ANZIAM J. 48(2007), 461–473

USING MODIFIED PROPER ORTHOGONAL DECOMPOSITION (MPOD) FOR REDUCING ECOSYSTEM MODELS

J. LAWRIE[✉ 1] and J. W. HEARNE[1]

(Received 31 May, 2006; revised 21 March, 2007)

Abstract

In this paper we consider simplifying a model of the nitrogen cycle in Port Phillip Bay, Victoria, Australia. The approach taken is to aggregate state variables that are linearly related using a projection in state space. The technique involved is a modification of proper orthogonal decomposition and was developed so that a resulting simplified model retains an ecological interpretation. It can be applied automatically, and enables insights into the system to be gained that were not obvious beforehand. In the case of the Port Phillip Bay model, we find that the variables representing water and sand are unaffected by the remaining variables, while only variables on the same trophic level can be grouped together. The validity of the aggregation under several nutrient loads is also discussed.

2000 *Mathematics subject classification*: primary 93B11, 92D40; secondary 93A30.
Keywords and phrases: Large ecosystem models, complexity, aggregating state variables, model order reduction.

1. Introduction

Environmental studies are often undertaken to investigate the dynamics of a certain variable of interest. In an aquatic setting this may be the number of a particular fish species or a certain measure of water quality. Consequently, these studies have largely focused on a single species ([6]). However, they have a limited capacity to describe the influence of other species or environmental factors on the variable of interest. For this purpose, the whole ecosystem must be considered.

A useful tool for understanding ecosystem dynamics is a whole-ecosystem model. Such models first appeared in the 1970s (for example [13]), and a plethora have since been developed. Examples include the forest model JABOWA ([2]) and Ecopath with Ecosim ([4]). However, these models have large data requirements and can be difficult

[1]School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia; jock.lawrie@gmail.com.

to calibrate ([1, 14, 25, 27]) and validate ([2, 12, 25]). More importantly, they do not necessarily perform better than simpler models ([7, 8, 9, 10]). Indeed, the sensitivity of the model output to the input tends to increase with complexity ([17, 26]). Aside from these problems, it may be difficult to identify the underlying cause of an observed model behaviour from a large number of possibilities.

On the other hand, models that are too simple may distort the importance of a possible cause of an observed behaviour ([28]), or not exhibit the behaviour at all ([7]). Hence it is desirable to reduce model complexity where possible, but not so much so that the model fails to serve its purpose. Thus this paper is concerned with simplifying an ecosystem model as much as possible while retaining adequate estimates of the given performance measures.

In particular, we are concerned with reducing the order of models described by ordinary differential equations via the aggregation of state variables. To this end there has been little work done in the context of complex ecological models. Indeed, the most common model structures examined are the linear ([3, 21]) and Lotka-Volterra forms ([18, 21]). Furthermore, most of the test cases have a small number of state variables. For example the model of Kooi *et al.* [15] has only one equation per trophic level. While these can be instructive in the construction of larger models (Murray and Parslow, [20]), their analysis is not a substitute for the analysis of the larger models.

Aggregation of variables in large complex models has largely been realized via systematic comparison of models with different groupings ([5, 7, 22]). While this approach is informative about the selection of an appropriate model structure, its implementation requires knowledge of the system. In this study, a method is presented for aggregating state variables that can be implemented automatically. It is based on the method of Proper Orthogonal Decomposition ([23, 24]), which is briefly described in Section 2. Although the method can be applied to a variety of complex systems models, it is demonstrated here on the model described in Section 3. The results and discussion comprise Section 4.

2. The method

2.1. Proper Orthogonal Decomposition (POD) - The basic idea The method of POD ([23, 24]) has its origin in data fitting, in particular, statistical regression. The basic idea is to identify a line of best fit to a given set of data. Here "best fit" refers to minimizing the sum of squares of the Euclidean distances between the data set $\{x_i\}$ $i = 1, \ldots, N$ and its estimate $\{\hat{x}_i\}$. In several dimensions, the "line" is a hyperplane. This concept can be extended to the reduction of dynamical systems models described by ordinary differential equations. After collecting data from simulated system trajectories and identifying the plane of best fit, we not only

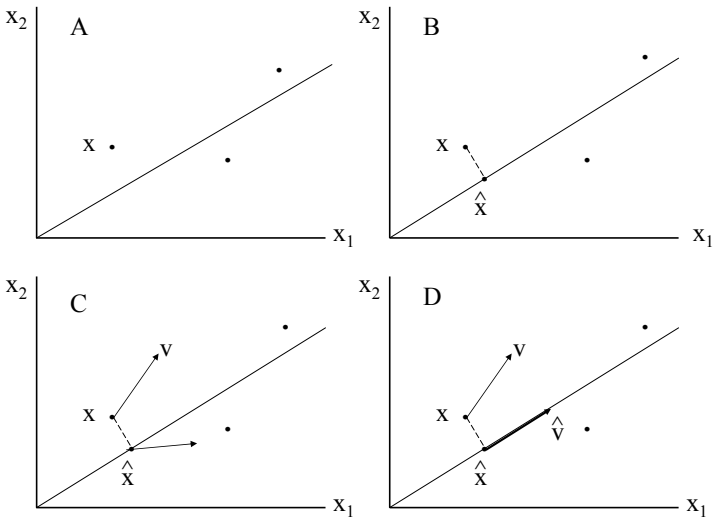

FIGURE 1. The idea of POD. A. Data with a line of best fit. B. The point x is orthogonally projected onto the line at the point $\hat{x}$. C. The derivative v at x is shown with the derivative at $\hat{x}$. D. The derivative at $\hat{x}$ is projected onto the line, resulting in $\hat{v}$, the approximation of v.

project the data onto the plane, but also the time derivative vectors of the projected points. Figure 1 illustrates this idea. More formally, suppose that we have a model

$$\dot{x} = f(x(t), u(t), t),$$

where $x(t)$ is the n-dimensional state vector, $u(t)$ is a vector-valued input function and $x(0) = x_0$. Its estimate of dimension $k < n$ is given by

$$\dot{\hat{x}}(t) = Pf(\hat{x}(t), u(t), t), \tag{2.1}$$

where $\hat{x}(0) = P(x_0 - x^*) + x^*$ and x^* is the origin of the plane of best fit (see [23] for details). In this study we shall refer to the original model as the *full* model, and its estimates as *reduced* models. In simple models x^* might be an equilibrium point, but since we are dealing with a non-autonomous system under a set of inputs $\{u_m(t)\}$, $1 \leq m \leq s$, we take x^* to be the overall mean $\bar{x} = \frac{1}{s}\sum_{m=1}^{s} \bar{x}_m$, where $\bar{x}_m$ is the mean under the m^{th} input $\{u_m(t)\}$.

The rank k matrix P is equal to $\rho^T\rho$ ([23]), where ρ is the $k \times n$ matrix whose rows are the k eigenvectors corresponding to the k largest eigenvalues of the covariance matrix cov of the model output. In this investigation we seek relationships between state variables that hold under all the $\{u_m(t)\}$, so we choose $cov = \frac{1}{s}\sum_{m=1}^{s} cov_m$, where

cov_m is the covariance (centred about $\bar{x}$) under the m^{th} forcing. The coordinates of the approximated data are then given by $\hat{x}(t) = P(x - \bar{x}) + \bar{x}$.

2.2. Modified Proper Orthogonal Decomposition (MPOD) The model described by Equation (2.1) can be represented in k-dimensional space (with origin at $\bar{x}$) using the transformation $z = \rho(x - \bar{x})$ ([23]). This alternative representation is as follows:

$$\dot{z} = \rho f(\rho^T z, u, t). \tag{2.2}$$

From Equation (2.2) we can see that the state variables of the reduced model are in fact linear combinations of the original state variables. Although this is mathematically sound, the ecological interpretation of such combinations is difficult. The interpretation of their interactions, as described by the right-hand side of Equation (2.2), is harder still.

In particular, the trophic structure is obscured because the projection matrix P (see Equation (2.1)) has in general nonzero entries. We can see from Equation (2.1) that each reduced model derivative contains terms that are not present in its full-model counterpart. That is, every variable *directly* affects every other variable, so that a graph representation of the network would be complete. This is not only difficult to interpret, it obscures insights into trophic dynamics and does not enhance our understanding of the system. Also, although the model order has decreased, the complexity of the derivatives has increased.

With this in mind, we seek to aggregate subsets of the variables, so that the reduced model retains an ecological interpretation, and so that the complexity of the derivatives is not increased. This approach also allows variables to remain unaggregated, so that variables in which we are particularly interested can remain explicitly represented. The reduced model is found by projecting each subset onto a separate one-dimensional line using POD. We shall refer to this version of POD simply as Modified Proper Orthogonal Decomposition (MPOD).

2.3. Choosing the aggregates for MPOD The projections of MPOD are based on minimizing the Euclidean distance between a trajectory $x(t)$ and its projection $Px(t)$. The error measure is given by

$$\frac{\int_0^T \|x(t) - Px(t)\|^2 \, dt}{\int_0^T \|x(t)\|^2 \, dt}, \tag{2.3}$$

where T is the simulation time frame and $\|v\|$ denotes the magnitude of the vector v. Since the vectors $x(t) - Px(t)$ and $Px(t)$ are orthogonal we have at any time t

$$\|x(t)\|^2 = \|x(t) - Px(t)\|^2 + \|Px(t)\|^2. \tag{2.4}$$

By integrating Equation (2.4) over time, dividing both sides by $\int_0^T \|x(t)\|^2\,dt$ and then substituting into (2.3), we see that the error measure (2.3) can be written as

$$1 - \frac{\int_0^T \|Px(t)\|^2\,dt}{\int_0^T \|x(t)\|^2\,dt}. \tag{2.5}$$

The second term in (2.5) ranges between zero and one. This is a measure of variance in the trajectory that is accounted for by the projected trajectory, and is thus analogous to the r^2 value used in regression analysis. We shall use it as a measure of "appropriateness" of the aggregate described by the projection axis.

For computational purposes, we note that it is equal to

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{n} \lambda_j},$$

where k is the dimension of the subspace, and the λ_i are the eigenvalues of the covariance matrix cov of the model output. They are ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ ([23]). For MPOD, we have $k = 1$ for each projection, so that the appropriateness measure is given by

$$A = \frac{\lambda_1}{\sum_{j=1}^{n} \lambda_j}. \tag{2.6}$$

If one of the variables in the aggregate is large compared to the others, then λ_1 will be much larger than the remaining λ_j. Thus the measure will be close to one, regardless of whether the variables really are linearly related. To avoid this problem, each state variable is standardized so that the effects of scale are removed. The covariance matrix of these standardized variables is the correlation matrix. Thus the metric A in Equation (2.6) is calculated using the correlation matrix rather than the covariance matrix.

For each variable x_i, the variables $\{x_j\}$ that can be grouped with it are identified by calculating the corresponding value of A (see Equation (2.6)). The aggregate initially includes only x_i, and is built by successively adding the variable x_j that gives the highest value of A. If a variable x_j is added that results in an aggregate with a nonsensical ecological interpretation, then x_j is removed from the aggregate and the process resumes. For example, suppose that sand is aggregated with phytoplankton. Since this aggregate has no ecological interpretation, sand is removed from the aggregate and the remaining variables are scrutinized for their suitability for aggregation with phytoplankton.

After all the suitable variables have been added, up to $n - 1$ values of A have been calculated. The aggregate corresponding to the highest value of A is then selected for the reduced model (note that A does not necessarily increase with the number of

variables in the aggregate). If a variable x_i can be a part of two different aggregates, it is assigned to the aggregate that results in the higher value of A.

Once a set of aggregates has been identified, they are ordered by appropriateness as measured by Equation (2.6). One by one they are implemented in the model and the results assessed. This is achieved by a POD projection of each aggregate using the covariance matrix of the variables in the aggregate. Thus the order of the reduced model is equal to the number of aggregates plus the number of unaggregated variables. Here we note that MPOD is equivalent to aggregating the variables in the same block. Indeed an aggregate is a linear combination of variables. Moreover, the linear relationship between the reduced-model estimates of an aggregate and its constituent variables is uniquely determined by the projection (that is, the line of best fit).

3. Model description

The MPOD method is demonstrated on a model of Port Phillip Bay, Australia - a semi-enclosed marine environment with an area of about 1930km^2 and a narrow opening to Bass Strait. It is a modified (and smaller) version of the Port Phillip Bay Integrated Model ([19]), which was developed by the CSIRO in the 1990s as part of the Port Phillip Bay Environmental Study ([11]). In our model the bay is divided into water column and sediment layers, and is subject to biological and physical processes. The main currency of the model is nitrogen in the form of nitrates (NO) and ammonia (NH), collectively called dissolved inorganic nitrogen (DIN). The state variables in our model are listed in Table 1, where the subscript "s" denotes a sediment variable. The values in Bass Strait are held constant.

TABLE 1. The state variables in the model, their units and their symbols used throughout the discussion. The columns labelled WC, Sed contain variables that appear in the water column and sediment layers respectively.

Variable name	Units	WC variables	Sed variables
Dissolved Inorganic Nitrogen	mg N m^{-3}	DIN	DIN$_s$
Phytoplankton	mg N m^{-3}	Phy	Phy$_s$
Zooplankton	mg N m^{-3}	Z	
Detritus	mg N m^{-3}	DET	DET$_s$
Water Column Sand	mg WCS m^{-3}	WCS	
Water Column Volume	m^3	WCV	
Sediment Water Volume	m^3		SWV
Sediment Solid Volume	m^3		SSV

The model equations are given next, with the terms briefly described in Sections 3.1

and 3.2. A more detailed description is available in [16]. Our notation is as follows.

$$
\begin{aligned}
N'(t) &= \text{breakdown}_{\text{D}} + \text{release}_{\text{Z}} - \text{uptake}_{\text{P}} + \text{exchange}_{\text{N}} - \text{settling}_{\text{N}} \\
&\quad + \text{resuspension}_{\text{N}} + \text{tide}_{\text{N}} + \text{load} \\
P'(t) &= \text{growth}_{\text{P}} - \text{grazing}_{\text{Z}} - \text{settling}_{\text{P}} + \text{resuspension}_{\text{P}} - \text{dilution}_{\text{P}} + \text{tide}_{\text{P}} \\
Z'(t) &= \text{growth}_{\text{Z}} - \text{mortality}_{\text{Z}} - \text{settling}_{\text{Z}} + \text{resuspension}_{\text{Z}} - \text{dilution}_{\text{Z}} + \text{tide}_{\text{Z}} \\
D'(t) &= \text{production}_{\text{Z}} - \text{breakdown}_{\text{D}} - \text{decay}_{\text{D}} - \text{settling}_{\text{D}} + \text{resuspension}_{\text{D}} \\
&\quad - \text{dilution}_{\text{Z}} + \text{tide}_{\text{Z}} \\
WCS'(t) &= -\text{settling}_{\text{WCS}} + \text{resuspension}_{\text{WCS}} - \text{dilution}_{\text{WCS}} + \text{tide}_{\text{WCS}} \\
N_s'(t) &= \text{release}_{\text{D}_s} - \text{exchange}_{\text{N}} + \text{settling}_{\text{N}} \\
P_s'(t) &= \text{settling}_{\text{P}} - \text{mortality}_{\text{P}_s} \\
D_s'(t) &= \text{mortality}_{\text{P}_s} - \text{breakdown}_{\text{D}_s} - \text{decay}_{\text{D}_s} + \text{settling}_{\text{D}} \\
WCV'(t) &= \text{resuspension}_{\text{SWV}} - \text{settling}_{\text{WCV}} + \text{inflows} + \text{tide}_{\text{Z}} \\
SWV'(t) &= \text{settling}_{\text{WCV}} - \text{resuspension}_{\text{SWV}} \\
SSV'(t) &= \text{settling}_{\text{part}} - \text{resuspension}_{\text{SSV}}
\end{aligned}
$$

3.1. Biological processes The biological processes involve the DIN content of the organic state variables. Here they are merely outlined; see Murray and Parslow [19] for details. The main biological processes for the living variables are growth, mortality and grazing. The corresponding rate parameters incorporate a time-dependent oscillation representing seasonal temperature variation. Phytoplankton growth rates involve limitations of nutrients and light, while that of zooplankton depends on how efficiently it can assimilate its prey. What is not assimilated is either excreted back to the water as DIN, or becomes detritus. Loss due to mortality also becomes detritus, which breaks down at a constant rate (r_{DL}) and is remineralized into DIN.

In the sediment, some DIN is denitrified to nitrogen gas and released from the bay altogether. This is the bay's most significant way of expelling nitrogen. In the model denitrification is represented empirically, and is zero when remineralization is below zero or above the level R_0. Between these values denitrification varies linearly with remineralization, with efficiency reaching a maximum when remineralization is R_D.

3.2. Physical processes Besides denitrification there are three main ways that the bay interacts with its surroundings. These are the tidal exchange of water between the bay and Bass Strait, freshwater fluxes from rivers, rainfall and evaporation, and nutrient inputs from Bass Strait, the rivers and the atmosphere. These processes are described by time-dependent functions obtained from auxiliary models ([29, 19]) and have units of $\text{m}^3\ \text{day}^{-1}$ for the flows and $\text{mg}\ \text{day}^{-1}$ for the nutrient inputs.

Within the bay all horizontal and most vertical movement of water is accounted for

by the aforementioned hydrodynamic processes. Some water is transferred between layers when sand and detritus settle to and are resuspended from the sediment. The constant settling speeds are given in m day^{-1}, while resuspension is dictated by the erosion rate also given in m day^{-1}. This is the product of a constant erosion rate, and a dimensionless time-dependent function representing stress on the bay floor. There is also an exchange of water between layers at a rate given in m^3 day^{-1}, though it causes no change in layer volume. Finally, detritus decays at a constant rate with units sec^{-1}. Full descriptions of these processes can be found in Walker and Sherwood [30].

3.3. The goals of this investigation There are two goals in this investigation. Firstly, we seek insights into the internal mechanisms of the Port Phillip Bay system. Secondly, we seek a smaller model than that described earlier that performs "similarly" in some sense.

In particular, we examine the response of the bay to increased nutrient loads over a decade or so. The simulation period is $T = 4384$ days (12 years) and the current input of DIN, referred to as the base load, is scaled by factors of 0.5, 1, 1.5, 2, 2.5, 3, 3.5 and 4. The response we are primarily interested in is the average annual phytoplankton production, which is expressed mathematically as

$$\text{production} = \frac{1}{12}\int_0^{4383} \text{growth}_{\text{Phy}}\, dt \tag{3.1}$$

where $\text{growth}_{\text{Phy}}$ denotes the growth rate of water column phytoplankton (there is no production in the sediment). The full-model values of production under each load are shown in Figure 2.

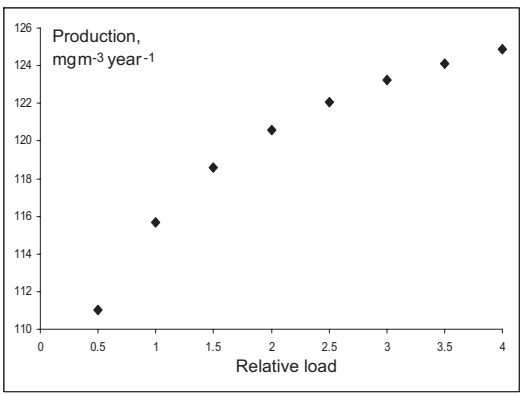

FIGURE 2. The values of production under each nutrient load predicted by the full model. Here load is expressed relative to the base load.

The adequacy of the reduced model is measured by its values of production, calculated from Equation (3.1) for each load. We are particularly interested in approx-

imating production to within a given tolerance $\alpha\%$ under all loads. In this study we arbitrarily choose $\alpha = 5\%$.

4. Results and discussion

Both goals prescribed in Section 3.3 were achieved. Regarding reduced models, the simplest model obtained using the MPOD procedure had four aggregates, each with two constituent variables. These are described in Table 2, which shows the A values (see Equation (2.6)) calculated for each aggregate. Since there are four aggregates and three unaggregated variables, the reduced model has order 7.

TABLE 2. Aggregates in the reduced model

Aggregates	A
DIN, DET	0.795
Phy, Phy_s	0.918
DIN_s, DET_s	0.903
SWV, SSV	1.000

The largest relative error between the full-model production and its reduced-model estimate was about 1.4% - well within the $\alpha = 5\%$ tolerance. Hence the reduced model is an adequate approximation of the full model for the purpose of investigating phytoplankton production.

Regarding our goal of gaining insights into the system, we notice that the biological variables within any given block are on the same trophic level. Indeed, DIN and DET are at the lowest level, while phytoplankton in the water column and sediment are obviously at the same level. This is consistent with Fulton's ([7]) conclusions that only variables on the same trophic level should be aggregated. Therefore, the dynamics of detritus and DIN need not be explicitly represented in the model. Rather, they can be aggregated in both layers.

Considerable insight can be gained by comparing the trajectories from the full and reduced models. In particular, we measure the similarity of the time series of a variable $x_i(t)$ to its reduced-model estimate $\hat{x}_i(t)$ by the correlation r^2 between them. Table 3 shows these correlations for each variable, averaged over the loads.

The most striking values in Table 3 are those of water and sand in the water column and sediment (WCS, WCV, SWV, SSV), and DIN. Since these variables are almost perfectly correlated we learn that they are not affected by being aggregated or by the approximation of the other variables. That is, they are robust to the simulated range of values of the other variables. Moreover, we learn that most of the top centimetre of the sediment is sand and not detritus. Indeed, the dynamics of water column sand

TABLE 3. The r^2 values for each variable averaged over the loads.

Variable	r^2
DIN	0.996
Phy	0.300
Z	0.394
DET	0.705
WCS	1.000
DIN_s	0.621
Phy_s	0.442
DET_s	0.736
WCV	1.000
SWV	1.000
SSV	1.000

(WCS) are completely unaffected ($r^2 = 1.0$, Table 3). Since WCS is resuspended from the sediment along with detritus, either WCS is unaffected by detritus, or the dynamics of detritus are unaffected by being aggregated. The latter possibility is not the case since the r^2 values for DET and DET_s are around 0.7 (Table 3).

On the other hand the aggregations have a significant effect on some of the other variables. For example, the r^2 value for zooplankton ($r^2 = 0.4$, Table 3) implies that estimates of the zooplankton values would have large variability and therefore be unreliable, even though zooplankton was not aggregated. Similarly, the phytoplankton (Phy, Phy_s) dynamics are different ($r^2 \leq 0.44$, Table 3), despite the production being accurately modelled.

The poor r^2 values for phytoplankton can be partly explained by the forms of their reduced-model derivatives. In particular, they are different to their full-model counterparts since they are weighted sums of the original derivatives (see Equation (2.1)). That is, the reduced-model derivatives contain terms that were not in their full-model counterparts.

This fact may also help explain the discrepancy in the r^2 values for variables within a given aggregate. For example, the derivative of DET in the reduced model has a scaled uptake term which the original DET derivative did not. Thus a link between detritus and phytoplankton has been artificially created in the reduced model. No such links are created for the DIN variable. It is possible that this is the cause of the poor DET r^2 value (0.7, Table 3) relative to that of DIN (0.99), though verifying this requires further investigation.

For comparative purposes, estimates of production were obtained using POD with $k = 7$. The relative errors were similar to those obtained using MPOD, and in fact were slightly inferior. Indeed, the largest error was about 1.9%, as opposed to 1.3% for

the MPOD model. Thus although the magnitude of the state vector is more accurately estimated using POD rather than MPOD (by construction), this is not necessarily the case for functions of the state vector, such as production. This is because there is no consideration of the *direction* of the state vector in the construction of the projection matrix P.

Finally, we note that a reduced model can only be a substitute for the full model under certain conditions. In particular, caution should be exercised when simulating scenarios that differ widely from those used in the reduction process. In this respect, the model reduction process is similar to that of validation. That is, confidence in the performance of a reduced model is strongest when the model is used within the range of scenarios and diagnostics considered in the reduction (or validation) process.

5. Conclusion

A model reduction method called Modified Proper Orthogonal Decomposition (MPOD) was proposed, which is a modification to the method of proper orthogonal decomposition. Its purpose is to identify meaningful linear relationships among variables in a dynamical system, and use them to create a reduced model with an ecological interpretation. In this regard, MPOD is more useful than POD. The method was applied successfully to an aquatic model with the aim of estimating phytoplankton production. Consideration of the results led to the following key findings:

- Significant reduction was possible. In particular, the model order was reduced from 11 to 7.
- The biological variables within any given aggregate were on the same trophic level.
- Most of the top centimetre of the sediment is sand and not detritus.
- The estimates of the variables in an aggregate are affected by differences in the forms of their original derivatives. In particular, the introduction of new terms into a derivative as a result of the projection can cause the errors of the variables in the same block to be notably different.
- It seems that the physical variables are independent of the biological variables, but not vice versa. However, verifying this requires further investigation.
- The performance of MPOD can be superior to that of POD, depending on the performance measure. This was the case when $production$ was the performance measure.

Aside from identifying simpler models and revealing insights into the system, the method allowed reduction of the original model without requiring a detailed understanding of the system *a priori*. However, caution should be exercised when modelling situations that differ widely from those used to obtain the reduced model. Neverthe-

less, MPOD is a useful method for reducing the complexity of ecological models without sacrificing their ecological interpretation. It is also useful for learning about the mechanisms built into such models.

Finally, we note that this method is intended to be applied to larger models than that examined in this study. The authors are currently investigating its applicability to a model with 29 state variables and 7 performance indicators. The investigation considers the effects of several different aggregations on the performance indicators, thus highlighting certain sensitivities between different components of the system.

Acknowledgements

This study was funded by the Australian Research Council and by the CSIRO Postgraduate Scholarship Program. The CSIRO (Commonwealth Scientific and Industrial Research Organisation) also supplied the model and data. The authors are grateful to Beth Fulton at CSIRO Marine and Atmospheric Research for her assistance with the model.

References

[1] M. B. Beck, "Hard or soft environmental systems", *Ecol. Model.* **11** (1981) 233–251.
[2] H. Bugmann, "A review of forest gap models", *Climatic Change* **51** (2001) 259–305.
[3] W. G. Cale, Jr. and P. L. Odell, "Behaviour of aggregate state variables in ecosystem models", *Math. Biosci.* **49** (1980) 121–137.
[4] V. Christensen and C. J. Walters, "Ecopath with ecosim: methods, capabilities and limitations", *Ecol. Model.* **172** (2004) 109–139.
[5] K. L. Denman, "Modelling planktonic ecosystems: parameterizing complexity", *Prog. Oceanography* **57** (2003) 429–452.
[6] I. Fazey, J. Fischer and D. B. Lindenmayer, "What do conservation biologists publish", *Biol. Conserv.* **124** (2005) 63–73.
[7] E. A. Fulton, "The effects of model structure and complexity on the behaviour and performance of marine ecosystem models", Ph. D. Thesis, School of Zoology, University of Tasmania, Hobart, Australia, 2001.
[8] E. A. Fulton, J. S. Parslow, A. D. M. Smith and C. R. Johnson, "Biogeochemical marine ecosystem models II: the effect of physiological detail on model performance", *Ecol. Model.* **173** (2004) 371–406.
[9] E. A. Fulton, A. D. M. Smith and C. R. Johnson, "Mortality and predation in ecosystem models: is it important how these are expressed", *Ecol. Model.* **169** (2003) 157–178.
[10] L. Håkanson, "Optimal size of predictive models", *Ecol. Model.* **78** (1995) 195–204.
[11] G. Harris, G. Batley, D. Fox, D. Hall, P. Jernakoff, R. Molloy, A. Murray, B. Newell, J. Parslow, G. Skyring and S. J. Walker, "Port Phillip Bay Environmental Study Final Report", Technical report, CSIRO, Canberra, Australia, 1996.
[12] P. S. Homann, R. B. McKane and P. Sollins, "Below-ground processes in forest-ecosystem biogeochemical simulation models", *Forest Ecol. Management* **138** (2000) 3–18.

[13] S. E. Jorgensen, "A eutrophication model for a lake", *Ecol. Model.* **2** (1976) 147–165.
[14] S. E. Jorgensen, S. Ray, L. Berec and M. Straskraba, "Improved calibration of a eutrophication model by use of the size variation due to succession", *Ecol. Model.* **153** (2002) 269–277.
[15] B. W. Kooi, J. C. Poggiale and P. Auger, "Aggregation methods in food chains", *Math. Comput. Model.* **27** (1998) 109–120.
[16] J. Lawrie, "Reducing the complexity of large ecosystem models", Ph. D. Thesis, School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia, 2006.
[17] K. E. Lindenschmidt, "The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling", *Ecol. Model.* **190** (2006) 72–86.
[18] N. K. Lukyanov, Yu. M. Svirezhev and O. V. Voronkova, "Aggregation of variables in simulation models of water ecosystems", *Ecol. Model.* **18** (1983) 235–240.
[19] A. Murray and J. Parslow, "Port Phillip Bay Integrated Model: Final Report", Technical Report 44, Port Phillip Bay Environmental Study, CSIRO, Canberra, Australia, 1997.
[20] A. Murray and J. Parslow, "The analysis of alternative formulations in a simple model of a coastal ecosystem", *Ecol. Model.* **119** (1999) 149–166.
[21] R. V. O'Neill and B. Rust, "Aggregation error in ecological models", *Ecol. Model.* **7** (1979) 91–105.
[22] J. K. Pinnegar, J. L. Blanchard, S. Mackinson, R. D. Scott and D. E. Duplisea, "Aggregation and removal of weak-links in food-web models: system stability and recovery from disturbance", *Ecol. Model.* **184** (2005) 229–248.
[23] M. Rathinam and L. R. Petzold, "A new look at proper orthogonal decomposition", *SIAM J. Numer. Anal.* **41** (2003) 1893–1925.
[24] C. W. Rowley, T. Colonius and R. M. Murray, "Model reduction for compressible flows using POD and Galerkin projection", *Phys. D* **189** (2004) 115–129.
[25] H. Scholten, M. W. Van Der Tol and A. C. Smaal, "Models or measurements", in *Meet. of the Int. Counc. for the Exploration of the Sea, Cascais, Portugal, 16–19 Sep*, (ICES, Copenhagen, Denmark, 1998).
[26] S. D. Snowling and J. R. Kramer, "Evaluating modelling uncertainty for model selection", *Ecol. Model.* **138** (2001) 17–30.
[27] M. E. Toal, C. Yeomans, K. Killham and A. A. Meharg, "A review of rhizosphere carbon flow modelling", *Plant Soil* **222** (2000) 263–281.
[28] E. H. Van Nes and M. Scheffer, "A strategy to improve the contribution of complex simulation models to ecological theory", *Ecol. Model.* **185** (2005) 153–164.
[29] S. J. Walker, "Hydrodynamic models of Port Phillip Bay", Technical Report 38, Port Phillip Bay Environmental Study, CSIRO, Canberra, Australia, 1997.
[30] S. J. Walker and C. R. Sherwood, "A transport model of Port Phillip Bay", Technical Report 39, Port Phillip Bay Environmental Study, CSIRO, Canberra, Australia, 1997.