monitoring programmes
  - Relationships between nutrient status and WFD status
  - Uncertainty in model outputs
  - BME – maps of iso-flushing contours

Bayesian Maximum Entropy Overview

- Mapping of environmental variables taking into account all available data
- Utilises uncertain data
- Kriging facilitates interpolation for mapping – not extrapolation
- BME – interpolation and extrapolation in space and time
- Posterior PDFs generated throughout spatiotemporal grid
- Posterior PDFs fully reflect underlying data – no assumed Gaussian normal shape – non linear estimator generated

Bayesian Maximum Entropy

- Varying data types:
  - Soft data: model output
    uncertain historic datasets
  - Hard data: recent monitoring data (EPA)
  - Background knowledge

- 3 clearly defined steps:
  - Prior: analysis of general data
  - Meta-prior: separation of available soft data
  - Posterior: integration of previous data

Bayesian Maximum Entropy

- Prior stage:
  - Produces a general pdf f_G
  - Shaped by constraints
- Meta-prior stage:
  - Soft probabilistic data
  - Soft interval data
- Posterior stage:
  - Update general pdf

Bayesian Maximum Entropy: Prior stage

- Prior PDF of the form $f_G = e^{\mu_0 + \mu^T g}$
- g refers to the vector of general knowledge equations
- μ_0 and matrix of μ values determined by constraints
- Constraints determined by:
  - Statistical moments (Mean, Covariance, Variogram)
  - Physical laws - site knowledge
- Solution to prior pdf:
  - Substitution of unsolved prior pdf equation into general knowledge equations
  - Solve for values of μ_0 and matrix of μ

Bayesian Maximum Entropy: Prior stage

- Prior knowledge – general knowledge equations
- Spatiotemporal random field theory:
  - Grid definition: $p_{map} = (p_1, p_2, \ldots p_m, p_k)$
  - Random variables: $x_{map} = (x_1, x_2, \ldots x_m, x_k)$
  - Realisations: $\chi_{map} = (\chi_1, \chi_2, \ldots \chi_m, \chi_k)$.

- General knowledge equations:

$$\overline{h_\alpha}(p_{map}) = G_\alpha[\chi_{map}, p_{map};\ f_G]$$

$$\overline{h_\alpha}(p_{map}) = \overline{g_\alpha(x_{map})}, \qquad \alpha = 0, 1, \ldots, N_c$$

$$\overline{g_\alpha(x_{map})} = \int d\chi_{map}\, g_\alpha(\chi_{map}) f_G(\chi_{map};\, p_{map})$$

Bayesian Maximum Entropy: Prior stage

- General knowledge equations:

$$\overline{h_\alpha}(p_{map}) = G_\alpha[\chi_{map}, p_{map};\ f_G]$$

  - h_α - terms representing statistical moments: mean, covariance, variogram, third order moments of data
  - 1st order statistical moment: mean $\overline{x_i}$
  - 2nd order statistical moment: covariance $\overline{(x_i - \overline{x_i})(x_k - \overline{x_k})}$

$$\overline{g_\alpha(x_{map})} = \int d\chi_{map}\, g_\alpha(\chi_{map}) f_G(\chi_{map};\ p_{map})$$

  - g_α - a function involving a realisation at the grid point in question
  - Realisation corresponding to 1st order moment χ_i
  - Realisation corresponding to 2nd order moment $\chi_i\ \chi_k$
- Process repeated for each constraint type at each spatiotemporal grid location

Bayesian Maximum Entropy: Entropy Maximisation

- Entropy = potential information
- Knowledge in prior PDF maximised
- Inverse relation between information content and probability

$$Info_G[\chi_{map}] = log\{Prob_G[\chi_{map}]\}^{-1} = -log\{Prob_G[\chi_{map}]\}$$

- Log scale limits extent of information measure to 10
- Maximisation of following expression:

$$M = -\int d\chi\, f_G(\chi)\, log\ f_G(\chi)$$

- Expression maximised by solving the Euler Lagrange equation:

$$\frac{\delta(-f_G(\chi)\, log\ f_G(\chi))}{\delta f_G} + \sum_{\alpha=0}^{N_c} \mu_\alpha \frac{\delta(g_\alpha(\chi) f_G(\chi))}{\delta f_G} = 0$$

Bayesian Maximum Entropy: Prior stage

- Empirical results or physical laws add to *a priori* knowledge
- Can provide additional structure to general knowledge equations
- Soil moisture X and rainfall Y given by:

$$\eta Z_r \frac{\partial}{\partial t} X(p) = -\eta X(p) + \kappa \nabla^2 X(p) + Y(p)$$

  - 1st order knowledge equation:

$$\overline{h_1} = \eta Z_r \frac{\partial}{\partial t} \overline{X(p)} \quad \overline{g_1} = \int d\chi(-\eta\chi + \kappa\chi\nabla_s^2) f_G(\chi) + m_y$$

  - 2nd order knowledge equation:

$$\overline{h_2} = \eta Z_r \frac{\partial}{\partial t'} C_x(p, p') \quad \overline{g_2} = \iint d\chi d\chi' \chi\chi' (-\eta + \kappa\nabla_s^2) f_G(\chi, \chi') + \iint d\chi d\psi' \chi\psi' f_G(\chi, \psi')$$

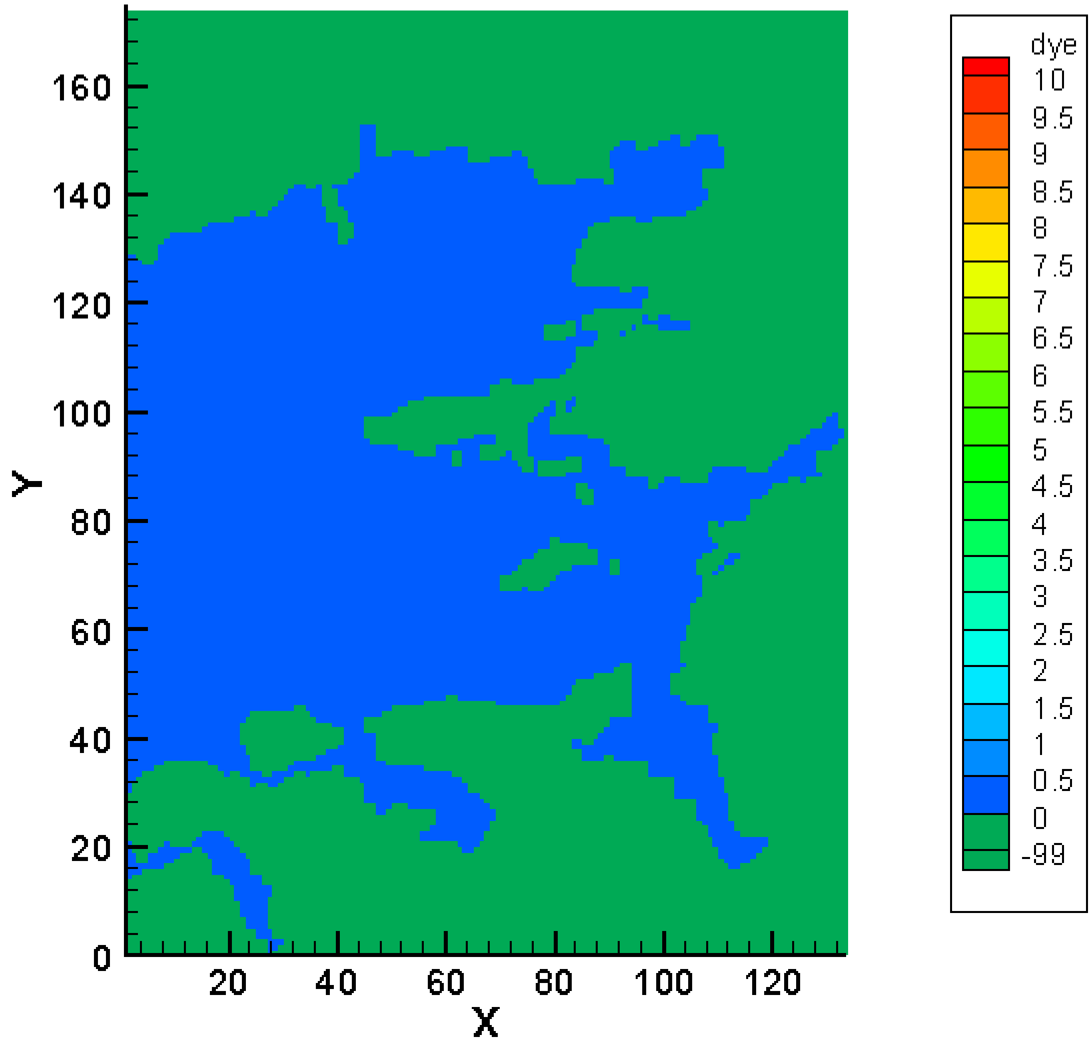
Bayesian Maximum Entropy: Meta-prior stage

- Available data: Hard or soft?
  - Hard data: Parameter values returned from lab analysis, using the best current practice
  - Soft data: Historic data (less accurate) model output data (interval data), probabilistic data (using probe measurements)

Bayesian Maximum Entropy & DIVAST

- Depth integrated velocity and solute transport model
- Developed by R.A. Falconer at the University of Bradford, U.K.
- Applicable to shallow well mixed coastal and estuarine water bodies
- 2-D finite difference model
  - Hydrodynamic module:
    - Navier-stokes equations
    - Yields water currents & elevations
  - Water quality & solute transport module:
    - Advection-diffusion equations
    - Salinity, BOD, organic, ammoniacal and nitrate nitrogen, DO, chlorophyll *a*, organic phosphorus and orthophosphate

Bayesian Maximum Entropy & DIVAST



Bayesian Maximum Entropy: Posterior stage

- General knowledge based prior PDF updated by Bayesian Conditionalisation
- PDF updated at each grid point using soft data from 3-5 adjacent grid points in space and time
- Posterior PDF given by the following:

$$f_K^{bc}(\chi_K) = A^{-1} \int_D d\Xi_S(\chi_{soft}) f_G(\chi_{map})$$

$$A = \int_D d\Xi_S(\chi_{soft}) f_G(\chi_{data})$$

- See Bayesian Conditionalisation:

$$Prob_K\{\chi_k | \chi_{data}(S)\} = \frac{Prob_K\{\chi_k \text{ and } \chi_{data}(S)\}}{Prob_K\{\chi_{data}(S)\}}$$

Bayesian Maximum Entropy: Spatiotemporal estimates

- BMEmode estimate - most likely value at grid pt. given by:

$$\frac{\delta}{\delta \chi_k} f_K(\chi_k)|_{\chi_k=\hat{\chi}_k} = 0$$

- BMEmean estimate – value minimises the mean square estimation error:

$$\hat{\chi}_k = \int d\chi_k \, \chi_k \, f_k(\chi_k)$$

BME: Uncertainty assessment

- Standard deviation of posterior PDF at each location is determined by:

$$\sigma_{k|K}^{bc} = \left[\int d\chi_k (\chi_k - \hat{\chi}_{k,mean})^2 f_K^{bc}(\chi_k)\right]^{1/2}$$

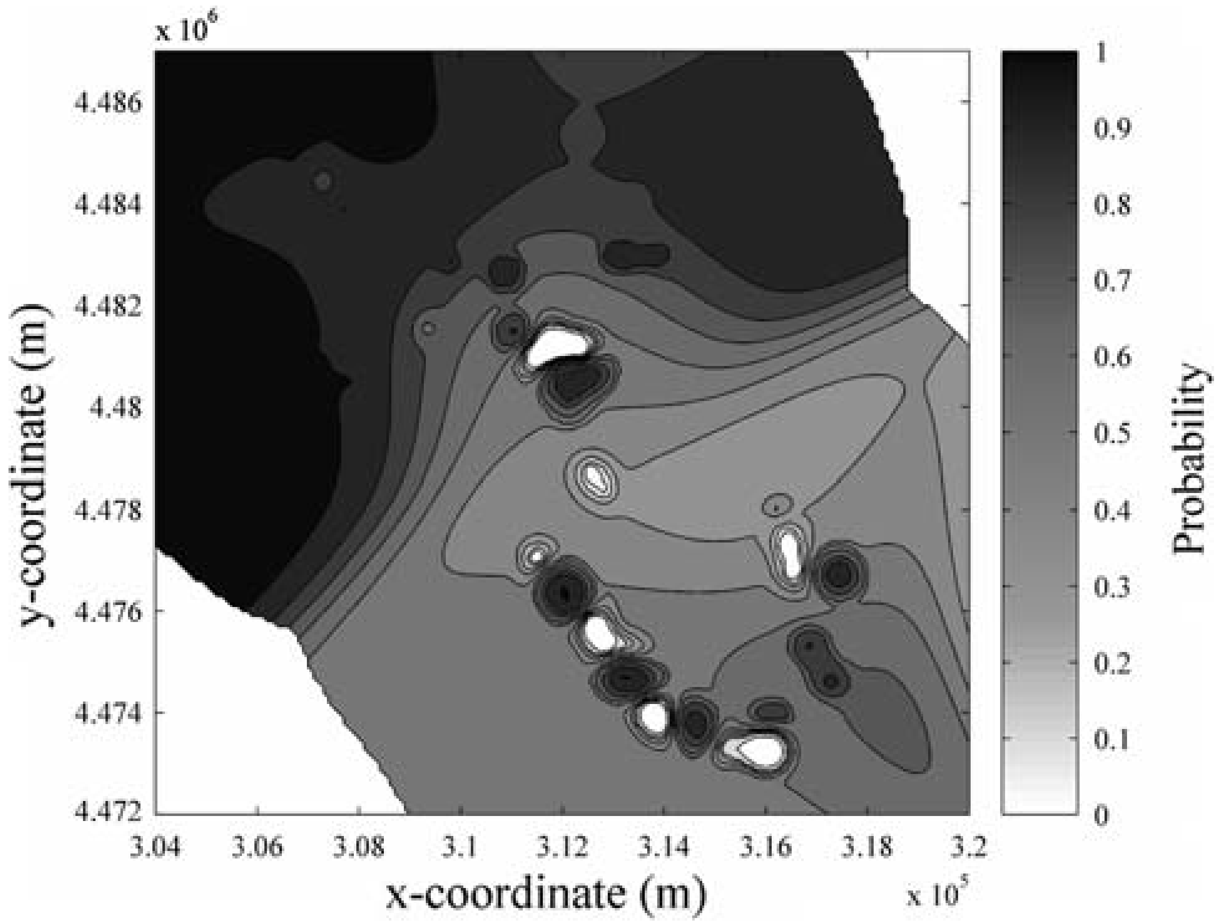
- Limits of confidence interval depend upon desired confidence level (mean+/- 1.96*std for 95% CI)
- CI centred about BME estimate
- Confidence Interval a suitable proxy for uncertainty assessment for monitoring purposes
- Uncertainty assessment of estimates – validation soft datasets

Interpretive maps for monitoring and assessment

- Risk assessment maps
  - Posterior PDFs throughout grid considered
  - Each location graded 0 – 1
    - 0 – no portion of PDF is above elected limit value
    - 1 – entire PDF occurs above the limit value
    - Intermediate values – grade depending upon proportion of PDF above limit
    - BME analysis carried out on grades – map generated from results.
  - See *Modis, K., Vatalis, K., Papantonopoulos, G., Sachanidis, Ch., (2010). Uncertainty management of a hydrogeological data set in a greek lignite basin, using BME. Stoch Environ Res Risk Assess 24:47-56.*

Interpretive maps for monitoring and assessment

- Modis *et al.* *(2010)* Risk assessment maps

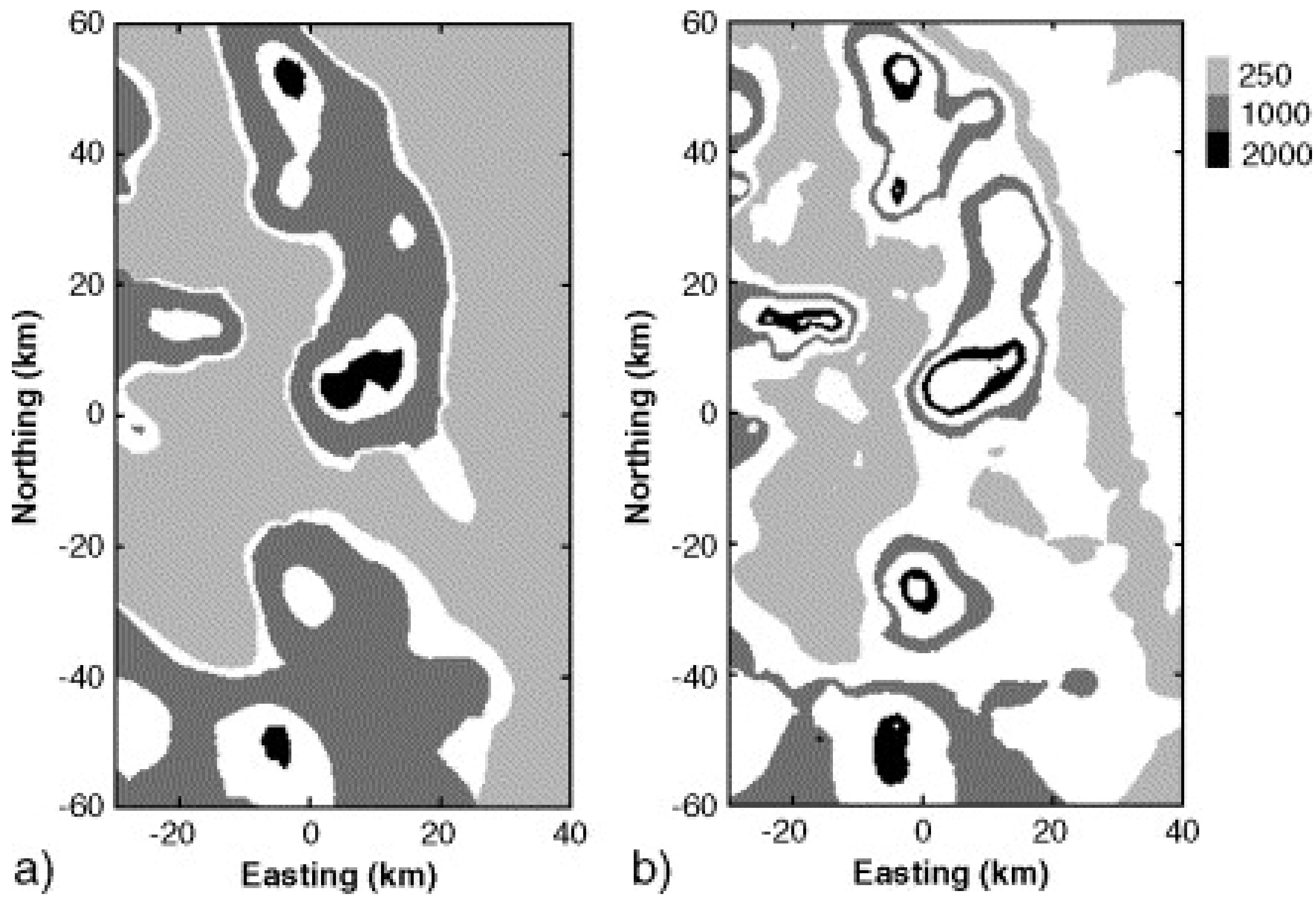


Interpretive maps for monitoring and assessment

- “Thick” contour maps
- Contours indicative of:
  - Parameter concentrations
  - Prediction uncertainty
- Map assembled for e.g. 90% Confidence
- Each thick contour - a zone which contains points who’s confidence interval contains the value of the contour
- Contour thickness indicative of uncertainty
- See *Savelieva, E., Demyanov, V., Kanevski, M., Serre, M., Christakos. G. , (2005). BME-based uncertainty assessment of the Chernobyl fallout. Geoderma 128:312-324.*

Interpretive maps for monitoring and assessment

- Savelieva *et. al. (2005)* Caesium 137 "thick" contours



BME: Relevance to monitoring and potential uses

- Estimation at unknown locations/instances
- Rigorously processes sparse datasets of varying quality
- Lowers estimation uncertainty given sufficient soft data
- Most probable value generated
- Mapping:
  - Optimise monitoring programmes
  - Guide advanced monitoring
- Cost of monitoring lowered
- Investigate low cost techniques

In conclusion…

- Limited WQ monitoring programme guidance on design and optimisation
- BME utilises all available data
  - Enhances understanding
  - Optimise monitoring
- BME routines central to objectives

- Thanks to the Irish EPA for funding provided under the STRIVE postgraduate research programme funded under the National Development Program 2007-2013