Seventh Framework Programme Theme 6 Environment

Collaborative Project (Large-scale Integrating Project)

Project no. **212085**

Project acronym: **MEECE**

Project title: **Marine Ecosystem Evolution in a Changing Environment**

D3.2 Common set of metrics

Due date of deliverable: 31.08.2011
Actual submission date: 31.08.2011
Organisation name of lead contractor for this deliverable: NERC-POL

Start date of project: 01.09.08 Duration: 48 months

Project Coordinator: Icarus Allen, Plymouth Marine Laboratory

Project co-funded by the European Commission within the Seventh Framework Programme, Theme 6 Environment		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission)	
RE	Restricted to a group specified by the consortium (including the Commission)	
CO	Confidential, only for members of the consortium (including the Commission)	

D3.2 Common set of metrics

Lead Contributor; Jason Holt, NERC-POL

Contents

1. Introduction

Central to any scientific analysis of environmental systems is the processing of the resulting information into a form that aids the interpretation of the analysis in the context of our empirical and theoretical understanding, and also aids the comparison between similar and contrasting sources of information. Essentially the 'raw' data must be processed before it is generally useful. In the case of spatially/temporally resolved model simulations this often involves aggregation of the five dimensional output (three space, time and constituent vectors) into a form that is tractable and more appropriate. It can also involve a combination of aggregation and sub-sampling to give a form that more closely matches our observational base. Beyond this the processing can involve a whole spectrum of more sophisticated techniques, such as spectral analysis, principle component analysis, wavelet analysis, cluster analysis etc. The definition of these 'metrics' is an important part of the scientific processes that shapes how our understanding of the environmental system evolves, and as such should be under constant scrutiny.

The model scenarios we are treating in MEECE are outlined in D1.5 and described in D3.1, and since a diverse range of regions and models is considered, a degree of commonality in the analysis of the results is highly desirable, to permit the inter-comparison between models and regions, to use this to both build our understanding of the system and to explore the uncertainties in the results and the reliability of the conclusions.

2. Overview

Defining a standard set of metrics is not necessarily a desirable process as it risks stifling the scientific exploration of the system and as such the standard set must be considered a minimal starting point. However, such an exercise does serve an important purpose that is particularly significant for MEECE. Namely, by defining a common approach to data analysis it allows the inter-comparison of data sets arriving from different sources. In the current context these are the different models from different regions and the corresponding observational data (here we include both *in-situ* and remote sensed data). Exploring the differences between these metrics then produces a new set of comparison metrics that has been extensively investigated in the context of model skill assessment (see Stow et al., 2009 and references therein), ranging from general difference measures such a root mean square error to metrics that focus on specific aspects of a model's performance (e.g. its reliability as an aid to decision making, such as the ROC test). In the MEECE context these have been considered in some detail in D2.7. While these approaches have generally been applied to comparing models with observations they are also appropriate for comparing models with each other, particularly with respect to an observation based reference. Finally, the processed data metrics and inter-comparison metrics can be related to 'expert' knowledge of the system, both singly and in groups, to define simple indicators that might be used to guide assessment and decision making processes. Examples, include thresholds of model skill assessment variables that indicate whether a model is 'fit for purpose' in a particular context and threshold based indicators used for simple mapping (e.g. OSPAR indicators) in either a uni-variant or multi-variant (e.g. hierarchal classifications) fashion.

Hence we arrive at four types of metrics:
1. Processed model and observational data
2. Model-observation comparison metrics
3. Model-model comparison metrics
4. Synthesised maps and indicators

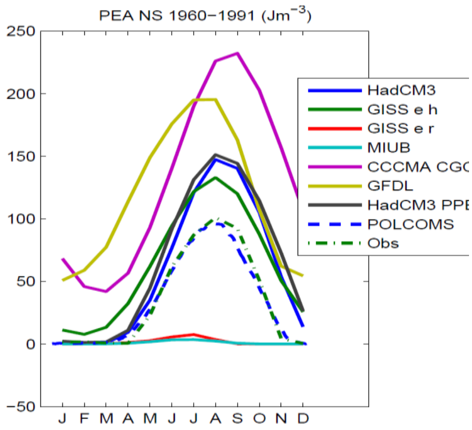


Figure 1. Example of Metric Type 1 (see text).

This can be contrasted with the metric's defined in the MERSEA project (Crosnier and Le Provost, 2007) for the assessment of operational physical ocean models. Here we are introducing a much broader definition that can encompass the scientific interpretation, the inter-comparison, and the assessment; three of the four metric classes defined in MERSEA are included here in TYPE 1. Examples of these types are shown in Figures 1, 2, 3 and 4. Figure 1 shows a derived quantity (potential energy anomaly) that has been temporally averaged for 8 models and observations to show a mean annual cycle at a single location (Holt et al., 2010). Note: while we can compare the models/observations by eye it lacks any quantitative information of the difference. Figure 2 shows a simplified Taylor diagram assessing the model skill between station observations of various variables and model output sub-sampled to match these (Allen et al., 2007a). While comparison statistics are provided it lacks any subjective comment on these (i.e. what is a 'good' value). Figure 3 shows the fractional difference in net Primary production between a future scenario and the corresponding control (Holt et al., 2011). Figure 4 shows a hierarchical map used for marine spatial planning (Connor et al., 2006). A set of model information has been categorised into seasonal maps based on 'expert' specified thresholds. In this case the information is readily accessible but most of the dynamical information and assessment of model quality has been lost.

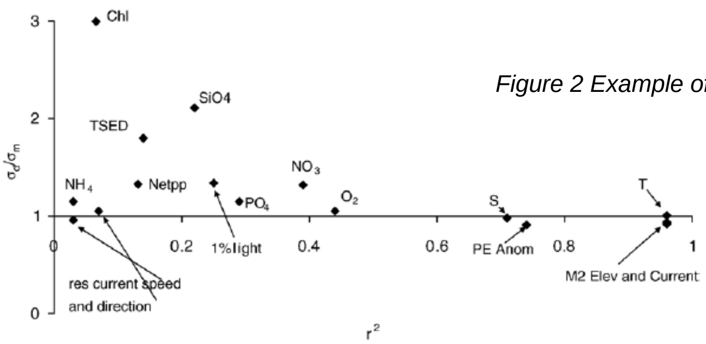
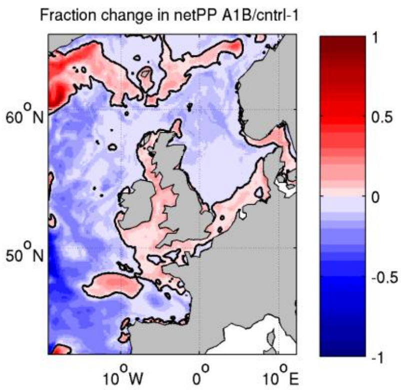


Figure 2 Example of metric type 2 (see text)



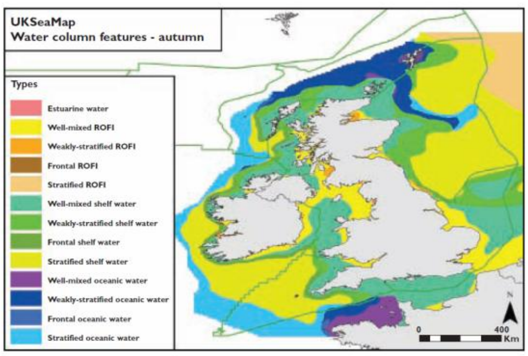


Figure 3 Example of Metric Type 4 (see text)

Figure 4 Example of metric Type 3 (see text)

2.1 Spatial and temporal scales

Critical to any definition of metrics is the consideration of the relevant spatial and temporal scales. These are set by the model type and configuration, often by technical limitations, such the amount of training data or the computational cost, by the pertinent bio-physical scales, and by the societal scales of interest.

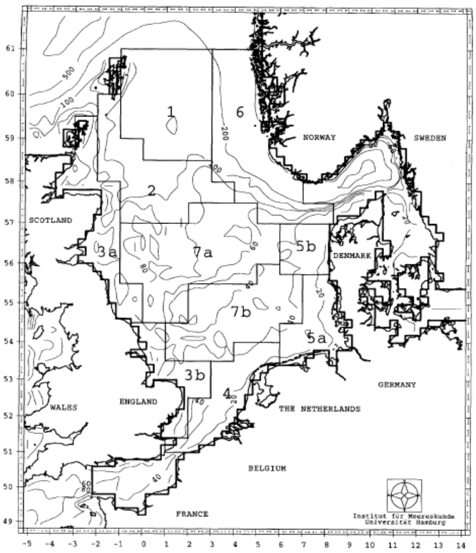


Figure 5. Modified ICES boxes (Lenhart and Pohlmann, 1997)

Horizontal spatial scales are particularly important in defining the areal aggregation of data. The smallest spatial scale of interest is the model grid scale (typically a few km's for regional dynamic models). Information at the model grid scale reflects the details of the solution approach and is generally less reliable than aggregation over a number of cells. The pertinent physical scale for coastal ocean dynamics relate to the natural response to forcing of these systems (Coastal Trapped waves, e.g. Holt et al (In Press), i.e. scales of $(gh)^{0.5}/f$ (~10^2-10^3 km) for barotropic processes and $(g'h)^{0.5}/f$ for baroclinc processes, ~1-10 km), and to the geographic scales of bathymetry and coastal geography. These scales are apparent in objective pattern recognition approaches such as self-organising maps (Allen et al., 2007b). More often a less analytic but much more straightforward approach is adopted such as ICES boxes. In this example (Figure 4), the North Sea is divided into a series of boxes on a 0.5^o latitude by 1^o longitude grid, properties are then integrated over these boxes. A finer resolution approach (e.g. Proctor et al., 2003) may be better at representing the different biophysical regions, but the definition of higher resolution regions has not been standardised.

At a national and international level the policy relevant scales are generally much coarser, such as EEZ's, LME'S and MSFD European regions (see below).

Temporal scales are important in defining the periods for temporal aggregation/filtering. The fastest processes of interest here are plankton growth times, coastal trapped waves, and mesoscale meteorological variability all with time scales of O(1 day). The slowest processes are the various modes of large scale climatic variability and the adjustment of the benthic system O(1-50 years). At mid-high latitudes, the seasonal cycle tends to dominate and many processes have a time scale of O(1 year), so this is an important scale to capture in any metric. The longer time scales tend to be dependent on experiment design. The timeslices considered in D3.1 are ~20 years and this limits the period of variability, which can be resolved to <~5yrs. An important distinction needs to be made between reanalysis forced simulations and coupled OA-GCM forced simulations. In the former case the forcing is constrained by observations, so the phase of the inter-annual variability in the forcing would be expected to be realistic (i.e. a model year can be compared directly with the corresponding real year). This is not the case for OA-GCM forced simulations, where the phase of natural variability is unconstrained to real conditions, and a particular model year cannot be directly compared with its real counterpart. In this case the variability needs to be treated in a statistical sense with the phase information removed (e.g. using spectral methods). A straightforward way to do this is to produce a mean annual cycle of monthly mean values (e.g. Figure 1). The interannual variability can then be described as ‘error bars’ on this cycle. This works well for (quasi) sinusoidal properties, as show here. However, care is needed with more ‘spiky’ time series, such as phytoplankton biomass, since the averaging will lead to an unrealistically smoothed time series. In such cases a ‘conditional sampling’ approach might be more appropriate. For example Figure 5 shows mean annual time series of depth integrated phytoplankton biomass in the central North Sea (from POLCOMS-ERSEM). It is apparent that, while the post-bloom conditions are reasonably well represented by the mean daily or monthly annual cycle, they are very poor at representing typical bloom conditions. Due to variations in the timing both mean cycles give values that are substantially less than the bloom in any one year. A similar effect is also apparent on spatial aggregation. However, practical considerations mean that a degree of spatial and temporal aggregation is required, so we much accept that the details of the high frequency/small scale variations will not be retained.

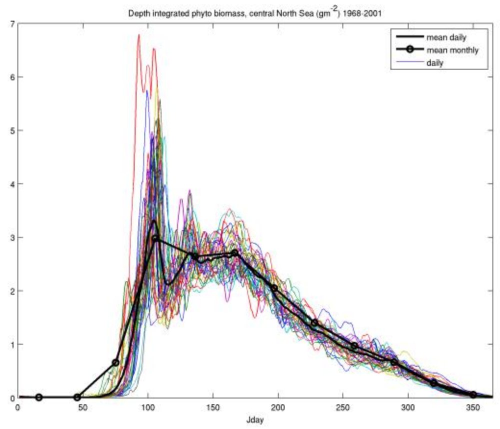


Figure 6. Annual time series of depth integrated phytoplankton biomass. Mean annual cycles based on daily and monthly means are also shown

3. Common metrics

Here we consider the common metrics for MEECE. As noted above, this set of ‘common metrics’ must be seen as a minimal starting point, and in no way hinder the creative exploration of the modelled and observed system. Because they need to be common across all regions they are necessarily the simplest approaches and more sophisticated analysis is considered on a region by region basis. As appropriate a sub set of these metrics will be used to generate the MEECE Atlas, whereas others will form the basis for analysis and intercomparison papers.

3.1 Spatial aggregation

The MSFD defines four European regions, two of which have four sub-regions:

- Baltic Sea
- Black Sea
- Mediterranean
  - The western Mediterranean
  - The Adriatic Sea
  - The Ionian Sea and the central Mediterranean
  - The Aegean-Levantine Sea
- North East Atlantic
  - The Greater North Sea, including Kattegat and English Channel

- The Celtic Seas
- Bay of Biscay and Iberian Coast
- Water surrounding Azores, Madeira and the Cannery Islands

MEECE covers most of these regions and in addition it also includes the Barents Sea, Benguela current region and a global perspective. Hence, MSFD regions provide a natural choice for large scale regional arrogation.

Of equal importance is the vertical representation of data. Spatially averaged vertical profiles are generally problematic owing to spatial heterogeneity of the properties and variably bathymetry. Hence depth resolved information is generally only presented as profile time series at a point or time-horizontal section averaged over a particular period. Horizontal maps generally require a specified vertical averaging (either distance or number of layers). Here we consider four layers:

- Surface
- Mixed layer integrated
- Depth integrated
- Near bed

The definition of these is context/model dependent and must considered separately for each case.

3.2 Temporal Aggregation

As noted above average monthly mean cycles provide a good approach to displaying time series, where inter-annual variability is less important. Seasonal means provide a good approach for displaying horizontal maps. In several cases annual and 'full-period' means are also appropriate. For example, annual means are suitable for multi-decadal timeseries where higher frequency data looses clarity.

We propose the following common metrics:

3.2.1 Type 1: Native and derived model variables, observational products

Models:
Data is provided as maps of monthly/seasonal/annual/full period means as appropriate and time series of area means (monthly means and mean annual cycles).

For hindcast (present day reanalysis forced), control (present day climate OA-GCM forced), and scenarios (climate and anthropogenic drivers) we propose:

1. Surface and near bed
   a. Temperature and salinity
2. Surface
   a. Chlorophyll
   b. Nutrients (NO3, PO4, Si Fe as appropriate)
3. Mixed layer depth
4. Depth integrated:
   a. Phytoplankton biomass (small and large)
   b. Zooplankton biomass (small and large)
   c. Net Primary production (Gross PP minus phytoplankton respiration)
   d. Net Community production (Gross PP minus community resp.)
   e. pH
5. Fish (abundance/size/biomass) as appropriate to model in question

6. Benthic biomass (as appropriate)
7. Time series of volume flux into/out of the region of interest

Observations

Mean monthly annual cycle of all available WODC data averaged onto model grid (i.e. no gap filling).

1. In-situ
    a. Surface T, S, N, P, Si, Chl, O (as appropriate)
    b. Near bed: T, S, N, P, Si, O (as appropriate)
    c. Regionally specific *insitu* validation data indentified in D1.3

2. Satellite
    a. SeaWiF/MODIS (monthly composites, mean monthly cycle)
    b. AVHRR (monthly composites, mean monthly cycle)
    c. PhySat phytoplankton functional types from D1.3. (monthly composites, annual cycle)

3.2.2 Type 2: Inter-comparison: model-observation metrics

For these we select from then methods recommended in D2.7, specifically: Univariate metrics of r^2, RMSE, Bias, Model Efficiency and Taylor diagram (showing r^2, RMSE and sd) and Target Diagram (showing rmse, BIAS). Figure 7shows an example of Taylor and Target diagrams from a comparison between a POLCOMS-ERSEM hincast and ICES data

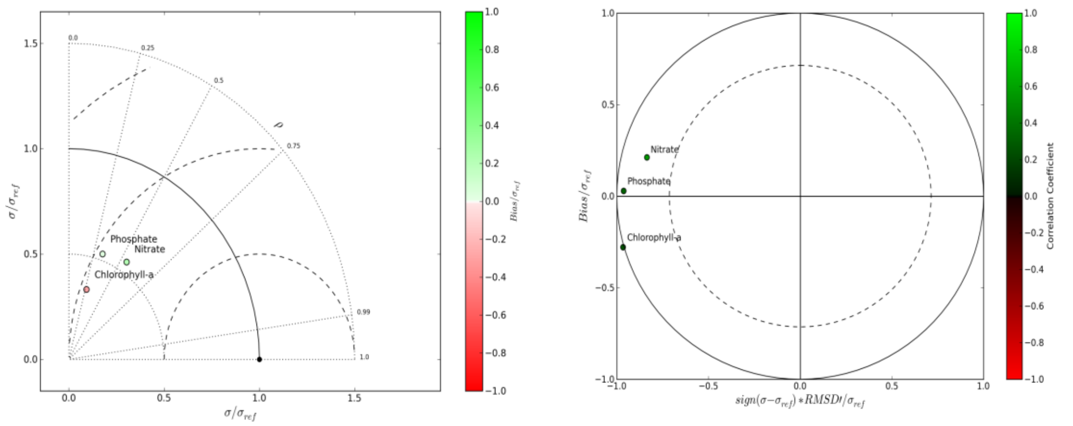


Figure 7. Outcomes of point to point validation of the hindcast on the entire domain vs. the in situ ICES database. On the left the Taylor diagram, on the right the target diagram.

3.2.3 Type 3: Inter-comparison: model-model metrics

These largely draw on type 2, specifically the Taylor and target diagram using a common observational reference to compare the different hindcasts. The Taylor diagram is not appropriate where the variability is not expected to match (i.e. for climate model forced simulations). In these cases, the absolute and fractional difference between different models is considered. Future scenarios are compared through the difference (absolute and fractional) against the control run and between control and hindcast. A significance test between scenario and control against inter-annual variability (e.g. T-test) is then used to assess whether a future climate signal is separable from the natural variability.

3.2.4 Type 4: Indicators

Here we focus on the MSFD descriptors of Good Environmental Status which are as follows

- Descriptor 1: Biological diversity
- Descriptor 2: Non-indigenous species
- Descriptor 3: Population of commercial fish / shell fish
- Descriptor 4: Elements of marine food webs
- Descriptor 5: Eutrophication
- Descriptor 6: Sea floor integrity
- Descriptor 7: Alteration of hydrographical conditions
- Descriptor 8: Contaminants
- Descriptor 9: Contaminants in fish and seafood for human consumption
- Descriptor 10: Marine litter
- Descriptor 11: Introduction of energy, including underwater noise

MEECE has undertaken to map potential model outputs for the range of MEECE models for each region onto potential indicators for each descriptor (Fig 8; D2.15). MEECE models were found to have outputs of relevance to descriptors 1-6 and 8. These are summarised in Table 1 in terms of the core MEECE metrics identified previously.

Table 1. Illustrating how core MEECE metrics map onto descriptors based on the MEECE model library http://www.meece.eu/Library.aspx and the MSFD Management group report (A.C. Cardoso 2010). X indicates a key indicator, (x) indicates of some relevance.

	Biodiversity	**Invasive Species**	**Commercial Fisheries**	**Food webs**	**Eutrophication**	**Seabed integrity**	**Contaminants**
Physio chemical							
Temperature	X			(x)		(x)	
Salinity	X						(x)
Nutrient Nitrate Phosphate Silicate	X			X	X	(x)	
pH	X			(x)			(x)
Biological Features							
Phytoplankton Small Large	X	(x)		x	X	(x)	X
Zooplankton Small Large	X	(x)		x	X	(x)	X
Fish	X	(X)	x	x		(x)	
Chlorophyll					X		
Net Primary Production	(X)			X	X	(X)	(X)
Community production	(x)			X			
*Bottom fauna**	X	(x)	X	x		x	X

*ERSEM NW European Shelf Only

Figure 8 Shows how the modelling effort in different regions maps onto the MSFD GES descriptors.

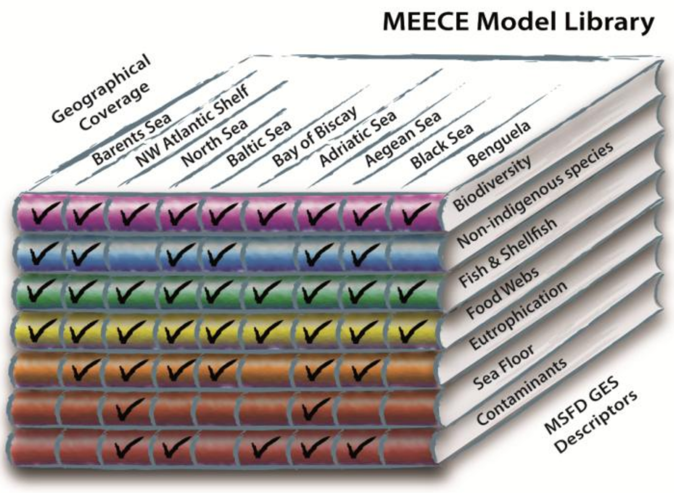


Figure 8. The modelling tools available by European Seas and per MSFD GES Descriptor. The Library will include additional models as documentation becomes available (from D2.15).

4. The MEECE ATLAS

Some of metrics defined here will be drawn together to form the MEECE Atlas (D3.5, & D4.4 due in month 48). The pages of the Atlas will take a 'factsheet' form, one for each of the variables listed above. These will show the areal average temporal variability in the different time slices (as described by D3.1), maps of indicative change between them and a Taylor and/or Target diagram indicating the reliability of the model results. They will also include an explanation of the data source, how to find further information and the important caveats and limitations.

5. References

Allen, J.I., Holt, J.T., Blackford, J.C. and Proctor, R., 2007a. Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM. Journal of Marine Systems, 68(3-4): 381-404.

Allen, J.I., Somerfield, P.J. and Gilbert, F.J., 2007b. Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models. Journal of Marine Systems, 64(1-4): 3-14.

Cardoso AC, S. Cochrane, H. Doerner, J.G. Ferreira, F. Galgani, C. Hagebro, G. Hanke, N. Hoepffner, P.D. Keizer, R. Law, S. Olenin, G. J. Piet, J. Rice, S.I. Rogers, F. Swartenbroux, M. Tasker, W. van de Bund 2010, Scientific Support to the European Commission on the Marine Strategy Framework Directive – Management Group Report Luxembourg: Office for Official Publications of the European Communities 57 pp. DOI 10.2788/86430

Connor, D.W. et al., 2006. UKSeaMap: the mapping of seabed and water column features of UK Seas. http://jncc.defra.gov.uk/page-3918#download, Joint Nature Conservation Committee, Peterborough.

Crosnier, L. and Le Provost, C., 2007. Inter-comparing five forecast operational systems in the North Atlantic and Mediterranean basins: The MERSEA-strand1 methodology. Journal of Marine Systems, 65(1-4): 354-375.

Holt, J., Butenschon, M., Wakelin, S.L., Artioli, Y. and Allen, J.I., 2011. Oceanic controls on the primary production of the northwest European continental shelf under recent past and potential future conditions. Biogeosciences Discussions, 8(4): 8383-8422 DOI: 10.5194/bgd-8-8383-2011.

Holt, J. et al., In Press. Coastal Ocean Circulation Dynamics. In: P. Craig, C. Mooers, Norden and N. Huang (Editors), Coupled Coastal Wind-Wave-Current Dynamics. Cambridge University Press.

Holt, J., Wakelin, S., Lowe, J. and Tinker, J., 2010. The potential impacts of climate change on the hydrography of the northwest European Continental shelf. Progress in Oceanography: doi:10.1016/j.pocean.2010.05.003

Lenhart, H.J. and Pohlmann, T., 1997. The ICES-boxes approach in relation to results of a North Sea Circulation Model. Tellus, 49A: 139-160.

Proctor, R., Holt, J.T., Allen, J.I. and Blackford, J.C., 2003. Nutrient fluxes and budgets for the North West European Shelf from a three-dimensional model. Science of the Total Environment, 313-316: 769-785.

Stow, C.A. et al., 2009. Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems, 76(1-2): 4-15.