ECO_2 project number: 265847

Deliverable Number D12.1
Report on model interfacing and evaluation strategy;
WP12/CCT2; lead beneficiary number 1

D12.1 – CCT2 Report on Model Interfacing and Evaluation Strategy

Introduction

Determining the risks and impacts of CO_2[1] leakage by numerical modelling is a multi-disciplinary problem and involves a variety of research branches. The aim of CCT2 (WP12) is to develop leakage scenarios, to define the model sequence and interfacing, and to coordinate the simulation of those scenarios. The simulated scenarios attempt to predict the environmental and economic impacts of leakage of CO_2 and formation water as it leaks from the storage unit, ascends through the sedimentary overburden, interacts with the sediments as well as benthic and pelagic ecosystems, disperses in the water column and ocean, and some fraction eventually re-emits back into the atmosphere. In order to ensure that leakage scenarios are modelled in a coherent manner, it is vital to pass on the required information correctly to subsequent models in the simulation chain as well as organize feedbacks between models.

This document defines the necessary model interdependencies and the nature of the interfaces, i.e. it lists details on information exchange between models, the required data input for individual models and the respective model output. The document will be updated continuously throughout the project and provides a foundation for inter-model discussions. In addition, a CCT2 workshop was held in December 2011, where CCT2 participants identified the interfaces (meeting minutes are stored in the project database: MS4).

Overview of model interfaces

The task in CCT2 requires the interfacing of 14 models from various disciplines (see outline in Fig. 1). In the numerical modelling strategy the way of the leaking CO_2 and formation water from the escape out of the storage unit, their migration through the overburden and surface sediments, and the emission and spread in the water column is followed.

Initial input information is provided by the geological models developed within WP1 (UiT, GEOMAR). The interpreted geology of the overburden, reservoir, and underburden includes the distribution of fractures and faults as well as available details on stratigraphy, physical and geochemical properties. This basic information is fed into several transport models for the overburden and reservoir. While DUMUX (U. Stuttgart, WP1) handles the transport of the multiple phases involved and calculates, for example, the pressure distribution in the reservoir. Based on this information the geomechanical models (TNO & UiB, WP1) evaluate the potential for mechanical failures of the caprock/overburden, i.e. formation of new or re-activation of existing faults and fractures, and their distribution. This, in turn, is coupled back to DUMUX to estimate fluxes of leakage from depth to the sediment surface. In the shallow surface sediments, the dispersion of the rising CO_2 plume or percolation of bubbles through the porous matrix, in addition to the fracture flow in DUMUX, is investigated by the HW-LBMC model (HWU, WP1).

---

[1] This includes the occurrence of CO_2 in all possible phases: gas, liquid, supercritical, dissolved, gas hydrate.

DUMUX and HW-LBMC will pass on their water/gas fluxes as well as other geochemical parameters to C.CANDI (GEOMAR, WP2), which specialises in biogeochemical transport-reaction processes occurring near the sediment-water interface and below (down to a depth of ~1000 m). Additional physical and sediment properties, such as porosity and permeability, will be obtained from the geological models (UiT & GEOMAR, WP1). C.CANDI will link the information flow from the subsurface models (WP1) to the water column (WP3) and biological models (WP4). Major processes of relevance that are solved in the model reaction network are, for example, the dissolution of CO$_2$ during its ascent and the induced weathering of the sediment, both of which determine the pH and distribution of the carbonate species. C.CANDI will also treat feedbacks of leakage on microbial POC remineralisation, nutrient fluxes, and biogenic transport processes (bioturbation, bioirrigation).

This information is then exchanged with the model SedROLM (NIVA, WP4), which focuses on redox geochemistry and respective fluxes in the bottom boundary layer, and the ERSEM ecosystem model (PML, WP4). In turn, the biological model provides feedback information on the activity and abundances of benthic organisms as well as solute/particle fluxes from the water column into the sediment (C.CANDI).

Based on the CO$_2$ flux emitted at the sediment surface, predicted by the models in WP1 + WP2, the models in WP3 will determine the fate of the gaseous and dissolved CO$_2$ as well as other solutes in the water column of the ocean. The DBM (GEOMAR, WP3) and the HW-TPPM (HWU, WP3) are specialized in the simulation of gas bubble dissolution and the spread of the CO$_2$ gas plume in the near-field of a release site. They will closely interact with each other and inform the BOM (UiB, WP3), which will then simulate the larger-scale distribution of the dissolved CO$_2$, low pH in the water column, based on information such as ocean currents. All three models will also estimate the possibility of any flux of CO$_2$ back to the atmosphere. Effects of the CO$_2$ dispersion on the water column biogeochemistry is assessed by the OxyDep-Carb model (NIVA, WP3), which will closely interface with the physical oceanographic model (BOM, UiB).

The ERSEM model (PML, WP4) will use the predicted distribution of the dissolved CO$_2$ plume in the water column and analyse the impact of the corresponding pH reduction on benthic and pelagic ecosystem structures and foodwebs. The inclusion of key species that may be used as leakage indicators for monitoring purposes will be investigated.

Besides the calculation of leakage fluxes and their impacts, all of the above physical, chemical, and biological models will also attempt to assess the likelihood of an impact event. To do so, model uncertainties will be analyzed based on an evaluation strategy outlined in section 4. The established general framework of Environmental Risk Analysis (ERA) will be adapted to fit the needs of offshore CO$_2$ storage sites in WP5. This will require a concise, high-level description of leakage and seepage scenarios together with their probabilities. It will further require description of the biotope receptors and their vulnerability with respect to exposure to leaked CO$_2$ and formation water, which is anoxic and can contain toxic substances. Although there exists currently no standardized numerical model to support this new ERA, work flows to perform this exist and will be used to establish an ERA model framework. Finally, the economic models in WP5 will utilize the above information to value the cost of leakage (CO$_2$ itself, mobilized toxic substances etc.). This includes factors, such as investment and operation of CCS infrastructure, clean-up of leakage, liability and insurance, social impacts, monitoring strategies. The models DEMETER (ECN, WP5) and DART (IfW, WP5) assess the economic implications of CO$_2$ leakage on a global as well as a regional scale, thereby also comparing the costs and effectiveness of CCS to other mitigation options. Through these models a close cooperation with the general risk assessment in WP5 is established.

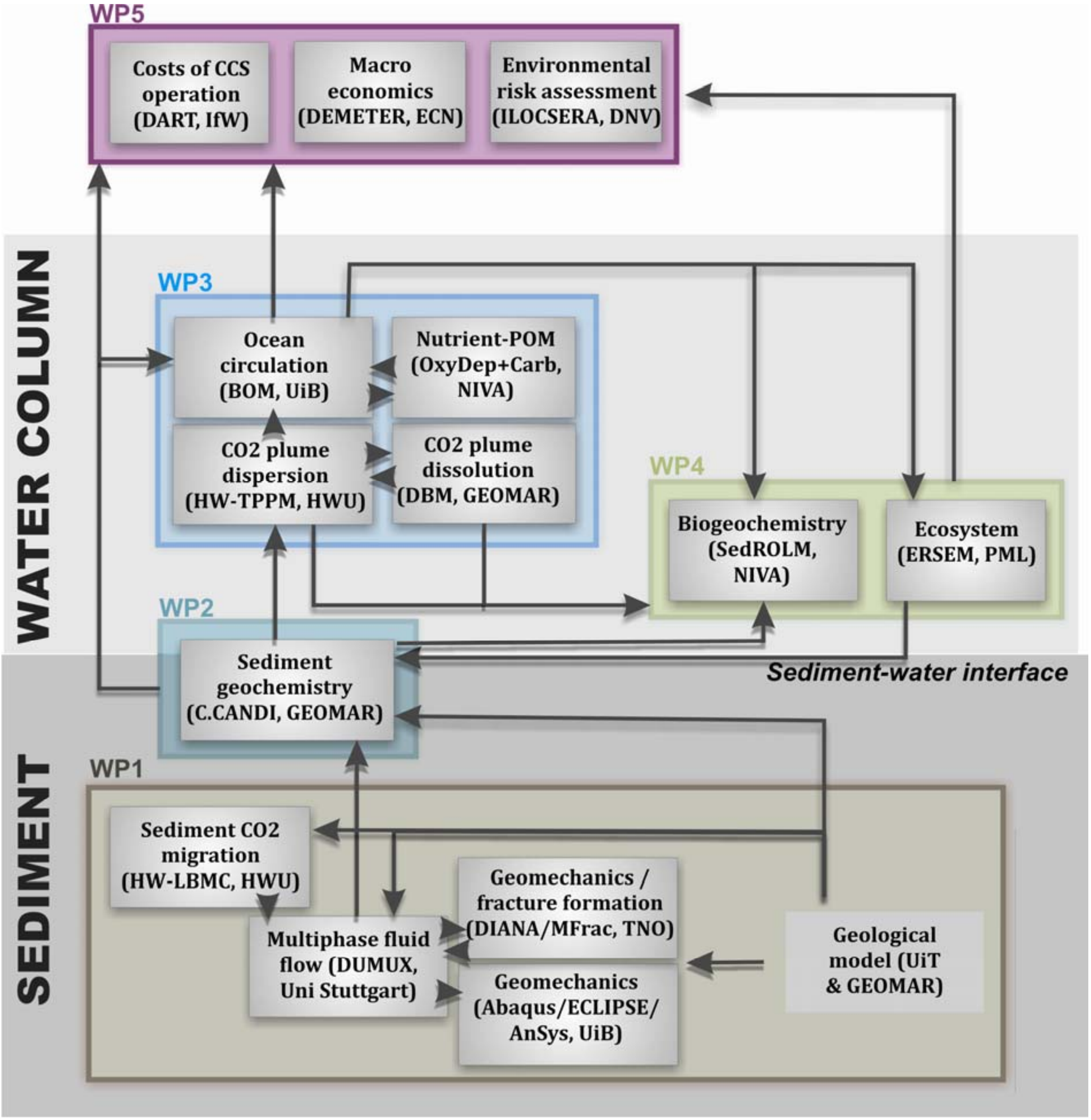


Figure 1: Sketch of main information flow between numerical models and summary of their tasks in the simulation of leakage scenarios. The tasks are in accordance with the respective work packages: WP1 – geomechanics of and multiphase fluid flow in the overburden; WP2 – biogeochemical reactions and transport in the sediments & fluxes across the sediment-water interface; WP3 – spread of CO$_2$ plume and gas bubble behaviour in the water column; WP4 – impact on benthic and pelagic ecosystems; WP5 – economic valuing, costs of leakage, and

environmental risk assessment. Arrows represent a model interface and indicate direction of exchange.

Summary of model interfaces

WP1 – DUMUX (University of Stuttgart, Germany)

Input

- Distribution of potential leakage pathways (i.e. fractures and faults) and information on permeability / porosity from the geological models (UiT & GEOMAR, WP1)
- CO_2 flux through fractures / faults from DIANA/MFrac (TNO, WP1) and AnSys (UiB, WP1)
- Information on CO_2 migration and dissolution rates as well as a hydrate formation model provided by HW-LBMC (HWU, WP1)

Output

- Fluid/gas/CO_2 flux to the seabed for C.CANDI (GEOMAR, WP2) and HW-LMBC (HWU, WP1)
- Pressure and CO_2 saturation distribution for geomechanical models (TNO, WP1)

WP1 – Abaqus / ECLIPSE / AnSys (University of Bergen, UiB, Norway)

Input

- Hydraulic data, like permeability and porosity, from earlier work on ECO2 sites
- Thermal properties and geology from UiT & GEOMAR (WP1) and earlier work on ECO2 sites

Output

- Migration velocity and deformation of overburden for DUMUX (U. Stuttgart, WP1)

WP1 – DIANA / MFrac (TNO, The Netherlands)

Input

- Information on geology (facies and other geological features) from UiT & GEOMAR (WP1)
- Pressure and temperature fields from DUMUX (U. Stuttgart, WP1)

Output

- Geomechanical behaviour, fracture formation of the overburden for DUMUX (U. Stuttgart, WP1)

WP1 – iMoss / Petrel (University of Tromsø, UiT, Norway)

Input

- Geological surfaces from 3D seismic images

Output

- Permeability, porosity and anisotropy to DUMUX (Stuttgart, WP1)
- Structural geological model to DIANA/MFrac (TNO, WP1)
- Information on biogeochemistry to C.CANDI (GEOMAR, WP2)
- Local models of relative permeability of CO_2 and water for HW-LBMC (HWU, WP1)

WP1 – HW-LBMC (Heriot-Watt University, HWU, Scotland)

Input

- Profiles of CO_2 concentrations or dissolution rates from C.CANDI (GEOMAR, WP2)

Output

- Information on CO_2 migration and hydrate formation for DUMUX (U. Stuttgart, WP1) and C.CANDI (GEOMAR, WP2)

WP2 – C.CANDI (GEOMAR, Germany)

Input

- Fluid/gas/CO_2 flux from depth to seabed from DUMUX (U. Stuttgart, WP1)
- Field data on porewater & sediment composition from cruises and the geological models (UiT & GEOMAR, WP1)
- Sedimentation rates, nutrient fluxes, microbial activity from cruises and experiments (WP2, WP4)
- Changes in activity, abundances & distribution of benthic organisms, nutrient fluxes, redox conditions and sedimentation rates from the water column to the sediment from ERSEM (PML, WP4), SedROLM (NIVA, WP3) and BOM (UiB, WP3)

Output

- Profiles of CO_2 concentrations and dissolution rates to HW-TPPM (HWU, WP3)
- CO_2 fluxes into water column for economic models in WP5 (DART, IfW; DEMETER, ECN)
- Fluxes of CO_2 and other solutes for BOM (UiB, WP3) and ERSEM (PML, WP4)
- Information on pH gradients, carbonate system parameters and metal concentrations of surface sediments to SedROLM (NIVA, WP3) and OxyDep+Carb (NIVA, WP4)

WP3 – 1-D plume model / DBM (GEOMAR, Germany)

Input

- Environmental conditions (pressure, temperature, salinity distribution) at leakage site from cruises (WP3)

Output

- Information on distribution of CO_2 gas plume, rise height, dissolution rates for BOM (UiB, WP3), HW-TPPM (HWU, WP3), ERSEM (PML, WP4)
- Fluxes of CO_2 gas into the atmosphere for economic models in WP5 (DART, IfW; DEMETER, ECN)

WP3 – HW-TPPM (Heriot-Watt University, HWU, Scotland)

Input

- Information on ocean currents (time series over several months), temperature, and sedimentation from BOM (UiB, WP3)
- CO_2 fluxes from C.CANDI (GEOMAR, WP2) and DUMUX (U. Stuttgart, WP1)
- Initial bubble size and dissolution rates from DBM (GEOMAR, WP3)

Output

- Information on CO_2 plume behaviour and distribution for BOM (UiB, WP3) and ERSEM (PML, WP4)

WP3 – BOM (University of Bergen, UiB, Norway)

Input

- Background measurements of currents, profiles of temperature and salinity, wind forcing and topography
- Fluxes of CO_2 dissolved in seawater from C.CANDI (GEOMAR, WP2)
- Dissolution rates and vertical distribution of CO_2 plume from DBM (GEOMAR, WP3) and HW-TPPM (HWU, WP3)

Output

- Background current conditions for HW-TPPM (HWU, WP3)
- Information on pressure temperature, salinity, distribution of dissolved CO_2 in the water column and exposure times of particles for OxyDep+Carb (NIVA, WP4) and ERSEM (PML, WP4)
- Fluxes of CO_2 into the atmosphere for economic models in WP5 (DART, IfW; DEMETER, ECN)

WP3 – SedROLM (NIVA, Norway)

Input

- Information on oxygen, carbonate system parameters, redox metals and hydrogen sulphide concentrations of surface sediments from C.CANDI (GEOMAR, WP2).

Output

- Fluxes of nutrients, oxygen, metals etc. from bottom boundary layer to sediments forC.CANDI (GEOMAR, WP2) and to the water column for ERSEM (PML, WP4) and OxyDep+Carb (NIVA, Norway)

WP4 – OxyDep+Carb (NIVA, Norway)

Input

- Fluxes of oxygen, nutrients and carbonate species into the water column from C.CANDI (GEOMAR, WP2) and SedROLM (NIVA, WP3)
- Oxygen, nutrient, carbonate species and plankton distribution from field data and ERSEM (PML, WP4)
- Environmental conditions from BOM (UiB, WP3)

Output

- Oxygen, nutrient, carbonate species and pH for BOM (UiB, WP3) and ERSEM (PML, WP4)

WP4 – ERSEM (Plymouth Marine Laboratory, PML, United Kingdom)

Input

- Environmental conditions and ocean currents from BOM (UiB, WP3)
- Distribution of nutrients in the water column from OxyDep+Carb (NIVA, WP4) and distribution of CO_2 from DBM (GEOMAR, WP3) and HW-TPPM (HWU, WP3)
- Fluxes of CO_2, pH, TA, nutrients from the sediment from C.CANDI (GEOMAR, WP2)

Output

- Impact of CO_2 leakage on ecosystems for economic models in WP5 (DART, IfW; DEMETER, ECN)
- Feedback of activity, abundances, and distribution of benthic organisms, fluxes of solutes and sedimentation rate from the water column to the sediment for C.CANDI (GEOMAR, WP2)

WP5 –DART (IfW, Germany)

Input

- Probability of leakage
- Re-flux of CO_2 into the atmosphere

Output

- Costs of leakage remediation
- Economic valuation of CCS operation

WP5 –DEMETER (ECN, The Netherlands)

Input

- Leakage rates (to the ocean and atmosphere) and damage caused by leakage (from pH lowering, released toxic substance), impacts on ecology

Output

- Assessment of global economic, climatic implications of leakage, effectiveness of CCS

WP5 – ILOCSERA[2] (DNV, Norway/UK + TNO, The Netherlands)

Input

- Site-specific inventory of discrete leakage and seepage pathways from target storage unit to seafloor
- Site-specific inventory of key biotope populations and their seasonal movements, variations in size, life-cycle phases and respective vulnerability profiles and potential for re-generation after impacts from exposure to leaked or seeped CO$_2$ or formation water
- Site-specific inventory of discrete leakage/seepage scenario descriptions with appropriate probability and rates (to the ocean and atmosphere) and damage caused by leakage (from pH lowering, released toxic substances), impacts on ecology

Output

- Assessment of site impacts and risks in the context of an Environmental Risk Assessment process for permitting and public acceptance

Model evaluation strategy

Numerical models generally simulate the complex network of natural processes in an abstracted form trying to reproduce individual processes and the entire system as accurately as possible. A final purpose of numerical simulations is to predict the reaction of a complex system to certain changes of the external forcing. Since there is no independent means of verifying the correctness of the model prediction, an evaluation strategy has to be developed.
This strategy involves several steps (see also Fig. 2): (1) verification of the numerical solver, (2) calibration of the model, (3) validation of model results, and (4) sensitivity analysis of model parameters. Since step 1, verification of solvers, is a task the model developers will have to perform each time a change is made in the code, it will not be discussed any further in this context.
Most models have also been calibrated by applications to various field data and environmental settings (step 2). However, through the field campaigns in ECO2 an opportunity to reduce uncertainties, in e.g. parameter estimates, will be given. Uncertainties in parameters will propagate through the model system and have direct influence on the quality of the output. The uncertainties in numerical simulations are often categorized as aleatory (stochastic), known uncertainties with a probability distribution, and epistemic, uncertainties due to lack of knowledge (Oberkamp and Roy, 2010). The latter can further be divided into known unknowns, i.e. we know that they exist but at best can provide a (rough) interval estimate on

---

[2] Integrated Life cycle Offshore CO$_2$ Storage ERA. This is more a methodology and work flow than a standardized numerical simulation model with standardized inputs and will be produced as part of the input from WP5 to WP14 (CCT4).

the uncertainty, and unknown unknowns, i.e. we do not know that our knowledge is incomplete. We should aim for bringing the epistemic uncertainties toward aleatory, and the ECO2 field campaigns will provide valuable data for this purpose. Hence, specific calibrations will be implemented by modelling field and experimental data that are collected within the course of the ECO2 project (see documents MS5 and MS6).

Step 3 comprises a model validation procedure or model skill assessment: How accurate are the model predictions? How large are the uncertainties? It intends to test each model's capabilities and shortcomings. Typically, this involves the comparison of measured responses from in situ field data, laboratory experiments or natural analogues with the output of the numerical models. The required data can also be acquired as part of the ECO2 field campaigns.

Care will be taken not to use the same data for model validation that were used for model calibration. While calibration experiments or measurements aim for a controlled environment and repeatability of the experiments to estimate internal model parameters, the scope of the validation experiments is to characterize the experimental system and surrounding. The emphasis is not on repeatability, but to precisely measure the conditions of an uncontrolled experiment. Variability in the surroundings of a validated experiment is not critical, as long as these are precisely measured. For experiments with uncontrolled conditions, a number of experimental realizations will be necessary in order to capture the variability of the system. It will provide boundary conditions and initial conditions for the model system, and highlights the data necessary for validating the model.

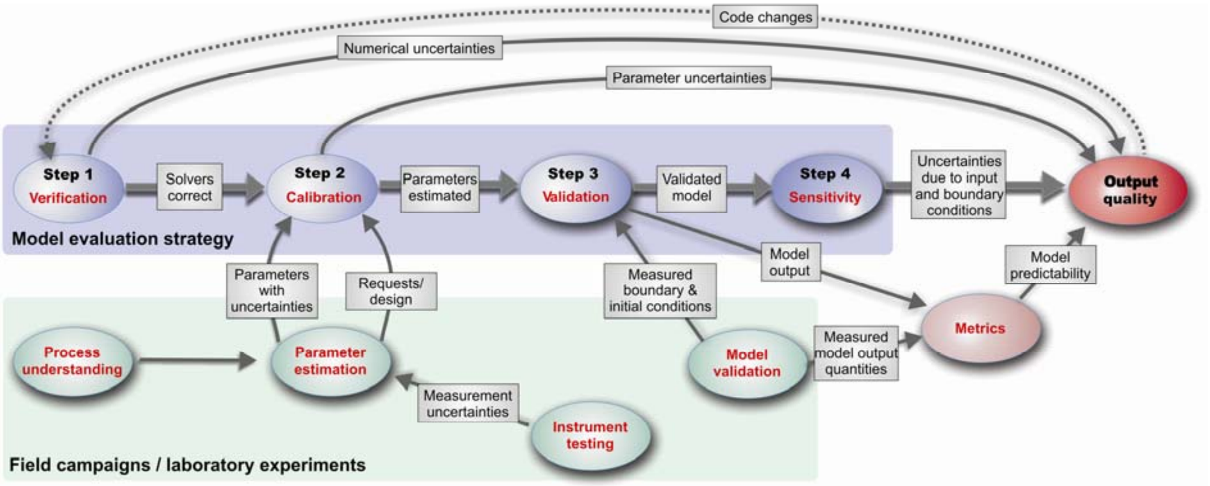


Figure 2: Overview of model evaluation strategy steps indicating the interrelationships between model-specific evaluation processes (blue box) and related measurements from field campaigns & laboratory experiments (green box) as well as statistical metric methods.

In order to quantify the (dis)agreement between model predictions and actual observations several statistical formulations, so-called metrics, exist. Some of these measures provide information on the general tendency of the model predictions to vary with the observations (correlation coefficient) or allow characterizing the misfit between observations and the model results, such as the cost function, root mean squared error, average error, or average absolute error. If one would like to further quantify the misfit of the model results, the skewness, modelling efficiency, percentage bias, or reliability index can be calculated. For example, they provide information on the model tendency to over- or underestimate with respect to observations. Aside from a means to classify the model performance, the metric procedures are useful in identifying when and where the model is less reliable, where errors propagate and also where improvements are required.

Such statistical evaluation of the model output can only be applied to simulations where data is available, e.g. background situation, lab experiments, or natural analogues. Hence, the metric analyses give hints towards of the general quality of model simulations that can be used to evaluate the leakage scenario simulations, but it cannot analyse the scenarios directly. Therefore, a fourth strategy is to analyse the sensitivity of the model result with respect to variations of relevant and important model parameters (step 4). During a sensitivity study all independent input parameters are varied over a range of values and the model run with each individual parameter setting, thus allowing a quantification of the parameters influences on the model outcome. This approach gives valuable insight into the robustness of each model and also allows assessing the impact of the uncertainty that is passed on from model to model on the simulated scenarios.

Lastly, the reliability of the model results as indicated by metric and sensitivity approaches can be increased by overlaps in the model domains. If two models show the same result for the overlapping domains, it is more likely that the respective model predictions are accurate. For example, the computation of the pH value is inherent to several models, such as DUMUX (Stuttgart, WP1), C.CANDI (GEOMAR, WP2), BOM (UiB, WP3), and ERSEM (PML, WP4), which also spatially overlap at their interfaces. Thus, if the computed pH values agree at those overlapping points in space and time, it is far more probable that the result is acceptably accurate. It should be noted, however, that this method is limited by the fact that the different models compute parameters and processes with varying complexity. Thus comparing the output of a model specialised in the computation of this particular parameter with the output of a model that has a different emphasis, will provide only a rough estimate of the models performance.

In conclusion, the evaluation strategy of the CCT2 models will focus on a combined metric and sensitivity study to rectify any known detectable flaws and to ensure internal model consistency. When this is achieved, the joint CCT2 model scenario can be said to be valid.

APPENDIX – Evaluation metrics

No one metric is able to give a complete picture of a model's performance. Therefore a suite of metrics is generally used, each of which extract a particular quality of the model – data comparison. The following metrics are based on Stow et al. (2009) and Allen et al. (2007).

Table 1: Summary of evaluation methodologies.

Metric		**Indicates**	**Ideal**
R/R^2	Correlation coefficient	Covariance	1
CF	Cost function	Goodness of fit	0
Pbias	Percentage model bias	Bias and direction	0
MEF	Nash & Sutcliffe Model Efficiency	Prediction relative to average	1
RMSE	Root mean squared error	Mean of discrepancies	0
RI	Reliability index	Average deviation	1
AE	Average error	Bias & direction	0
AAE	Average absolute error	Absolute bias	0
Skew	Estimate of skewness	Asymmetry c/w data	0
RSD	Ratio of model std to data std	Comparison of variance	1

Some metrics, such as the correlation coefficient, the reliability index and the model efficiency, are dimensionless or scaled by the magnitude of the particular variable. Other metrics, such as the average error and root mean squared difference, score relative to the mean and variance of the data and must be interpreted as such. Example: a RMSE score for dissolved inorganic carbon (average value of DIC in the surface ocean is around 2050.0 μmol/kg) cannot be directly compared to the score for pH (average value in the surface ocean is around 8.1, and a log-scale).

The principle metrics are described in more detail below using the following notations:

n = the number of observations
O_i = the ith of n observations (data)
P_i = the ith of n predictions (model output)
$\overline{O}$ = the average of the observations
$\overline{P}$ = the average of the modelled predictions

A1 Correlation coefficient of the model predictions and observations (R/R^2)

The correlation coefficient is the tendency of modelled predictions and observations to vary together. An absolute value near one is considered a close match (negative values indicate inverse variation). However, this metric does not penalise a constant offset, so even if the correlation is near one, the predicted and observed values may not match each other. Additionally, the correlation coefficient is sensitive to small numbers of extreme values that may not reflect the behaviour of the bulk of the data. Further, comparisons between similarly behaving but out of phase data sets would score poorly.

$$r=\frac{\sum_{i=1}^{n}(O_i-\overline{O})(P_i-\overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i-\overline{O})^2\sum_{i=1}^{n}(P_i-\overline{P})^2}}$$

A2 Cost function (CF)

The cost function indicates a 'goodness of fit' between model predictions and data (OSPAR, 1998), measuring the ratio of the model – data misfit to the variance of the data. A value near 0 indicates a good fit. Some debate in the literature exists regarding interpretation, however Radach and Moll (2006) suggest that values below 1 are very good, between 1 and 2 are good, between 2 and 3 reasonable and exceeding 3 poor. σ_O is the standard deviation of the data.

$$CF=\frac{1}{n}\sum_{i=1}^{n}\frac{|O_i-P_i|}{\sigma_O}$$

A3 Bias (Pbias)

The percentage bias measures whether the model is either systematically over or under estimating with respect to observations. A value near 0 is optimal. Marechal (2004) defined scores of under 10 as excellent, 10-20 as very good, 20-40 as good and over 40 as poor.

$$bias=\frac{\sum(O-P)}{\sum O}*100$$

A4 Modelling efficiency (MEF)

The MEF measures how well the model predicts relative to the average of the observations (Nash & Sutcliffe, 1970). A value near one indicates a close match. A value of zero indicates that the model predicts individual observations no better than the average of the observations. Values less than zero indicate that the observation average would be a better predictor than the model results. However, since the MEF as defined here is influenced by bias, model fields with a bias might have an MEF score below 0, whilst still reproducing the patterns and variability of the observations well.

$$MEF=\frac{\left(\sum_{i=1}^{n}(O_i-\overline{O})^2-\sum_{i=1}^{n}(P_i-O_i)^2\right)}{\sum_{i=1}^{n}\left(O_i-\overline{O}\right)^2}$$

A5 Root mean squared error (RMSE)

RMSE is a measure of the mean of discrepancies between observed and predicted values. A value near zero represents a close match. However it fails to tell you in which direction the error lies.

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(P_i-O_i)^2}{n}}$$

A6 Reliability index (RI)

The RI is the average factor by which the model deviates from the observational results. A value near one indicates a close match, a value of two indicates that the average model result is twice the value of the data. However this metric doesn't assess patterns in the model or data.

$$RI = \exp\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log\frac{O_i}{P_i})^2}$$

A7 Average error (AE)

AE is another indicator of bias and measures the size of discrepancies between observed and predicted values. A value near zero represents a close match, although can be misleading because negative and positive discrepancies can cancel each other.

$$AE = \frac{\sum_{1=1}^{n}(P_i - O_i)}{n} = \overline{P} - \overline{O}$$

A8 Average absolute error (AAE)

AAE is a measure of the size of discrepancies between observed and predicted values. A value near zero represents a close match. This overcomes the problem of the Average Error statistic of possible cancelling of negative and positive discrepancies, but gives no indication of the direction of discrepancy.

$$AAE = \frac{\sum_{1=1}^{n}\left|P_i - O_i\right|}{n}$$

A9 Skewness (Skew)

The skewness metric characterises the degree of asymmetry of a distribution around its mean. A positive value indicates that the model tends to make more under estimations, a negative value indicates that the model is prone to producing a large number of over estimations. Skewness values of 2 standard errors or more (1 SE can be approximated as $(6/n)^{0.5}$) can be interpreted as substantially skewed.

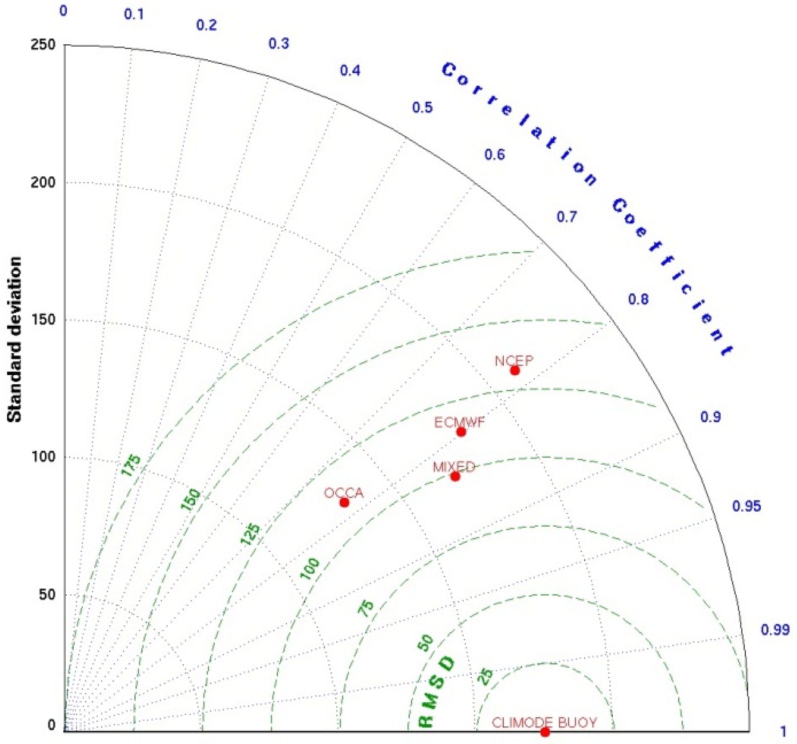
$$Skew = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{(O_i - P_i) - \overline{(O_i - P_i)}}{\sigma_O}\right)^3$$

A10 Ratio of standard deviations (RSD)

This metric gives a simplified comparison of how variance in each data set compares.

A11 Taylor Diagram

The Taylor diagram (Taylor, 2001) is a method of graphically summarizing how well model predictions match observations, based on correlation, the centred root-mean-square difference and the ratio of their standard deviations. These diagrams are a useful tool with which to compare performance of several variables within a complex model or to gauge the relative performance of a suite of model systems. The optimal position is marked on the x-axis as 'Climode buoy'.



References and some more literature on the topic

Allen J.I., Holt J.T., Blackford J., Proctor R. (2007) 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. (2009) A multivariate approach to model skill assessment. Journal of Marine Systems 76, 83-94.

Arhonditsis G.B., Brett M.T. (2004) Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Marine Ecology Progress Series 271, 13-26.

Marechal D. (2004) A soil-based approach to rainfall-runoff modelling in ungauged catchments for England and Wales. PhD thesis, Cranfield University. 157 p.

Nash J.E., Sutcliffe J.V. (1970) River flow forecasting through conceptual models: Part I – A discussion of principles. Journal of Hydrology 10, 282-290.

Oberkampf W.L., Roy C.J. (2010) Verification and validation in scientific computing. Cambridge University Press. ISBN 978-0-521-11360-1.

Radach G., Moll A. (2006) Review of three-dimensional ecological modelling related to the North Sea shelf system. Part II: Model validation and data needs. Oceanography and Marine Biology: an Annual Review 44, 1-60.

Stow C.A., Jolliff J., McGillicuddy D.J., Doney S.C., Allen J.I., Friedrichs M.A.M., Rose K.A., Wallhead P. (2009) Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems 76, 4-15.

Taylor K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research 106(D7), 7183-7192.