

Ten Steps Applied to Development and Evaluation of Process-Based Biogeochemical Models of Estuaries

Barbara J. Robson^a, Ian T. Webster^a, David P. Hamilton^b, Terence Chan^c, Teemu Kokkonen

^a CSIRO Land and Water, GPO Box 1666, Canberra ACT 2601, Australia. barbara.robson@csiro.au;

^b Centre for Biodiversity and Ecology Research, University of Waikato, Private Bag 3105, Hamilton, New Zealand

^c Centre for Water Research, University of Western Australia, Stirling Hwy, Crawley, WA, Australia

Abstract: The procedures involved in model development may be set out as a ten-step process, beginning with defining the purpose of the model and ending with evaluation of the appropriateness and utility of the completed model. This process, recently outlined by Jakeman et al. [2006], is often iterative as model development is a continuous process that refines and improves the intended capacity of the model. Here, we show how the ten steps of model development are relevant to process-based biogeochemical modelling of aquatic systems, using examples from two case studies: a model of phytoplankton succession and nutrient concentrations in the Swan-Canning Estuary (Western Australia) and a model of sediment and nutrient transport and transformation in the Fitzroy Estuary and Keppel Bay (Queensland).

Keywords: biogeochemical model, ten steps, model development, process-based, water quality

1. MOTIVATION

Jakeman et al. [2006] recently outlined ten steps underpinning best practice model development to support natural resource management. This paper shows how these ten steps are relevant to process-based (mechanistic) water quality and biogeochemical modelling, using examples from two biogeochemical modelling studies of estuaries. Biogeochemical modelling, as used here, refers to modelling the biological and chemical processes affecting nutrients (usually nitrogen and phosphorus) and primary production.

2. TEN STEPS

1.1 Define model purpose

Our goals in modelling are shaped by the value attributed to the system, its anticipated problems or changes, and what can be done with the available resources, data, and technology. Here, we draw on case-studies of biogeochemical models applied to two Australian estuaries; namely, the Swan-Canning Estuary (Western Australia) and the Fitzroy Estuary (Queensland), including Keppel Bay and its major tidal creeks.

The Fitzroy contaminants project is a multidisciplinary, multiorganisation project conducted as one of the focal studies of the Coastal

Zone Cooperative Research Centre. The aims of the study were to understand sediment, nutrient and primary production processes in the macrotidal, tropical Fitzroy Estuary and Keppel Bay (near Rockhampton, Queensland, Australia) and to understand how the system responds to changes in sediment and nutrient loads, particularly how it intercepts sediments and nutrients generated by the catchment and delivers them to the Great Barrier Reef Lagoon. The study included collection of physical, geochemical and biological data in the field, laboratory studies, and the development of mechanistic hydrodynamic, sediment and biogeochemical models. Here, we largely confine our discussion to the development and application of the biogeochemical model.

The purpose of the model was, in combination with field and laboratory studies, is to:

- i. Understand how nutrients are transported through and transformed within the Fitzroy Estuary and Keppel Bay; and
- ii. Predict how this will change in response to changes in flows and nutrient loads resulting from changes in catchment land use.

In general terms, the aim of the model was to provide information to facilitate management of the estuary and its catchment in the context of

ameliorating the impact of catchment-derived materials on the Great Barrier Reef Lagoon.

The Swan-Canning Estuary modelling study, focussing on a moderately polluted urban estuary in Perth, Western Australia, had different, though somewhat related aims. This study relied largely on interpretation and incorporation of data from an ongoing monitoring programme, rather than gathering new data. A model was required to:

- i. Predict phytoplankton concentrations and blooms of cyanobacteria (blue-green algae) and other phytoplankton; and
- ii. Predict nitrogen and phosphorus concentrations in the estuary and the likely effects of changes in catchment management and use.

Again, this project had a broader underlying goal of providing information and a tool to help inform decisions regarding catchment and estuary management and to evaluate the likely effectiveness of plans to improve the health of the estuary through improvements in catchment management.

1.2 Specify modelling context: scope and resources

- What resources are available?

Both the Fitzroy Contaminants programme and the Swan-Canning modelling project were relatively large projects, with funding that supported the modelling for approximately three years. They included modelling teams of hydrodynamic (and in the case of the Fitzroy, sediment dynamic) modellers as well as biogeochemical modellers. Sufficient personnel time was available to develop and calibrate complex models which were custom-adapted from models previously applied to other systems. The Fitzroy model was based on the model of Murray and Parslow [1999b] for Port Phillip Bay, while the Swan project used the ecological model CAEDYM [Hamilton and Herzfeld, 1999] and was the first application of CAEDYM in conjunction with a three-dimensional hydrodynamic model.

The Swan-Canning project drew on an extensive dataset from an ongoing weekly monitoring program by the Water and Rivers Commission [Hamilton and Turner, 2001], while the Fitzroy Contaminants project relied on data from a field programme designed in conjunction with the modelling project. Each of these situations had certain advantages: the more extensive dataset

available for the Swan-Canning estuary allowed more detailed 'validation', whereas the flexibility of field and laboratory studies conducted in cooperation with the modelling project allowed knowledge gaps that affected model setup, e.g. properties of local sediments, to be specifically targeted.

- What outputs are required?

To achieve the goals outlined in step 1, we must first specify precise objectives in terms of what outputs are required. The development process then focuses on providing a path to provide these outputs.

For the Fitzroy/Keppel Bay biogeochemical model, required outputs included a) predicted concentrations of nitrogen in the water column and b) an estimate of how much nitrogen is exported to the Great Barrier Reef Lagoon. Also important, but secondary to these factors, were concentrations and exports of phosphorus, an estimate of primary production, and an understanding of the relative importance of processes such as settling and resuspension, nitrification and denitrification, nitrogen fixation, remineralisation, benthic and pelagic primary production.

For the Swan-Canning modelling project, the most important outputs were a) predictions of relative concentrations of different groups of phytoplankton and b) absolute concentrations of chlorophyll *a*. Also important, but secondary, were predictions of concentrations of various species of nitrogen and phosphorus, concentrations of dissolved oxygen, and an understanding of the relative importance of sediment nutrient exchanges, inputs from groundwater and urban drains and inputs from major tributaries in determining phytoplankton community composition and chlorophyll *a* concentrations.

In both projects, the ability to predict higher-level ecological effects such as effects on fish and shellfish populations might have been desirable, but was not considered practicable within the scope of the projects given knowledge and resources available.

- What is the timescale relevant to model output?

In both the Fitzroy and Swan-Canning projects, the emphasis was on prediction of seasonal patterns and trends, ideally with a sufficient resolution to simulate the effects of events such as storms that may have immediate influences on a daily to weekly scale. At the same time, we wanted to be

able to conduct scenarios to predict responses on a scale of one to three years.

- What is the relevant spatial scale?

In the Fitzroy project, the area selected for modelling included the Fitzroy Estuary below the barrage at Rockhampton and much of Keppel Bay (Figure 3).

In the Swan-Canning project, the domain included the estuary from its upstream tidal limit of the estuary to the mouth of the estuary at Fremantle. The extent of the selected domain was influenced by considerations of where the greatest management problems were observed, by the spatial extent of regular monitoring data, and by the traditional definition of the estuary.

- Who will use the model?

This question is important because it affects the implementation of a model (e.g. whether an elaborate user interface is necessary) as well as how complex it can be. Mechanistic biogeochemical models are usually complex and, when designed to produce quantitative results, are often accessible only to trained modellers. For both the Fitzroy and Swan-Canning modelling projects, the models used were designed to be used by experienced modellers only. The models were designed to be used to run scenarios relevant to management questions and stakeholder concerns. Model output was then analysed to provide scientific and plain English interpretations for stakeholders, including regional environmental managers.

1.3 Conceptualise the system

Conceptualisation of an estuary usually begins with flows of water. Fresh water flows in to the estuary from rivers, drains and groundwater, and is mixed with salty water from the ocean. Exchanges of water are also exchanges of nutrients and sediments. At the interface, fresh water tends to flow over salt water because it is less dense. Other important physical factors are surface elevation changes with tides, mixing and transport of water within the estuary, settling and resuspension of particles, and changes in density due to changes in salinity and temperature. Salinity and temperature are influenced by boundary conditions, mixing, evaporation, precipitation and thermodynamics.

Within this physical context, a great many chemical and biological processes occur. Figure 1 illustrates just a few of the animals, plants and chemical components of a typical estuary. The relationships between these components can be thought of in a number of ways, but the conceptualisation represented by Figure 1 is a “stocks and flows” model, with nitrogen as the currency. Each component or stock (such as seagrass) is considered in terms of how much nitrogen it contains, and each process (represented as an arrow) is represented as a flow of nitrogen from one stock to another. Processes include growth of green algae (which transfers nitrogen from ammonium and nitrate stocks to the green algae stock), grazing of green algae by shrimp (transferring nitrogen from the green algae stock to the shrimp stock) and mortality of shrimp (transferring nitrogen to a stock of detritus). Similar diagrams can be drawn to represent stocks and flows of phosphorus, carbon, or even energy in a system.

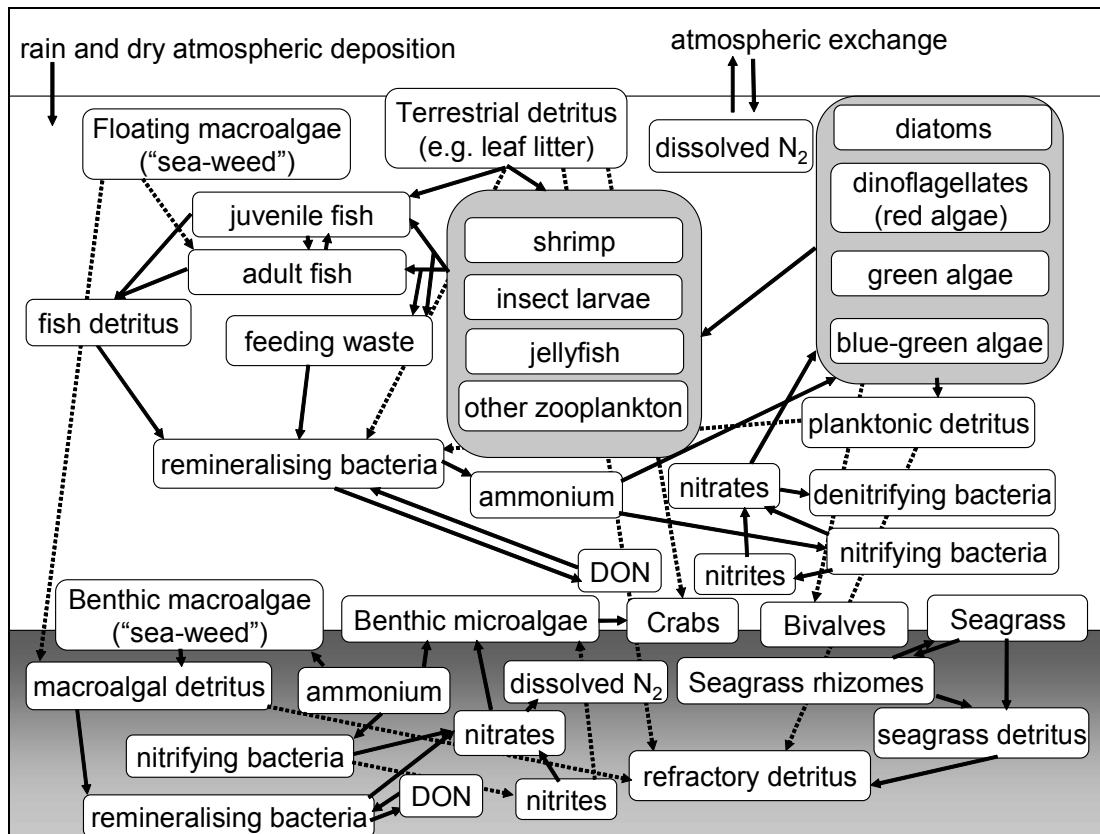


Figure 1 A conceptual diagram illustrating some of the many nitrogen stores and pathways in an estuary (not all links shown). “DON” refers to dissolved organic nitrogen. Sediment stores are shown in the darker section at the bottom.

It is rarely desirable to include explicitly in a biogeochemical model all of the components shown in Figure 1. Which components and which processes should be included depends on what information is available and what outputs are required.

Physical and chemical data to help define biogeochemical models for particular systems came from a variety of sources including regular monitoring by various government agencies, input from other models where measured data are not available, and field and laboratory studies designed to fill knowledge gaps [Robson and Hamilton, 2003; Radke et al., 2005].

1.4 Select model features (form of model)

a) Modelling approach

Several general approaches to modelling aquatic systems are possible and each has advantages and disadvantages. Statistical approaches such as neural network models sometimes prove very accurate in predicting variations in, for example, phytoplankton biomass, and do not rely on any

preconceived notion of how the system functions. If the system changes beyond the range for which a neural network model was trained, however, their predictions are unlikely to be valid, and statistical models are not designed to provide insight into the internal dynamics of a system. Maier and Dandy [2000] provide a review of artificial neural networks for prediction and forecasting of water resources variables.

Bayesian models [e.g. Borsuk et al., 2006] are becoming increasingly popular. These models have the advantage that they can draw together knowledge in different forms, which may or may not be quantitative, and which may vary in certainty and accuracy. They can also be designed to provide relatively easily an estimate of uncertainty as part of their predictions, and can be flexibly altered to take into account input from stakeholders. Bayesian models, however, do not directly incorporate a biophysical understanding of individual processes within the system, which may have implications for interpretation of the results. Bayesian models are rarely quantitative in the sense of being able to provide precise time-series

predictions of changes in concentrations and nutrient stores.

Process-based (or “mechanistic”) biogeochemical models have a few advantages over other approaches for the two applications described here:

- i. They explicitly represent our understanding and functionality of the system, that is, not only its responses, but also its internal dynamics.
- ii. They allow us to quantitatively simulate the current behaviour of the system in detail.
- iii. They provide a means to predict responses to changes, even (with caution) when those changes take the system beyond its historical variability.

Process-based models also have disadvantages: they tend to have high data input requirements and high complexity, can have high computational costs, and it can be difficult to quantitatively estimate the uncertainty of the predictions. Nonetheless, for the purposes of the Swan-Canning and Fitzroy/Keppel Bay modelling projects, process-based biogeochemical models were considered the best option given the available resources and expertise.

Process-based models may be implemented with either a traditional approach or an agent-based modelling approach. Agent-based modelling is often particularly appropriate when the components are readily conceptualised as individual agents; for example, individual fish and animals in an ecosystem model [e.g. Gribble, 2004], or individual investors in an economic model [e.g. Panzarasa et al., 2001]. For the cases considered here, most system components (e.g. dissolved nitrogen) are more readily conceptualised in terms of mass and concentrations, so a more traditional (non agent-based) approach was taken.

b) Conceptual model

In both of the case studies here, an estuary was conceptualised in terms of stocks and flows of nitrogen and phosphorus, and this conceptualisation is reflected in the structure chosen for the models. The system illustrated in Figure 1 was in each case simplified to include only those processes that we believed to be most important to the desired outputs, on the basis of previous scientific literature describing the estuaries in question and similar estuaries as well as our past experience in biogeochemical modelling of estuaries.

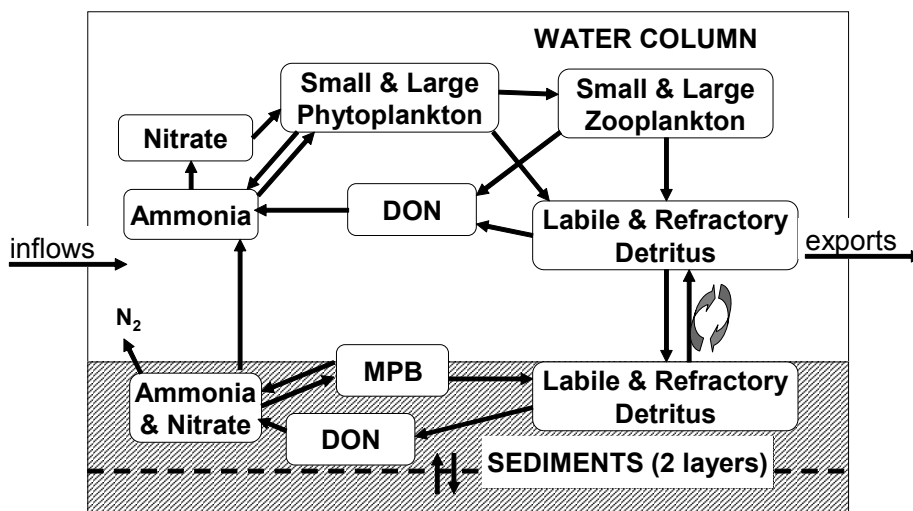


Figure 2 Major nitrogen pathways and stores included in the Fitzroy/Keppel Bay model (c.f. Figure 1). Groups of biota represented include small phytoplankton, large phytoplankton, small zooplankton, large zooplankton and benthic microalgae (MPB). Other nitrogen stores included in the model include ammonium, nitrate, dissolved organic nitrogen (DON), labile detritus and refractory detritus. Each living and non-living nitrogen store is represented in the water column and in two sediment layers, and each process in the biogeochemical model represents a transfer between two or more nitrogen stores.

Major nitrogen pathways included in the Fitzroy model are shown in Figure 2. The chemical and biological processes of interest here include growth, respiration and mortality of phytoplankton, benthic microalgae and zooplankton, grazing, settling, and transformations of nitrogen and phosphorus from one form to another (remineralsation of organic material, nitrification and denitrification, nitrogen fixation, nutrient uptake by primary producers, adsorption and desorption of phosphorus from sediment surfaces, flocculation and aggregation of fine particles).

In the case of the Swan-Canning project, a similar stock and flow model was developed, but zooplankton was not explicitly simulated, but grazing by zooplankton was represented in a more abstract sense and benthic microalgae was not considered important, and hence was not included in the model [Robson and Hamilton, 2004]. The Swan-Canning estuary is microtidal and therefore does not have extensive intertidal mudflats on which benthic microalgae might dominate. Nor is the water sufficiently clear to encourage extensive growth of algae on the estuary bed.

A further simplification for the Swan-Canning model was to represent sediments as a “black box”, with exchanges between sediments and the water column controlled by shear stress, dissolved oxygen concentrations and nutrient concentrations water overlying the bottom sediments, but without sediment stores explicitly represented [Robson and Hamilton, 2004]. This simplification may have been appropriate for a model of the Swan-Canning estuary, but would clearly not have been appropriate for a model of the Fitzroy Estuary, in which the sediments are much more dynamic due to strong tidal currents and much of the total nitrogen store is in the sediments.

Phytoplankton, by contrast, was represented in much more detail in the Swan-Canning model than in the Fitzroy model, because one of the major goals (see step 1) was to predict phytoplankton succession and blooms. The Swan-Canning model therefore included up to five distinct taxonomic groups of phytoplankton (compared with the two size classes simulated in the Fitzroy Estuary and shown in Figure 2), with each group differentiated by growth and respiration rates, size, and responses to changes in nutrient concentrations, salinity and temperature.

c) Spatial and temporal scales

Features of the Swan-Canning Estuary include strong along-estuary salinity and nutrient gradients, persistent vertical stratification, and spatially patchy phytoplankton dynamics. The Fitzroy Estuary is vertically well-mixed with respect to temperature for most of the year, but retains strong vertical variations in sediment concentrations. During flood events, a plume of fresh water extends over the surface from the mouth of Fitzroy Estuary and into Keppel Bay. These three-dimensional dynamics could be reproduced only with a three-dimensional hydrodynamic model.

There are often tradeoffs between the time scale and spatial scale used to resolve an estuary for biogeochemical modelling. Three-dimensional hydrodynamic modelling usually implies a large number of grid cells and high computational demands. Coupling with a biogeochemical model increases these computational requirements. For the Swan-Canning model, a “straightened” [Hodges and Imberger, 2001] three-dimensional grid of cells of 1000 m length, 100 m width and 0.6 m depth was used after a model with 100m x 100m cells was found to be unacceptably slow. For the Fitzroy/Keppel Bay model, a curvilinear grid (Figure 3) was applied, with grid resolution varying from 200 m in Fitzroy Estuary to 2 km at the seaward boundary of Keppel Bay and vertical layers varying from 0.5 m at the surface to 2 m at the bottom near the maximum depth of 18 m [Herzfeld et al., 2005]. For numerical stability, hydrodynamic models on these scales require integration with time-steps on the order of 20 seconds.

The real-time to run-time ratios for the coupled Swan-Canning model at this resolution on a desktop computer circa 2002 was about 35:1 (i.e. a single one-year simulation took almost two weeks of computer time to complete), while the real-time to run-time ratio for the coupled Fitzroy model on 12 processors of a high-performance scientific computer in 2005 was about 70:1 (allowing a one-year simulation to be completed in a little under one week). Because calibration in each case required many partial and complete model runs, achieving a fine temporal resolution necessarily limited the spatial resolution achievable.

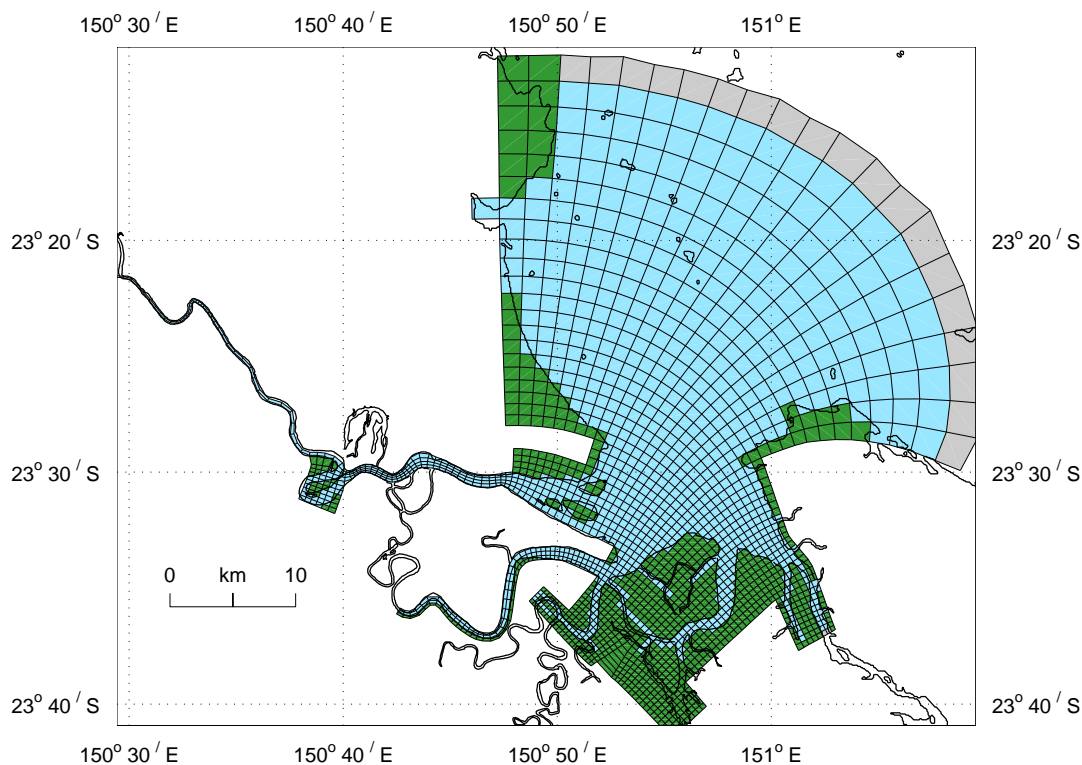


Figure 3 A two-dimensional view (looking down from above) of the three-dimensional model grid used for the Fitzroy/Keppel Bay model [(from Herzfeld et al., 2005)]. Green cells represent land and grey cells represent the ocean boundary, where Keppel Bay meets the Great Barrier Reef Lagoon.

1.5 Determine how model structure and parameter values are to be found

In both case studies, we embedded the biogeochemical model within a hydrodynamic model (and in the case of the Fitzroy/Keppel Bay study, with a sediment dynamic model). We structured the models as “stock and flow” models, with nitrogen and phosphorus as the models’ currency; i.e. most biogeochemical processes were represented as transformations between one form of nitrogen or phosphorus and another. Changes in concentrations of each pool were calculated through numerical solutions to systems of simple partial differential equations.

We aimed wherever possible to base functional algorithms for each process as well as parameter values on our biophysical understanding of how the estuaries functioned, though the level of abstraction varied depending on the complexity and importance of each process. In the case of the Fitzroy/Keppel Bay model, for instance, we employed a physiologically realistic mechanistic model for nutrient uptake and light interception by phytoplankton cells [Baird et al., 2003], based on cell size. In the case of the Swan-Canning model,

phytoplankton were differentiated not by cell size but by taxonomic grouping, and the use of more abstract Michaelis-Menten algorithms for nutrient uptake and phytoplankton growth allowed us to draw upon an extensive literature of laboratory studies and previous modelling to define parameter ranges for each phytoplankton group.

Algorithms and parameter values for process-based models can be determined from:

Laboratory and field studies relating directly to the system being modelled. In the case of the Swan-Canning project, parameters defining the response of the cyanobacterium, *Microcystis aeruginosa* to different salinities were determined through observation of growth rates *M. aeruginosa* isolated during a bloom in the Swan-Canning estuary and grown in water at several different salinities in the laboratory [Robson and Hamilton, 2003]. In the Fitzroy Contaminants project, sediment particle size ranges were set partly from field observations of settling rates [Margvelashvili et al., 2005].

Scientific literature describing previous laboratory, field and modelling work. Both the Swan-Canning project and the Fitzroy Contaminants project drew

heavily on such literature to define ranges for many parameter values. Sources of parameter values for the Swan-Canning project are listed by Robson and Hamilton [2003], while ranges for most parameters in the Fitzroy model were taken from Murray and Parslow [1997].

Calculations based on underlying physical properties. For example, rates of nitrogen and phosphorus uptake by phytoplankton are a function of the gradient between extracellular and intracellular nutrient concentrations and the size of phytoplankton cells. This relationship is used in the Fitzroy model to determine nutrient limited phytoplankton growth rates, using an algorithm developed by Baird et al. [2003].

Because of the inherent variability of biological processes, many parameters will be defined only within quite broad ranges; for instance, observed half-saturation constants for uptake of nitrate by diatoms range from $0.4 \mu\text{g L}^{-1}$ to $5.1 \mu\text{g L}^{-1}$ [Raymont, 1980]. Parameter values therefore needed to be calibrated within these ranges to arrive at values appropriate to the study sites.

1.6 Choose performance criteria

Performance criteria for environmental models must reflect the overall aims and specific objectives of the modelling activity. For the Fitzroy Contaminants project, our aim was to develop a model that would allow us to improve our understanding of the system dynamics and predict the effects of changes in loads and flows on primary production, water column concentrations of sediments and nutrients, and exports to the Great Barrier Reef Lagoon over a timescale of a year or more. It follows that the model could be assessed as performing well if it could reproduce observed variations in salinity, sediment concentrations, nutrients (especially total nitrogen and dissolved inorganic nitrogen) and chlorophyll *a* along the length of the estuary over a 12-month period that includes a wet season and a dry season, using process-based algorithms, a consistent set of parameter values and realistic inputs.

In the case of the Swan-Canning project, a major aim was to predict the effects of changes in loads and flows on phytoplankton succession, nutrient concentrations and algal blooms. Minimal performance criteria therefore include the ability to predict which phytoplankton group is dominant at any given time, to predict the approximate timing and magnitude of phytoplankton blooms, and to predict approximate concentrations of total nitrogen, dissolved oxygen and dissolved inorganic nitrogen. The ability to predict the

occurrence, timing and taxonomic grouping of blooms is more important for management of this estuary than the ability to predict concentrations of chlorophyll *a* during non-bloom periods. Prediction of chlorophyll *a* is probably more important than prediction of nitrogen, but in a nitrogen-limited system, a biogeochemical model that cannot reproduce observed concentrations of dissolved inorganic nitrogen would lack credibility. As for the Fitzroy/Keppel Bay project, it was important that the model used process-based algorithms, a consistent set of parameter values and realistic inputs.

More generally, criteria set for both models were that the predicted responses were *plausible* in light of our biophysical understanding of the system; that the models correctly reproduced the observed *ranges* of nutrient and phytoplankton concentrations; that the simulated *median* concentrations were close to the observed medians; that they successfully reproduced typical *spatial and temporal patterns* in the data (such as the distribution of suspended sediments in Keppel Bay and the seasonal pattern of phytoplankton succession in the Swan Estuary); and that they were able to reproduce the approximate *timing of major events* (such as the timing of dinoflagellate blooms in the upper Swan Estuary and the duration of elevated nutrient concentrations after a flood in Fitzroy Estuary).

These performance criteria were known in qualitative terms but were not specified in quantitative terms in advance.

1.7 Identify model structure and parameters

Processes included in the biogeochemical model for the Fitzroy Contaminants project included remineralisation of organic material, growth and mortality of benthic microalgae, growth and mortality of three phytoplankton groups (small phytoplankton, large phytoplankton and *Trichodesmium*), growth and mortality of two zooplankton size-classes (including grazing of phytoplankton), nitrogen fixation by benthic microalgae and *Trichodesmium*, nitrification, denitrification, phosphorus adsorption and desorption, exchanges between sediment layers and between sediments and the water column, growth and mortality of seagrasses and macroalgae, and physical processes affecting hydrodynamics and sediment dynamics. The algorithms used to define each of these processes are described by Robson et al. [2005] and Murray and Parslow [1997].

Processes included in the biogeochemical model for the Swan-Canning modelling project included growth and mortality of five to six phytoplankton groups (affected by nutrient uptake, salinity and temperature), a loss term for grazing by zooplankton, nitrification, denitrification, decay of biological oxygen demand (equivalent to remineralisation of detrital material in the Fitzroy model), oxygen exchanges with the atmosphere at the surface, sediment oxygen demand and sediment nitrogen and phosphorus releases, as well as physical processes relating to hydrodynamics and settling and resuspension of particles. The algorithms are described in detail by Robson and Hamilton [2004] and Hamilton and Herzfeld [1999].

Water column and sediment-bed nutrient and sediment concentrations in the Fitzroy Estuary and Keppel Bay were initialised with observations from the first intensive field campaign in the Fitzroy Contaminants project, in September 2003. This survey occurred during the long dry season, when concentrations were not changing rapidly. The model was then run from September 2003 through to the end of February 2005. This allowed a run-up of almost a year between the start of the simulation and the time of the next major field campaign (August 2004), and allowed comparison of model results with observations during the two major seasons (a second dry-season field campaign was conducted during August 2004, and a wet season campaign was conducted in February 2005).

Water column nutrient and phytoplankton concentrations for the Swan-Canning model were initialised with observational data from 9 sites in December 2004. The model was then run over a five-year simulated period, over which weekly monitoring data allowed an almost continuous comparison of observational results with model predictions.

The models used in both of the case-studies discussed here are relatively complex and require a large number (> 100) of parameter values to be set. Most process-based biogeochemical and ecological models are mathematically overparameterised at first glance, but in practice, tight limits on physically reasonable parameter ranges and the use of laboratory and field observations to further constrain these ranges greatly reduce this problem.

Formal calibration and parameter estimation procedures such as Monte Carlo optimisation and variants such as the Hornberger-Spear approach [Hornberger and Spear, 1983] and genetic algorithms [e.g. Mulligan, 1998; Ng and Perera, 2003] can be successfully applied to process-based biogeochemical models in some circumstances, but it is not yet feasible to apply these techniques to coupled three-dimensional hydrodynamic and biogeochemical models at reasonable resolutions because of the computational costs of running large numbers of simulations with such models. As mentioned previously, each one-year model run took one to two weeks of computer time to complete. Calibration runs over shorter periods were used in both cases, but year-long calibration runs were also required to ensure that the model could adequately reproduce patterns on a seasonal time-scale. Computational constraints seriously limited the number of calibration runs possible in both cases, and hence limited the techniques that could be applied.

In practice, complex biogeochemical simulation models are most commonly calibrated by trial and error: an expert modeller with an understanding of both the biophysics of the system and the structure of the model compares model results with field data either by eye or with the aid of some measure of goodness-of-fit, and adjusts parameter values by hand within literature ranges. This was the approach taken in both of the case studies presented here. Approximately 65 model runs were required to achieve a satisfactory calibration of the Fitzroy/Keppel Bay model, although most of these were for short (one- to three-month) simulations. The number of calibration runs required for the Swan-Canning model was not recorded, but probably ran to hundreds, due to the greater density of field data to be matched. In both cases, there was more than one variable to be optimised, and calibration efforts aimed to produce an acceptable fit between model results and observations of total phosphorus, total nitrogen, dissolved inorganic phosphorus and nitrogen, dissolved oxygen and phytoplankton concentrations. Figure 4 compares concentrations of each of four phytoplankton groups in the Swan-Canning estuary as estimated from field observations during 1995 with concentrations simulated by the calibrated model.

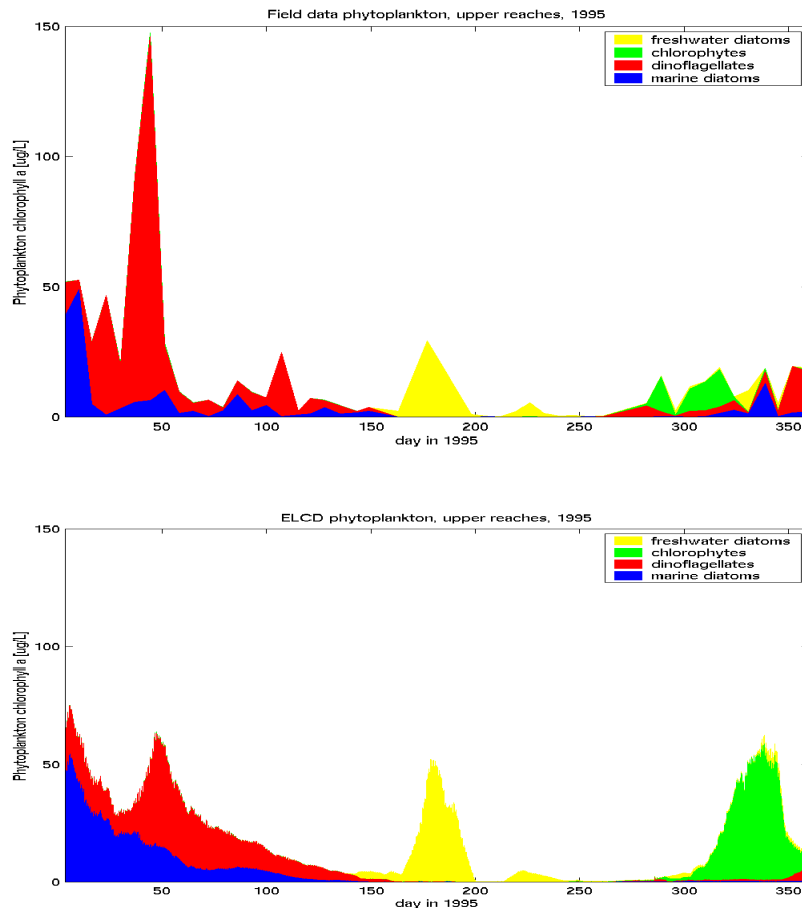


Figure 4 Chlorophyll *a* concentrations in the upper Swan-Canning estuary in 1995, averaged over the six upstream sites. Total chlorophyll *a* is given by the total height of the shaded areas; colours indicate different phytoplankton groups; (a) in the field; (b) as simulated by the Swan-Canning model.

More rigorous approaches to sensitivity analysis and parameter estimation for complex mechanistic simulation models have been proposed and may be advisable [Brun et al., 2001], but still require a relatively large number of runs, bearing in mind that both models included over 100 parameters. Combining a complex or high-resolution model with a simpler or lower-resolution model of the same system can be another useful way to reduce the computational demands of calibration, as initial, broad-level calibration and selection of algorithms can sometimes be performed using the simpler model [Murray, 2001]. This was attempted with the Fitzroy/Keppel Bay model, using a much lower resolution version of the final model for initial calibration. This was not particularly successful in this case, however, as the low-resolution model (which used only 23 vertically well-mixed boxes to represent the entire area) was not able to capture the tidal dynamics that were so important to sediment (and hence also nitrogen and phosphorus) dynamics in this system.

1.8 Verification and diagnostic testing

In the case of the Swan-Canning modelling project, the majority of calibration was conducted using observational data for 1995, with additional fine-tuning using 1996 and 1997 observations. This allowed model verification against observational data from 1998 and 1999. Validation of the Swan-Canning model against weekly surface and near-bed observations of salinity, temperature, dissolved oxygen, total nitrogen, total phosphorus, nitrate, phosphate, ammonium and chlorophyll *a* concentrations at nine sites in the estuary is described in detail by Chan [2006]. Temporal comparisons of estimated field concentrations with simulated concentrations of four groups of phytoplankton in 1995 and 1997 are reproduced from Chan [2006] in Figure 4 and Figure 5. The model predicted nutrient concentrations and phytoplankton succession very well for the calibration year (1995) and acceptably well for the verification year (1997) given the large uncertainties inherent in the input and observational data, although it was not always able

to reproduce accurately the balance of the two marine groups, dinoflagellates and marine diatoms (Figure 5). Among the sources of error discussed by Chan [2006] was the limited temporal

resolution of data to specify seaward boundary conditions.

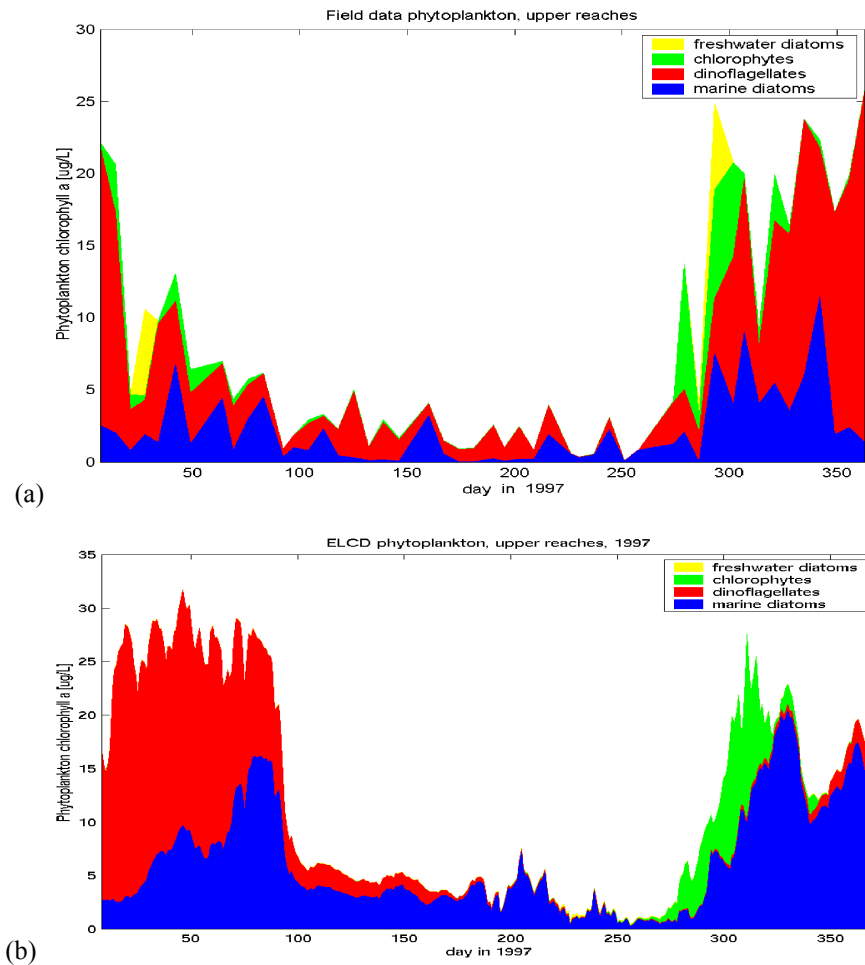


Figure 5 Chlorophyll a concentrations in the upper estuary in 1997, averaged over the six upstream sites. Total chlorophyll a is given by the total height of the shaded areas; colours indicate different phytoplankton groups; (a) in the field; (b) as simulated by the Swan-Canning model.

In the case of the Fitzroy Contaminants project, model calibration was conducted primarily against observational data from the August 2004 (dry-season) field campaign, leaving only the February 2005 (wet-season) field campaign for independent verification. Unfortunately, the calibration period did not include a significant inflow event and thus parameters relating to freshwater nutrient loads were not adequately calibrated in the first instance. Comparison of model results with field observations for February 2005 indicated a need to increase the value of the parameter defining the breakdown rate of detrital material in incoming freshwater to reflect that this material was more readily bioavailable than “old” detrital material in the estuary. Hence, there was no completely independent verification dataset for the Fitzroy/Keppel Bay model. As is often the case,

resource constraints did not allow additional field campaigns.

Spatial comparisons of wet- and dry-season observational data and simulated concentrations of total nitrogen, total phosphorus, dissolved organic nitrogen, dissolved inorganic nitrogen, dissolved organic phosphorus, dissolved inorganic phosphorus, chlorophyll *a* and dissolved oxygen are given by Robson et al. [2005]. Spatial comparisons of wet-season and dry-season total nitrogen (Figure 6 and Figure 7), dry season chlorophyll *a* (Figure 8) and wet-season dissolved inorganic nitrogen (Figure 9) are reproduced below.

In general, Robson et al. [2005] concluded that the Fitzroy/Keppel Bay model simulates dry-season

concentrations of nitrogen and phosphorus (as DON, DOP, DIN, DIP, PN and PP) well. Simulation of wet-season conditions is reasonable, with some reservations: wet-season particulate nitrogen and dissolved organic nitrogen and phosphorus are underestimated by the model. This may be due to the difficulty of estimating inflow nutrient concentrations under high flow conditions or to over-estimation of the rate of deposition of fresh particulate material in the Fitzroy Estuary. Analysis of sediment modelling results [Margvelashvili et al., 2005] found that suspended solids were underestimated during high flow events. Possible reasons discussed by Marvelashvili et al [2005] include insufficient data

to prescribe the inflow boundary condition and gaps in our understanding of sediment processes.

A quantitative estimate of overall model performance can be obtained by comparing model predictions with field observations interpolated to the same grid for both the wet-season and dry-season campaigns. This comparison indicates reasonable spatial agreement for most variables: $r^2 = 0.49$ for DIN, 0.58 for DON, 0.37 for DIP and 0.41 for O_2 , indicating that the model explains 40-60% of the observed variability. Agreement is a lower for particulate materials, resulting in an r^2 of 0.37 for TN (with a 35% relative error) and an r^2 of 0.19 for TP (with an 83% relative error).

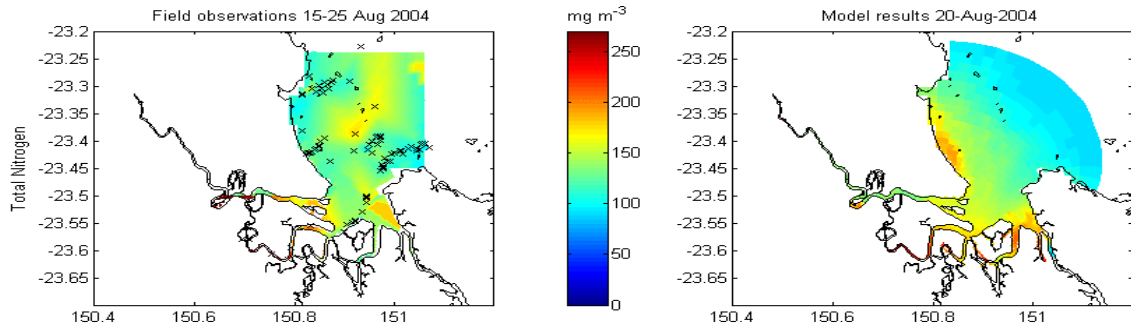


Figure 6 Total Nitrogen from field data (left) and from model output (right), August 2004. Crosses indicate tidally corrected locations of field observations.

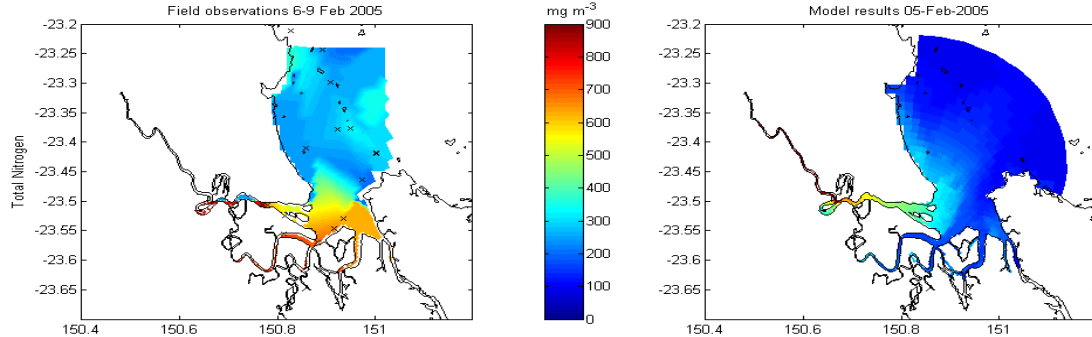


Figure 7 Total Nitrogen from field data (left) and from model output (right), February 2005.

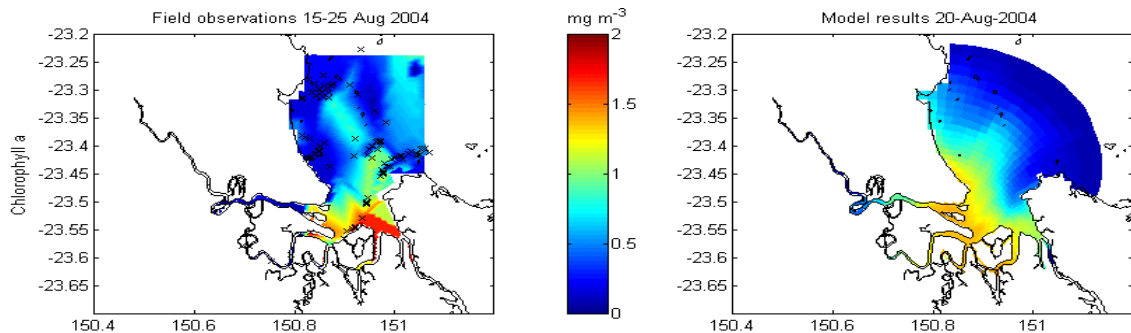


Figure 8 Chlorophyll a from field data (left) and from model output (right), August 2004, omitting a suspect point at (151.0280, -23.3471).

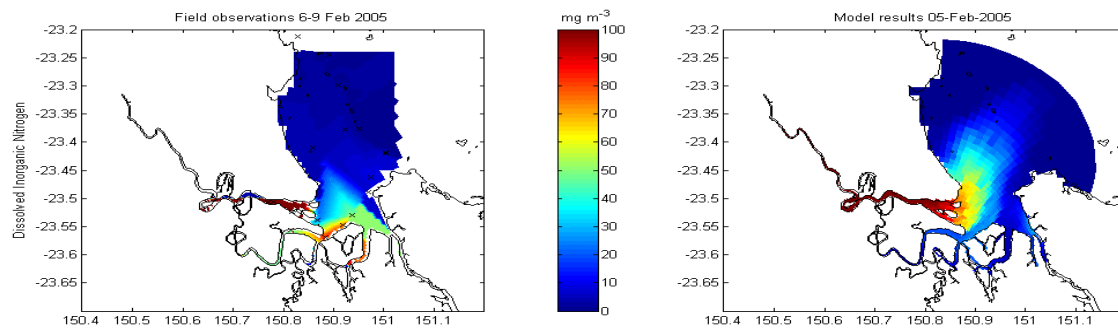


Figure 9 Dissolved Inorganic Nitrogen from field data (left) and from model output (right), February 2005.

1.9 Quantification of uncertainty

Quantification of uncertainty was perhaps the weakest component of the ten steps for both the Swan-Canning and Fitzroy/Keppel Bay modelling projects. Uncertainty in the predictions of both models is relatively high. Nitrogen and sediment budgets calculated using the Fitzroy/Keppel Bay model are probably accurate to within $\pm 200\%$ of the values calculated [Robson et al., 2005], while the timing of the start and finish of phytoplankton blooms predicted by the Swan/Canning model is generally accurate to within less than one month, but is heavily dependent on upstream boundary conditions.

Sources of uncertainty and error in process-based biogeochemical models include:

- Uncertainties and errors in input data, including sampling and measurement errors, the need to correct for tidal skewing of sample locations, limited spatial and temporal resolution of input data specifying boundary conditions and data for verification, missing data, and uncertain conversions (e.g. conversions between cell counts and chlorophyll *a* attributable to each phytoplankton group for the Swan-Canning model; conversion between chlorophyll and nitrogen content and between sediment concentrations and turbidity for the Fitzroy/Keppel Bay model).
- Limited spatial resolution of the models (spatial resolution affects the accuracy of hydrodynamic modelling and limits the ability of the model to reproduce small-scale patchiness in chlorophyll and nutrient concentrations).
- Uncertainty in the choice of which processes to include and which to omit. Processes included were chosen on the basis of our biophysical understanding of the estuaries at the start of model development and limited by the availability of input data (e.g. the effect of iron

concentrations on processes such as nitrogen fixation and adsorption of phosphorus onto sediment surfaces could not be modelled in the absence of data describing iron concentrations).

- Uncertainties in the algorithms chosen to represent processes. Murray and Parslow [1999a] explore the impact of the choice of algorithm for zooplankton mortality on model results for Port Phillip Bay.
- Uncertainty in the parameter values. Parameter ranges taken from the scientific literature are not always relevant to the conditions at a particular site, and calibration “by eye” is not mathematically optimal.

Uncertainty in process-based models often increases with complexity, as the various uncertainties in different parts of the model may combine in non-linear ways. The accuracy with which the model reproduces observational data provides a guide as to how much confidence should be placed in the results when the model is used to make predictions, although errors are likely to increase when the range of conditions to which the model is applied is extended.

As with formal parameter estimation techniques, formal sensitivity analyses to determine sensitivity to different parameter values or different algorithms are often problematic with models of this type. For example, a straightforward analysis to test second-order interactions in a model with 100 parameters, using 3 possible values for each parameter would require around 87,300 model runs, and even this would take into account only one source of uncertainty. Formal sensitivity analyses on more limited subsets of parameters are possible, but were not undertaken for the Swan-Canning or Fitzroy/Keppel Bay biogeochemical models.

1.10 Model evaluation

The success of a model must ultimately be assessed in terms of how well it fulfils the purpose specified in step 1. In this regard the key questions to be addressed include:

- *How well does the model reproduce an independent data set?*

This question was addressed in step 8. Both of the models discussed above performed reasonably well in reproducing observations, though with a relatively large margin of error.

- *How well does the model perform under unusual conditions?*

One advantage of a process-based model is that it may be able to predict responses to changes that bring the system outside its normal range under current conditions. A good test of a model is to apply it to a period outside the calibration period when the system is stressed or forced by unusual conditions. This opportunity arose for the Swan-Canning model with the occurrence of an extraordinary summer flow event in February 2000, which resulted in an unprecedented bloom of the freshwater cyanobacteria, *Microcystis aeruginosa* [Robson and Hamilton, 2004]. The model in its original form did not include a cyanobacterial phytoplankton group and so was not able to reproduce the bloom event, however it performed well with no other modifications when an additional phytoplankton group was added to allow simulation of *M. aeruginosa*.

- *Is the model better than a simpler model?*

This question is a particularly important one for mechanistic biogeochemical and ecological models, which are often complex and often demanding in terms of input data and resources. When a model is not performing as well as hoped, it is always tempting to include additional processes or ecosystem components, or to attempt to include more detail and complexity in the way existing processes are simulated. This modification is generally at a cost of greater complexity and data requirements, and there is also a risk of over-parameterisation (i.e. inclusion of so many calibrated parameters that the model can be fit to observed data despite inaccurate parameter values [Friedrichs et al., in press]). Furthermore, increasing the complexity of a model does not always improve its performance [Arhonditsis and Brett, 2004]. Fulton et al. [Fulton et al., 2003] show that the effectiveness of well designed ecological models in prediction may increase with

complexity to an intermediate level, but thereafter declines as complexity continues to increase. Perrin et al. [Perrin et al., 2001] consider this issue in mathematical terms, again concluding that an intermediate level of complexity is likely to have more predictive power. Arhonditsis [2004] reviewed 153 published aquatic biogeochemical modelling studies and found no improvement in model predictions with increased complexity. Murray [2001] considers the relationship between simple and more complex biogeochemical models from another point of view, illustrated by the use of a simple model of Port Phillip Bay to facilitate calibration and design of a more complex model.

The Swan-Canning model was found to be very sensitive to upstream boundary conditions [(Chan et al., 2002)]. Given this sensitivity, it is possible that a simpler model, such as the box model used by Robson and Hamilton [2003] may equally well achieve some of the modelling objectives. Nonetheless, our experience with the more complex, fully coupled model was invaluable in the design of the simpler model to address a specific question, as it highlighted the most important processes and parameter values.

In the case of the Fitzroy Estuary/Keppel Bay model, we were not able to achieve satisfactory results using a simple, depth-averaged box model rather than a three-dimensional model. Inclusion or omission of macroalgae and seagrasses, and inclusion or omission of nitrogen fixation by benthic microalgae, on the other hand, was found to make little difference to the simulation results. Such informal exploration of variations in model complexity are common, but are not often discussed in the final reporting of the model results; nor have they been discussed in detail here. The concise nature of journal papers generally does not allow for this type of detail to be expounded. More formal evaluation of the effects of different levels of model complexity in process-based biogeochemical and ecological modelling might well prove worthwhile.

- *Can we improve understanding of underlying system function?*

Models of most varieties allow improvements in understanding of system function by highlighting inter-relationships among variables. Mechanistic models are also a means of formalising and testing current understanding of the way a system functions. These models can be used to quantify the roles of different processes and system components, enabling questions to be addressed such as “how important is the sediment bed in controlling water column nutrient concentrations

in Keppel Bay and exports to the great Barrier Reef Lagoon?" In this case, the answer was "very important", according to the simulation output of the Fitzroy/Keppel Bay model [Robson et al., 2005]. An important question for the Swan-Canning Estuary was "were the unusual physical conditions of high temperature combined with low salinity associated with the February 2000 flow event more important than the associated nutrient influxes in triggering the 2000 *M. aeruginosa* bloom?" For this question, the answer was "resoundingly so", according to simulation output from the Swan-Canning model for this time period [Robson and Hamilton, 2003].

- *Finally, and most importantly, does the model help to answer the questions we set out to answer and make predictions about the future?*

Following the development of the two models discussed in this paper, each was applied to a series of scenarios designed to explore how the estuaries might behave under different conditions. Robson et al. [2005] describe how the Fitzroy/Keppel Bay was used to simulate the response of the Fitzroy Estuary and Keppel Bay to the changes in sediment and nutrient loads expected to occur if catchment land use changes from its present use (around 60% vegetation cover) to either 30% or 70% vegetation cover for grazing. Chan et al. [2002] describe the use of the Swan-Canning model to simulate possible responses to management changes aimed at reducing nitrogen loads to the Swan River. The model was also used to explore how the system might have changed since European settlement of the Swan Coastal Plain. In both case studies, the results are of direct relevance to ongoing management of the systems.

3. CONCLUSIONS

The "ten steps" are readily applicable to process-based aquatic biogeochemical modelling and a

development process that considers each of these steps in turn has the potential to improve modelling standards by enforcing an explicit consideration of the scientific and policy context of the model, achievable goals and alternative approaches as well as an honest appraisal of model performance in meeting goals that have been determined in advance.

4. Acknowledgements

Many organisations and individuals contributed funding, data, or collaborated in the two projects described. The Fitzroy Contaminants Project was funded by the Coastal Zone CRC. Hydrodynamic modelling was conducted by Mike Herzfeld and sediment modelling by Nugzar Margvelashvili. Phillip Ford (CSIRO Land and Water), Lynda Radke (Geoscience Australia), Andy Revill (CSIRO Marine and Atmospheric Research) and others were involved in collection and interpretation of field data. Some chlorophyll *a* data was provided by Miles Furnass at the Australian Institute of Marine Science.

The Swan-Canning Modelling Project was funded by the Western Australian Estuarine Research Federation and conducted while B.R. was a Research Associate at the Centre for Water Research (CWR), University of Western Australia. Data for the Swan-Canning project was provided by the Water and Rivers Commission (WRC), the Western Australian Environmental Protection Authority (now the Department of Environment and Conservation) and the Western Australian Department of Transport. Hydrodynamic modelling was conducted by Ben Hodges (then at CWR). Aditya Jha (WRC) assisted with formatting of field data and preparation of input files for scenarios. Jörg Imberger (CWR), Peta Kelsey (WRC) and Malcom Robb (WRC) provided critical commentary and useful discussions during model development.

5. References

- Arhonditsis, G. B. and Brett, M. T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modelling. *Marine Ecology Progress Series*, 271: 13-26.
- Baird, M. E., Walker, S. J., Wallace, B. B., Webster, I. T., and Parslow, J. S. 2003. The use of mechanistic descriptions of algal growth and zooplankton grazing in an estuarine eutrophication model. *Estuarine, Coastal and Shelf Science*, 56: 685-695.
- Brun, R., Reichert, P., and Künsch, H. R., 2001. Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, 37: 1015-1030.
- Chan, T. 2006. Phytoplankton dynamics in a seasonal estuary. University of Western Australia.
- Chan, T. U., Hamilton, D. P., Robson, B. J., Hodges, B. R., and Dallimore, C. J., 2002. Impacts of hydrological changes on phytoplankton succession in the Swan River, Western Australia. *Estuaries*, 25: 1406-1415.

- Friedrichs, M. A. M., Hood, R. R., and Wiggert, J. D., (in press). Ecosystem model complexity versus physical forcing: quantification of their relative impact with assimilated Arabian Sea data. *Deep Sea Research part II*.
- Fulton, E. A., Smith, A. D. M., and Johnson, C. R., 2003. Effect of complexity on marine ecosystem models. *Marine Ecology Progress Series*, 253: 1-16.
- Hamilton, D. and Herzfeld, M. 1999. Computational Ecosystem Dynamics Model (CAEDYM). An ecological water quality model designed for coupling with hydrodynamic drivers. Nedlands, WA, Centre for Water Research, University of Western Australia.
- Hamilton, D. P. and Turner, J. V. 2001. Integrating research and management for an urban estuarine system: the Swan-Canning Estuary, Western Australia. *Hydrological Processes*, 15: 2383-2385.
- Herzfeld, M., Andrewartha, J. R., Sakov, P., and Webster, I. 2005. Numerical Hydrodynamic Modelling of the Fitzroy Estuary. Australia, Cooperative Research Centre for the Coastal Zone, Estuary and Waterway Management.
- Hodges, B. R. and Imberger, J., 2001. Simple curvilinear method for numerical methods of open channels. *Journal of Hydraulic Engineering-ASCE*, 127: 949-958.
- Hornberger, G. M. and Spear, R. C., 1983. An approach to the analysis of behaviour and sensitivity in environmental systems: New York, Springer-Verlag, pp. 101-116.
- Jakeman, A.J, R.A. Letcher, and J.P. Norton, Ten iterative steps in development and evaluation of environmental models, *Environmental Modelling & Software*, 21(5), 602-614, 2006. *Environmental Modelling and Software*.
- Margvelashvili, N., Herzfeld, M., and Webster, I. 2005. Modelling of Fine Sediment Transport in Fitzroy Estuary and Keppel Bay. Australia, Cooperative Research Centre for the Coastal Zone, Estuary and Waterway Management.
- Mulligan, A. E., 1998. Genetic algorithm for calibrating water quality models. *Journal of Environmental Engineering*, 124: 202-211.
- Murray, A. G., 2001. The use of simple models in the design and calibration of a dynamic 2D model of a semi-enclosed Australian bay. *Ecological Modelling*, 136: 15-30.
- Murray, A. G. and Parslow, J. S., 1999a. The analysis of alternative formulations in a simple model of a coastal ecosystem. *Ecological Modelling*, 119: 149-166.
- Murray, A. G. and Parslow, J. S., 1999b. Modelling of nutrient impacts in Port Phillip Bay - a semi-enclosed marine Australian ecosystem. *Marine and Freshwater Research*, 50: 597-611.
- Murray, A. and Parslow, J. 1997. Port Phillip Bay Integrated Model: Final Report. Canberra, Australia, CSIRO Division of Marine Research.
- Ng, A. W. M. and Perera, B. J. C., 2003. Selection of genetic algorithm operators for river water quality model calibration. *Engineering Applications of Artificial Intelligence*, 16: 529-541.
- Perrin, C., Michel, C., and AndréAssian, V. , 2001. Does a Large Number of Parameters Enhance Model Performance? Comparative Assessment of Common Catchment Model Structures on 429 Catchments. *Journal of Hydrology*, 242: 275-301.
- Radke, L., Ford, P., Webster, I., Atkinson, I., and Oubelkheir, K. 2005. Keppel Bay: Physical Processes and Biogeochemical Function. Australia, Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management.
- Raymont, J. E. G., 1980. Plankton and productivity in the oceans. Volume 1 Phytoplankton: Oxford, Pergamon Press.
- Robson, B. J. and Hamilton, D. P., 2003. Summer flow event induces a cyanobacterial bloom in a seasonal Western Australian estuary. *Marine and Freshwater Research*, 54: 139-151.
- Robson, B. J. and Hamilton, D. P., 2004. Three-dimensional modelling of a Microcystis bloom event in the Swan River estuary, Western Australia. *Ecological Modelling*, 174: 203-222.
- Robson, B. J., Webster, I. T., and Rosebrock, U. 2005. Biogeochemical Modelling and Nitrogen Budgets for the Fitzroy Estuary and Keppel Bay. Australia, Cooperative Research Centre for the Coastal Zone, Estuary and Waterway Management.