NERC SCIENCE OF THE ENVIRONMENT

Meeting Report: NERC Integrated Environmental Modelling Workshop

Held at the British Geological Survey, Keyworth, 4-5th February

Editors: KR Royse and AG Hughes

BAS BGS CEH NCAS NCEO NOC

Introduction

This report describes the results of the NERC Integrated Environmental Modelling (IEM) workshop that was held at BGS Keyworth on the 4th to 5th February, 2014. The meeting brought together some 35 scientists from the NERC Science Office and five NERC centres and surveys. The workshop builds on previous meetings which have been convened over a number of years; such as the International summit on Integrated Environmental Modelling, Washington (2010), the AGU fall meeting, San Francisco (2009), and the international congress on Environmental Modelling and Software (2010). From these meetings it was recognised that there were many communities involved in developing IEM and that the two main groups of the US and Europe needed to work together to create an open community for all. Out of these meetings it was clear that there was a need to; provide accessible linkable components, to address uncertainty, to professionalise the development of integrated models, to engage with the user community (particularly decision and policy-makers) and to develop a community of practise to aid the development and uptake of IEM.

The aim of this meeting was to:

- Develop a strategy for Integrated Environmental Modelling (IEM) within NERC
- Develop an Integrated Environmental Modelling Methodology that can be easily adopted both inside and outside NERC
- Establish several exemplar projects that are cross-cutting (institutional and discipline) and also deliver the key goals of IEM and the new NERC strategy

The outcomes from the Washington meeting in 2010 have been considered and further developed. From this, the proposed NERC IEM strategy has been broken down into 4 key areas which formed the topics for each of the breakout groups and are: -

1. The premise that eventually all models will at some point in their life-cycle need to be linked to other models. What needs to be done to enable this to happen?
2. To encourage the development of modelling platforms. What needs to be done to enable this to happen?
3. To assess and quantify uncertainty arising from integrated modelling. What needs to be done to enable this to happen?
4. To develop specific exemplar projects. How should this be done within NERC and with NERC's partners?

The report outlines the strategy and vision for Integrated Environmental Modelling in NERC and then gives a description of the main findings from each of the 4 main topics areas addressed within the NERC IEM workshop

Vision:

'combine data, models and visualisation techniques with an assessment of uncertainty in an integrated, flexible and reusable manner using a multi-disciplinary approach to inform policy decisions and communicate environmental problems and solutions to society'

NERC IEM Strategy

The environment is too complex for us to reason through the needs of policy without models. To answer the deceptively simple 'what if' questions, such as how could a tsunami affect transport and energy infrastructure? We need to be able to improve our predictive modelling capabilities. These questions are often the result of complex interactions between the natural and managed environment, engineered structures and humans. It is increasingly recognised that different sectors are interlinked; thus energy policy interacts with the water, land use (and hence food production), climate mitigation and industrial policies. To make evidence based and holistic assessments for risk assessment, policy and development we need an integration of our models within an Integrated Environmental Modelling (IEM) framework.

Integrated Environmental Modelling provides a means of achieving this by making models more easily interoperable, increasing access and allowing for the formation of a consultative infrastructure. These advances enable the application of multi-disciplinary knowledge to explain, explore and predict the Earth's response to natural and human induced stressors and provide the answer to the 'what if' questions facing policy-makers today.

IEM provides a science based structure to assimilate and organise multidisciplinary knowledge and information. It provides a way of applying this knowledge to explain, explore and predict an environmental system response to natural and human induced stressors. Its structure serves as a unifying vehicle of communication among stakeholders' diverse perspectives, values and priorities. IEM serves decision-makers' needs to understand the dynamic coupled systems involving socio-economic and environmental components, compare impacts of several scenarios, analyse trade-offs and adapt strategies based on ongoing monitoring and observation of the system and respond to possible unintended consequences.

The NERC IEM strategy addresses these challenges by creating a community that prioritises modelling needs and a modelling framework to facilitate greater sharing and linking of data and models. The overall objective is to promote a change in culture towards greater collaborative working, improved accessibility and effective use of existing models and tools developed over many years and also to encourage a move away from 'silo management' where each problem is addressed in isolation.

The key challenges to be addressed within IEM are:

- Development of workable standards that define model input and output unambiguously and provide a description of the model and the assumptions it is based on.
- Understand and quantify uncertainty, so that we can understand how uncertainty is propagated within a linked modelled system and adopt user specific communication methods.
- Provide long term support for infrastructure, which is future proofed and supported by the IEM modelling community.
- Develop easy to use model integration platforms appropriate to the scale of the problem being addressed, with defined standards for the interaction between models and their data.
- Understand the impact of temporal and spatial scales on model operation and processing. Design tools that can explore and mitigate the impacts of changing scales on linked model systems.
- Develop framework independent model libraries, which would facilitate the exposure of models and data as web-services.
- Develop our capabilities within Integrated Environmental Modelling through a set of real world case studies and problem solving by developing generic solutions.

These will be addressed by:

To address this, a programme of work will be developed through four exemplar projects which will highlight the breadth and depth of scientific interdisciplinary research which can be achieved through integrated modelling methods (see Figure 1). The infrastructure programme will develop the supporting technology to sustain the IEM modelling communities and enable the science user to be able to find linkable components and use them. The exemplar projects will address one or more real-world problems and have been chosen to take IEM from being a concept to an operational reality.

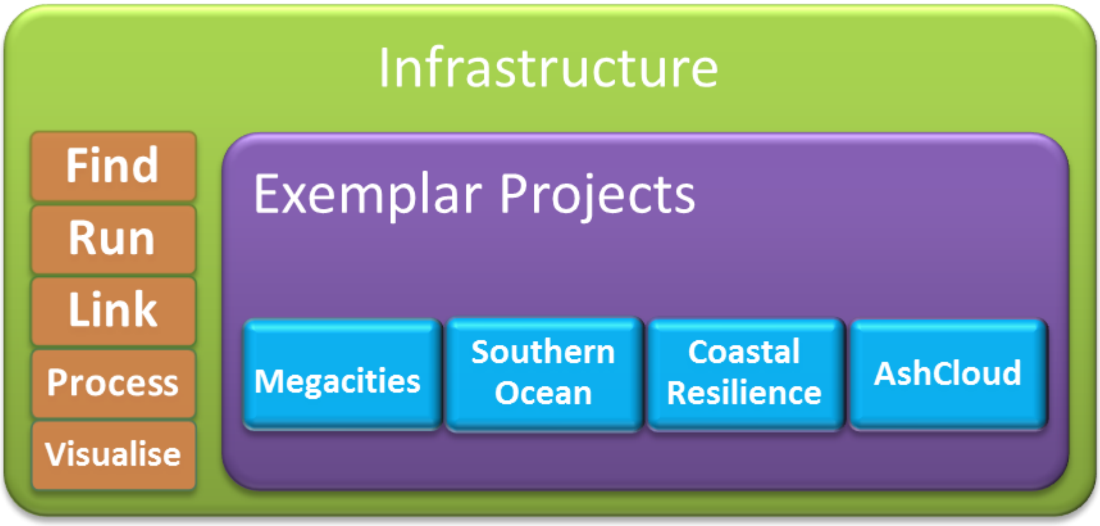

Figure 1. Diagram showing relationship between different activities within any programme

In the future most models will at some point in their life-cycle need to be linked to other models. What needs to be achieved in order for this to happen?

Standards are critically important to enable integrated modelling to happen. Linked to this, is a need to motivate and incentivise researchers to adopt standards during the modelling process. The reasons given for this were:

- For all models, researchers need to be able capture the scientific accuracy of the final model results
- There is a danger that by re-using model components; users can make inappropriate links and be using model components beyond the design limitations
- Quantification and assessment of uncertainty, necessitates the use of a standard set of metadata to describe the model, its inputs and outputs and assumptions used
- Need to be able to bench-mark model performance
- Need to encourage multiple agencies to use the models whose development is funded by NERC workflows and frameworks
- Once we have a defined set of standards linking models becomes much easier

It was clearly recognised that we need a 'Taxonomy' of standards which would allow researchers to select and adopt those standards that best suited the modelling work being undertaken. There is a need for a minimum set of standards which must be light touch and simple to implement and standard for all components within a linked modelled system.

To be included within an IEM standard:

A clear ontology defining types, properties and inter-relationships of the modelling processes and data manipulation undertaken, i.e. what has been inputted and outputted by the models and what operations (modelled processes) have been performed.

The necessary information required to understand a model, such as a description of the models and the assumptions used, and the parameter types that the model can link to. This will enable future users to more easily understand the scientific accuracy that the final model can produce.

Another element was, understanding what would be gained from linking models together. Exemplars should therefore examine the benefits and time required for linking and we should clearly be able to demonstrate the advantage in dynamic linking. Technology platforms that will be generated need to be appropriate to the scale of the problem we are looking at.

There will be a need to generate a critical mass of linkable components (both models and the data used to serve them) in order to gain momentum. By delivering the infrastructure development through high profile exemplar projects, it should be possible to generate the momentum required.

We should incentivise by creating a culture where the individual contributions of researchers are rewarded i.e. citations given to modellers when their models are re-used and at a corporate level; work with RCUK/NERC to encourage the use of a limited number of standards and frameworks.

IPR and Corporate Liability will need to be addressed, to ensure models can be made available to be linked. For the IEM community to grow, an open source and open access ethos will need to be adopted wherever possible. The IPR value will be in being the expert in your model component and data, not actually in the possession of it. If models are linked together to provide solutions to decisions-makers, who is it that assumes liability if something goes wrong?

There are significant skill shortages in staff with expertise to manage linking technologies, fitting models to standards, assisting with metadata capture and developing automated metadata capture systems.

NERC SCIENCE OF THE ENVIRONMENT

The development of modelling platforms

Not all scientific questions requiring modelling need researchers to use an integrated modelling approach. We must be led by the research requirements not the technology. Currently in the UK we do not have an readily accessible infrastructure that is suitable for linking and running models in a "plug and play" style that is accessible and user friendly. The development of modelling platforms should therefore be concurrent with research being undertaken and the technical solutions to the research problems being solved should as much as possible be generic. Platforms have the ability to open up our models to a wider community, but need to be future proofed as much as possible in terms of models and data structures.

What do we mean by a platform and what does it do?

A basic framework is essentially an environment in which the user can explore data using a selection of data manipulation tools, models and visualisation methods.

A platform provides a structure which facilitates the integration of model components and environmental datasets in a way that can deliver the outputs required by users; including an objective assessment of the uncertainties associated with the output.

A platform, therefore, consists of an infrastructure consisting of hardware and software with a defined set of standards for interaction between code and datasets.

What do we need a platform to be able to do?

The platform needs to be flexible so that the software and hardware can be changed easily to answer different questions using the same platform. The scale of the questions that you are trying to answer will ultimately drive what hardware you need i.e. a laptop to solve a dredging problem or HPC to solve UK flooding problem.

No single platform would be able to answer all questions posed; a small number of platforms would provide a much smarter and cost effective approach.

The platform environment does not need to be too complex. We already have tools that if combined could become a platform.

It is essential within a linked modelling environment that we develop generic standards that are suitable for a number of communities that describe the models, data inputs and outputs, and the computing environment. Metadata must be created in order to provide information to non-specialist users of the platforms. This would have to be tailored to different communities (households/policy makers). Areas that would need to be addressed are:

- Different layers of metadata
- Metadata is essential if impacts are unexpected, automating the construction of a model-chain
- Platform needs to generate its own metadata
- Version data/library is important
- Liability and traceability are important (litigation drives things in the US)
- We need to be more rigorous with linked components

Scaling is also important, as different properties may emerge at different scales. For example, gridded data is available at different resolutions, but it may not be meaningful to use data generated at one resolution to drive models at a finer scale. These problems may be unexpected where multiple systems are interacting therefore:

- Platforms should have the tools to explore impacts
- Platforms could be used as a means of bridging scales
- Scale bridging needs to be done in an intelligent way
- Built-in checks are important to tell you when something is not scaled properly

Assessing and quantifying uncertainty

The ability to share data and models in order to base rational decisions and polices on the outcomes requires a rigorous assessment of uncertainty. It is necessary to develop mechanisms, standards and tools to enable uncertainty to be managed in a simple but robust way. An assessment of current uncertainty tools, both from within the environmental science sector and from other sectors such as the medical and engineering communities is required.

When assessing and quantifying uncertainty, it is clear that it isn't the uncertainty that is reduced but the user's confidence in the modelled outputs. In general, communication of uncertainty is very much user context dependent and therefore, it is the end-user that will determine what they need and how the information should be displayed and/or communicated. We have to be clear about our assumptions and the use of metrics of uncertainty when communicating our results to end-users.

Within a linked modelled system as well as understanding the parametric and structural uncertainty, there is a need to understand the component uncertainty across interfaces in order to understand the uncertainty within a linked modelled system. We therefore, need to develop tools that will track parametric uncertainty through each component of the linked modelling process. There is a need to understand the sensitivity of model parameters to the problem being posed, how it can vary depending on the question being asked and to understand how this affects the overall uncertainly of the linked model results.

Work that needs to be done includes:

1. Develop tools and standards that will allow for the management of uncertainty within linked modelled systems
2. Develop robust and repeatable methods to validate models and understand how observations bias results
3. Develop tools that can understand and track how uncertainty is propagated through linked components and how parametric uncertainty may change depending on the question being asked or the components that are being integrated
4. Be clear as to the limitations of models and look towards codifying the known limitations of models to ensure that their use is fit for purpose

Exemplar and Infrastructure projects

Introduction

Discussions during the workshop and immediately afterwards identified four "exemplar" projects: megacities and their resources; southern ocean and krill populations; coastal resilience with reference to erosion; and volcanic ash cloud and their effects on aviation. To enable these projects and others to be undertaken, it was recognised that an overarching infrastructure needs to be developed to facilitate this. Therefore, a separate infrastructure project has been scoped. A summary of the infrastructure scoping study and the four exemplars is provided below,

Infrastructure

Name	**Infrastructure**
Location	National/ international
Purpose	To put in place key items of scientific infrastructure that will enable NERC researchers and their partners to exploit the world leading position they have achieved in integrated environmental modelling (IEM).
Organisations/individuals involved	BGS, CEH, NOC, PML, NCAS, BAS
Objective	NERC, along with its European partners, has played a leading role in transforming IEM into a useful and useable technology. The challenge now is to put that technology into the hands of NERC scientists in a form which they can all use. If we can do so then we will greatly increase their ability to address the challenges of our time and raise the value of our contribution to society.
Approach adopted	1. Assemble/build from existing material on-line best practice

	guides for: a. Developing linkable models and integrated modelling b. Integrated modelling 2. Establish or adopt a basic metadata standard and related search tools for models and associated modelling components that will enable modellers to: a. Catalogue linkable modelling components (and any other models or modelling components) b. Discover and decide the appropriateness of linkable modelling components 3. Adopt, adapt or develop a basic set of controlled vocabularies for naming and defining the variables exchanged between models for use in relation to 2a 4. Make information available on and encourage the use of the more widely used interface standards. E.g.: ESMF, OpenMI, OMS, etc. 5. Develop 'adapters' to enable modelling components using different interface standards to communicate. 6. Make available or secure access to one or more modelling platforms and their related tools for: a. Making models linkable b. Linking models into a composition (i.e. creating an integrated model) c. Running the linked model d. Logging linked model runs 7. Acquire a basic set of reformatting and visualization tools for analysing the output of linked model runs.
Notes	The platforms should cover the widely used modelling environments, e.g. NET, java, Linux, etc.

Megacities

Name	**Megacities: Critical catchment capital coastal environmental observatory (Megacities C^4EO)**
Location	London the Thames catchment and the Thames Estuary
Purpose	A large fraction of the Earth's population lives in megacities therefore to protect these cities and their inhabitants there is a need to increase the resilience of "coastal megacity catchments" to future environmental and societal change
Organisations/individuals involved	BGS, CEH, NOC, PML, NCAS.
Question(s) being addressed	How will environmental change affect the availability of resources to London's population? What physical, biological and chemical impact will a change in the magnitude and frequency of weather extreme events, and how will these things impact public health? How much will climate change cost in terms of protecting vital

	infrastructure, housing and ecosystems?
Approach adopted	The approach will comprise four closely related and essentially parallel components, each of which will help inform the design and next-steps of other components. These are: 1. **Spatial environmental observations**: Set up and further develop a networked, physical environmentalobservatory (similar in style to the US Critical Zone Obs). Environmental observations would include: sediment turbidity sediment flux, auto sampling of sediment for sediment quality, sediment coring for temporal record of sediment quality and flooding frequency, metals and nutrients within sediments and adsorbed onto riverine sediments, CO_2 evasion from river waters, soil erosion, DOC, DO, pH, T, etc. 2. **Temporal environmental observations**: Develop a record of antecedent conditions (e.g. flood frequency, channel development) to serve as a baseline to which current behaviour can be compared and from which we develop the baseline recurrence behaviour. 3. **Develop an IEMP in the context of a C^4EO** in order to enable software tools to be integrative and communicative. The IEMP is the software partner to the physical environmental observatory. 4. **Run scenario and sensitivity testing** using high-resolution environmental response models as constrained by observational data from this observatory. Modelling of this sort will also help design the setup of the environmental observatory.
Models to be linked	Need holistic understanding of the water and sediment routing environment in the first instance, so require linkage of climate, surface processes, river, estuarine, marine and groundwater models.
Stakeholders involved	Many and varied: Port of London Authority, Environment Agency, Defra, Association of Business Insurers, GLA, City of London Cabinet Office, NHP, Thames Water et al, Crown Estate Energy companies, DoT and Coastal communities > SMPs

Southern Ocean

Name	**Integrated Modelling of the Southern Ocean ecosystem**
Location	Southern Ocean
Purpose	To assess the response of the Southern Ocean marine ecosystem to environmental change by developing an integrated modelling framework.
Organisations/individuals involved	BAS, NOC, PML, NCAS, Met Office
Question(s) being addressed	1. How can we sustainably manage the Southern Ocean krill fishery in the context of increased demand and climate change? The underlying models required to answer Q1 will also be able to

	contribute to the following questions: 2. What is the size of the Southern Ocean carbon sink and how is it changing? 3. What is the role of the Southern Ocean in the global ocean circulation?
Approach adopted	Antarctic krill is a key species in the Southern Ocean marine ecosystem; it is central to the Antarctic marine food web, is important in biogeochemical cycling in the ocean and is the target of a commercial fishery that is likely to be increasingly exploited as pressure for food grows. Antarctic krill is found throughout the Southern Ocean and is associated with a range of habitats including shelf environments, the open ocean and the sea ice zone. **Approach:** Understand what drives the present day distribution of Antarctic krill on regional and circumpolar scales using time-varying physical and biogeochemical data and model output linked to a biological krill model. Use modelled scenarios of future change as drivers of 1. to examine the impacts of environmental change on the ecosystem. Outputs from 1. and 2. will form a major input to the development of food web and fishery-based models for the Southern Ocean. Provision of the input and output data at coherent temporal and spatial resolutions, as dictated by each particular model, is key to successfully integrating the data sets and models, and to validation and assessment of the sensitivity of the constituent components. Therefore, a central aim of this project is to develop the integration system using standard, generic, shared IEM infrastructure, and will involve the coordination of activities across the UK environmental modelling community.
Models to be linked	Atmosphere-ocean-sea ice, biogeochemical, life cycle, food web, fishery and climate models and datasets at varying temporal and spatial resolutions.
Stakeholders involved	Foreign and Commonwealth Office, South Georgia Government, Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR), fisheries.
Notes	This approach can be applied to other marine ecosystems by changing the key species at the centre of this structure.

Coastal Resilience

Name	**Coastal Resilience**
Location	Coastal region of the UK
Organisations/individuals involved	NCAS, BAS, NOC, PML, BGS (Mike Ellis)

Question being addressed	How can the resilience of coastal regions to erosion and flooding be improved? How can we best assess the sensitivity of UK's coasts to changes in future forcing (e.g. wave climates, sea-level rise, storm surges) at the necessary spatial and temporal scales that are useful to policy and decision makers?
Approach adopted	We aim to tackle this in two related stages. The first is to quantitatively assess the sensitivity of coasts to possible changes in external forcing (wave climate, sea-level rise, storms) and possible management scenarios over decadal to centennial time-scales. Such modelling requires an ensemble approach and is most efficiently accomplished by so-called reduced complexity (exploratory) models. This approach allows the relatively rapid identification of coastal reaches that are highly vulnerable to likely future changes. The second stage targets these highly vulnerable sites with more detailed models to provide focused solutions. All models will require training from existing datasets.
Models to be linked	Climate, ocean circulation, wave generation, coastal morphology change models (there are a few), CFD models.
Stakeholders involved	Defra, Environment Agency, Natural England, Crown Estate

AshCloud

Name	**AshCloud**
Location	International
Organisations/individuals involved	NCAS, BGS
Question being addressed	How can the effects of volcanic ash on aircraft in the atmosphere be predicted and avoided?
Approach adopted	By developing an understanding and prediction of volcanic ash generation and combining this with models of particulate movement in the atmosphere, then a better prediction of the path of volcanic ash clouds can be made. This can be used in combination with aero engine manufactures and plane operators to ensure safer air travel during ash events.
Models to be linked	Ash generation, atmospheric models
Stakeholders involved	Cabinet Office, Rolls Royce, Airline representatives

Acknowledgements

The participcants of the workshop (see Appendix 2) are gratefully acknowledged for their contribution to the workshop and this document. Andrew Barkwith, BGS, provided review comments. The NERC Centre/Surveys are also thanked for providing funding with BGS providing the facilities, accommodation and refreshments.

NERC SCIENCE OF THE ENVIRONMENT

Appendices

NERC SCIENCE OF THE ENVIRONMENT

Appendix 1 – Agenda

Day 1, Tuesday 4 February

10:00 Registration and Coffee

10:30 - 12:00 Introductory talks to set the scene and outline the vision

1. Welcome to BGS (Katherine Royse)

2. Previous activity within NERC in relation to IEM: Stephen Mobbs

3. Context for NERC: Robert Gurney

12.00-13.00 Lunch

13.00-13.15 Briefing session setting out the purpose of breakout groups

13:15-14.45 Breakout sessions on Topics 1 and 2. Each break-out group to cover both.

A. Breakout one 1:15 - 2:00pm
B. Breakout two 2:00 - 2:45pm

14:45-15.00 Tea

15.00-16.00 Presentations of Exemplar Projects (10 minutes each):-

a. Hydrology/Catchment exemplar (Andrew Hughes/Richard Harding)
b. Coast Zone exemplar (Jason Holt)
c. Ice2Sea exemplar (Rob Arthern)
d. Krill exemplar (Tony Phillips/Sally Thorpe)
e. Volcanic Hazards (Sue Loughlin/Stephen Mobbs)

16.00-17.30 Breakout sessions on Topics 3 and 4

C. Breakout three 16.00-16.45
D. Breakout four 16.45-17.30

17.45 BGS Reception, coach to Jury's Inn Hotel, Waterfront Plaza, Station Street, Nottingham NG2 3BJ

19.00 for 19.30 Memsaab Restaurant, Maid Marian Way, Nottingham: dinner

Day 2, Wednesday 5 February

08.30	Jury's Inn Hotel, Nottingham: coach to BGS Keyworth
09.00-09.15	Recap of Day 1
09:15-10:00	IEM in NERC and possible funding models: Ned Garnett
10:15-10:45	Plenary Session on Breakout Groups
10.45-11.15	Coffee
11:15-12:30	More Breakout sessions on Topic 4: showcase projects

At the end of these sessions groups should have well-formed ideas about how showcase projects could be taken forward, including how they will be brought to the next workshop with NERC's stakeholders.

12.30-13.30	Lunch
13:30-15.00	General Discussion: What happens next?
15.00	Close

Appendix 2 – Attendees

Surname	First Name	Organisation
Amoudry	Laurent	NOC
Archibald	Alex	NCAS
Arthern	Rob	BAS
Barkwith	Andy	BGS
Bell	Vicky	CEH
Black	Emily	NCAS
Bricheno	Lucy	NOC
Collier	Chris	NCAS
Cosby	Jack	CEH
Crummy	Julia	BGS
Dijkstra	Tom	BGS
Ellis	Mike	BGS
Fry	Matt	CEH
Garnet	Ned	NERC
Gurney	Robert	NERC
Harding	Richard	CEH
Harpham	Quillon	HR Wallingford
Holland	Paul	BAS
Holt	Jason	NOC
Hughes	Andrew	BGS
Hughes	Ruth	NERC
Jackson	Chris	BGS
King	John	BAS
Kingdon	Andy	BGS
Lawrence	Bryan	NCAS
Ludden	John	BGS
Mobbs	Stephen	NCAS
Moore	Roger	OpenMI Association
Phillips	Tony	BAS
Rees	John	RCUK
Reis	Stefan	CEH
Royse	Kate	BGS
Stone	Sally	BGS
Swift	Jackie	BGS
Thorpe	Sally	BAS
Torres	Ricardo	PML
Vaughan	David	BAS
Vosper	Hayley	NERC
Ward	Rob	BGS
Wolf	Judith	NOC
Yang	Xin	BAS

Appendix 3 – Summary of breakout groups

Q1: In the future most models will at some point in their life-cycle need to be linked to other models. What needs to be achieved in order for this to happen?

Group 1

- The idea that platforms are heading towards a system was introduced.

- Qu. - What is your definition of a platform?

    - Basis/framework starting point
    - An environment in which one can explore data/using models and other tools
    - Common systems designed to link model components
    - Lots of models, wires for linking up visualising results, platform upon which to play with models
    - A modelling platform is a structure that allows model components to hang together, communicate, the sum of which is a recognisable tool to tackle a Dig problem
    - A system that facilitates the integration of model components and environmental datasets into a framework that can deliver the outputs required by users including an objective assessment of the uncertainties associated with the output
    - Hardware and software infrastructure with defined set of standards for interaction between code and datasets. Could also be populated with models and data required
    - <u>Infrastructure</u> - servers web access (cloud) accessibility
    - <u>Software</u> – Operating systems security, interoperability, connectivity, data
    - <u>Definitions</u> – (ontology & semantics) workflow
    - <u>User interface</u> – Inputs, outputs adaptability
    - A (computer-based) infrastructure for users to engage with models, data and tools for the processing and visualisation of results…and/or…a toolbox

How do we get to the next stage in order to achieve the vision?

- Platform needs to solve a problem
    - Without a problem is difficult to construct
    - Do we need several platforms?
    - Are Met Office models platforms?
    - Not necessarily one all signing/dancing platform – principle is not affected by the scale of the problem. Current infrastructure is not there for linking/running models in a plug-play style
- Not only need a framework but also a set of tools/visualisation

- Platforms open up modelling to the wider community (social sci/economist/non-specialist)
- There is a list of simple questions that will require a lot of work to answer
- We are not the only people looking at this – Medical
- Generic platform with different views
- Provision of metadata must exist to provide info to non-specialist users. This would have to be tailored to different communities (Households/policy makers).
    - Different layers of metadata
    - Metadata is essential if impacts are unexpected
    - Can then construct a model-chain
    - Platform needs to generate its own metadata
    - Version data/library is important
    - Liability and traceability are important (litigation drives things in the US)
    - We need to be more rigorous with linked components
- Artificial Intelligence for metadata
- Quality and Uncertainty must form a part of the platform
- Scaling is important as different properties may emerge at different scales. These may be unexpected where multiple systems are interacting
    - Platform should have the tools to explore impacts
    - Platform as a means of bridging scales
    - Scale bridging needs to be done in an intelligent way
    - Built in checks are important to tell you when something is not scaled properly
- NERC or EPSRC needs to stat a research program into integrated modelling
- Just because two components are the same – should they be linked?
    - Probably not
- We need to have a conceptual framework behind the platform (a set of rules and a checklist)
- CSDMS has a tool where you pick a component and that guides you through the process
- OpenMI does not impose any constraints on the modelling you do, however real world links need to be checked by the user
- Requirement for a research program where a working group should establish some of the biggest/best questions to be answered
- Is a platform liked by/useful for scientists?
    - Scientist benefits from platform by developing the conceptual model
    - Scientists can free up time to spend on other research
- We are at the start of a large learning curve – IEM is difficult
    - Platforms help bridge the gap and bring down the required skill levels
    - E.g. Google earth is easy – the early GIS on which they are based are difficult
- If barriers are in place for scientists to make their models platform compliant, they won't
- Computing infrastructure has matured to a point where we can make better progress
    - We do however need to make things slick so that it is used
- Culture barrier in academia to develop something original – are platforms original?
    - Find it had to publish platform output
    - Find it difficult to get funding – seen as risky
    - Not really platform specific
    - How do you QA model output from platform

    - Role for research centres
    - NERC Training up users
- Why are you building a platform/community?
    - Need to feel like you are part of something
    - Scientists want their output used
- We have to do the boring things to get things working

Group 2

- Qu - What is your priority in getting a platform together?
- Dog wagging tail, i.e. questions drive system not a system developed for its own sake
    - Needs to be a science problem
    - Forums for linking scientists – a forum with model scientists
- Need to define users and what they need – should this be left to NERC?
    - Links to needing a strong science-policy interface
    - Should they be specially trained?
    - Policy makers don't have the time to pose the best questions
    - Scientists don't know the pressure on policy makers to answer questions
    - Feeds in to defining your question
- Need to set an output that is attainable and need a question that brings people together
- Assumption we are making is that the questions need IEM to solve
- If the question is not defined – then difficult to answer
- Flexibility is key – use of software/hardware can be changed easily to answer different questions using the same platform.
- Must supply as much information as possible to the user
    - Allows the user to trust the model
    - Security is important
    - Needs to be efficient – changing who is in control
    - Learning from other users
- How do you motivate scientists to document it properly – strengths and weaknesses
    - Have to change the reward system for scientists to do this?
    - Version control software (got to trust system)
    - Education is required
    - Can cause damage if this is not done
- IPR – are all models going to be open source
    - Could just make the output available and not necessary the input/model data
- Access to the code makes you feel better
- Commercial side may not allow you publish/release data/code
- A universal way of providing feedback for a model
- A way of versioning so that mistakes in the code don't make their way back into the code-proper. Maybe have a trunk version?
- Academic –Commercial interface is a difficult area in which to operate.
    - If one part of the model chain is commercial does that impose IPR issues to the rest of the platform output
    - Encryption on links – e.g. the insurance industry

- The industry is moving towards open source – the data that drives the models will however remain hidden
- Problems with data access/licensing (strategic UK datasets especially)
    - But there are some areas where access is good (e.g. WRF)
    - Variability in data access is a problem
- Can we scope what resources will be needed
    - High end – Airbus has whole teams
    - Low end – couple of people
    - This QA is undertaken by making code open source
- There is a spectrum of users/developers that code with good-will

Group 3

The previous groups' findings was introduced [listed above] and asked the question – what do you think?

- Scale of question can drive what hardware you need (e.g. laptop to solve dredging problem and HPC to solve UK flooding problem)
- Who are the modellers and who are the scientists – mostly researchers are both
- Danger of underestimating the answer to the question
    - You can end up with a large modelling platform to answer a small problem which wastes resources
    - A small number of platforms would be a better approach
- What we can do now with technology has not been clearly described – policy makers don't know what is possible
- Strengths of this group – Atmosphere/Earth surface modellers in one group
- Two drivers
    - Scientific Problems that can't be solved (need big spending on resources and less on people)
    - End user problems (needs people and less on resources)
    - Both important but separate and require different approaches
    - Need to be clearer about what the driver is guiding
- We now have a set of tools and if we combine them we may have a platform
- If the NERC definition of platform (Ship/Plane/etc.) be used here
    - Large piece of infrastructure that only works on the bigger scale
- People resources is the bottleneck that is holding thing up
- We could end up with some new platforms if we brought stuff we already have together
- Should we working more closely with other organisations or use the competition to drive innovation
- What turns a workflow into a platform?
- Who is the user?
    - Science – need a good question
    - Corporate – need a good partner
- Governance

- Platforms not Platform is important
- Modular structure is important – how do we get missing model components?
- Is IEM value for money – in terms of people or resources
- Needs to be future proofed in terms of modelling and data structures

NERC SCIENCE OF THE ENVIRONMENT

Q2: Encourage the development of modelling platforms. What needs to be done to enable this to happen?

There are number of challenges to achieve this goal:

- A significant amount of work is required!
- Consideration should be given as to whether all models need to be made linkable
- For example, some should be used "stand alone"

But – modular approach is a good approach

- Start with new framework
- Conceptual & predictive –may need different frameworks

The next steps are as follows:

- Categorise models
- Focus on questions to solve
- Perhaps choose a generic framework/questions (flooding)
- Standards need to be developed and/or existing ones used for the following:
- described models, i.e. metadata
- data in/out (flat files as well as runtime linking)
- computing environment
- Motivation for adopting standards – this needs to be considered
- Become market driver – can also hold back.
    - Open standard! (community contribute) – these need to be clearly documented
    - CF standard – example
    - Categorise standards
    - Level of integration

- Standards at each level – taxonomy of standards – processes in 1 model can provide the input parameter of another
- Models to exchange processes – time consuming
- Metadata standards are including assumptions
- Think about how & what models are exchanging (map/flow diagram)
- Financial – bar code idea (useful?)
- Work on different areas & harmonise
- How to find common denominate
- Realistic standards that number of communities can meet
- Functional difference – how much – very different problems
- Standardisation vers variety
- Common group? – time stepping
- Break off 'big' categories

- Mesh/grid independent
- Spatially resolved – geographically rep in space
- e.g. climate system - but what else?
- Decision support – env/economic/social
- Where feedback – more complex
- Understand where existing models failing
- Demonstrate better model
    - e.g. Somerset levels
- Linking more systematic & robust
- Repeatability
- Validation
- Interoperability
- Extensibility
- <u>Model Metadata</u>
- Allows taking out tight coupling – for example
- Model description (assumptions, emulators etc.)
- Is model right for right reasons?
- Effort to describe & pop metadata – time
- Some standards available e.g. ISO 19115
- Drivers for metadata
- Legal
- Funding
- Citation
- Community models
- High level s/w for metadata capture
- <u>Skills</u> software eng. required
- Traceability of incorporated models
- Most models simple in essence
- History
- Can data pick up metadata as it goes through chain?
- S/w should maintain audit trail
- Useful to have

<u>Summary:</u>

- Resource
- Incentivise
- Taxonomy of Standards
- Metadata
- Automatic capture
- Skills

<u>Group 1, comments on previous discussion:</u>

- Overlap with frameworks
- Platform/interface distinction

    ↓ Data exchange (model irrelevant)

    ↘ What needed in interface

- Parameters/plumbing – horses for courses
- Parameters – in metadata
- Clear ontology
    - What in/out
    - What operation performed
- Coding expert knowledge in to get meaningful result
- Incentivisation
    - Funders to require compatibility
    - Establish critical mass – NERC/EPSRC e.g.
    - Encourage linkages between platforms
- Description of model
- Look at other systems – e.g. human body
- Similar discussion in medical model – link
- Need RCUK level input
- Skills
    - S/W expertise to assist metadata
    - How to support legacy models of E.O.
    - E.O. has international input
    - Need strategic investment

    → Mapping – example – link GIS to social data

    → Evidence of usefulness

    ↓ Google – build up

    → How to navigate transition to strong take up

    → 'Sell' benefits

    → Huge potential

- Metadata
    - What do you need for standards to communicate
    - IPR issues
    - Open data to solve?

- Free financial models
- But input data may have different IPR for linked models
- Conflict with commercial exploitation
- Issue of comm. Use data/models
- Political problem – address
- Description – taxonomy of standards
    - Simple linkage
    - Complex feedback
    - Span platforms & levels – do things simply
- Make standards work at particular level – not big ahead of taxonomy
- Parameter types (A.B.C…) – what dataset/models can I link to
- ESMF – e.g.
- What is standard for
- Investment in current standard – period of stability

<u>Summary from Group 1 & 2 on Question 1:</u>

- Need for minimum standard – model metadata – (plus understanding model)
- 'Taxonomy' of standards – need to adopt/meet that which best 'fits'
- Link model standards
- <u>Focus on interface</u>
- Multiple agencies using same 'framework'
- Incentivisation (RCUK)
- Critical mass
- IPR
- Skills
- Gain momentum

<u>Standards:</u>

- Bench – marking – model performance (e.g. water & energy balance)
- Capturing scientific reality/accuracy of final model – getting it right
- Danger of 'plug & play' – users can make inappropriate links
- Avoid 'bolting together' – interactive process
- Include users experience of using version 'x'
- Do you always need to fully couple – effort worthwhile
- IPR – value still in being expert on your model
- Analysis of pros/cons of open access
- 'Idiots' guide to standards
- Horizon 2020 – funding for standards
- Exemplars – could examine benefits & time for linking – advantage in dynamic linking?

Q3: Assessing and quantifying uncertainty. What needs to be done to enable this to happen?

<u>Assessing & Quantifying Uncertainty in IEM:</u>

Ensembles & Models (ISIMIP)

Climate & Impact Models
Range of uncertainties for predictions
One component of Uncertainty

- Structural also parameters
- Users need different measures
- Can be changed parameters or input data ranges
- Hydromodels – flood forecasting/drought
- Scenario uncertainty (boundary conditions)
- Need to track uncertainty
- Comparison with observation (real-time or historic)
- Must occur at each model
- Especially at interfaces
- Extreme events

Different Models will be responsive to mean or extremes: Smoothing determines lack of

- Coupled models scales
- Baysian
- Model assumptions
- Uncertainty measures
- Metrics relevant to end-users,
- 2 types of structural uncertainty
  i. Spatial configuration
  ii. Time-series variability
- Kinds of structural uncertainty
- States that particular models can reach depending on input parameters
- Changes to equations

- <u>Presenting Uncertainty</u> Should we present it, without devaluing values

Coupling of many models makes understanding "modelling chain" a largely redundant concept

- Will ensembles help
- Incorporating model use risks massive uncertainty

- Remediated by collaboration
- Open source models facilitates checking, but also miss-use
- Uncertainty as a flag for maturity
- Metadata must include boundary conditions & fundamental limitations of models
- Codifications of other peoples' model limitations to ensure models used is fit-for-purpose
- Are their problems that can be solved better if more models included?
- People will assess their own uncertainty

- Uncertainty understanding required by model developers rather than end-users
- Uncertainty used to assess usefulness of models

Data assimilation for models not communicating to end-users

- Statistical assessment of models required
- Model outputs need to be tested
- Can we have scientific (e.g. geophysical) interface to allow us to test components?
- Lack of "geophysical zippers" as scale determines how this can be done

Baysian Uncertainty requires understanding of priority weighting/rating the "arbitrariness"

<u>Coupling uncertainty</u> in how you couple models as will otherwise get too complex

- Coupling different time??
- Bias corrections

Observation System simulation experiments

- How do observations bias results?
- Observation biases underplayed
- Difficulties measuring extremes
- Time stops can multiply uncertainty

<u>Principals</u>

1. User context dependent: End-users will determine what they need
2. Effect of uncertainty calculations on interactions
3. Effect of interactions on uncertainty
4. Exponential increase in work
5. Assumptions & metrics of uncertainty
6. Not uncertainty that's improved, but confidence
7. To what extent do people write extra applications to manage uncertainty around core equations?
8. Tracking parameter uncertainties with parameters through each element of modelling process
9. Sensitivity – Is uncertainty of parameter important to problem being tested (e.g. relevant scales)
10. Limits where model results cannot be exceeded
11. Need component uncertainty to understand linked system

12. Scenario uncertainty/intrinsic chaos/how well does model represent processes?
13. Emulators to assess uncertainty

Issues to address:

1. Integral assumptions can be documented in metadata
2. document uncertainties in results separate – dynamic
3. Understanding of statistics in deterministic models – mismatch between statistics with modeller
4. Explore workflow engines – identify best fit-for-purpose
5. Knowledge of tools from other sectors
6. Do uncertainties compound? Feedback
7. Communication to end-users
8. Some uncertainties cannot be meaningfully calculated

Confidence – has uncertainty been minimised in sense that best estimate been made rather than variability of output parameters.

- End-users want to know that underpinning data is included, not ignored, of that range is within same limits
- Clear vocabulary of uncertainties
- Process of assessing uncertainty more important than value
- Known unknowns
- Statistical distribution rather than statements
- Can interfaces cope with additional uncertainty information
- Computational environment creating uncertainty through calculation changes or precision
- Metrics of how well modelled
- Observational evidence –to understand system & deal with extremes
- Mismatch between observers & modellers
- Need to look at uncertainties between components across interfaces, within interfaces & whole system
- Training to understand uncertainties as important as training in the IT
- Make clear that quantifying models does not increase its value
- Need to understand users perception of risk