Available online at www.sciencedirect.com
ScienceDirect
Environmental Modelling & Software 23 (2008) 369–384
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

Ten steps applied to development and evaluation of process-based biogeochemical models of estuaries

Barbara J. Robson [a,*], David P. Hamilton [b], Ian T. Webster [a], Terence Chan [c]

[a] CSIRO Land and Water, GPO Box 1666, Canberra ACT 2601, Australia
[b] Centre for Biodiversity and Ecology Research, University of Waikato, Private Bag 3105, Hamilton, New Zealand
[c] Water Studies Centre, Faculty of Science, Monash University, Clayton, Victoria 3800, Australia

Received 6 September 2006; received in revised form 4 May 2007; accepted 6 May 2007
Available online 19 September 2007

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. [Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling and Software 21, 602–614], is often iterative as model development is a continuous process that refines and improves the intended capacity of the model. Here, the ten steps of model development are critiqued and applied using a process-based biogeochemical model of aquatic systems, with 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

Software availability

Contact address for: ELCOM-CAEDYM, Centre for Water Research, University of Western Australia, http://www2.cwr.uwa.edu.au/~ttfadmin/model/elcomcaedym/
Program language: Fortran 95.
Availability: Executable file available as freeware, source code through collaborative arrangements since 2000.

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.

Two case studies are considered: biogeochemical modelling of the Swan-Canning Estuary and of Fitzroy Estuary and Keppel Bay. It is not the intention here to fully describe the models and algorithms used or the conclusions drawn from the modelling results. For these details, the reader is referred to published reports (Hamilton and Herzfeld, 1999; Robson et al., 2006a,b; Herzfeld et al., 2006) and papers (Chan et al., 2002, 2003; Robson and Hamilton, 2004; Bruce et al., 2006), with further papers currently in preparation. Rather, the purpose here is to explore how well the Jakeman et al.'s ''ten steps'' approach relates to this type of modelling, and how well mechanistic models stand up to an evaluation using the ten steps.

* Corresponding author. Tel.: +61 2 6246 5614; fax: +61 2 6246 5560.
E-mail address: barbara.robson@csiro.au (B.J. Robson).

1364-8152/$ - see front matter
doi:10.1016/j.envsoft.2007.05.019

2. Ten steps

Each of the following sections (Sections 2.1–2.10) corresponds to one of the ten steps defined by Jakeman et al. (2006).

2.1. Define model purpose

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. The present article draws 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 was a multidisciplinary, multiorganisation project conducted as one of the focal studies of the Cooperative Research Centre for Coastal Zone, Estuary & Waterway Management. 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, the discussion will be confined to the development and application of the biogeochemical model.

The purpose of the model, 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 is 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 land 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.

2.2. Specify modelling context: scope and resources

2.2.1. 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 3 years. They included modelling teams of hydrodynamic (and in the case of the Fitzroy, sediment dynamics) 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 and incorporated the three-dimensional hydrodynamic model SHOC (CSIRO Marine and Atmospheric Research), 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, ELCOM.

The Swan-Canning project drew on an extensive data set from an ongoing weekly monitoring programme 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 data set 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.

2.2.2. What outputs are required?

To achieve the goals outlined in Section 2.1, it is necessary first to 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 of 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.

2.2.3. 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, it was desirable to have the ability to conduct scenarios to predict responses on a scale of 1–3 years.

2.2.4. What is the relevant spatial scale?

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

In the Swan-Canning project, the domain included the estuary from its upstream tidal limit to the mouth 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 region.

2.2.5. 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 expert modellers. For both the Fitzroy and Swan-Canning modelling projects, the models were designed to be used by experienced modellers only. The models were designed 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.

2.3. Conceptualise the system

Conceptualisation of an estuary usually begins with flows of water. Freshwater flows into the estuary from rivers, drains and groundwater, and is mixed with salty water from the ocean. Exchanges of water also involve exchanges of nutrients and sediments. At the interface, freshwater tends to flow over salt water because it is less dense. Other important physical factors include surface elevation changes due to tides; mixing and transport of water within the estuary, settling and resuspension of particulates, 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 the physical domain of the estuary numerous chemical and biological processes occur. Fig. 1 illustrates just a few of the flora, fauna 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 Fig. 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.

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

For the applications described in this study, physical and chemical data to help define biogeochemical models were obtained 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).

2.4. Select model features (form of model)

2.4.1. 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 is trained, however, the predictions are unlikely to be valid. Furthermore, statistical models are not usually 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 resource variables.

Bayesian network 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 a quantitative estimate of uncertainty as part of their predictions, and can be flexibly altered to take into account input from stakeholders. Bayesian network models, however, do not directly incorporate a biophysical understanding of basic physical processes within the system, although higher-level relationships resulting from these underlying processes are represented. For example, it is known that dissolved inorganic nutrient concentrations in an aquatic system affect the likelihood of algal

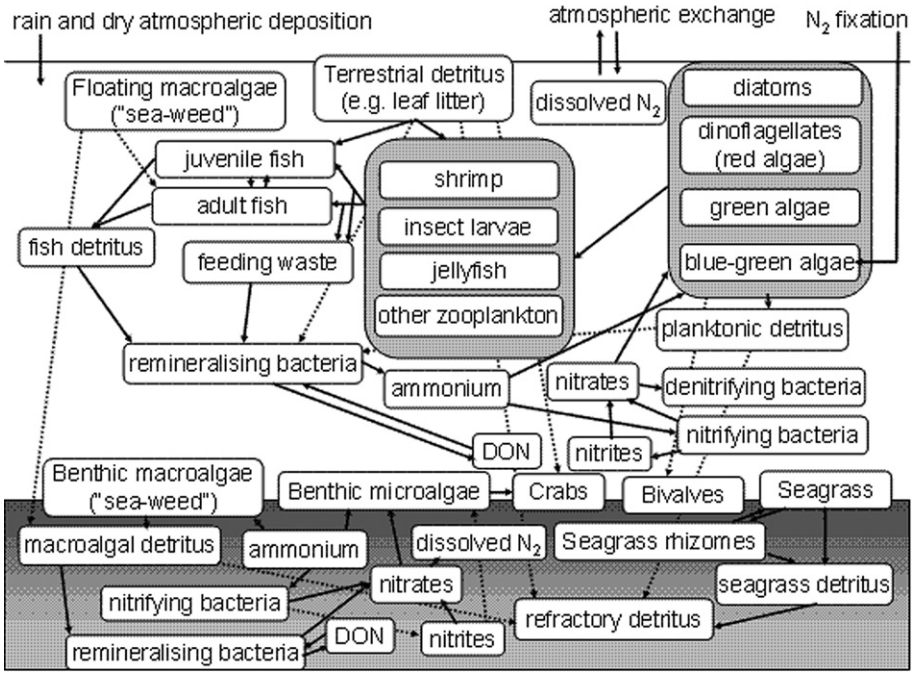

Fig. 1. Conceptual diagram illustrating some of the nitrogen stores and pathways in an estuary (not all links shown). Arrows represent flows of material from one nutrient store or ecosystem component to another. Some arrows are shown with dashed lines simply for visual clarity. ''DON'' refers to dissolved organic nitrogen. Phytoplanktons (diatoms, dinoflagellates, green and blue-green algae) are grouped together, as are zooplankton groups (shrimp, insect larvae, jellyfish and other zooplankton). Sediment stores are shown in the darker section at the bottom of the diagram.

blooms. Where a Bayesian model may include a relationship between observed inorganic nutrient concentrations and probability of an algal bloom, a mechanistic model will instead include a representation of the process of uptake of nitrogen from the water column by phytoplankton cells, allowing the higher-level relationship to emerge from process relationships represented in the model.

Process-based (or ''mechanistic'') biogeochemical models have a few key characteristics that make them particularly suitable for the two applications described here:

i. They explicitly represent understanding and functionality of the system, that is, not only its responses, but also its internal dynamics, and allow this understanding to be tested.
ii. They allow for a detailed, quantitative simulation of the current behaviour of the system.
iii. They provide a means to predict responses to changes, even (with caution) when those changes take the system beyond its measured historical variability.

Process-based models also have disadvantages: they tend to have high data input requirements and a high level of complexity, and they may have high computational costs, all of which can make it 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 (nonagent-based) approach was taken.

2.4.2. 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 Fig. 1 was in each case simplified to include only those processes that were considered to be most relevant to the desired outputs, on the basis of previous scientific literature describing the estuaries in question and similar estuaries, as well as past experience in biogeochemical modelling of estuaries.

Major nitrogen pathways included in the Fitzroy model are shown in Fig. 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 (remineralisation 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).

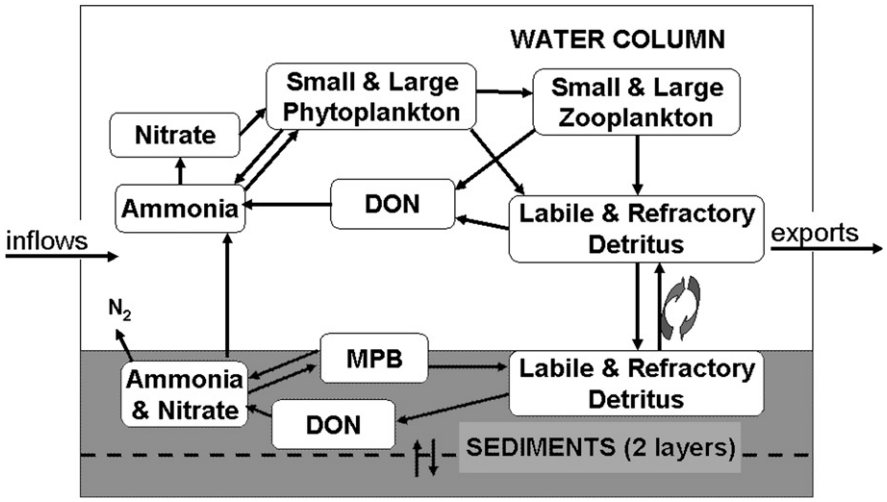

Fig. 2. Major nitrogen pathways and stores in the Fitzroy/Keppel Bay model (cf. Fig. 1.). Groups of biota represented in the model include small phytoplankton, large phytoplankton, small zooplankton, large zooplankton and benthic microalgae (MPB). Other nitrogen stores included in the model are ammonium, nitrate, dissolved organic nitrogen (DON), labile detritus and refractory detritus. Each living and nonliving 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.

In the case of the Swan-Canning project, a similar stock and flow model was developed. Zooplankton was not explicitly simulated, but grazing by zooplankton was represented in a more abstract sense as a component of losses of phytoplankton biomass. Benthic microalgae was not considered important in the upper estuarine reaches, and hence was not included in the model (Robson and Hamilton, 2004). Although there is reasonable microphytobenthic activity in the lower estuary (Masini and McComb, 2001), the Swan-Canning estuary is microtidal and therefore does not have extensive intertidal mudflats like those of the Fitzroy Estuary, where benthic microalgae might be expected to dominate.

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 in 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

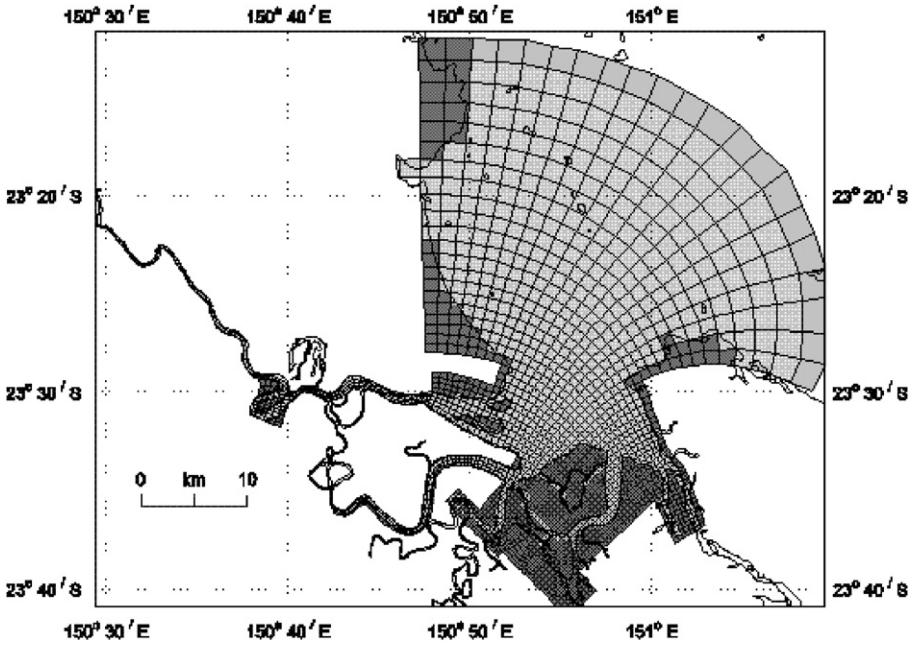

Fig. 3. Two-dimensional view (looking down from above) of the three-dimensional model grid used for the Fitzroy/Keppel Bay model (from Herzfeld et al., 2006). Green cells represent land and grey cells represent the ocean boundary, where Keppel Bay meets the Great Barrier Reef Lagoon. Axis labels show latitude and longitude in degrees and minutes.

have been appropriate for a model of the Fitzroy Estuary, in which the sediments are much more dynamic due to strong tidal currents and where there is such a large quantity of the total nitrogen store 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 Section 2.1) was to predict phytoplankton succession and blooms. The Swan-Canning model therefore included four distinct taxonomic groups of phytoplankton (compared with the two size classes simulated in the Fitzroy Estuary and shown in Fig. 2.).

For the Swan-Canning model, the biogeochemical model, CAEDYM (Hamilton and Herzfeld, 1999), was applied, and modified as necessary. CAEDYM is particularly suited for this application because it allows representation of multiple phytoplankton groups, differentiated by growth rates, zooplankton food preferences, ability to fix nitrogen (i.e. cyanobacteria), and responses to variations in temperature, salinity, and inorganic nitrogen and phosphorus concentrations. Many other biogeochemical models omit one or more of these features – particularly the salinity responses which have been shown to be important in the Swan-Canning Estuary (Chan and Hamilton, 2001).

For the Fitzroy/Keppel Bay model, the CSIRO Environmental Modelling Suite (EMS) was applied, again with some minor customisation. This model was appropriate to this application because it included a detailed representation of sediment dynamics and interactions between sediments, nitrogen and phosphorus in a varying-salinity environment. For the highly turbid Fitzroy Estuary, these features were deemed essential but they are often not included in biogeochemical models.

2.4.3. 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 freshwater extends over the surface from the mouth of Fitzroy Estuary and into Keppel Bay.

Although two-dimensional models are appropriate for some systems (e.g. Ebrahimi et al., 2007), the mixing and transport dynamics in these two estuaries could be reproduced adequately only with a three-dimensional hydrodynamic model. For the Swan-Canning model, the three-dimensional hydrodynamic model, ELCOM (Hodges et al., 2000), was used, while for the Fitzroy/Keppel Bay model, the three-dimensional hydrodynamic model, SHOC (Herzfeld et al., 2005), was selected as the basic tool to simulate transport and mixing for the biogeochemical variables. In both cases the selection of a suitable hydrodynamic model among others, similar three-dimensional hydrodynamic models, was largely pragmatic: these models included all relevant (baroclinic and barotropic) processes, were available to the authors and were designed for application to estuaries and for coupling with appropriate biogeochemical models.

There are often tradeoffs between the timescale 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. The spatial resolution chosen for each model was the finest resolution possible within the computational and time constraints. 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 100 m × 100 m cells was found to be unacceptably slow. This (1000 m × 100 m × 0.6 m) resolution was not sufficiently fine to reproduce phytoplankton patch dynamics in the estuary over scales of tens of metres, but was sufficient to allow overall spatial patterns to be simulated.

For the Fitzroy/Keppel Bay model, a curvilinear grid (Fig. 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., 2006). This grid was chosen because it allowed adequate representation of both the 200 m wide Fitzroy Estuary and the broader Keppel Bay, and allowed wetting and drying of intertidal mudflat areas in the estuary mouth and along the shore. A grid that used coarser cells in outer Keppel Bay may have been sufficient for our biogeochemical modelling objectives, however, a separate component of the project was to examine in detail the hydrodynamics of the system (Herzfeld et al., 2006), and this would have been compromised by a coarser grid.

For numerical stability, hydrodynamic models on these scales require integration with time-steps on the order of 20 s.

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 1-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 1-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 that was achievable.

2.5. Determine how model structure and parameter values are to be found

In both case studies, the biogeochemical model was embedded within a hydrodynamic model (and in the case of the Fitzroy/Keppel Bay study, within a sediment dynamic model). The models were structured 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.

The aim wherever possible was to have functional algorithms for each process, as well as parameter values, based on biophysical understanding of how the estuaries function, 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, a physiologically realistic mechanistic model was employed for nutrient uptake and light interception by phytoplankton cells (Baird et al., 2003). In the case of the Swan-Canning model, phytoplankton was differentiated not by cell size but by taxonomic grouping, and the use of more abstract Michaelis–Menten kinetics in rate equations for nutrient uptake and growth allowed the use of extensive literature from 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 observations of growth rates of *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 physiological processes, many parameters will be defined only within quite broad ranges. For example, observed half-saturation constants for uptake of nitrate by diatoms range from 0.4 to 5.1 $\mu g\,L^{-1}$ (Raymont, 1980). Parameter values were therefore calibrated within these ranges to produce values appropriate to the study sites.

2.6. 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, the objective was to develop a model that would allow improved understanding of the system dynamics and give predictions of 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 included 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 and blooms, as well as nutrient concentrations. Minimum performance criteria therefore included the ability to predict which phytoplankton group was dominant at any given time, the approximate timing and magnitude of phytoplankton blooms, and approximate concentrations of total and dissolved inorganic nitrogen, and dissolved oxygen. 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 nonbloom 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 for the model to use 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 systems; that the models correctly reproduced the observed *ranges* of nutrient and phytoplankton concentrations; that the simulated *median* concentrations were close to the observed medians and that there was successful reproduction of 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 the models were able to reproduce the approximate *timing of major events* (such as the occurrence of dinoflagellate blooms in the upper Swan Estuary and the periods 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.

2.7. 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, 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, as well as hydrodynamics and sediment dynamics. The algorithms used to define each of these processes are described by Robson et al. (2006b) and Murray and Parslow (1997).

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

Nutrient and sediment concentrations in the water column and bed sediments in the model of 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 dry season, when concentrations were relatively constant. The model was then run from September 2003 through to the end of February 2005. This extended period allowed a run-up time of almost 1 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 (from a second dry season field campaign in August 2004, and a wet season campaign in February 2005).

Water column nutrient and phytoplankton concentrations for the Swan-Canning model were initialised with observations from nine sites in December 2004. The model was then run over a 5-year simulated period, over which weekly monitoring data allowed an almost continuous comparison of observations against 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 may initially be perceived to be mathematically over-parameterised, 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. Parameter estimation techniques that do not merely find ''the best'' set of parameter values that fit the model to the data, but instead seek to determine *all* sets of parameter values that adequately fit the data (e.g. Malve et al., 2007) provide a means to test the specificity of the model and estimate the uncertainty inherent in scenario results.

Formal calibration and parameter estimation procedures such as Monte Carlo optimisation and variants such as the Hornberger–Spear approach (Hornberger and Spear, 1983), which identifies a range of parameters yielding acceptable results, 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 highly temporally and spatially resolved models. As mentioned previously, each 1-year model run took 1–2 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 observations at seasonal timescales. 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 trial and error 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 (1–3 month) simulations. The number of calibration runs required for the Swan-Canning model was not recorded, but probably ran to several hundred, 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. Fig. 4 shows a comparison for the Swan-Canning estuary of concentrations of each of four phytoplankton groups from field observations during 1995, against concentrations simulated with the calibrated model.

More rigorous approaches to sensitivity analysis and parameter estimation for complex process-based simulation models have been proposed (Brun et al., 2001), but still require a relatively large number of simulations, bearing in mind that both estuarine 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 calibration and refinement of algorithms can sometimes be performed using the simpler model (Murray, 2001). This procedure was attempted with the Fitzroy/Keppel Bay model, using a much lower resolution for initial calibration than that used for the final model. 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.

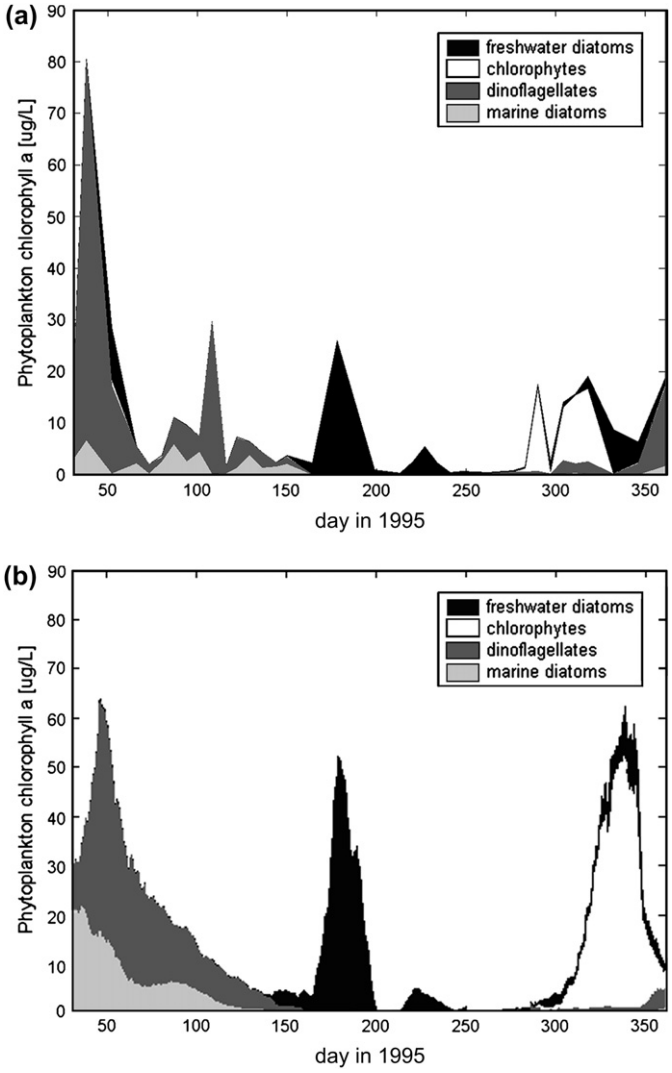

Fig. 4. Chlorophyll *a* concentrations in the upper Swan-Canning estuary in 1995, averaged over the six upstream sampling sites. Total chlorophyll *a* is given by the total height of the shaded areas; colours indicate contributions due to different phytoplankton groups; (a) in the field and (b) as simulated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

2.8. Verification and diagnostic testing

In the case of the Swan-Canning modelling project, the majority of calibration was conducted using observations for 1995, with additional fine-tuning using 1996 observations. Verification of the model was the undertaken with observational data from 1997. Validation of the Swan-Canning model against weekly surface and near-bed observations of salinity, temperature, dissolved oxygen, total nitrogen, total phosphorus, nitrate, phosphate, and ammonium concentrations at nine sites in the estuary is described in detail by Chan (2006) and will not be discussed further here. Instead, the focus here will be on phytoplankton succession both spatially and temporally. Temporal comparisons of estimated field concentrations with simulated concentrations of four groups of phytoplankton in 1995 and 1997 are reproduced from Chan (2006) in Figs. 4 and 5. The model simulations captured the observed phytoplankton succession well for the calibration year (1995), correctly reproducing which of the four phytoplankton groups dominated 67% of the time and predicting whether a marine or freshwater species would dominate 89% of the time, but less well for the verification year (1997). In 1997, the model simulations correctly predicted the dominant phytoplankton group only 27% of the time with consistent overprediction of marine diatom populations throughout the year (Fig. 5), but predicted whether a marine or freshwater species would dominate 76% of the time. Omitting marine diatoms from the analysis (not shown), the model correctly predicted which of the three remaining groups was present in greatest abundance 80% of the time.

Manipulation of parameter values defining marine diatom growth and mortality rates did not improve the simulation.

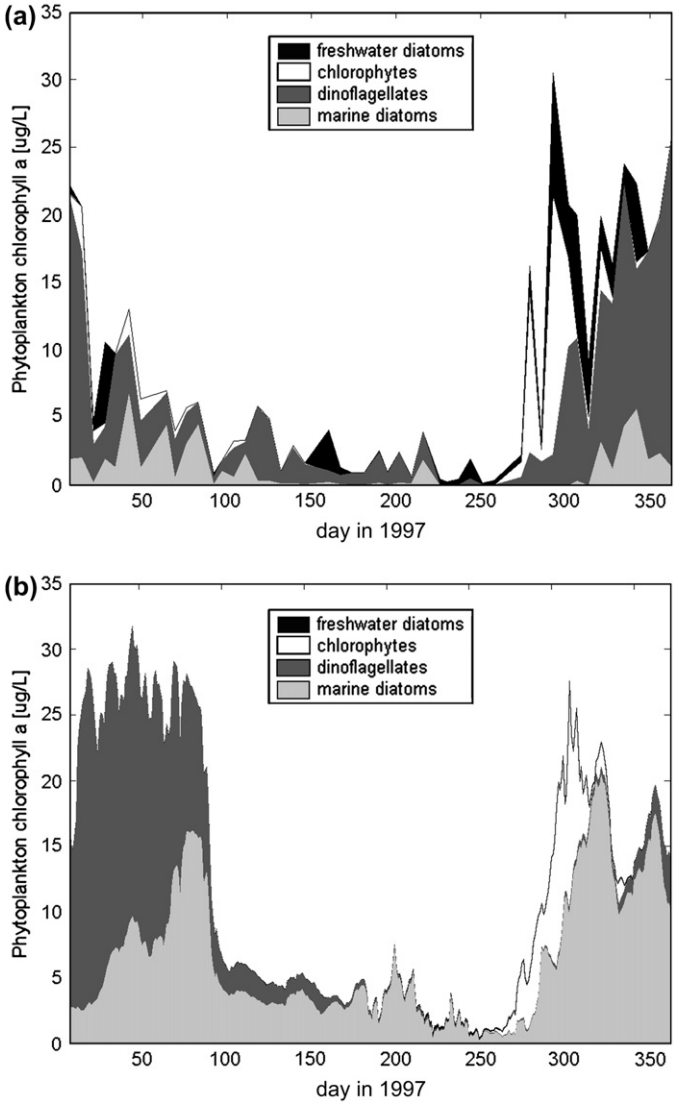

Fig. 5. Chlorophyll *a* concentrations in the upper estuary in 1997, averaged over the six upstream sites in the Swan River. Total chlorophyll *a* is given by the total height of the shaded areas; colours indicate contributions due to different phytoplankton groups; (a) in the field and (b) as simulated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

Simulation of marine diatoms might be improved with a better representation of zooplankton grazing, or perhaps with a better specification of marine diatom concentrations at the seaward boundary.

The fraction of variation (r^2) explained by model simulations for chlorophyll *a* in the upper estuary is low in both years (0.24 for 1995 and 0.22 for 1997), but this is not surprising given the large uncertainties inherent in the input and observational data, particularly the need to convert between phytoplankton cell counts to equivalent chlorophyll *a* content for each of the modelled phytoplankton groups. Among other sources of error discussed by Chan (2006) was the limited temporal resolution of data to specify seaward boundary conditions. The model provided a basis, however, to distinguish between "bloom" and "no bloom" conditions, where "bloom" conditions are defined as chlorophyll *a* concentration $>10\ \mu g\ L^{-1}$ at the time of observation. In the validation year 1997, the model simulations correctly predicted bloom conditions with a probability of detection of 0.87 and a false alarm ratio of 0.18. For 1995, the equivalent figures were 0.64 and 0.19, respectively. In the case of the Fitzroy Contaminants project, model calibration was primarily conducted against observational data from the August 2004 (dry season) field campaign, leaving 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 (i.e. labile) than "old" (i.e. refractory) detrital material within the estuary. Hence, there was no completely independent verification data set 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. (2006b). Spatial comparisons of dry season total nitrogen, dissolved inorganic nitrogen and chlorophyll *a* are reproduced in Figs. 6, 7 and 8, respectively.

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. Arhonditsis and Brett (2004) reviewed the performance of 153 published mechanistic aquatic biogeochemical modelling studies. For the purposes of the current discussion, "reasonable performance" will be defined as a coefficient of determination (r^2) better than the 40th percentile of these studies. By this measure, the Fitzroy/Keppel Bay model achieved reasonable spatial agreement for dissolved inorganic nutrients: $r^2 = 0.49$ for DIN (Arhonditsis and Brett report 40th percentiles averaging 0.42 for the DIN components ammonium and nitrate) and $r^2 = 0.37$ for DIP (Arhonditsis and Brett report a 40th percentile of 0.30 for phosphate). The coefficient of determination for DON in the Fitzroy/Keppel Bay model was 0.58. Arhonditsis and Brett do not provide statistics for comparison for dissolved organic nutrients.

Reproduction of spatial variability of dissolved oxygen (DO) was poor by the coefficient of determination measure ($r^2 = 0.41$, compared with a 40th percentile value of 0.62 reported by Arhonditsis and Brett), however, DO was not a primary focus of the simulations. Of more concern was the level of agreement between model simulations and observations for particulate materials, with an r^2 of 0.37 for TN (with a 35% relative error) and an (poor) r^2 of 0.19 for TP (with an 83% relative error). Arhonditsis and Brett do not provide statistics for comparisons for TN or TP.

In general, Robson et al. (2006b) concluded that simulation of dry season concentrations of nitrogen and phosphorus (as DON, DOP, DIN, DIP, PN and PP) was satisfactory. During the wet season, the model underestimated particulate nitrogen and dissolved organic nitrogen and phosphorus. Analysis of

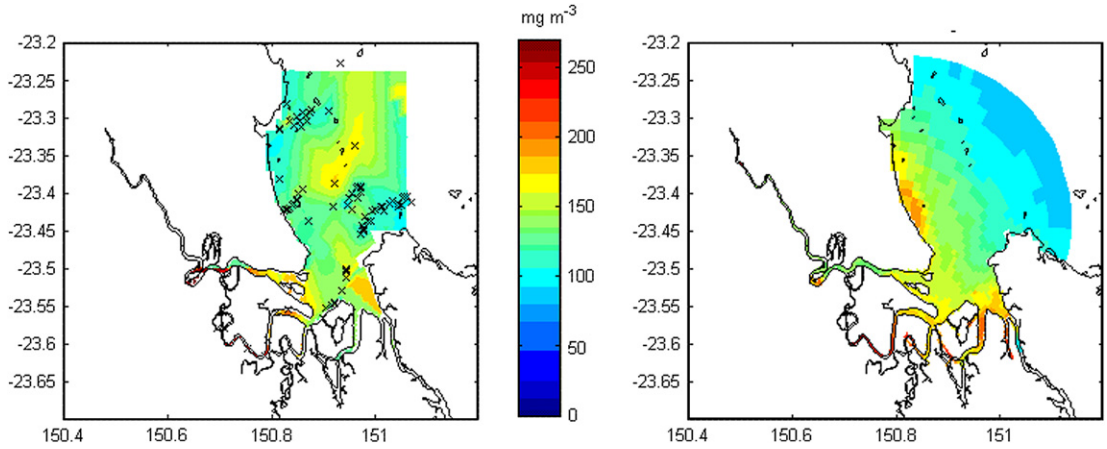

Fig. 6. Field data (left) and model simulation output (right) of total nitrogen concentrations in August 2004 (dry season) in the Fitzroy Estuary and Keppel Bay. Crosses indicate tidally corrected locations of field observations – contouring of concentrations beyond the spatial limits of the crosses is based on extrapolation of field data, and is unreliable. Axis labels show latitude and longitude in degrees.

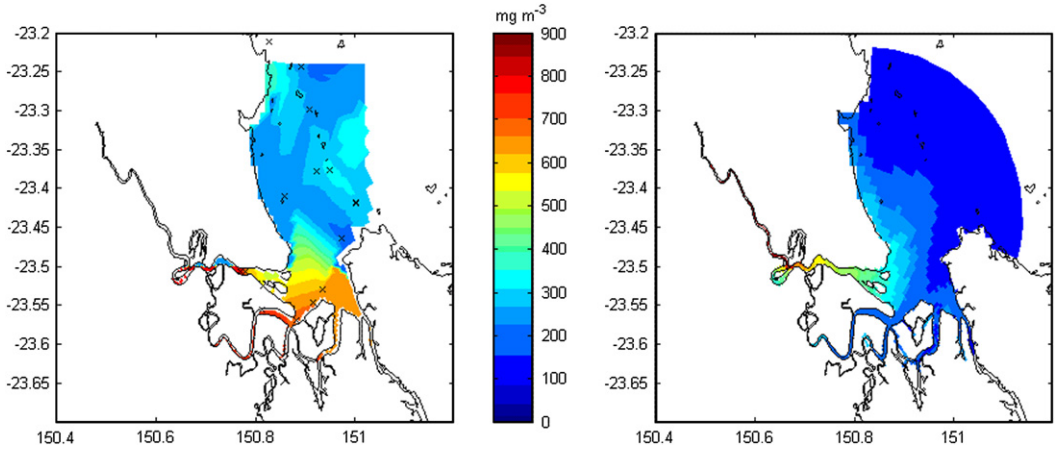

Fig. 7. Dissolved inorganic nitrogen concentrations from field data (left) and from model output (right), for August 2004 (dry season) in Fitzroy Estuary and Keppel Bay. Crosses indicate tidally corrected locations of field observations – contouring of concentrations beyond the spatial limits of the crosses is based on extrapolation of field data, and is unreliable. Axis labels show latitude and longitude in degrees. Concentrations of up to 40 $\mu g\,L^{-1}$ occur in the loop section of the estuary in both the model and field observations.

sediment modelling results (Margvelashvili et al., 2005) showed that suspended solids' concentrations were underestimated during high flow events. Possible reasons discussed by Margvelashvili et al. (2005) and Robson et al. (2006b) include insufficient data to prescribe the inflow boundary condition (i.e., concentrations and characteristics of sediments and nutrients in freshwater inflows during flood events) and gaps in understanding of sediment processes.

The r^2 results are generally slightly below the medians reported by Arhonditsis and Brett (2004) for aquatic biogeochemical models. This is not surprising given the strong tidal variability of the study site, the number of state variables simulated, and the length of the simulation. Arhonditsis and Brett (2004) report a negative correlation between simulation period and model performance in published studies, with the best predictability obtained for models with simulation periods of up to several days. Few if any of the modelling studies included in the review were applied to turbid macrotidal systems such as the Fitzroy Estuary. Total suspended solids concentrations at some sites in Keppel Bay were observed to vary by two orders of magnitude within the space of a few hours, due to tidal advection and resuspension (Margvelashvili et al., 2005), and the hydrodynamic model indicates that parcels of water in parts of Keppel Bay move by up to 20 km between low and high tide (i.e. the tidal excursion in some parts of Keppel Bay is around 20 km) during spring tide events. Nutrient concentrations in this system are strongly tied to transport and degradation of suspended particulate materials (Robson et al., 2006b), so accurate representation of spatial and temporal variations in nutrient concentrations depends on accurate simulation of hydrodynamics and suspended sediments.

2.9. Quantification of uncertainty

Quantification of uncertainty was perhaps the weakest component of the ten steps for both the Swan-Canning and

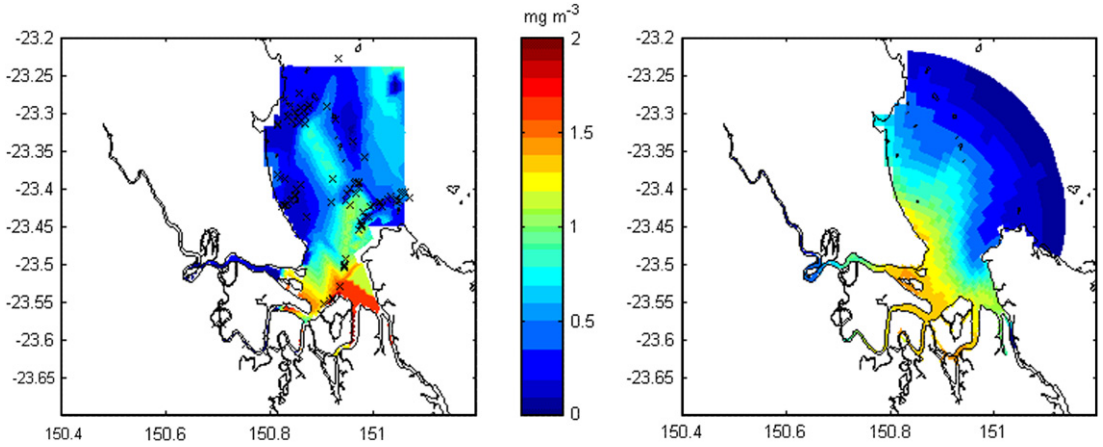

Fig. 8. Chlorophyll *a* concentrations from field data (left) and model output (right), for August 2004 (dry season) in Fitzroy estuary and Keppel Bay, omitting a suspect point at (151.0280°W, 23.3471°S). Axis labels show latitude and longitude in degrees. Crosses indicate tidally corrected locations of field observations – contouring of concentrations beyond the spatial limits of the crosses is based on extrapolation of field data, and is unreliable. The localised high chlorophyll *a* concentrations in the lower right of the field data diagram (left) are attributable to extrapolation from a single data point.

Fitzroy/Keppel Bay modelling projects. Uncertainty in the predictions of both models is relatively high, but remains unquantified. Nitrogen and sediment budgets calculated using the Fitzroy/Keppel Bay model are probably accurate to within a factor of two (Robson et al., 2006b), while the timing of the onset of phytoplankton blooms predicted by the Swan/Canning model is generally accurate to within 1 month, but is heavily dependent on upstream boundary conditions. This level of accuracy is likely sufficient to assist with assessing effects of long-term catchment planning, but not sufficient to aid in management of bloom conditions as they occur.

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 for specifying boundary conditions and for verification, missing data, and uncertain conversions (e.g., conversions between cell counts and chlorophyll *a* concentrations 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*: this affects the accuracy of hydrodynamic modelling and limits the ability of the model to reproduce small-scale patchiness in chlorophyll *a* and nutrient concentrations.
- *Uncertainty in the choice of which processes to include and which to omit*: processes included in the models were chosen on the basis of biophysical understanding each estuary at the start of model development when there was limited information (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 algorithms chosen to represent processes*: Murray and Parslow (1999a) explore the impact of the choice of algorithm for zooplankton mortality on simulation of phytoplankton and nutrient concentrations in Port Phillip Bay.
- *Uncertainty in parameter values*: parameter ranges taken from the scientific literature are not always relevant to the conditions at a particular site, and calibration "by eye" does not lead to reproducibility of parameter values and 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 nonlinear 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 complex, spatially resolved, process-based models. The computational cost of sensitivity analyses tends to increase exponentially, rather than linearly, with the number of parameters. Both the models described here employ >100 parameters and the heavy computational demands of the models make a complete assessment of parameter sensitivity on the timescale of these simulations impossible. A sensitivity assessment over a much shorter timescale would not have been particularly relevant when most variability in the state variables occurred at seasonal or interannual timescales. 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. The models were not structured to allow subdivision into submodels for partial sensitivity assessment, though this is an area of research that is likely to yield a more complete understanding and predictive capability for biogeochemical models.

Despite the lack of formal sensitivity assessment, calibration of the models involved informal assessment of model sensitivity to various parameters and boundary conditions, adjusting individual parameters to provide improvements to the visual comparison of model simulations against field data.

The Swan-Canning model was found to be sensitive to several phytoplankton physiological parameters (optimum and maximum salinity, nitrogen half-saturation constants and maximum growth rates) and to parameters affecting sediment nutrient releases (maximum nutrient release rates and parameters influencing dependence of release rates on dissolved oxygen), suggesting that the simplistic representation of sediments in this model may not be justified and the model's response to alternative sediment nutrient algorithms (such as are now available in newer versions of CAEDYM) should be explored. An earlier exploration using a mechanistic approach to describe sediment oxygen demand in the Swan River (Herzfeld et al., 2001), however, revealed the complexities of dealing with rapid biogeochemical transitions associated with redox fronts at the sediment–water boundary.

The Swan-Canning model was found to be very sensitive to upstream boundary conditions (Chan et al., 2002). Despite weekly resolution of measurements at the upstream boundary there was considerable variability at the upstream boundary and also uncertainty in converting from observed phytoplankton cell counts to the amount of chlorophyll *a* contained in each phytoplankton group. In future modelling efforts for the Swan-Canning estuary, moving the present upstream boundary further upstream into Avon River (where phytoplankton concentrations are lower) may reduce the importance of this source of error.

The Fitzroy/Keppel Bay model was found to be sensitive to remineralisation rates of dissolved and particulate organic material, light extinction coefficients of dissolved and particulate substances in the water column, and concentrations of dissolved organic nutrients at the open boundary. This sensitivity highlighted the importance of particulate organic material in the nutrient cycle in this system, as well as the interplay between light-limited primary production in the turbid estuary

mouth and nutrient limited production in the clear waters of outer Keppel Bay. Work is currently underway to improve specification of the open boundary condition through assimilation of remote sensing data to provide greater temporal and spatial resolution of total suspended solids, chlorophyll *a* and coloured dissolved organic matter.

The sensitivity of the Fitzroy/Keppel Bay model to inclusion or omission of some model components was also considered, and will be discussed briefly in the following section.

2.10. Model evaluation

The success of a model must ultimately be assessed in terms of how well it fulfils the purpose specified in Section 2.1, i.e. the original objectives and purpose of the modelling. There appears to be some overlap between this step in the 10 steps framework of Jakeman et al. (2006) and Section 2.8 (verification and testing, which is also part of evaluation). However, a complete model evaluation goes beyond simply testing its efficacy in reproducing or predicting field conditions and should address the question of how *useful* the model is and how well it fulfils the purpose for which it was developed. Questions to consider include the following.

2.10.1. How well does the model reproduce an independent data set?

This question was addressed in Section 2.8. The Swan-Canning model was tested with data from 1997 and it was shown that the simulations captured the timing of phytoplankton blooms and provided useful information on phytoplankton species succession, but did not reproduce observed chlorophyll *a* concentrations as well as desired. The Fitzroy/Keppel Bay model performed reasonably well (by the standard defined in Section 2.8) in reproducing observed dissolved inorganic nutrient concentrations, but less well in reproducing particulate nutrient concentrations. Due to limited availability of additional data, this model has not been tested against an independent data set.

2.10.2. 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 in a system that are outside of those captured in the initial hindcasting simulations. A good test of the performance 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 cyanobacterium, *M. aeruginosa* (Robson and Hamilton, 2004). The model in its original form did not include a cyanobacterial phytoplankton group and was therefore 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* (see Robson and Hamilton, 2004).

2.10.3. Is the complex model better than a simpler one?

This question is a particularly important one for mechanistic biogeochemical and ecological models, which are often both complex and demanding in terms of input data and resources. When a model is not performing as well as anticipated, it is tempting to include additional processes or ecosystem components, or to include more detail and complexity in the way existing processes are simulated. This modification is generally at a cost of greater complexity and increased data requirements, and there is also a risk of over-parameterisation (i.e. inclusion of so many calibrated parameters that the model can be fitted to observed data despite inaccurate parameter values (Friedrichs et al., 2006)). Furthermore, increasing the complexity of a model does not always improve its performance (Arhonditsis and Brett, 2004). Fulton et al. (2003) show that the effectiveness of well-designed ecological models may increase with complexity to an intermediate level, but thereafter declines as complexity continues to increase. 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. Arhonditisis (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, and showed how a simple model of Port Phillip Bay could be used to facilitate calibration and design of a more complex model. Although the simpler model was not a realistic model in itself, it could be used as an ‘‘analysis tool to [predict] the effects of changes to model formulation, parameter values and external forcing on the full model’’.

Given the sensitivity of the Swan-Canning model to upstream boundary conditions, it is likely that a simpler model, such as the box model used by Robson and Hamilton (2003), may equally well achieve many of the modelling objectives. Nonetheless, our experience with the more complex, fully coupled model was invaluable as it highlighted the most important processes, parameter values and interactions amongst state variables.

In the case of the Fitzroy Estuary/Keppel Bay model, a simple depth-averaged box model was tested as a possible substitute for the slower three-dimensional model, but the simpler model did not give satisfactory results. Inclusion or omission of macroalgae and seagrasses, or of nitrogen fixation by benthic microalgae, on the other hand, was found to make little difference to the simulation results. Such informal explorations 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. More formal evaluation of the effects of different levels of model complexity in process-based biogeochemical and ecological modelling might well prove worthwhile but reporting of these outcomes is often limited by the need for brevity in scientific papers and in reporting methods.

2.10.4. Can the model be used to improve understanding of underlying system function?

Most models allow improvements in understanding of system function by highlighting relationships amongst variables. In addition, mechanistic models are 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., 2006b). 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 *M. aeruginosa* bloom?" For this question, the answer was "resoundingly so", according to simulation output from the Swan-Canning model when applied to this time period (Robson and Hamilton, 2003).

2.10.5. Finally, and most importantly, does the model help to answer questions about the system function and can it be used to 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 respective estuaries might behave under different conditions. Robson et al. (2006a) describe how the Fitzroy/Keppel Bay model was used to simulate the response of the system to expected changes in sediment and nutrient loads (Dougall et al., 2006) if catchment land use changes from its present predominant form (around 60% vegetation cover) to (a) 30% and (b) 70% vegetation cover for pastoral grazing. Model simulations were also used to estimate total export of nitrogen and sediments across the outer boundary of Keppel Bay into the Great Barrier Reef Lagoon; scenarios which would not have been possible with, for example, a simple heuristic or statistical model relating nitrogen and sediment concentrations to flow and meteorological conditions.

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.

The relatively high uncertainty of model predictions may limit the confidence with which results can be used in making management decisions. High uncertainty comes at an economic cost, complicates decision-making processes and may suggest a need for a precautionary approach in management (Gollier and Treich, 2003), particularly when a valuable environmental resource is at stake.

The Swan-Canning model has not been fully successful in providing the information needed by managers, and the Western Australian Department of Water is currently planning development of a new hydrodynamic model of the Swan-Canning Estuary and will consider avenues for further biogeochemical modelling when this is complete (Malcolm Robb, personal communication).

The Fitzroy Basin Association (FBA), which manages the catchment of the Fitzroy Estuary, has welcomed the Fitzroy/Keppel Bay model and considers its output to be the best available information regarding the likely impact of changes in sediment and nutrient loads on in-stream conditions in the estuary and a useful indication of export of material to the Great Barrier Reef Lagoon. The FBA is currently planning to invest in this use of the model, together with the SedNet/ANNEX catchment model (Dougall et al., 2006), and an ecological risk assessment model that is currently in development, to evaluate a series of additional land management scenarios.

3. Conclusions

The "ten steps" of Jakeman et al. (2006) that underpin best practice in development of models for natural resource management are readily applicable to process-based aquatic biogeochemical modelling. 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, setting achievable goals and alternative approaches, as well as providing an honest appraisal of model performance in meeting goals that have been determined in advance.

This retrospective analysis of two case studies in terms of the "ten steps" approach to model development has helped to identify weaknesses in the modelling procedures applied. At the same time, this exercise invites a closer evaluation of the ten steps approach itself. System conceptualisation may be out of place at Section 2.3: on one hand, definition of modelling goals (Section 2.1) to some extent requires that an implicit conceptualisation of the system already exists. Conversely, the type of conceptual model employed and the level of detail incorporated into the system conceptualisation depends on the model approach, which is not selected until Section 2.4. While the mechanistic biogeochemical models discussed here required that the system be conceptualised in terms of stocks and flows of nutrients and the chemical processes controlling these flows, another modelling approach might conceptualise the system in terms of statistical relationships that might or might not be causal, or from the point of view of factors affecting individual agents within the system. In practice, system conceptualisation probably needs to be revisited at each step of the modelling procedure.

Section 2.7 lumps together two components that come together naturally in some modelling approaches, but not all. In the cases considered here, the model structure was largely fixed at Sections 2.4 and 2.5, with only limited and informal exploration of variations in this structure occurring subsequently. Process-based biogeochemical models of aquatic systems might benefit from more formal consideration of optimal structure, and perhaps breaking Section 2.7 into two parts would serve to encourage this.

Finally, Section 2.10 can be seen to have a number of components. The first and most straightforward component of model evaluation is verification and diagnostic testing, which is covered separately as Section 2.8. Additional questions to consider during this step, relevant particularly when a process-based model has been chosen, have been suggested in the subheadings of Section 2.10; namely:

- Can the model be used to improve understanding of underlying system function?
- Does the model help to answer specific questions about system function?
- Can it be used to make predictions about the future?
- Is it better than a simpler model of the system?; and
- How well does it perform under unusual conditions?

The two case studies discussed here are of complex, process-based biogeochemical models, but have been applied with different aims and therefore emphasise different aspects of the biogeochemical cycles simulated. Weaknesses of many of the current generation of complex mechanistic aquatic models, including those presented here, are that parameter estimation often does not follow a robust statistical process and uncertainties can be difficult to estimate. These uncertainties suggest a cautious approach in applications of these models outside of the domain for which they were calibrated and validated.

Because of the inherent complexity of biological systems, the quantitative performance of biogeochemical models of aquatic systems is not as good as that of models of the physics of these systems. The predictive capacity of mechanistic biogeochemical models for management purposes is therefore not up to the standard set by hydrodynamic models. Where prediction of nutrient concentrations or chlorophyll *a* within the range of historical variability is the primary goal, process-based biogeochemical models may not be the most appropriate modelling tool. Particular strengths of process-based models, however, are that they embody an understanding of system function (e.g. allowing a calculation of nitrogen fluxes from Keppel Bay to the ocean) and that they allow tentative predictions of responses to changes that may be beyond the range of variation in historical input data. Such predictions must always be used with caution: bioinvasions that radically alter or 'shift' ecosystems would confound biogeochemical predictions.

The current generation of process-based biogeochemical models could benefit from work aimed at improving estimations of uncertainty, identifying the most appropriate level of model complexity, and improving the flexibility of models to allow them to be broken into submodels or models of varying resolution to facilitate calibration and modern parameter estimation techniques. Wherever possible, model development should not be separated from data collection, but rather, coupled with targeted field-based process studies.

Both the estuarine model application case studies discussed here were successful in improving understanding of system function. Both models also allow alternative management scenarios to be simulated speculatively, providing quantitative output.

Acknowledgements

Many organisations and individuals contributed funding, data, or variously collaborated in the two projects described. The Fitzroy Contaminants Project was funded by the Coastal Zone CRC. Hydrodynamic modelling for this project 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 several other scientists were involved in collection and interpretation of field data. Some chlorophyll *a* data were 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 were provided by the Water and Rivers Commission (WRC), the Western Australian Environmental Protection Authority (now the Department of Water) 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.

We thank Tony Jakeman and Teemu Kokkonen for convening the workshop on ''Creative and Rigorous Model Development and Evaluation'' and Teemu Kokkonen for presenting this work on our behalf at the third Biennial Meeting of the International Environmental Modelling and Software Society (iEMSs) in Burlington, Vermont, USA in July 2006.

References

Arhonditsis, G.B., 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., 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.

Borsuk, M.E., Reichert, P., Peter, A., Schager, E., Burkhardt-Holm, P., 2006. Assessing the decline of brown trout (*Salmo trutta*) in Swiss rivers using a Bayesian probability network. Ecological Modelling 192, 224–244.

Bruce, L.C., Hamilton, D., Imberger, J., Gal, G., Gophen, M., Zohary, T., Hambright, K.D., 2006. A numerical simulation of the role of zooplankton in C, N and P cycling in Lake Kinneret, Israel. Ecological Modelling 193, 412–436.

Brun, R., Reichert, P., 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. Ph.D. thesis, University of Western Australia, Perth, Western Australia.

Chan, T.U., Hamilton, D.P., 2001. The effect of freshwater flow on the succession and biomass of phytoplankton in a seasonal estuary. Marine and Freshwater Research 52, 869–884.

Chan, T.U., Hamilton, D.P., Robson, B.J., 2003. Modelling phytoplankton succession and biomass in a seasonal West Australian estuary. Verhandlung Internationale Vereingung de Limnologie 28, 1086–1088.

Chan, T.U., Hamilton, D.P., Robson, B.J., Hodges, B.R., Dallimore, C.J., 2002. Impacts of hydrological changes on phytoplankton succession in the Swan River, Western Australia. Estuaries 25, 1406–1415.

Dougall, C., Packet, R., Carroll, C., Sherman, B.S., Read, A., Chen, Y., Brodie, J., 2006. Sediment and nutrient modelling in the Fitzroy NMR region. In: Cogle, A.L., Carroll, C., Sherman, B.S. (Eds.), The Use of SedNet and ANNEX Models to Guide GBR Catchment Sediment and Nutrient Target Setting, vol. 5. Department of Natural Resource, Mines and Water. QNMRM06138.

Ebrahimi, K., Falconer, R.A., Lin, B., 2007. Flow and solute fluxes in integrated wetland and coastal systems. Environmental Modelling and Software 22, 1337–1348.

Friedrichs, M.A.M., Hood, R.R., Wiggert, J.D., 2006. Ecosystem model complexity versus physical forcing: quantification of their relative impact with assimilated Arabian Sea data. Deep Sea Research Part II: Topical Studies in Oceanography 53, 576–600.

Fulton, E.A., Smith, A.D.M., Johnson, C.R., 2003. Effect of complexity on marine ecosystem models. Marine Ecology Progress Series 253, 1–16.

Gollier, C., Treich, N., 2003. Decision-making under scientific uncertainty: the economics of the precautionary principle. Journal of Risk and Uncertainty 27, 77–103.

Gribble, N.A., 2004. A spatially explicit multi-competitor coexistence model of Penaeid (shrimp) distribution on the Australian Great Barrier Reef. Ecological Modelling 177, 61–74.

Hamilton, D., Herzfeld, M., 1999. An ecological water quality model designed for coupling with hydrodynamic drivers. Scientific manual. Computational Ecosystem Dynamics Model (CAEDYM). Centre for Water Research, University of Western Australia, Nedlands, WA.

Hamilton, D.P., 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., Webster, I., 2006. Numerical hydrodynamic modelling of the Fitzroy Estuary, Australia. Cooperative Research Centre for the Coastal Zone, Estuary and Waterway Management, Technical Report No. 38

Herzfeld, M., Hamilton, D.P., Douglas, G.B., 2001. Comparison of a mechanistic sediment model and a water column model for hindcasting oxygen decay in benthic chambers. Ecological Modelling 136, 255–267.

Herzfeld, M., Waring, J., Parslow, J., Margvelashvili, N., Sakov, P., Andrewartha, J., 2005. SHOC: Sparse Hydrodynamic Ocean Code Scientific Manual. CSIRO Marine and Atmospheric Research, Hobart, Australia.

Hodges, B., Imberger, J., Saggio, A., Winters, K.B., 2000. Modeling basin-scale internal waves in stratified lakes. Limnology and Oceanography 45, 1603–1620.

Hodges, B.R., Imberger, J., 2001. Simple curvilinear method for numerical methods of open channels. Journal of Hydraulic Engineering – ASCE 127, 949–958.

Hornberger, G.M., Spear, R.C., 1983. An approach to the analysis of behaviour and sensitivity in environmental systems. In: Beck, M., van Straten, G. (Eds.), Uncertainty and Forecasting of Water Quality. Springer-Verlag, New York, pp. 101–116.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling and Software 21, 602–614.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling and Software 15, 101–124.

Malve, O., Laine, M., Haario, H., Kirkkala, T., Sarvala, J., 2007. Bayesian modelling of algal mass occurrences – using adaptive MCMC methods with a lake water quality model. Environmental Modelling and Software 22, 966–977.

Margvelashvili, N., Herzfeld, M., 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. Technical Report No. 39

Masini, R.J., McComb, A.J., 2001. Production by microphytobenthos in the Swan-Canning Estuary. Hydrological Processes 15, 2519–2535.

Mulligan, A.E., 1998. Genetic algorithm for calibrating water quality models. Journal of Environmental Engineering 124, 202–211.

Murray, A. Parslow, J., 1997. Port Phillip Bay integrated model: final report. Canberra, Australia, CSIRO Division of Marine Research. Technical Report No. 44.

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., 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., 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.

Ng, A.W.M., Perera, B.J.C., 2003. Selection of genetic algorithm operators for river water quality model calibration. Engineering Applications of Artificial Intelligence 16, 529–541.

Panzarasa, P., Jennings, N.R., Norman, T.J., 2001. Going public and the sale of shares with heterogeneous investors: agent-based computational modelling and computer simulations. Group Decision and Negotiation 10, 423–470.

Perrin, C., Michel, C., 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., Oubelkheir, K., 2005. Keppel Bay: physical processes and biogeochemical function. Australia, Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management. Technical Report No. 34.

Raymont, J.E.G., 1980. Plankton and Productivity in the Oceans. In: Phytoplankton, vol. 1. Permagon Press, Oxford.

Robson, B.J., 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., 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., Margvelashvili, N., Herzfeld, M., 2006a. Scenario modelling: simulating the downstream effects of changes in catchment land use. Indooroopilly, Qld, Australia, Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management. Technical Report No. 41.

Robson, B.J., Webster, I.T., Rosebrock, U., 2006b. Biogeochemical modelling and nitrogen budgets for the Fitzroy Estuary and Keppel Bay. Australia, Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management. Technical Report No. 40.