The Art of Modelling

Solving problems in research and applications

Notes to accompany lecture course on "The Art of Modelling" by Professor Ian Townend.

Contents

1. Introduction

This white paper considers the art of problem solving. This might entail deriving a solution to a set of mathematical equations (analytical or numerical), testing some hypothesis, or the development of a model and its subsequent application to a real world problem. Whatever the problem, the steps are very similar and the issues that need to be addressed have much in common. We therefore begin by considering the overall process and then examine some of the more important components in some detail.

Even at the most basic level of hypothesis testing, some form of ‘model’ is involved. This might be a logical expression or statement, some form of empirical relationship (often derived from data), or a set of mathematical equations that have to be solved to obtain a solution. We therefore explore the art of modelling, examining issues of conceptualisation, abstraction, assumptions, constraints, conditions, calibration and validation. Where there are many lines of evidence or a number of different models being used, there is a need to interpret the results through a process of synthesis, taking due account of uncertainties in all the various methods. This highlights the issue of what constitutes robust evidence and how we compare information of different quality.

Finally, the white paper concludes with a brief consideration of the need to communicate the research or study findings. This can have a significant influence on the uptake of the research, or acceptability of a study (such as an impact assessment), and often reflects the amount of effort that has been put into distilling the findings through the processes of interpretation and synthesis. In the author’s experience, the process of distillation can take as long as, or longer, than all of the earlier steps to obtain a solution to the problem (steps 1-3 in Figure 3).

As already noted, this white paper aims to provide a general summary of the art of problem solving, which is as applicable to those working on research in an academic environment, as those working on bespoke applications or projects in a commercial environment. Where there is a need to make a distinction in the text, research programmes and studies are referred to as ‘research’, whereas applications on commercial projects are referred to as ‘projects’ for brevity.

2. Solving a Problem

A classical introduction to solving mathematical problems is provided in the delightful book "How to Solve It" by George Pólya (1985). This outlines a set of strategies to do this based on four steps, as summarised in Table 1.

Table 1 – Steps to solve a mathematical problem.

1. Understanding the problem	What is unknown? What are the data? What is the condition?
2. Devising a plan	Has it been solved before? Is there a related problem?
3. Carrying out the plan	Check each step. Is each step correct? Can you prove it is correct?
4. Looking back	Can you check the result? Can you derive the result differently?

This has clear parallels with the widely adopted 'hypothetico-deductive' approach to scientific endeavour. The main difference being that a logical proof in mathematics has some sense of finality about it (as long as there are no flaws in the solution) whereas science, and the application of science to projects, has no such finality. As noted by Karl Popper (1959), "Science is not a system of certain, or well-established, statements: nor is it a system which steadily advances towards a state of finality. Our science is not knowledge: it can never claim to have attained truth, or even a substitute for it, such as probability."

A hypothesis is an imaginative preconception of what might be true in the form of a declaration with verifiable deductive consequences (Medawar, 1967). It is an anticipation of a principle that is to be accepted or rejected. This is essentially a problem solving process, often starting with an existing but inadequate theory (Deutsch, 1998). Such inadequacy may arise because there are a number of competing or conflicting theories, or as a result of new observations that cannot be explained by existing theories. The driver is always to try and improve the explanatory power of our theories.

Given some particular problem we must try and devise a better explanation in the form of a theory (this can be a structured explanation of events or just an informed guess at why something happens). We do this by formulating the theory (explanation) as a set of one or more postulates, propositions or hypotheses in a form that can be falsified and tested. The testing then takes various forms of *criticism*. Some testing can be done by simply examining the hypothesis for its logical and internal consistency. This ensures that:

(i) the hypothesis is not tautological, which would mean that it was by definition true and therefore not testable; and

(ii) that any conclusions drawn from the hypothesis do not conflict and are therefore logically consistent amongst themselves.

A comparison with other competing theories should first consider whether the new hypothesis adds to the explanation of the existing theory. If it does not, then the new theory is likely to be of limited merit. If it does, then there may be scope to establish areas of conflict between the competing theories/hypotheses and embark on a programme of testing that will corroborate one and falsify the other. Even in a situation where there are no alternative propositions, it is necessary to formulate tests that will potentially falsify the hypothesis. Repeated verification has only limited additional value, as explained further below. These tests need to be repeatable by others and this is usually the function of publication and peer review, where the researcher makes their hypothesis and the method of testing available to others. If after a period others have reproduced the tests, or have undertaken their own criticism, perhaps using quite different tests then the theory progressively becomes accepted as part of the existing knowledge base; although always susceptible to further testing and falsification. The process is thus cyclic as set out in Figure 1.

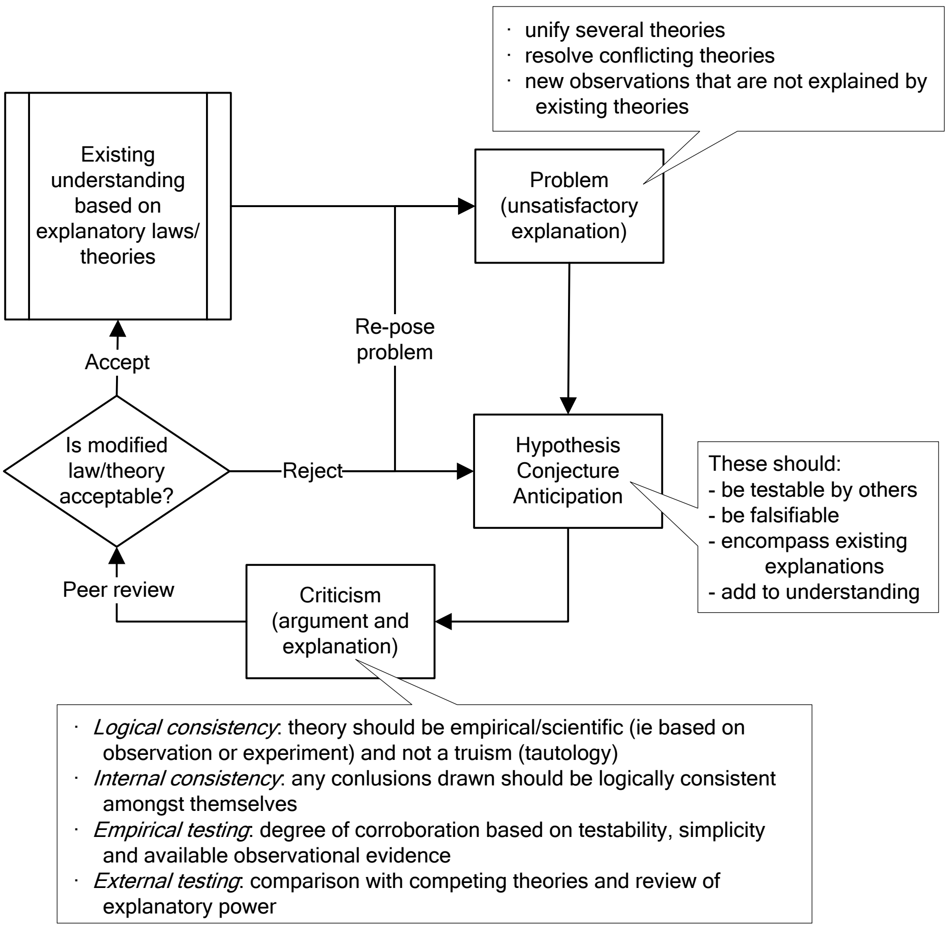


Figure 1 – Process of hypothesis led deduction

The scientific and philosophical literature contains a wealth of debate about the nature of the scientific method. Much of this revolves around the relative merits of 'induction' (inferring some general principle or law from specific instances) and 'deduction' (conclusion or inference drawn by reasoning from the general to the specific, as outlined above). However, despite all the debate, the scientific method has proved difficult to define in a formal sense. Indeed Medawar (1967) concludes that it is only possible to discuss the 'gist' of the method because there is no abstract formal framework. He attributes this to the rapid iterative process that often occurs between the acts of mind involved in discovery ('having an idea', which is not a logical process) and in developing a proof (which is a logical process). This melding of discovery and proof is sometimes purported to entail inductive reasoning, an approach which dominated much of the early thinking on the subject. In this, one proceeds from observations to some form of generalisation. Further observations may then serve to substantiate the initial generalisation, Figure 2. The problem is that repeated observations can never justify or "prove" the generalisation (or theory). This is because each observation serves only to verify a singular statement about the observed process. However, unless we have all possible observations, we have no way of knowing that there is not one or more observations that will conflict with the "theory". Hence the "theory" cannot be verified without an infinite number of observations.

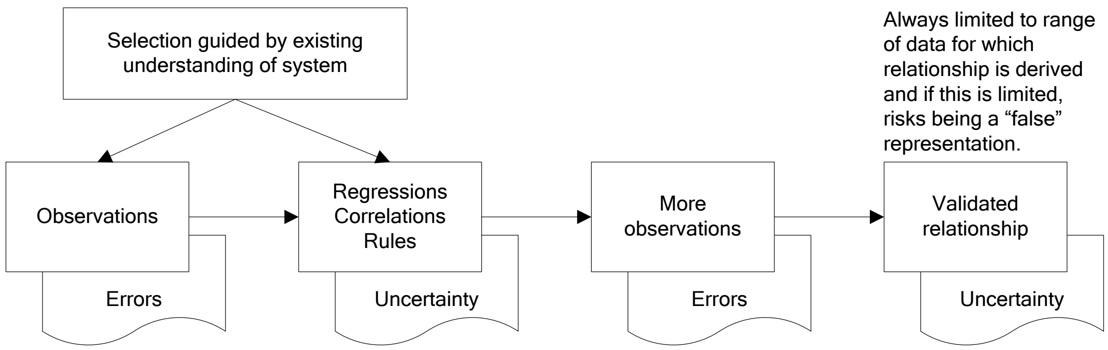


Figure 2 – Process of induction (sometime called a "black-box" model)

Whilst accepting that certain or "true" scientific theories are not attainable, whatever method is adopted, the hypothesis-led-deductive system is preferred because it provides an explanatory framework. Furthermore, in seeking to continuously test and refute our theories, the hypothesis-led-deductive system subjects theories to variation and selection, establishing a form of natural selection for scientific theories (Popper, 1959; Deutsch, 1998).

In seeking to make use of science to solve real world problems, we are invariably faced with limited understanding. Indeed, in some cases, all we may have is a set of limited observations and the present body of knowledge. We are also usually seeking to apply what is known to an immediate problem and are not able to wait for some new scientific theory to be conceived, formulated and tested (in some cases assessments have to be made using the information 'to-hand' and even further data collection may be heavily constrained). Extensive use is therefore often made of collection (when possible), collation and classification of data and subsequent search for patterns and regularities. To some this is a process of induction (Coolican, 1996), whereas to others the process of pattern recognition, or the identification of a rule set, may provide the basis for the formulation of an *explanatory* theory. This reflects the fact that "no general statement …. can arise merely from the conjunction of raw data. The mind always makes some imaginative contribution" (Medawar, 1967). Medawar and others therefore cast this initial process of hypothesis formulation as one of telling a story: "Scientists are building explanatory structures, telling stories, which are scrupulously tested to see if they are stories about real life" (Medawar, 1967).

The overall approach which is now accepted across scientific disciplines (eg. Coolican, 1996) can be summarised as follows:

1. Observation gathering and ordering of data
2. Pattern detection, regularities and generalisation (sometimes called induction)
3. Development of explanatory theories
4. Deduction of hypotheses to test theories
5. Testing of the hypotheses
6. Support or adjustment of theory

Steps one and two are essentially just one way of identifying and defining a problem (ie how do we explain what we have observed). The remaining steps then follow the hypothesis-led-deductive approach as outlined above. A simple example is given in Box 1.

A similar procedure is equally applicable to a wide range of problem solving activities. For example, these steps have been adapted to clarify the process of identifying change in estuaries (anthropogenic or natural) and are equally applicable to research studies and commercial projects (Townend, 2004, Figure 4). Similarly in the field of policy analysis, a somewhat expanded list, with "ten commandments" has been proposed (Granger-Morgan *et al*, 1990). Although this has ten steps many of these simply unpack the steps outlined above in a little more detail, with particular emphasis given to the more open-ended nature of many policy questions.

Comparing these closely related frameworks for solving a problem, it is clear that they all follow a similar process, as summarised in Table 2. This comparison also points to an equivalence between a scientific hypothesis and a conceptual model describing how we think a system behaves, both *tell a story* which can be tested and refined.

Box 1 - A simple example of hypothesis-led-deduction

The process of a hypothesis-led-deduction could be considered in simplified form as follows:

Theory: The mid-tide cross-sectional area (A) at the mouth of tidal inlet depends on the total discharge (ie the tidal prism, P) of the inlet according to the relationship $A=10^{-4}P$; a simplified version of the O'Brien equation (Obrien, 1931). (Note: this is a very weak theory because it has only limited explanatory power).

Hypothesis: The ratio of A/P on any particular estuary should be 10^{-4}.

Test: Measure the cross-sectional area and tidal prism of the River Orwell

- If the ratio is 10^{-4}, the theory is supported (measurement errors could provide a basis for accepting a value that is close).
- If the ratio is not close to 10^{-4}, then the theory, as posed, appears wrong.

The River Orwell estuary has a value of $1.15\text{x}10^{-4}$, which supports the hypothesis as posed. Clearly we can go on to measure the area-prism ratio in other estuaries and this may provide further support for the theory but it does not *prove* the theory or make it *more true*. What is really damaging is when we find a result that does not support the theory (for instance the River Yealm has a value of 10^{-3} and the River Tees has a value of 10^{-5}). One way forward is to sub-divide the population of estuaries in some way, so that the sub-groups continue to support a modified version of the theory. For instance, by noting the extent to which inlets have responded to sediment infilling over the Holocene, a set of three relationships was derived for geological (no infilling), partially infilled and sedimentary systems (Townend, 2005). Note how additional information is being introduced and how it seeks to explain the sub-division proposed. This limits the exceptions and increases the explanatory power and so improves the theory.

An alternative approach, is to move away from simple regressions and derive the relationship from existing physical principles (Kraus, 1998). This (a) links the “theory” into the existing framework of understanding and (b) greatly improves the explanatory power. The work of Kraus suggested that the regression equation of O’Brien could be derived from consideration of the equilibrium conditions. Subsequent “criticism” of this work established that the derivation of an equation similar to O’Brien’s was fortuitous, as it depends on the choice of a particular approximation[1]; the form of the bottom friction coefficient (Townend, 2005).

In a purest sense it might be argued that the process should have started with the derivation from existing physical principles as undertaken by Kraus. However this overlooks the fact that this analysis was only undertaken because the relationship had been observed (O’Brien, 1931) and Kraus sought to explain it. This therefore follows the mixed process outlined, that uses pattern detection (induction) as *one* means of identifying and defining a problem and then proceeds to solve the problem by posing an explanatory theory, deducing a hypothesis and testing it.

[1] An assumption is in effect an auxiliary hypothesis. It should increase the ability to falsify the theory and hence make it more testable, which is true in this case because the resultant equation is highly dependent on the form assumed for the bottom friction coefficient.

Table 2 – Comparison of frameworks for problem solving

Mathematics	Research	Projects
Define problem	Explanatory theory	Issues to be examined
Devise plan	Hypothesis to be tested	Conceptual model
Implement plan	Criticism (testing)	Work programme
Look back	Review (update theory)	Synthesis
Solution	Accepted theory	Conclusions/Solution

As there are lots of texts explaining how to solve mathematical problems (eg. Polya, 1985), or how to formulate and test a hypothesis(eg. Coolican, 1996; Mayo, 1996), we focus here on the more general problem solving skills needed to understand the behaviour of complex dynamic systems, in order to identify and predict the system response to defined changes (e.g. some form of modification to the system itself, or a perturbation in the dynamic forcing conditions). The steps in developing the evidence needed to draw some conclusions, or propose an effective solution, are those outlined in the third column of Table 2. Hence where the development of a conceptual model is discussed, it would be equally valid to consider the framing of a hypothesis to be tested. Similarly, the process of synthesis is essentially one of criticism through a process of reasoned argument, explanation and interpretation of the evidence, as used when evaluating a hypothesis (Figure 1).

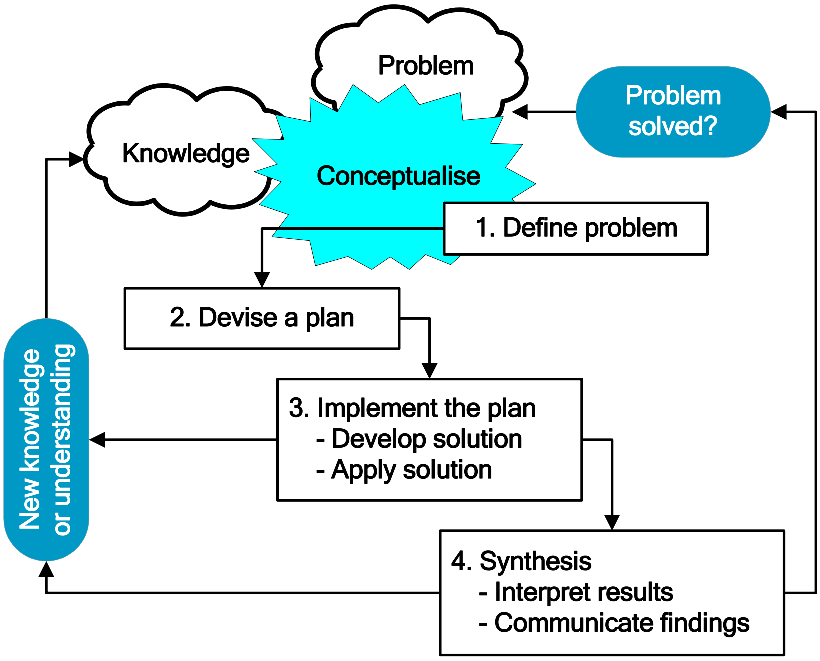


Figure 3 – Steps to solve a problem

The steps to be followed are set out in Figure 3. The first step in understanding the problem is the need to relate what is already known to the particular problem and conceptualize a way forward. With some form of conceptual understanding of the problem, it is possible to devise a plan (step 2). The third step of carrying out the plans is here sub-divided into two parts. The first is to develop the solution. Where this entails, say, developing a numerical model, this refers to the design, code writing and unit testing of the software. The second part involves the application of the solution to some real-world problem. The fourth step requires the distilling and synthesising of the results (often from many different sources, or lines of evidence) which is invariably needed when studying complex multi-faceted problems. Here again two sub-steps are identified, the first being the interpretation of the results and the second the appropriate communication of the findings.

2.1. Define the Problem

The first step is always to develop a sound understanding of the problem. This necessarily builds on existing knowledge (e.g. from past experience, literature, review etc) but also the context of the problem. In research this might be the relationship of this problem to a broader research agenda, or the objectives of a managed, or strategic, research programme. For projects this will include the client requirements, legal constraints, and may extend to taking account of other interests (e.g. stakeholders).

The output of this step should be an agreed definition of the problem. For research this may take the form of one or more hypotheses of be tested, whereas for projects it is more likely to be some agreed aims and objectives, or even a set of performance targets for the product that is being delivered (such as development study, design, construction, or manufactured device).

2.2. Devise a Plan

The initial conceptualisation will have set-out some basic understanding of the problem but to determine how to solve it we need to establish the “art of the possible” and this may often be constrained by what is economically viable (or most cost-effective) even in a research environment. Developing a plan will need to draw on information about:

- what data already exists;
- previous research and projects; and the
- measurements, methods of analysis and modelling techniques that could be deployed.

In research this information will be used to first define one or more hypothesis that set out the principles to be tested. The nature of the testing will then determine much of the remaining plan of work, to check for consistency and explore the explanatory power of the arguments posed.

Problems concerning the natural environment invariably entail complex systems, which may be only partially understood. It is also often the case that no one method of analysis, or model, will describe the "problem" (especially for projects such as environmental impact assessments) and a number of techniques will need to be used and the results combined to produce the final outputs (Townend *et al*, 2007).

For these reasons a 'conceptual model' that describes the system and how it is believed to behave can provide a valuable framework to guide both the development of a work plan and the subsequent synthesis of the results. The basis of the conceptual model and the process of synthesis are explained in more detail in Section 3.3.

> The outputs from this step will define the work needed to produce a solution to the defined problem. This will typically involve identifying the theory to be exploited, and the requirements for data collection, field or laboratory experiments, analysis and modelling. It will also determine the most logical order for the work to be undertaken. Other requirements, especially for projects, include iterating the programme so that the time and budget are as efficient as possible and consistent with the client's needs and that the appropriate skills and resources needed are identified. A good plan will set out the phasing of the work and identify clear milestones or targets.

2.3. Implement the Plan

The balance between developing and applying the solution may differ substantially between research and projects. Nevertheless, both will generally entail some elements of these two steps.

2.3.1. Develop Solution

When research demands a new method of analysis, or model, then this often requires the writing and testing an algorithm, usually by means of some software. Similarly, when an existing model is to be deployed on a project, this will require the real world conditions to be idealised, so that they can be represented in the model set-up, which will then need calibrating and validating. In either case there may be additional needs for field or laboratory work and supporting data analysis, in order for the most appropriate 'abstraction' of the system to be selected and represented (whether this is using a new or existing model). The issue of abstraction goes hand-in-hand with the development of a conceptual model of the system and is discussed further below.

The output will be a means of solving the problem that has been adequately tested to confirm that the solution gives sensible results. This may be the result of structured unit tests (for the software development), testing against known analytical solutions or idealised problems, or calibration and validation using measured data.

2.3.2. Apply Solution

Whilst some demonstration of the sufficiency of the new method or model is always needed, the replication of known test cases may be an end in itself as a piece of research. However, most of the time, the method or model is a means to an end and allows real world processes or behaviour to be examined in more detail, or in some cases identified for the first time (e.g. identification of flow reversals on the flood tide in Southampton Water, which was subsequently confirmed by field measurements (Collins and Ansell, 2000). For projects, the main focus is usually the bespoke application of data analysis techniques and numerical or physical models, to develop some specific understanding of real world behaviour. When applying models it is good practice to carry out some sensitivity tests. These might check whether the temporal or spatial resolution is sufficient to give stable and consistent results, or the sensitivity to particular parameters applied within the model domain, or at the boundaries. Increasingly, models are being configured in a probabilistic framework, enabling Monte Carlo simulation to be carried out to investigate individual and multiple parameter sensitivity. This improves understanding of the range of applicability of the model and the uncertainty in the outputs. The model(s) can then be applied to specific scenarios, or "what if" test cases, with due regards to inherent uncertainties. The nature of the scenarios to be considered will reflect the hypothesis being tested, or the aims and objectives of the project.

The main output will be the results for the desired application. This should however be supported by as much information as possible relating to assumptions and simplifications made, particular sensitivities and where practicable, some quantification of the uncertainties in the results.

2.4. Synthesis

Simply presenting the output of a data analysis, or series of model runs, is generally not very informative. The results need to be put in context and presented in a way that explains the methods used, the issues that influence the results and whether the results deliver the desired solution (be this testing some hypothesis, or assessing the impact of a proposed development). The synthesis step therefore comprises interpretation of the results and then working out how best to communicate the findings. As already noted, this step can be particularly demanding.

2.4.1. Interpret the results

Interpreting the results usually combines:

- a process of checking that each step is correct and that the final results make sense; with
- the use of the results to extend understanding by answering specific questions.

If the results can be derived in a different (perhaps simpler) way, or compared to the results from previous studies, this can provide a valuable cross-check. Where different lines of evidence agree or reinforce each other, so the uncertainty is reduced. Conversely, where information conflicts, this can help identify discrepancies, errors, or areas where further data or research may be needed to reconcile the differences identified. The use of a 'conceptual model' was introduced in the Planning Stage (2.2) and such a model can also provide a useful framework for both testing and refining understanding of the system behaviour in the light of the results (see Section 3.3).

> The output will be answers to the problem posed, founded on a robust assessment of the results against other sources of information, worked-up into clear and defendable lines of evidence.

2.4.2. Communicate the Findings

A common mistake is to simply present a mass of data plots or model output as the results. In some cases, this may be necessary supporting material but, in almost all cases, the raw output has little explanatory power. A thorough interpretation of the results usually helps to distil out the most salient contributions to the solution. If these are put together in a logical order, they can be used to "tell a story". This should aim to build a clear picture of the evidence that supports the conclusions, whilst reflecting on the inherent uncertainties and discussing any discrepancies that have become apparent as a result of the interpretation analysis.

> The output is a clear and concise summary of the findings that addresses the problem posed and allows the reader to follow the lines of argument and evidence used to arrive at the stated conclusions.

3. Modelling

3.1. The Art of Modelling

The art of modelling is to develop new insight, or understanding, that is not available at the outset

This often entails an advance in science or a bespoke application of existing models and will always seek to resolve some uncertainty, even though many uncertainties may still remain.

However, modelling does not necessarily entail running some sophisticated software program. Modelling can be thought of as any representation of what we know, such as an empirical or statistical representation of some data (e.g. a linear regression of the form $y = a \cdot x + b$, where x and y are variables and a and b are fitting parameters), or some qualitative relationship, such as a descriptive rule set defining some aspect of system behaviour. What the model represents is based on a set of choices that we make, which in turn reflect the purpose of the model. Typical uses of modelling include:

- Interpolating and interpreting data;
- Simulating dynamic behaviour of processes and systems;
- Predicting change or making forecasts;
- Formalising knowledge and testing ideas;
- As an aid to understanding and communicating the structure, functioning and behaviour of systems (both within and between interacting systems);
- Providing evidence to support decision making that is robust, by which we mean credible and well founded;
- To both reduce uncertainty, as well as identifying and, where possible, quantifying outstanding uncertainties.

In any of these applications, the choices made will largely determine what can be achieved and the process of model abstraction therefore merits further consideration. Many complex environmental systems cannot be addressed using a single model and it is necessary to make use of several models and, as such, a number of abstractions may need to be made, each addressing specific aspects of the problem. To overcome, or at least manage, this complexity, a conceptual model can help frame current understanding and aid the interpretation of new information that becomes available as a result of the studies or modelling undertaken. The conceptual model also provides a useful framework for making clear the assumptions and constraints adopted, the conditions imposed (e.g. at the boundaries) and how uncertainties are propagated through the modelling process.

3.1.1. The Appropriate Complexity Method (ACME)

Building on various approaches to model development in the fields of ecology and geoscience, Larsen et.al. (2016) have proposed the Appropriate-Complexity Method (ACME). This guide sets out one way of developing and implementing models of complex environmental systems. The method emphasises the need to understand the dominant factors responsible for the system dynamics and how they respond to change in forcing conditions or constraints. The approach comprises a number of steps which are summarised by Larsen et al (2016) as follows:

(i) Model objectives are broken down into objectives and classified within a hierarchy. This sets the levels of interest and appropriate detail;

(ii) The emergent properties of the system, that the model should be able to reproduce, are identified. This is made explicit by developing a conceptual model of the processes and variables thought to be responsible for the observed properties;

(iii) The model is then implemented and evaluated to determine whether the model adequately reproduces the system's behaviour;

(iv) These steps are then repeated to add more detail and the model is expanded to progressively add detail and/or become location-specific.

3.2. Model abstraction

The way in which a model is framed or constructed, requires the modeller to describe how the particular problem should be represented. This usually requires an 'abstraction' of the real world or, more usually, the system within the real world that we seek to represent. Most systems can be decomposed into a hierarchy of component sub-systems that are active across a range of different spatial and temporal scales (Carter and Woodroffe, 1994). For a given problem, attention will be focussed on a particular part of this hierarchy – the level of interest – at which there may be a number of inter-acting sub-systems. Some of these we may choose to represent in detail. Others, notably those at a lower level, may be represented by some simplification, by emulation of a more complex sub-system, or by a separate model that operates at different space and time scales, Figure 4. The level of interest (Level 'n' in Figure 4) may itself be a sub-system for some higher (or larger scale) level of abstraction. Information from the higher level often determines boundary or forcing conditions for the system at the level of interest.

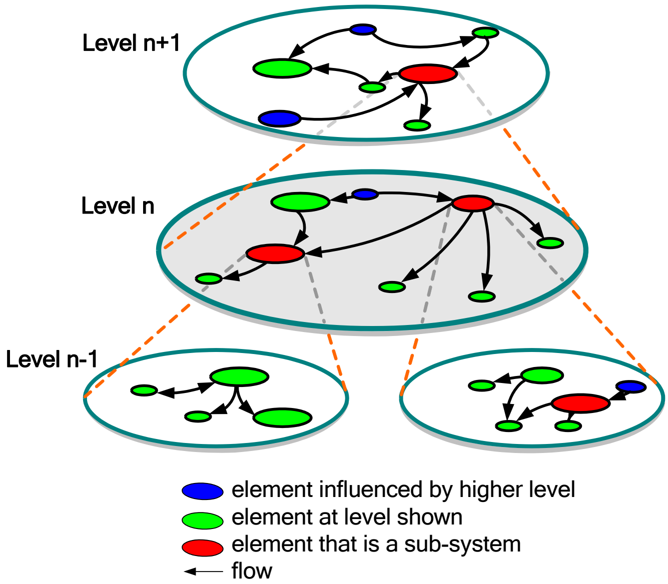


Figure 4 - System hierarchy, showing nested systems (based on Huggett, 1980)

Example of abstraction at different levels for the case of an estuary:

Higher level:	Global tidal dynamics and meteorological forcing operate at much larger spatial scales and would be typically prescribed as boundary conditions for the model.
Level of interest:	**Estuary system to predict water levels, flows and pollutant dispersion**
Lower level:	Variations in the character of the bed represented by a “simplification” in the form of a friction factor and turbulence in the flow structure represented by some suitable simplified formulation (turbulence closure).

The interconnectivity of systems across space and time scales means that the “whole” system is never represented. There will always be aspects that are external to the “modelled system” and have to be prescribed (e.g. using data to formulate boundary conditions). Similarly, there will be aspects that are internal but are simplified because their influence on the results has been determined to be small, the simplification is sufficient for the model’s purpose, or this is the only option given the current level of knowledge about the particular sub-system (e.g. the use of a phenomenological relationship to define turbulence closure in a hydrodynamic model). This system construct then guides the formulation of the model, Figure 5.

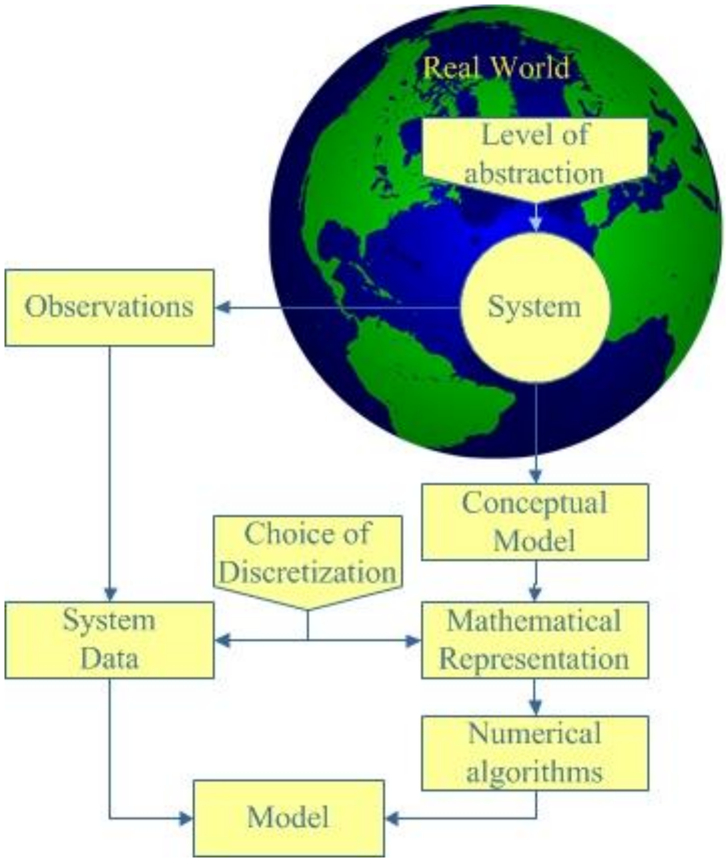


Figure 5 – system abstraction for model development

When developing a model, the abstraction is first posed as a conceptual construct, such as the representation of flow in a pipe, dispersion of a plume from an outfall, the seasonal variation of benthic communities in response to temperature, or the over-wintering mortality of bird populations. Such outline statements then need to be elaborated in detail, often through a process of iteration, to establish a well posed conceptual model of what is required. Whilst a physical model, or the use of some analogue, may be an option, in most cases the next step is develop a suitable mathematical representation. In some cases the resulting equations will be explicit (e.g. a direct relationship between two variables x and y), or an analytical solution can be derived – again allowing a direct solution to be obtained. However, in many cases of interest, a direct solution is not possible (especially when the problem is highly non-linear) and some form of numerical scheme is needed involving some form of iterative process. The most effective numerical scheme to solve a particular set of mathematical equations has been extensively studied (eg. Abbott and Basco, 1989) and is itself something of an art form. Careful examination of the equations to be solved and the schemes available for their solution can be time well spent as this can significantly reduce run-times and re-coding required in order to achieve the desired performance. This is because for some equations, even seemingly simple formulations (e.g. the equation for a one-line beach plan shape model; Roelvink and Reniers, 2012) can embody complex numerical stability issues (a subtle combination of advection and diffusion in the case given).

Based on the level of abstraction chosen, observations of the system are required to define the system data, Figure 5. These will be a combination of boundary and forcing conditions (the higher level in the system hierarchy), some data to determine aspects of model parameterisation (e.g. bed friction, or diffusion rates) and information that allows the model to be tested (e.g. time series measurements of the parameters of interest, such as flow velocity, or population biomass). Whether these data are being obtained from other sources or collected as part of the overall study programme, it is important to recognise that models and measurements are often based on quite different representations of space and time.

The way in which space and time are divided into discrete intervals (the model discretization) is determined by the need for:

- the numerical scheme to have enough information to be able to converge to a solution;
- a spatial and temporal resolution that is sufficiently refined to ensure that dynamic features of interest are reproduced and altering the resolution does not alter the solution.

The latter can be critical when trying to characterise some highly variable dynamics, such as eddy shedding in the flow around a structure. This often necessitates some sensitivity testing of the grid, to ensure that the model resolution is sufficiently refined to give stable results.

Representations in space and time are also important when trying to compare model outputs to measurements of a particular process or phenomenon. In most types of time stepping model the output is taken to be at an instant in time and representative of a grid cell (i.e. a type of spatial average). Traditionally, observations have been very different, usually measured at a discrete point and taking a time average of some rapidly fluctuating signal. With the development of remote sensing and satellite data there can be greater space and time consistency between model and measurement, although it must also be recognised that such measurements are themselves a model output based on the algorithms used to convert what the device actually measures to the parameter of interest. In all such comparisons, care is needed to ensure that the way model and measured data are represented is matched as closely as possible and that any differences are understood and taken into account when interpreting the results.

The process of abstraction for a single model, Figure 5, is equally applicable to more complex problems that may require an array of techniques and models in order to arrive at a satisfactory solution, Figure 6. This is particularly the case for projects that are trying to assess changes to the natural environment. Whilst for research one or more models may be written for the purpose at hand, it is increasingly common to use existing models where the software has already been developed and tested on a range of relevant problems. This is generally the norm for commercial projects. However a process of abstraction is still required and consists of selecting the most suitable models for the problem and, crucially, deciding how to apply them. As before this can entail choices about model discretization and model input data (boundary conditions). There is also a need to critically examine the applicability of the model. When developing a new model, determining the simplifications and assumptions to formulate the model and derive a solution are part of the process. However, this is not necessarily the case when applying an existing model to a new location, or in a novel way. It is important to consider carefully how the model is being applied and whether this is consistent with the assumptions made in the model's formulation. A number of other issues that may need to be considered when using models to address applied problems include:

- extent of the model domain;
- data requirements for model calibration and validation;
- how to combine data analysis and modelling to understand system behaviour;
- cross-checking using different formulations (both quantitative and qualitative);
- need for static or dynamic exchange of information between models that interact or are linked in some way;
- benefits of data assimilation (especially for forecast models);
- use of ensemble modelling or multiple simulations (Monte Carlo) to examine uncertainty.

Whereas the output from a single model is all that is sought, when applying many techniques there is a need to integrate the various outputs into some sort of coherent view. Mapping the individual outputs onto a conceptual model of the system and testing for inconsistencies provides a logical framework for this process of synthesis. This in itself can be an iterative process as new evidence throws up inconsistencies which themselves may need to be resolved in order to arrive at a satisfactory solution, Figure 6.

As already noted, a model abstraction represents what we know (or think we know), what we can solve and, importantly, how we simplify what we know to make the problem tractable (Smith and Stern, 2011). All of these conditions are further conditioned by the prevailing scientific and engineering ways of doing things (so called paradigms). For a more detailed discussion of this point see the first chapter of "Order out of Chaos" by Prigogine and Stengers (1985).

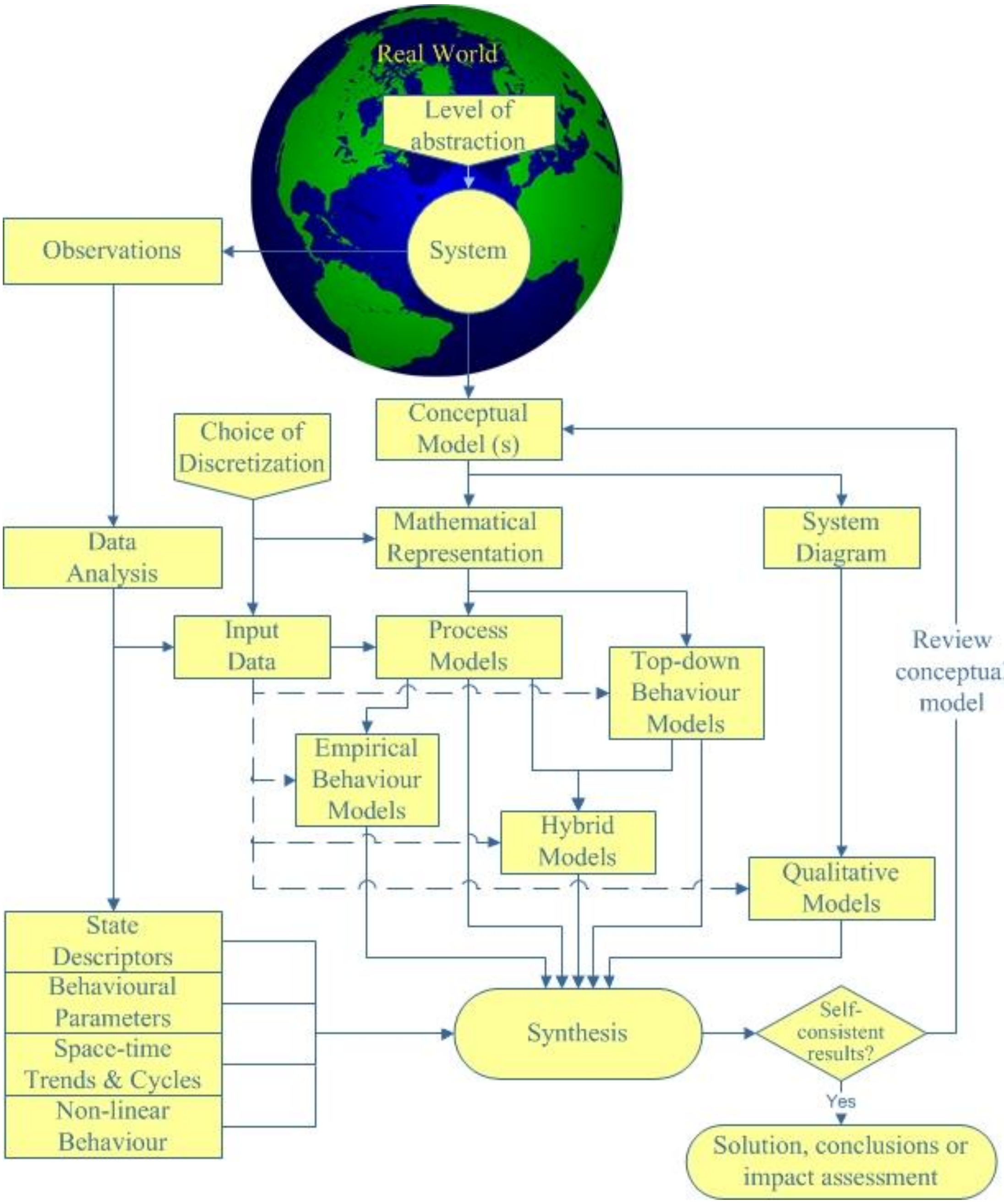


Figure 6 – System abstraction(s) required when solving problems for real world complex environmental systems

Thus even when a number, or ensemble, of different models are used to provide a range of abstractions on the real world, Figure 7, we need to be aware of events that are known about but are not captured by any of the abstractions. This is not too difficult to acknowledge because at least these are known about and can usually be ascribed to the abstractions adopted (i.e. the scope of each of the models has been deliberately limited for some reason), or are a consequence of the simplifying assumptions used in to order to develop solutions. Of greater concern are the events that are simply not known about: the unknown unknowns. These form part of the inherent scientific uncertainty that a modeller needs to be aware of and keep in mind when weighing the validity of the model outputs. One of the particular strengths of using a conceptual model to frame the interpretation of one or more models is that, as new information comes to light, it can be tested against what is known in the broadest sense (rather than the confines of a particular model). The merits of different information sources can be weighted and a consensus developed (see section on Synthesis below). This may require the conceptual model to be adapted to reflect new insights, which may in turn help to explain observations that are not adequately represented in the output from specific models.

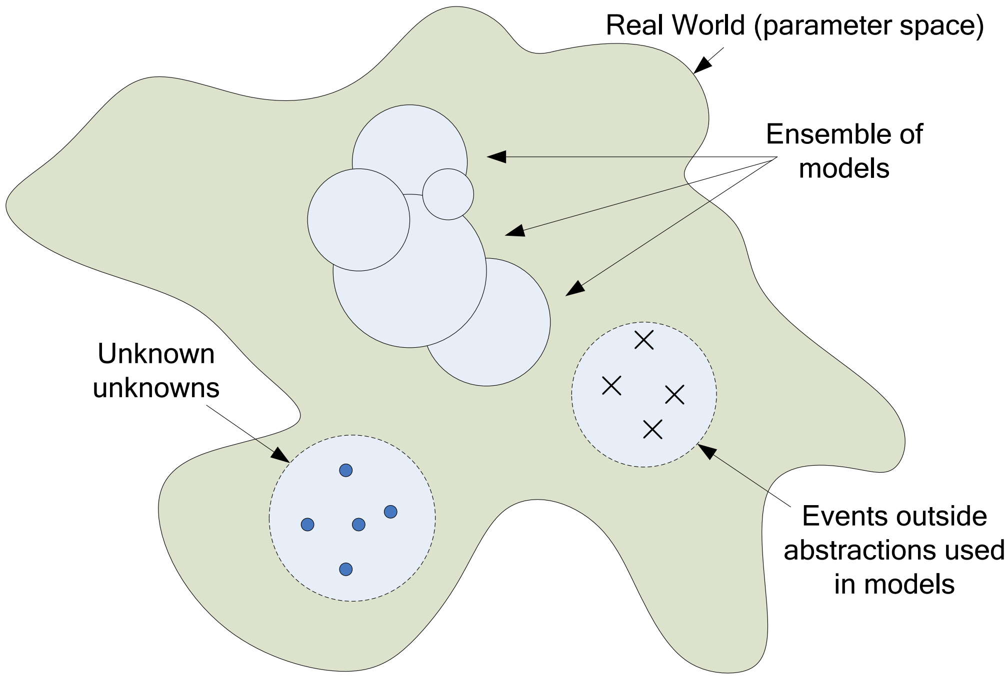


Figure 7 – Schematic to highlight the limited "completeness" and hence limited validity of models due to limits of knowledge, simplifying assumptions and approximations made in model abstractions. The bounding shape represents the complete parameter space of the real world – this is the coverage we would need to capture to represent all possible events. The circles within it are our more limited model abstractions and events that occur outside the coverage of our models, or that we do not even know about.

Hence models are only ever a way of formalising what we know (or think we know) in a form that is reproducible. To this extent they are a tool to help solve problems, not a solution in their own right. An understanding of the assumptions that have been made in deriving a particular model, the limits of applicability of the model, and the inherent uncertainties in both modelled and measured outputs are all essential for the responsible use of models.

3.3. Conceptual model

The conceptual model endeavours to present the key components of system behaviour, in a way that allows the likely responses to change to be readily evaluated. An example of a simple conceptual model is given in Box 2. Even when all the relevant elements and mechanisms have been identified, this is not an easy task for two reasons:

(i) the complexity of interactions taking place at a range of spatial and temporal scales, with the dominant influence often depending on the nature of the change and the specific characteristics of the system;

(ii) the limits to current understanding of behaviour. This gives rise to a level of uncertainty, which must be acknowledged when presenting the model.

Hence, great care is needed in the way in which the system is represented.

There is, therefore, a need to be able to characterise the interactions between elements of a system and also the dynamic response or behaviour that takes place. One way of representing the elements and how they interact is to use system diagrams (Huggett, 1980; Odum and Odum, 2000). Such an approach is reductionist in character, showing the components that comprise the system. For systems that are essentially linear this provides a clear map of cause and effect. However for dynamic non-linear systems the resultant "form and function" is often more difficult to discern. Such systems are unlikely to be in equilibrium and any steady-state will be a dynamic one, changing as the imposed constraints and energy flows vary. Given that we have variable climate and variable constraints (such as a non-homogeneous erodible geology) it is inevitable that the target steady-state, or states, will be continuously changing.

The time taken to respond to a given perturbation will also vary for individual features and this will tend to introduce lags into the system. Consequently, it is more probable that the system will be in transition, moving towards a steady-state, rather than in a steady-state condition. This juxtaposition of:

(i) the behaviour of the component parts, each seeking their own target state; and

(ii) the mix of space and time scales

often confounds any overly simplistic statements about the state of the system. There is thus a need to be very clear when discussing equilibrium (of whatever form: steady, dynamic, quasi, etc) about which parts of the system are involved and over what timescale the state is determined.

There are a number of behavioural models that summarise steady-state conditions, or explain transitional behaviour. These provide an indication of the likely mode of change. Typically, they are dissipative processes in which the system seeks to do as little work as possible. However, because of the open nature of many systems (exchanging matter and/or energy with the external environment), these dissipative processes give rise to inherent structure. With an appreciation of these phenomena, it is possible to examine how the system may evolve. Given changing forcing conditions and/or variations in the constraints imposed, the system may:

(i) alter its rate of transition;

(ii) switch to a different but similar state (form);

(iii) switch to a different state (form) altogether.

Identifying the modes of change and the likely outcome (or range of outcomes) enables a qualitative assessment to be made, even when it is not possible to make quantitative predictions.

Box 2 - A simple conceptual model: the sediment budget for an estuary

One of the simplest ways to encapsulate the behaviour of an estuary is through a sediment budget. This describes the main sediment exchanges, both in and out of the estuary and internally, Figure A1. Further refinement can add detail to the description of the various sediment transfers. It must be recognised that this says very little about the current state of the estuary or how this state may change in the future.

In many cases it can be difficult to identify, let alone quantify, all the mechanisms that give rise to sediment transfers. It may however be possible to derive approximate estimates of the amounts moving to and from sources and sinks based on measures such as transport potential and sediment demand. In effect one establishes an account and as with any account, the prime requirement is that it balances.

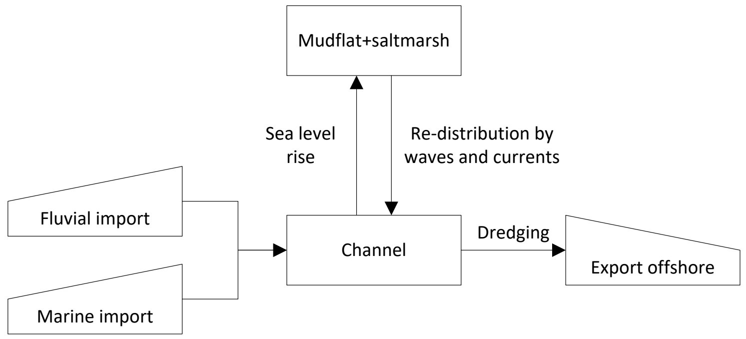


Figure A1: Simple example of sediment budget concept

An example of a net sediment budget for the Humber estuary, UK is illustrated schematically in Figure A2 (Townend and Whitehead, 2003). This does not describe the gross movements of sediment (e.g. the amount moving back and forth on every tide) but focuses on the net exchanges between key features of the system.

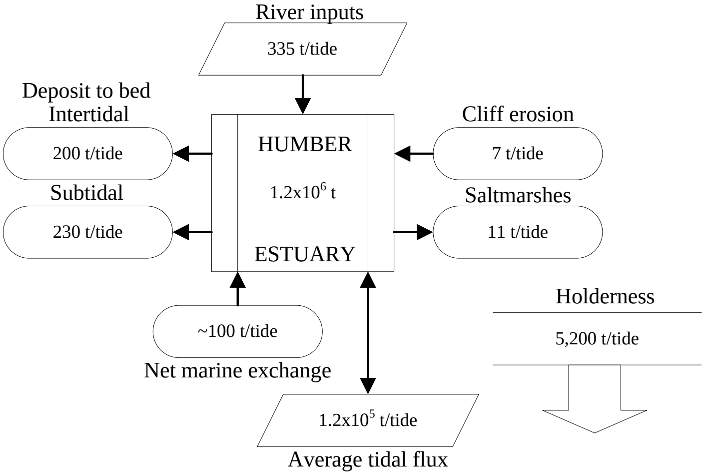


Figure A2: Schematic for net sediment budget for the Humber

There is no accepted formal method for producing a conceptual model. The purpose is to provide a means of rationalising current knowledge and understanding of how a system functions and being able to communicate this to others. The conceptual model must therefore be coherent and provide a self-consistent summary of how the system functions. At its most basic, the model may be no more than an energy or mass budget. However this says little about the system behaviour, and so, a more well developed conceptual model will identify both the mechanisms and the behavioural interactions that are likely to give rise to particular forms and, or states.

3.4. Assumptions, constraints and conditions

The derivation that forms the basis of a model or proof will invariably make use of some presumptions or conditions that restrict the basis of the model in some way. These may be logical constructs in the sense that they are preconditions for the truth of a particular statement (necessary condition) or guarantee the truth of the statement (sufficient condition). Equally they may be statements that are used as the premise for a logical argument (or model construct) which are simply taken for granted but may not necessarily be accepted, or be anything more than a simplified representation of reality.

Such ***assumptions*** and simplifications occur throughout the modelling process. Invariably we start with some form of theoretical idealisation (e.g. by adopting the Navier-Stokes equations as the basis for deriving a solution to a fluid flow problem). In order to derive a solution that is tractable, the equations are often simplified in some way. For example the fluid may be assumed to be incompressible, or the full 3-D equations may be reduced to 2 or 1 dimension for computational ease (and in many cases to establish a model that provides sufficient “insight” without undue complexity). Some additional assumptions or approximations may also be needed to develop a solution. A process that is not well understood may for example be represented by some simple empirical relationship usually derived from data of the process of interest, or a, so called, phenomenological parameterisation, which provides a much simplified description of the process, often based on some form of scaling of equations that provide a far more detailed description. A good example of a phenomenological parameterisation is the turbulence closure scheme used in many flow models.

An emerging issue is the user’s ability to keep track of the assumptions that have been adopted in creating the model (or in some cases even finding out about them). The increasing complexity of model systems, even well documented ones, is making this ever more difficult. However, it is incumbent on the modeller to find out! This may be demanding but without a proper understanding of the basis of the model, the risk and uncertainty embedded when applying the model increase dramatically because the user simply does not know whether they are respecting the limits of the model or not. For a simple system, this might be like agreeing to ride a bicycle having never ridden one before – you might be able to do it but you might fall off. For a more complex system it could be like agreeing to fly a helicopter without reading the manual, or receiving any training – you will almost certainly crash! The more complex the model, the greater the need to have a sound understanding of how it works. In the modellers case, it is essential to know how the assumptions adopted in the model limit what the model can reasonably be expected to represent.

A model is usually constrained to run within prescribed bounds. This is not to say that the prescription will necessarily be sufficient to derive a unique solution, or a solution at all (i.e. the model may crash if poorly constrained). Boundary conditions, in the form of the definition of the model domain (e.g. the bathymetry) and the forcing conditions (e.g. changes in water level or flow at open boundaries) need to be consistent with the state (conditions – see below) being modelled. It is very easy to get a sign wrong in the specification and have the flow at the boundary going the wrong way, or to specify conditions in a model with multiple boundaries that are inconsistent in some way, so producing spurious results. However, ***constraints*** can take other forms, which may or may not be represented in the model (an example is given in Box 3). It is important to identify all such constraints to ensure that the model results are interpreted in the correct context.

Box 3 – Example of a constraint that influences how model results are interpreted

In a model of sediment transport under waves and tides, it is quite common to assume a homogeneous sea bed with an infinite extent below the bed. This in itself ignores the 3-D variations in the sediments, however it also takes no account of geologically non-erodible, or hard, surfaces that may:

(i) limit the ability of the bed to erode and hence make sediment available; and
(ii) fix the morphology in one area, so inducing quite different changes elsewhere in the system.

Whilst fixity introduced by humans as part of a development is often represented in a model this more complex 3-D structure of the underlying geology is not and hence presents a challenge to any interpretation of likely morphological change.

There is also a need to give careful consideration to the state or ***conditions*** to be represented. This may be limited by the data that is available to define such things as the boundary conditions, or by what is known about a particular system. However, it can also be limited by what is feasible. For instance, computing power may mean that it is not possible to run a model for all possible combinations of the dominant forcing parameters (e.g. water levels and wave conditions). In such circumstances it may be necessary to choose a subset of conditions that can be interpreted to meet the needs of the particular problem. It is then important to consider whether the choice adopted can influence the outputs to any significant degree. For example, in a model of sand moving along a beach: is it sufficient to select a representative set of wave conditions and sum the results, or is the problem sensitive to the sequence of events?

3.5. Calibration and Validation

One of the thorny issues when setting-up and using a model is to determine whether the model is giving sensible answers. This may be in terms of the behavioural patterns reproduced by the model but also the performance of the model outputs against measurements, or known (often analytical) solutions when developing a new model.

Models derived from data using data fitting techniques, such as regression analysis, are often referred to as "black box" models. These define a relationship between the variables and, whilst there may be some underlying knowledge that led to the choice of the variables for which data were collected, the resulting empirical relationship has only weak explanatory power. It could easily be modified by new data, or some other un-measured variable may over-ride the observed relationship. Most importantly, the relationship can only be said to be valid for the range of the measured data used to derive the relationship in the first place. For this reason, extrapolation outside the measured range has to be done with extreme caution (Cunge, 2003).

More complex models, derived from the known science, can be considered to be more general and applied to any problems for which the science is thought to be valid. However, they also have a greater number of ways in which the model can be set-up. The user has to define the boundary conditions and various parameters so that the model becomes specific to a particular location. If we take a simple tidal flow model of an estuary as an example, Table 3 summarises some of the things the modeller will need to define.

Table 3 - Illustrative summary of boundary conditions and model parameters for a tidal model

The user has to define:	*Possible "adjustments" are then:*
- the shape of the sea bed (bathymetry);	this may need to be smoothed or adjusted locally so that the gridded data provides a suitable representation.
- the water levels at the open boundary need to be specified;	these may vary along the boundary but only be defined by data at a limited number of points. Some spatial adjustment may be needed to get a satisfactory representation of the forcing conditions.
- the model parameters such as the sea bed roughness or the turbulent diffusion;	it may be possible to collect data to support the values selected, or they may be set based on the literature or established practice.

For some models with many model variables (e.g. flow, salinity, suspended sediment concentration, etc) there will also be an expanded list of model parameters. One of the problems that can then arise is that there can be many different combinations of parameters that seemingly result in an acceptable representation of the conditions being studied (at least to the extent that we are usually able to verify the model outputs). Where there are multiple sets of parameters that lead to acceptable results it may simply not be possible to define an optimum model set-up (this is the problem of equifinality discussed by Bevan, 2002). Equally, in such circumstances, there may be a temptation to “tune” the model parameters to achieve an acceptable fit to measured conditions, without proper regard to reality.

The commonly accepted “good practice” is to use a sub-set of the field data to ***calibrate*** the model. In this step some of the model parameters may be adjusted to achieve a better to fit to the observations. This is followed by a subsequent ***validation*** of the model, during which some different data is used to test the model but this time keeping all parameter settings fixed. If this gives acceptable results, it is considered acceptable to apply the model to the particular problem.

During the calibration, any of the parameters used to set-up the model might be adjusted. However, in order to provide a representation that holds for the conditions to be examined, it is essential that the parameters adjusted during calibration are invariant in the proposed application (i.e. they do not change for the range of conditions, or scenarios, to be studied using the model). For example, calibrating the friction of the bed may be acceptable if the bed is unlikely to change. However, if the scenario to be examined is, say, a channel deepening, exposing a different bed formation or inducing a fluid mud layer, then the character of the bed and hence the friction will change. The initial calibration is then not very useful and may even misrepresent the situation (see Box 4).

In his excellent discussion on data and models, Jean Cunge goes further and argues that one should ideally do away with the calibration step altogether (Cunge, 2003). In this paradigm for good practice, one sets-up the model using the best information available. Parameters are defined using field data, values from similar studies in the past, or elsewhere, and expert judgement based on an understanding of what the particular parameter is supposed to represent. Where parameters are based on empirical relationships, consideration should be given to how these need to be constrained if the model conditions extend beyond what has been measured. In all of this expert judgement serves to synthesise what is known in a form that constrains the range of acceptable parameter values. The idea is that we use our best understanding of the physics, chemistry and biology to define what we believe to be a realistic representation. The model is not calibrated. Instead it is run for some measured conditions and validated using some or all of the available measured data. This provides a much more severe test of the model.

This approach to validation (with either no calibration, or highly constrained calibration) provides a test of all that is known against the observations that are available. Furthermore, an important purpose of the validation is to study the reasons why there are differences.

Understanding the reasons for any differences then provides a much more credible basis for interpreting the model results when applied to “predictive” scenarios.

Embedded in the resultant error statistics for the model performance (see section on uncertainty) is the inherent uncertainty in the application of the model to the particular problem, given what we think we currently know and understand. To some extent this overcomes Bevan’s issue of equifinality because the problem is constrained to parameter settings which we believe we can justify, rather than the much wider range of those that give rise to an acceptable fit to the calibration data. In practice there is usually a compromise. As far as possible one constrains the parameter settings to what is known and reasonable. A limited calibration may then be undertaken to check the performance and adjust parameters or boundary conditions for which there may be well understood uncertainty, allowing adjustments to be made in a way that reflects the underlying processes, as illustrated by the example in Box 4.

Box 4 – Calibrating the right parameters based on what is already known

The case of modelling an estuary illustrates this well. The information required to set-up the model is as defined in Table 3. It is possible to set-up the model using tidal elevations at the open sea boundary and calibrate the model by adjusting the friction within the model domain. However, Abbott and Cunge explained why this was inappropriate some 40 years ago. In this case, minor changes in the boundary forcing are much more significant than even quite large changes in roughness (Abbott and Cunge, 1975). Given that there is usually considerable uncertainty in the spatial variation of water levels along the open boundary, it therefore makes much more sense to set the roughness using accepted engineering values and to adjust the boundary conditions.

Where the model is to study a change within the estuary this is more acceptable because the boundary conditions will be invariant provided the boundary is sufficiently far from the changes. In contrast, the bed roughness may not be invariant, depending on the changes being considered.

Where external changes, such as sea level rise, are being examined, it may be necessary to develop a more sophisticated basis for adjusting the boundaries that can be replicated in a consistent way for each scenario. One approach commonly used is to set up a far field model to provide the boundary conditions for the more detailed model. Ultimately this leads to some form of downscaling from Global Climate Models to local models at the scale of interest.

As the foregoing discussion seeks to make clear there is a degree of uncertainty in the application of a model and even when following “good practice”, with or without the calibration step, there is a need to think carefully about what is being represented, what is known and how the model outputs might be misleading. Sound documentation of all steps is a prerequisite, followed by careful interpretation, synthesis and presentation of the results, as explained further below.

4. Synthesis

Bringing the findings of the various studies together involves the process of interpretation that leads to a synthesis of the findings.

Synthesis is the process of combining objects or ideas into a complex whole, as distinct from *analysis* where one seeks the division of a physical or abstract whole into its constituent parts to examine or determine their relationship or value – see box 5.

To some extent, this is a subjective matter and each individual will go about it in a slightly different way. However, where the conclusions have to be presented to a range of users (perhaps as part of a consultation process), or submitted for peer review, the basis for any conclusion will need to follow logically from the factual information and be as transparent as possible. A framework, which is as objective as possible, is therefore desirable.

Box 5 – Synthesis

Synthesis is not about finding a singular solution but about making sense of multi-dimensional and multi-faceted sources of data, information and knowledge. There is an element of Nietzsche's 'Will to Power' in the process, as an individual imparts their version of reality or truth in arriving at some conclusions. When there are multiple players the synthesis will be dynamic and consensus may ebb and flow. In this process, the Chinese idiom "to know what is useful, one must first know what is useless" (Zhang Zi) comes into play and can be sharpened by having methodologies in place that can identify errors, fallacies and misplaced conceptions. In this way the problem is better constrained and the potential for reaching wrong conclusions or decisions minimised. Even so, this still does not imply correct decisions in any absolute sense because it will always be conditioned by the prevailing cultural and societal paradigms of the time.

The process is aided by developing a conceptual model, which provides both a framework and a test bench to help explore the meaning and consistency of the various study outputs. Even when researching some detailed aspect of a system, this wider view is essential to give a context to the research effort. This will usually need to consider a number of spatial and temporal scales and may address the response of specific features or characteristics of the system, as well as the system as a whole.

Each of the various outputs is mapped onto the conceptual model. For some aspects there will be only limited information. For others there may be several sources. In each case it will be necessary to evaluate the uncertainty and, if necessary, consider what further information would help to reduce the level of uncertainty. Where different sources are incompatible, or conflict, the uncertainties need to be clearly identified and, where possible, resolved. It may be that the conceptual model can help to resolve the differences by indicating which source is most consistent with the overall picture. The aim is to establish a description of how the system works and how it will be affected by particular changes. For assessment type projects, the findings of the synthesis and, in particular, the conceptual model can then form the basis for assessing the future changes and the resultant impacts.

The various studies provide factual information for some combination of field data collection, laboratory testing, data analysis and modelling. Either individually or collectively these establish or refine the understanding of particular mechanisms in the system. However for the reasons already given, these individual contributions may not be sufficient to define the overall behaviour (particularly in the long-term). The task of synthesising the results therefore aims to formulate a behavioural summary of the system as a whole. The steps outlined in Table 4 helps to clarify and weigh the different lines of evidence, in a way that is both transparent and robust.

Table 4 – Process of synthesis to interpret multiples lines of evidence

→ Document factual outputs
→ Test against initial *conceptual model*
→ Modify *conceptual model* as necessary
→ Document interpreted outputs* and test against *conceptual model*
→ Consider what other information would help reduce uncertainties
→ Document *conceptual model* and resultant interpretation of change
* noting assumptions and uncertainties

An initial process of testing the conceptual model against factual data helps to eliminate any obvious misrepresentations and identify any areas of uncertainty. It is important that this is not done using interpreted results, as these may depend on assumptions that underpin the particular outputs and this may mask the real behavioural response. Most computational modelling results fall into this category. At this stage the conceptual model is founded on established behavioural concepts (as documented in the literature) and the factual information for the particular system. The next step is to introduce the additional information from the various analyses and modelling studies. These again provide a basis for evolving the conceptual model but this should now be done with much greater caution; recognising that the study results and the underlying assumptions may be the cause of any discrepancy, rather than some aspect of the conceptual model. Resolving such discrepancies is often a matter of judgement, underpinned by criticism (or hypothesis testing) through a process of reasoned argument, explanation and interpretation of the evidence (Figure 1).

Uncertainties in the conceptual model should now be much clearer. In some instances, there may be the opportunity to consider further modelling or data collection that might help reduce the level of uncertainty. Sensitivity studies, in particular, can help to determine bounds for the level of change associated with particular variables and so focus attention on the more significant areas of uncertainty.

4.1. Uncertainty

It is not uncommon for the output from a model to be taken as the "answer". Those used to working with data may recognise some uncertainty as a result of, say, measurement error. However the uncertainties are always more extensive than this and extend to what we do and do not, or can and cannot, know. These uncertainties apply to all aspects of study from data measurements, through laboratory experiments to analysis and modelling. They can be thought of as two types:

- Natural variability, which is variously referred to as aleatory (meaning to 'gamble'), random, stochastic, irreducible, inherent, external or real world uncertainty. This type of uncertainty is an important consideration in model applications of the natural environment; particularly for the temporal forcing conditions and the spatial variability in element properties (e.g. ground conditions, sediment distribution, bed roughness, etc).
- Knowledge uncertainty refers to the state of our knowledge of a system and our ability to measure and model it. It is often referred to as internal, functional, epistemic or subjective uncertainty, or as incompleteness.

To give a more complete representation of the "outputs" there is a need to say something about how these uncertainties might affect the results and hence influence the interpretation of them. For data measurements this is usually achieved by recording the instrument tolerance and, where replicates are available, estimating some measure of the scatter (e.g. mean and variance). In analysis and modelling more sophisticated statistical techniques are often used to estimate how the estimates, or "modelled" values, differ from either what is expected (i.e. some assumed pattern of variability such as a particular theoretical probability distribution), or from observations (eg. Bevington and Robinson, 2002), including:

- Fitting known distributions using techniques such as least squares or maximum likelihood;
- Testing the goodness of the fit;
- Propagation of errors to take account of all the contributing errors;
- Sampling techniques such as Monte Carlo simulation.

For example, for a simple comparison of a measured time series against model output one might consider the root mean square (RMS) error. If one is interested in the absolute error (including phasing errors) one would estimate the RMS of the difference between the observations and model output, whereas if one is only interested in seeing whether the relative magnitudes are being estimated adequately, one would use the difference between the RMS of the observations and the RMS of the model output. For variables that vary in space and time one would need to use more sophisticated methods, such as the Brier Skill Score (Sutherland *et al*, 2004).

More formally, the aim of statistical inference is to test a particular hypothesis. This requires a null hypothesis to be defined, which usually corresponds to the default 'state of nature' (e.g. assume no change from the initial state). The speculative hypothesis concerns whether an observed phenomenon can be supported. For example: the model represents the observations. By statistical convention, the speculated hypothesis is assumed to be wrong. The null hypothesis is that the phenomena occur by chance (ie that any agreement between model and observations is random).

However, because of the inherent uncertainty any statistical test may lead one to the wrong conclusion, Figure 8. The statistical test may indicate that the agreement is random when in fact it is not. This is a Type 1 error, which occurs when the wrong decision is made to reject the null hypothesis when it is in fact true. Equally the statistical test may lead one to conclude that the agreement between model and observations is not random, when it is in fact random. This results in a Type 2 error, when it is decided to accept the hypothesis when it is in fact false.

Null hypothesis is:	Accepted	Rejected
True	**Correct decision**	**Type 1 error**
False	**Type 2 error**	**Correct decision**

Figure 8 – Truth table for a statistical test of the null hypothesis. A severe test is one that is unlikely to be accepted if the hypothesis is false.

Thus, when seeking to test a particular hypothesis it is usual to consider the probability of Type 1 and Type 2 errors. A severe test (Mayo, 1996) is one that is unlikely to be accepted if the hypothesis is false (minimises the likelihood of a Type 2 error). Designing the test to minimise Type 2 errors for a given level of Type 1 error (say 5%) is the usual statistical practice and should lead to the identification of a severe test. The power function is the complement of the probability of a Type 2 error and provides an alternative measure. Other options such as the maximum likelihood ratio, serve to test the truth of the hypothesis but provide no measure of severity (because it is still necessary to assume some value of the maximum likelihood ratio as being acceptable or not).

There have also been attempts to capture the uncertainty in the way results are reported. One of the best known is the so called NUSAP method (Funtowicz and Ravetz, 1987). This is a notation that was developed to express the uncertainty attached to an estimate, such as say the numerical estimate of a parameter from a set of experiments (Appendix A). The components of the notation are the <u>N</u>umber, the <u>U</u>nits, the <u>S</u>pread, the <u>A</u>ssessment and the <u>P</u>edigree. The first two are self-evident. The spread is some measure of the variability such as the standard deviation, and the assessment is some measure of the confidence limits, or any systematic error. Finally, the Pedigree is a qualitative classification (typically assessed using a score from, say, 0-4 or High, Medium, Low) with four contributions:

- Data input – was the data used reliable and carefully scrutinised for error or bias?
- Theory or method – does the result conform to accepted theory?
- Peer acceptance – is the result accepted by specialists in the field?
- Consensus – is there agreement amongst stakeholders?

The full notation has not been widely adopted, as such, but is increasingly being used in one form or another to capture some important aspects of the uncertainty in a decision making context. When evaluating the results of an experiment, or model, one may be able to make some quantitative estimates of errors as explained above. Such estimates are of course a valuable contribution to the synthesis process, allowing due weight to be given to different estimates, or proper regard to the potential variability and hence uncertainty attached for the results. However, such quantitative estimates are not always possible. In addition, as modelling studies become more and more complex they may provide a necessary but not sufficient measure of the associated uncertainty. For this reason various forms of Pedigree scoring are emerging as a way or evaluating different strands of evidence.

For example, a study may have made use of more than one "model" to estimate some change. Where the results are different, how can one judge between them? Where possible, one might seek to undertake a 'severe' test of the two models, to see if either can be falsified. When this is not possible, the Pedigree provides a way of introducing some additional information to help evaluate the relative merits of the two models, as illustrated in Figure 9.

Pedigree		Method 1		Method 2
Data input		High		Low
Theory or method		High		Low
Peer acceptance		Medium		Medium
Consensus		Low		Medium

Figure 9 – Comparison of alternative methods using Pedigree classification

In recent applications, this form of assessment has often been reduced to a measure of confidence in the result and the degree of consensus attached to it. For example the guidance for the consistent treatment of uncertainties for the fifth assessment report of the Intergovernmental Panel on Climate Change, characterises the 'confidence' in a particular line of evidence in terms of how robust the evidence (type, amount, quality and consistency) and the degree of agreement there is regarding the evidence, as summarised in Figure 10.

The top right hand corner represents strong robust evidence and the bottom left corner is where not much is known, and there is little consensus of opinion.

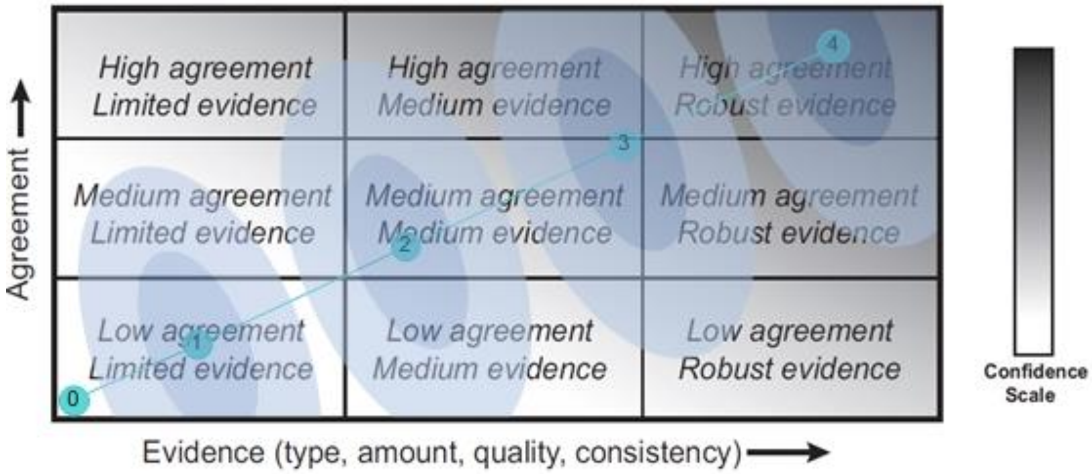


Figure 10 – a mapping of evidence and degree of agreement statements to reflect the level of confidence (grey scale), with a single mapping of confidence (0-4) superimposed and defined in Figure 11. Based on IPCC Assessment Report 5 guidance (2010).

Also shown on Figure 10 is how this can be related to an even simpler confidence scale, ranging from 0-4, where the scale is defined in Figure 11 and draws on the four contributions to the pedigree, noted above. This type of approach has the merit that it provides a consistent framework for the inter-comparison of different lines of evidence but is also sufficiently flexible to be able to accommodate widely differing types of evidence, from the highly numerate through to subjective estimates, or even guesses.

4	**Very high**	Comprehensive evidence using the best practice and published in the peer reviewed literature; accepted as a sound approach.
3	**High**	Reliable analysis and methods, with a strong theoretical basis, subject to peer review and accepted as 'fit for purpose'.
2	**Medium**	Estimates grounded in theory, using accepted methods and with some agreement.
1	**Low**	Expert view based on limited information, e.g. anecdotal evidence.
0	**Very low**	Non-expert opinion, unsubstantiated discussion with no supporting evidence.

Figure 11 – Classification of confidence based on subjective assessment of the quality of the evidence and degree of agreement in the literature and amongst experts.

4.2. Robust evidence

Having completed the interpretation of the results how can we be sure that they provide a sound basis on which to draw conclusions, or make decisions? Although a distinction is often drawn between the needs for policy decisions and business or engineering, decisions, they are all seeking to make decisions based on limited or incomplete information.

An examination of what makes evidence-based policy making robust identified five components of evidence robustness (from a policy-making perspective); credibility, transferability, reliability, objectivity and well founded (Shaxson, 2005). These measures of robustness are summarised in Table 5 and explained in more detail below.

Table 5 – Questions to test evidence robustness

Credible/valid - *sound line of argument?*
Transferable - *can the specific be generalised?*
Reliable - *can the evidence be depended upon?*
Objective - *has residual bias been acknowledged?*
Well founded - *have the right question been posed?*

- *Credibility*: This relates to the processes of analysing and synthesising information. Credible evidence relies on a strong and clear line of argument, tried and tested analytical methods, analytical rigour throughout the processes of data collection and analysis, and on clear presentation of the conclusions.
- *Transferability* (or generalisability): This refers to the way in which we make inferences and the ease with which it would be possible to take the evidence which has been collected for a specific purpose and use it in a different context, or to answer a different question. In some cases, this will refer primarily to sampling techniques, however in others it will refer to the broader framing of the issue and the policy question.
- *Reliability*: This relates to whether or not we can depend on the evidence for monitoring, evaluation or impact assessments.
- *Objectivity*: No evidence is bias-free. Bias can be introduced in the ways that information is gathered as well as in the ways that it is analysed. Questioning the bias in the evidence base deepens understanding of how it conditions our interpretation of the evidence for policy.
- *Well Founded* (or authenticity): This implies more than context, process, bias and the quality of information. Rather, it is about understanding the nuance of the evidence, exploring assumptions with an open mind, encouraging others to question the status quo, and thinking about who uses what evidence for what purpose.

These measures provide a means of assessing different lines of evidence, different proposals or, for the modeller, the results obtained from different models applied to the same problem. A similar approach to the one adopted for assessing pedigree (Section 4.1 and Figure 9) can be used to provide a comparison of how the different methods perform on each of the measures, as illustrated in Figure 12. The two approaches can also be used together and provide a powerful way of communicating the pros and cons of different.

Evidence Base		Method 1		Method 2
Credible		High		Medium
Reliable		Medium		Low
Objective		High		Medium
Well founded		High		Low
Transferable		High		Low

Figure 12 – Comparison of two alternative methods using the attributes that underpin robust evidence

There are also a number of formal methods for comparing evidence. There is, for example, an established framework for appraising the quality of qualitative evaluations (Spencer et al., 2003). It was developed for evaluations concerned with the development and implementation of social policy, programmes and practice. More generally, it is common to have a mix of qualitative and quantitative information and strategies for synthesising such information range from techniques that are largely qualitative and interpretive through to techniques that are largely quantitative and integrative (Dixon-Woods et al., 2005). Some further information on these methods is provided in Appendix B.

4.3. Communication

It is accepted good practice to keep notes, or a journal may kept to record what is being done, assumptions adopted, data used, changes to the model, and detail the cases examined as the work progresses. These notes then provide a valuable resource for the subsequent synthesis of the findings. Writing up these findings provides the next layer of documentation but is often still not the final output. Documenting the synthesis allows the different lines of evidence to be drawn together and, if the approach outlined above is adopted, there will be a need to explain the conceptual model and make clear the assumptions and uncertainties. Somewhat surprisingly, the act of writing up the study findings usually helps to clarify what is and is not known and identify where further work may be needed to improve aspects that are found to be weak, or poorly resolved.

This documentation of the synthesis phase is usually still an output for the benefit of the individual, or the team, doing the research, or project. Only when this is at least in draft is it possible to start thinking about how best to summarise the findings for the target audience; whether this is a thesis, or a report for a client. In planning the final write-up, which we will call the study report, there is a need to understand clearly the purpose of the report and the level of understanding of the target audience. Whilst some reports may simply be a factual record of the work undertaken, most study reports are to answer a research question, or address some project objectives. As such the study report needs to explain what has been done, the key findings and what this means in terms of the intended purpose. In doing this, the author will usually have to decide what to include and at what level of detail. As important, is deciding what to omit, so that the arguments presented are clear but at the same time do not obfuscate, distort, or misrepresent matters. It is common to think of this as telling a story. It will have a beginning, middle and end but more than anything it will be coherent and lead the reader to the key points that are being made without distractions, or excessive digressions.

Writing up the conceptual model can be a particular problem because it is invariably multi-dimensional in space and time and with numerous, often competing, lines of evidence. There is no established way of doing this and the approach is likely to vary from one situation to another. In most cases the conceptual model will comprise a written description, illustrated with results, schematic figures and flow diagrams. The complexity arises because one is often trying to present a number of behavioural concepts that interact over different space and time scales. Without any clear hierarchy this can be difficult to communicate. Whilst there are various types of system diagram that allow the linkages to be identified, such techniques often fall short of encapsulating the behaviour of the system.

One way of communicating this complexity to the user, is to describe the behaviour in a series of space-time intervals. Using a table, where the columns define time intervals and the rows various spatial scales or elements that make up the system, an explanation of the mechanisms at work and the resultant behaviour can be given, as illustrated in Figure 13. Individual cells in the table can then link to sections in the explanatory text (or separate reports) that provide a more complete explanation. In addition the same format can be used to summarise the existing situation and the behaviour predicted for a given set of imposed changes (e.g. sea-level rise, new development, etc). This is particularly valuable if there is a need to use the results in some form of impact assessment.

Good graphics can also be used to help explain complex study findings. This rarely if ever means raw model output. Good graphics invariably combines a number of variables – again to tell a story (Yau, 2011)! In his now classic book "The visual display of quantitative information", Edward Tufte exposes how data can be conveniently misrepresented in graphical form, in order to highlight the importance of graphical theory and good design (Tufte, 1983). Such graphical excellence often brings together a number of variables but always represents them faithfully and Tufte's book presents a number of fine examples.

		Time scale		
		Short	Medium	Long
Spatial scale	Micro	text	text	text
	Meso	text	text	text
	Macro	text	text	text

Figure 13 - Table to summarise space-time changes. Each text box would contain a short explanation for the associated space and time scale

The starting point for well-targeted graphics is to be clear about the purpose of the graphic, so that you know: the audience; how it will be used; what the graphic is trying to do; and any challenges in achieving this (Frankel and Depace, 2012). The graphic will then need to do the following (Tufte, 1983) (Tufte, 1983) (Tufte, 1983) (Tufte, 1983) (Tufte, 1983) (Tufte, 1983):

- display the data;
- lead the viewer to the intended meaning;
- avoid distorting the data;
- present numerical information in a compact and coherent manner;
- facilitate inter-comparisons of information;
- reveal any structure in the data;
- link closely with the description in the text.

Different authors have endeavoured to classify the different types of visualisation that can be deployed (eg. Yau, 2011; Frankel and Depace, 2012). Broadly speaking these reflect patterns over time, which usually characterises time varying processes; patterns in space of objects with varying proportions, which exhibit some form of structure; and finally relationships and differences, which provide a basis for comparing and contrasting information.

Once the purpose is clear a few doodles, a sketch, or story board, often help to clarify how the graphic should work. There are then various graphical elements to consider when developing the image. Tufte suggests that above all else one should show the data but then seek to maximise the ratio of data-ink to other content (by data-ink he means the amount of line work devoted to communicating the data) and erase all redundant information.

The aims of good graphics are to communicate complex ideas clearly, precisely and efficiently; and do this in as compact a form as possible. In order to ensure the integrity of the graphic it is necessary to ensure that (Tufte, 1983):

- data and graphics scale in a similar manner to avoid distortion and misrepresentation;
- labels clearly define all the key attributes again to counter distortion and also remove ambiguity (e.g. a north point on a map or a gradation scale for a colour contour plot);
- data are suitably standardised (e.g. a time series of some monetary measure may need to take account of ongoing changes, such as inflation and population growth, to give an unbiased representation of change);
- the number of information carrying graphical dimensions should not exceed the number of dimensions in the data; and
- the graphic only uses data in an appropriate context.

Preparing the data invariably involves some manipulation or rationalisation to suit the purpose of the graphic. There are then various aspects to be considered in developing the graphic itself (Frankel and Depace, 2012):

Composition - the organisation of the various elements that make up the graphic and their relationship with each other. This is particularly important on more complex figures that seek to use multiple images (e.g. several inter-related graphs, perhaps combined with photographs or 2/3-D images).

Abstraction - how to define and represent essential qualities and intended meaning of the information (e.g. using cartoons or insets to help explain specific details).

Colour - careful selection of colour schemes to support the intended meaning by drawing attention, showing relationships, or indicating scale.

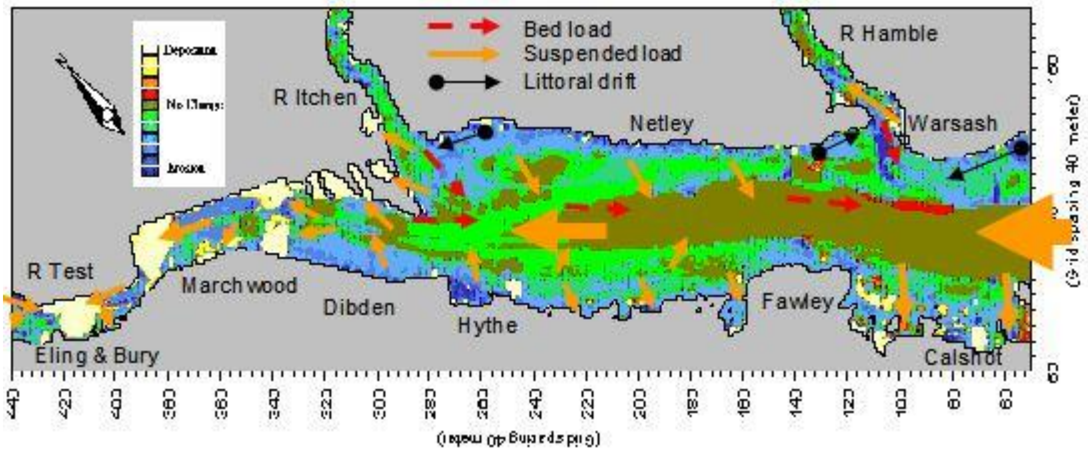
Layers - addition of layers to combine variables and create relationships in physical space.

In developing a good graphic there is almost always a need to revise, edit and simplify. Remember the text and graphics together have to tell a convincing story and convey a clear message. Two examples are given in Boxes 6 and 7.

Box 6 - A good explanatory figure usually needs more than just the model output

The following figure was developed to show the main sediment transport pathways in Southampton Water (Townend *et al*, 2007). The starting point was the model output showing the areas of erosion and accretion in the estuaries a coloured contours. The gird spacing marks have been retained to give an indication of scale. To this has been added a North point and a contour key. A base map could have been added (such as a local map or a satellite image) but this would make the figure very "busy". So that the user can easily locate themselves and to pick up references made in the text that describes the figure, a number of place names have been added.

To provide the reader with an appreciation of the sediment movement that is giving rise to the erosion and accretion, three types of transport vectors were added, with the size of each vector illustrating the relative magnitude. Again a key is added to explain the different arrows. Below is a thumbnail of the final figure, which was designed to be printed landscape on an A4 page.



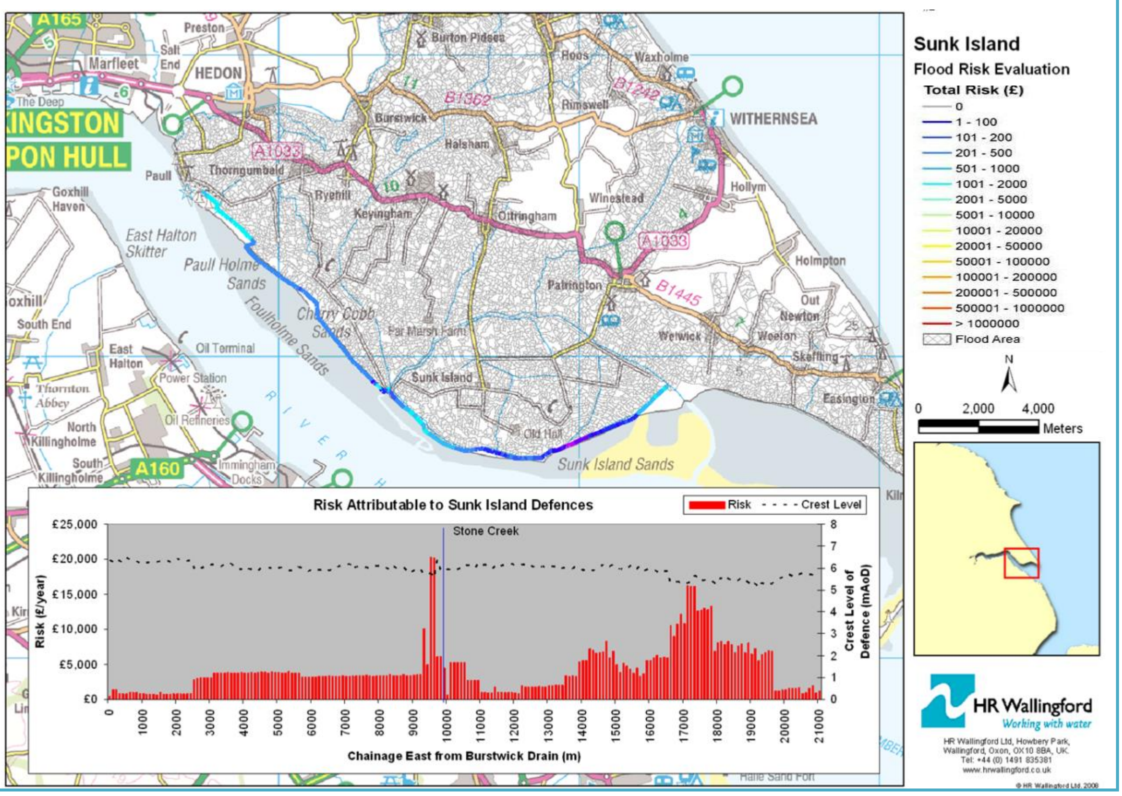
Box 7 – Combining different sources of information to highlight the main areas at risk

The figure shows the results from a detailed risk analysis and uses the output from a Monte Carlo analysis using a complex system model of flooding to estimate the risk to areas behind the flood defences and attribute the level of risk to each flood defence.

The base map is taken from a raster image of a UK Ordnance Survey map. This shows an area of land alongside the River Humber estuary. A thick line of various shades of blue has then been superimposed to illustrate some model results. The line represents the different sea defences and the level of risk that each one protects. There is a key to right. Also to the right of the main image is a north point, a scale bar and thumbnail image showing the location on the east coast of the UK of the main image.

Superimposed on the map is a graph that shows the variation in risk, as a monetary value, along the length of the defences. Included on the graph is the crest level of the defences. This highlights how locations of increased risk are associated with lengths of defence where the crest level is low. The combination of map and graph provide an easy means of interpreting the model results, which greatly simplifies the task of explaining the potential cause of variations in the level of risk that was identified for these flood defences.

To complete the figure, the title is given in the top right corner and details of the originator and copyright are included in the bottom right corner.



References

Abbott MB, Basco DR, 1989, Computational fluid dynamics: an introduction for engineers, Longman Scientific & Technical, Harlow.

Abbott MB, Cunge JA, 1975, Two-dimensional modelling of tidal deltas and estuaries, case study: River Seine Estuary Coarse Grid Model, In: Mahmood K, Yevhevich V (Eds.), Unsteady Flows in Open Channels, Water Res. Publications, pp. 795-799.

Beven K, 2002, Towards a coherent philosophy for modelling the environment, Proceedings of the Royal Society of London.Series A: Mathematical, Physical and Engineering Sciences, 458(2026), 2465-2484.

Bevington PR, Robinson DK, 2002, Data reduction and error analysis for the physical sciences, McGraw-Hill Science.

Carter RWG, Woodroffe CD (Eds), 1994, Coastal Evolution: Late Quaternary shoreline morphodynamics, Cambridge University Press, Cambridge, UK, pp. 1-32.

Coolican H, 1996, Introduction to research methods and statistics in psychology, Hodder & Stoughton, London.

Cunge JA, 2003, Of data and models, Journal of Hydroinformatics, 5(2), 75-98. see also reply by Jim Hall in 6,1,75-81,2004.

Deutsch D, 1998, The fabric of reality, Penguin Books, London.

Ellis EC, Li RG, Yang LZ, Cheng X, 2000, Long-term change in village-scale ecosystems in China using landscape and statistical methods, Ecological Applications, 10(4), 1057-1073.

Frankel FC, Depace AH, 2012, Visual strategies: A practical guide to graphics for scientists and engineers, Yale University Press.

Funtowicz S, Ravetz J, 1987, The arithmetic of scientific uncertainty, Phys.Bull., 38, 412-414.

Granger-Morgan M, Henrion M, 1990, Uncertainty, Cambridge University Press, Cambridge, UK.

Huggett R, 1980, Systems analysis in geography, Clarendon Press, Oxford.

Kraus NC, 1998, Inlet cross-sectional area calculated by process-based model, In: 26th International Coastal Engineering Conference, American Society of Civil Engineers, 3, pp. 3265-3278.

Larsen L G, Eppinga M B, Passalacqua P, Getz W M, Rose K A and Liang M, 2016, Appropriate complexity landscape modeling. Earth-Science Reviews, 160, 111-130, doi: 10.1016/j.earscirev.2016.06.016.

Mayo DG, 1996, Error and the growth of experimental knowledge, The University of Chicago Press, London.

Medawar PB, 1967, The art of the soluble, Methuen, London.

O'Brien MP, 1931, Estuary tidal prism related to entrance areas, Civil Engineering, 1(8), 738-739.

Odum HT, Odum EC, 2000, Modelling for all scales: An introduction to system simulation, Academic Press, London.

Popper KR, 1959, The logic of scientific discovery, Routledge, London.

Pólya G, 1985, How to solve it, Penguin Books, London.

Prigogine I, Stengers I, 1985, Order out of chaos, Flaimingo, HarperCollins Publishers, London.

Roelvink D, Reniers A, 2012, A guide to modelling coastal morphology, World Scientific Publishing, Singapore.

Shaxson L, 2005, Is your evidence robust enough? Questions for policy makers and practitioners, Evidence & Policy: A Journal of Research, Debate and Practice, 1(1), 101-112.

Smith LA, Stern N, 2011, Uncertainty in science and its role in climate policy, Phil Trans R Soc Lond A, 369, 1-24.

Sutherland J, Peet AH, Soulsby R, 2004, Evaluating the performance of morphodynamic modelling, Coastal Engineering, 51(8-9), 917-939.

Townend IH, Whitehead PA, 2003, A preliminary net sediment budget for the Humber Estuary, The Science of the Total Environment, 314-316, 755-767.

Townend IH, 2004, Identifying change in estuaries, Journal of Coastal Conservation, 10, 5-12.

Townend IH, 2005, An examination of empirical stability relationships for UK estuaries, Journal of Coastal Research, 21(5), 1042-1053.

Townend IH et al, 2007, The Estuary Guide, www.estuary-guide.net ed. ABPmer and HR Wallingford, Southampton, UK.

Tufte ER, 1983, The visual display of quantitative information, Graphics Press, Cheshire, Connecticut.

van der Sluijs JP, Craye M, Funtowicz S, Klorogge P, Ravetz J, Risbey J, 2005, Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: The NUSAP system, Risk Analysis, 25(2), 481-492.

Yau N, 2011, Visualise this: The flowing date guide to design, visualisation and statistics, Wiley Publishing, Indianapolis, Indiana.

Appendix A - Pedigree score

A form of 'pedigree scoring' has developed based on Numerical Spread Assessment of Uncertainty Pedigree, NUSAP (Ellis *et al.*, 2000; Van Der Sluijs *et al.*, 2005). The purpose of the scoring is to record and 'carry through' information on the 'strength of evidence' associated with a line of evidence. The concept behind NUSAP is to provide a more comprehensive description of a numerical result in a way that captures aspects of error and uncertainty. NUSAP stands for **N**umber: **U**nit: **S**pread: **A**ssessment: **P**edigree and was the notation proposed by (Funtowicz and Ravetz, 1987). This string is composed as follows:

- 'Number' is the value being reported (highlighting the active significant figures):
- 'Unit' records the system being used:
- 'Spread' may be the standard deviation, the range, or minimum and maximum values:
- 'Assessment' provides an indication of the confidence limits, or may indicate any systematic error making use of previous results:
- 'Pedigree' is the most subjective element of the description. For this values are assigned against four attributes as detailed in Figure A1.

Using this approach the results of an experiment to define a physical constant with a recorded value of 137.0360m would be reported as:

137+360 : E-4 m : ±1 : ±2.6 : (4,4,3,4)

Number : **U**nit : **S**pread : **A**ssessment : **P**edigree

Quite often in an assessment there will be a mix of quantitative and qualitative evidence. However, for the assessment process, the primary interest is on a narrative description as provided by the pedigree component of the notation (Figure A1).

	Score	Information or data	Theory and Method	Peer Acceptance	Consensus
"strength of evidence" – scientific pedigree" →	4	**Comprehensive information** with sound data and good quality control	**Best available practice** and well established theory	**Absolute** – peer reviewed evidence from research literature.	**Accepted as 'an ideal approach.'**
	3	**Reliable analysis** of the available data	**Reliable method** commonly accepted	**High** – peer reviewed evidence	**Accepted as 'fit for purpose.'**
	2	**Calculation or estimation** of values	**Accepted method**, partial theory but limited consensus	**Medium** – some agreement accepting that there are some contradictory views	**Some consensus** but different 'schools of thought'
	1	Education opinion. **Expert view** based on limited information	**Preliminary method** unknown reliability	**Low** – no agreement	'**New approach'** un-tested
	0	Non-expert view/**guess**	**Crude speculation**/No discernable rigour	**None**	**None** – inappropriate use of data/information/ modelling

Figure A1 - Pedigree scoring guidance

Appendix B -Methods for comparing evidence

There are now a number of frameworks established for synthesising both qualitative and quantitative information, ranging from techniques that are largely qualitative and interpretive, through to techniques that are largely quantitative and integrative (Dixon-Woods *et al.*, 2005). The approach referred to as the Weight of Evidence method, recognises this full range and provides a useful basis for classifying the various approaches. A summary follows, together with brief overviews of two methods that have been extensively applied to-date; *comparative risk assessment* seeking to compare risks in a formal but largely qualitative framework and *multi-criteria decision analysis* which adopts a more quantitative approach.

B.1 Weight of Evidence

In this approach a range of methods are recognised to include both qualitative and quantitative considerations. This ranges from qualitative methods such as simple listing of the evidence or Best Professional Judgement (BPJ) though logic and causal criteria which use formal decision criteria to fully quantitative methods that involve indexing, scoring or the full probabilistic quantification of risk and formal decision-theory tools. A useful classification has been proposed by Linkov *et al*. (2009), as illustrated in Figure B1.

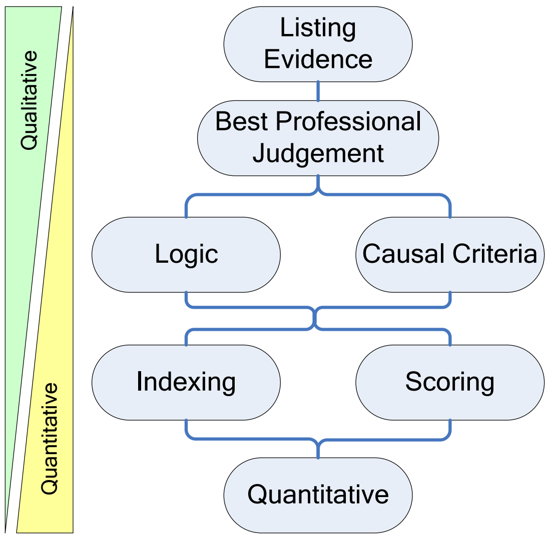


Figure B1 – Weight of evidence classification system (after Linkov et al., 2009)

- *Listing Evidence*: evidence is presented without any attempt to integrate different lines of evidence to synthesise a more in depth understanding;
- *Best Professional Judgement*: this is essentially a synthesis of the available evidence, testing for consistency amongst different lines of evidence to form a consensus (Spencer *et al.*, 2003; Townend, 2004; Shaxson, 2005);

- *Causal Criteria*: provides a framework for examining cause and effect relationships, which seeks to document cause-process-consequence links. This then provides a structured underpinning to the evidence based on identified relationships. The method still relies on BPJ to synthesise the evidence;
- *Logic*: this approaches uses previously defined methods or guidance as a basis for making an assessment (e.g. Defra, 2000 for one such method and some case studies). The method still relies on BPJ to synthesise the evidence;
- *Scoring*: various scoring methods exist which assigns weights to different lines of evidence, although typically this still uses BPJ to assign weights. This has been formalised in methods such as NUSAP (Funtowicz and Ravetz, 1987) – see Appendix A;
- *Indexing*: this combines the scores (with or without some weighting) to derive a single value that allows options or in this case risks to be ranked;
- *Quantitative*: uses formal decision theory methods (Benjamin and Cornell, 1970), such as Multi-Criteria Decision Analysis (MCDA) to weight the evidence and to integrate the findings, taking account of non-linearity and correlations (Linkov *et al.*, 2006). This approach also provides a reproducible method for integrating scientific results with expert opinion and decision maker's preferences.

It is quite common for the various methods identified in the classification to be combined for any given analysis. This will reflect the nature of the question being addressed, the data available and the ability to carry out a particular level of risk analysis. Many of the methods and applications relate to decisions that require selection from a number of options (e.g. scheme design options or regulatory choices). A slightly different problem is posed where a deeper understanding is sought to inform the decision-making process, based on a synthesis of the available information. A useful summary of the approaches available is provided Dixon-Woods *et al*. (2005). The underpinning methods are similar to those identified in the Weight of Evidence classification (see Table 1 in the paper) but a distinction is made between *integrative* and *interpretive* syntheses. The former is where the focus is on summarising data, and where the concepts (or variables) under which data are to be summarised are assumed to be largely secure and well specified. This type of summary may be achieved through pooling of the data, perhaps through meta-analysis, or less formally, perhaps by providing a descriptive account of the data. In contrast, the defining characteristic of an interpretive synthesis is its concern with the development of concepts, and with the development and specification of theories that integrate those concepts. Interpretive syntheses can be carried out on all types of evidence, both qualitative and quantitative. Interpretive synthesis is achieved by subsuming the concepts identified in the primary studies into a higher-order theoretical structure. Often this may take the form of a conceptual model against which different lines of evidence can be tested (see Section 3.3).

B.2 Comparative Risk Assessment (CRA)

Comparative risk assessment provides a systematic way of looking at problems that pose different types and degrees of risk. This approach tends to make use of the more qualitative methods in the Weight of Evidence classification, above, but often goes as far as some form of scoring and weighting. The aim is to compares risks fairly and meaningfully to see how they are different. It is very easy to compare risks badly, in a way that is meaningless because it does not relate to the management decision at hand or unfair because risks are framed inconsistently. It is therefore necessary to be careful to:

- frame the comparison in a way that applies equally to all the risks being considered;
- decide which are the qualities that matter most to the decision at hand.

To do this a four step process has been proposed (E.A., 2007a) as summarised in Figure B2.

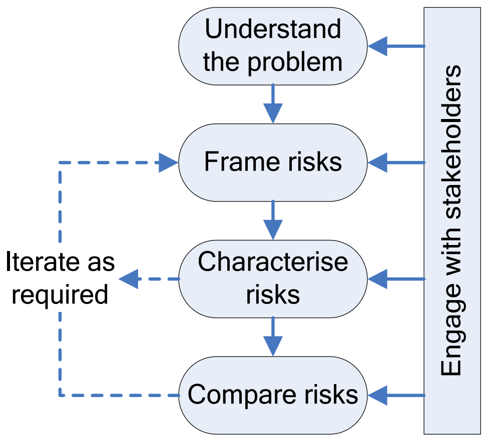


Figure B2 - EA framework for comparative risk assessment (EA, 2007)

1. Understand the problem

Prepare a plan that defines the question(s) to be answered, framed in comparative terms, defines the context for the decisions to be made, notes any constraints and obligations that should be considered in the analysis, and sets out how the analysis will be done.

2. Frame the risks

Identify the management goals and objectives as a structured hierarchy and a list of assessment endpoints for those objectives that cannot be measured directly and for each risk of interest.

3. Characterise the risks

Collect and collate the evidence about risks so that they can be evaluated against the defined objectives or endpoints.

4. Compare the risks

This may involve the use of scores and weightings to aid the process of comparing and contrasting the risks. However, the synthesis and interpretation requires subjective judgement (ie. BPJ).

Central to this process is the framing of the question to be addressed and then framing the risks against the question. A properly framed comparative risk assessment question should:

- refer to identified risks;
- make the point of comparison clear;
- clearly relate to the decision being made;
- identify any spatial, temporal or other boundaries.

To avoid wasted effort, it is also worth thinking about how detailed the analysis needs to be. It may be appropriate to begin with a simple, qualitative assessment of the problem and only move to more detailed techniques if the situation demands, and time and resources permit.

For this type of analysis goals, whilst useful, are a little vague, whereas objectives tend to be more specific and comprise a decision context, an entity and a preferred state for the entity (US EPA, 2001). Where it is not possible to measure objectives directly, assessment endpoints can be developed to act as surrogate measures. These are typically defined in terms of the entity and attributes of interest. This enables the analysis to focus on very specific formulations of the risks, tailored to the specific question and desired end objectives.

It is common to adopt some form of evaluative scale, chosen to suit the decision at hand:

- the maximum and minimum scores for each objective should be the same (because they are weighted separately);
- a zero score should indicate no change;
- the number of subdivisions should allow meaningful discrimination of outcomes and can vary between objectives;
- positive and negative scores should be consistently associated with positive and negative outcomes.

Where there is a wide range of possible outcomes, it may be simpler to adopt an exponential scale where outcomes are grouped by order of magnitude.

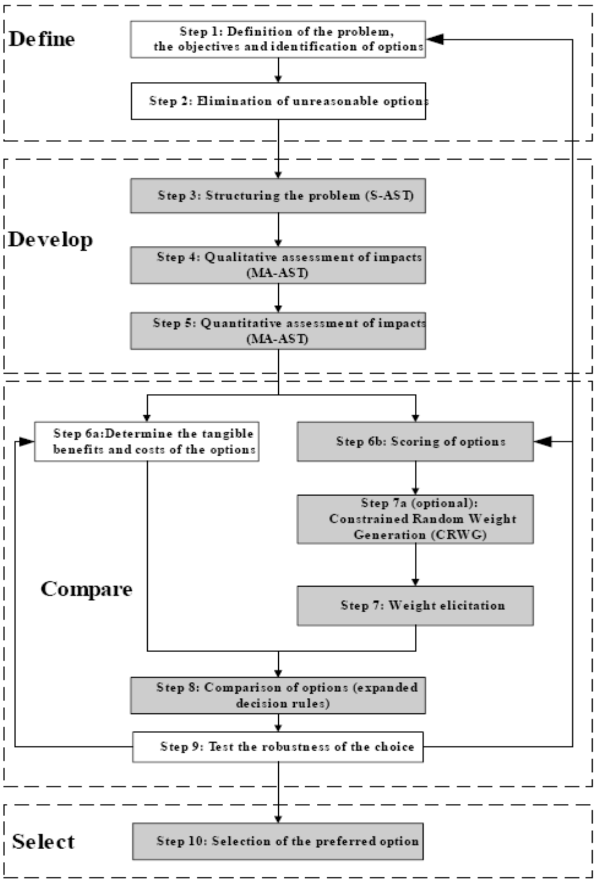
There is often a need for a participatory process to engage stakeholders in the scoring and weighting of risks. Whilst it is possible to then formalise these inputs using various decision-theory techniques (e.g. Morcom-Harneis *et al.*, 1998; Micallef *et al.*, 2001), it is more usual for this to be a reasoned evaluation. More sophisticated methods make use of ideas such as Utility theory and Bayesian updating and form part of the tool set that is now used to support multi-criteria decision analysis.

B.3 Multi-Criteria Decison Analysis (MCDA)

Multi-criteria decision analysis (MCDA) covers a range of techniques for assessing decision problems characterised by a large number of diverse attributes. These do not need to be expressed in money terms and MCDA techniques differ as to the characteristics of the options and measurement scales that they can handle, the decision rule that they apply, and the way scores are standardised (EA, 2007). This approach is the most extensively used quantitative Weight of Evidence methodology and is seen by some as an advance on comparative risk assessment methods (Linkov *et al.*, 2006). A recently completed literature review identified the strengths and weaknesses of the approach and noted the trend towards the greater use of participatory approaches for the development of alternatives, criteria selection and weight elicitation (E.A., 2007b).

The approach has been extensively applied in decision making related to a range of sectors including flooding, water resources, agriculture, environment, energy and transport. The steps are broadly similar to those proposed for Comparative Risk Analysis, Figure B2, and the EA guidance on framing the question and the attributes of most relevance to frame the risks remains pertinent. In the UK, the technique has been used to support decision making in both transport and flooding and in both cases, is an integral part of method used for economic appraisal (Defra, 2004; Dft, 2010).

For example, the project appraisal process for flood management and coastal defence involves four discrete stages - Define, Develop, Compare and Select - within which are included various procedural steps, as illustrated in Figure B3. This methodology uses a standard set of pre-defined impact categories throughout the process and allows defence options to be evaluated in cost-benefit terms but with stakeholder elicitation of the weighting given to the different impacts. More recent developments include the addition of the time dimension to the decision-making process acknowledging that some decisions can be left to the future, when more information will be available, providing decisions made in the present do not unduly constrain or eliminate possible future decisions. This approach is encapsulated in the Real Options method developed to investigate planning options for protection London from flooding (Treasury, 2009), and the close integration with fully probabilistic risk analysis methods (e.g. RASP Gouldby *et al.*, 2008) is currently the subject of ongoing research (Woodward *et al.*, 2010).



Impact categories used for assessment of flood management and coastal defence:

Economic Impacts

- Assets
- Land use
- Transport
- Business development

Environmental Impacts

- Physical habitats
- Water quality
- Water quantity
- Historical Environment
- Landscape and visual amenity

Social Impacts

- Recreation
- Health and safety
- Availability and accessibility of services
- Equity
- Sense of community

Figure B3 - Approach to Economic Appraisal including the MCA (Defra, 2004)

B4. Additional References

Benjamin J R and Cornell C A, 1970, Probability, statistics and decision for Civil Engineers, McGraw Hill Book Company, New York.

Defra, 2000, Guidelines for Environmental Risk Assessment and Management.

Defra, 2004, Evaluating a multi-criteria analysis (MCA) methodology for application to flood management and coastal defence appraisals: Guidance for the MCA-based element of the current approach to appraisal, Prepared by Risk and Policy Analysts Ltd (RPA) for the Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme, Report No: R&D Technical Report FD2013/TR, pp. 1-125, Defra, London.

Dft, 2010, Transport Appraisal and the New Green Book.

Dixon-Woods M, Agarwal S, Jones D, Young B and Sutton A, 2005, Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy, 10 (1), 45-53.

E.A., 2007a, Comparative Risk Assessment, Report No: Science Report - SC050030/SR, pp. 1-32, Environment-Agency, Bristol.

E.A., 2007b, Review of multi-criteria analysis applied to flood risk appraisal, Report No: Science Report SC050013/SR1, pp. 1-24, Enviornment-Agency, Bristol.

Ellis E C, Li R G, Yang L Z and Cheng X, 2000, Long-term change in village-scale ecosystems in China using landscape and statistical methods. Ecological Applications, 10 (4), 1057-1073.

Funtowicz S and Ravetz J, 1987, The arithmetic of scientific uncertainty. Phys.Bull., 38, 412-414.

Gouldby B, Sayers P B, Mulet-Marti J, Hassan M and Benwell D, 2008, A methodology for regional-scale flood risk assessment. Proceedings of the Institution of Civil Engineers, Water Management, 161 (3), 169-182.

Larsen L G, Eppinga M B, Passalacqua P, Getz W M, Rose K A and Liang M, 2016, Appropriate complexity landscape modeling. Earth-Science Reviews, 160, 111-130, doi: 10.1016/j.earscirev.2016.06.016.

Linkov I, Loney D, Cormier S, Satterstrom F K and Bridges T, 2009, Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. Science of The Total Environment, 407 (19), 5199-5205.

Linkov I, Satterstrom F K, Kiker G, Batchelor C, Bridges T and Ferguson E, 2006, From comparative risk assessment to multi-criteria decision analysis and adaptive management: Recent developments and applications. Environment International, 32 (8), 1072-1093.

Micallef A, Williams A T and Cassar M, 2001, Environmental risk assessment: application to a proposed beach nourishment, Malta. Shore & Beach, 69 (3), 13-17.

Morcom-Harneis A F, Townend I H, Kenny A J and Pontee N I, 1998, Environmental risk assessment and its application in managing coastal and estuarine developments, 1998 ICES Annual Science Conference Handbook, pp. T:14:11-T:14:10, International Council for the Exploration of the Sea, Cascais, Portugal.

Shaxson L, 2005, Is your evidence robust enough? Questions for policy makers and practitioners. Evidence & Policy: A Journal of Research, Debate and Practice, 1 (1), 101-112.

Spencer L, Ritchie J, Lewis J and Dillon L, 2003, Quality in qualitative evaluation: a framework for assessing research evidence, Government Chief Social Researcher's Office, London.

Townend I H, 2004, Identifying change in estuaries. Journal of Coastal Conservation, 10, 5-12.

Treasury H, 2009, Accounting for the Effects of Climate Change, Supplementary Green Book Guidance issued by HM Treasury and Defra, Report No.

Tufte E R, 1983, The visual display of quantitative information, Graphics Press, Cheshire, Connecticut,.

Van Der Sluijs J P, Craye M, Funtowicz S, Klorogge P, Ravetz J and Risbey J, 2005, Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: The NUSAP system. Risk Analysis, 25 (2), 481-492.

Woodward M, Gouldby B, Kapelan Z, Khu S-T and Townend I H, 2010, The use of real options in optimum flood risk management decision making, 1st European IAHR Congress, 4 - 6 May 2010, Edinburgh, UK.