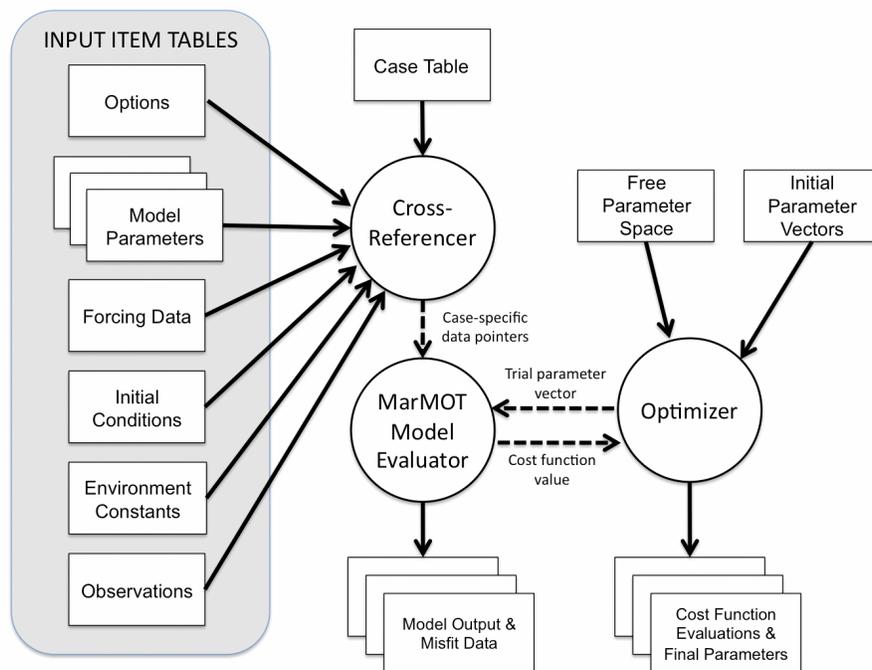


The Marine Model Optimization Testbed (MarMOT)

John Hemmings

National Oceanography Centre, UK



v1.1 System Overview

21 Nov 2013

The MarMOT testbed, introduced by Hemmings & Challenor (2012), provides a common physical and computational environment in which different plankton ecosystem models can be calibrated and compared with each other and with biogeochemical data. It is designed especially for computationally intensive experiments in which models are run in a 1-D framework many times with different input data. Features are provided to support multi-site ensemble water column simulations in a realistic 3-D context, using external environmental data from analyses of 3-D simulations and/or real-world observations.

A particular simulation case is defined by a combination of instances of each input item, selected via a simple case configuration record. Multi-record *case tables* can be used to

set up ensemble runs for sensitivity and uncertainty analyses. Individual cases are identified by site and ensemble member labels. Adjustable model parameters can be optimized over all cases in a given case table to set up multi-site and/or multi-member calibration experiments.

The core of the MarMOT system is a *model evaluator* that returns a misfit-based cost function value for one or more simulation cases. It is implemented as a specific application within a system referred to as the Generic Function Analyzer (GFAn). GFAn provides a generic data management framework that adapts to the requirements of the application, together with an optimizer for function minimization and a flexible experiment control interface. Input and output files use a standard table format with ASCII and binary versions. A separate set of programs called Table Toolbox is available for working with tables in this format. GFAn and Table Toolbox are written in C. The MarMOT Model Evaluator is written in C (user interface and I/O management) and FORTRAN (simulator procedures and plankton model interface). The software is designed for a UNIX/LINUX system environment.

Ref: Hemmings, J. C. P. and Challenor, P. G., 2012. Addressing the impact of environmental uncertainty in plankton model calibration with a dedicated software system: the Marine Model Optimization Testbed (MarMOT 1.1 alpha). *Geosci. Model Dev.* 5, 471-498.

MarMOT 1.1 Software Features

MarMOT Model Evaluator

- 1-D water column framework with implicit horizontal fluxes due to vertical divergence in the flow (zero horizontal gradients assumed); parameterization of explicit horizontal flux divergence effects by mean and/or autoregressive random perturbations for individual tracers; support for dependency of perturbations on local tracer concentrations (using tracer transformations); options for relaxation towards reference fields and/or direct replacement of individual tracers by prescribed external data
- Photosynthesis light limitation options: light attenuation coeff. from Anderson spectral model approximation with depth & pigment dependency (Anderson, T.R., 1993, *Limnol. Oceanogr.* 38, 1403-1419) or simple linear pigment dependency, optional noon path-length adjustment; chlorophyll-specific absorption from Anderson 1993 spectral model approximation or constant; photosynthetic response from Evans and Parslow daily mean model (Evans, G.T. & Parslow, J.S., 1985, *Biol. Oceanogr.* 3, 327-347) or Platt et al. daily mean model (Platt, T., Sathyendranath, S. & Ravindran, P., 1990, *R. Soc. Lond. Ser. B* 241, 101-111) or point-in-time photosynthesis model for resolving diel cycle
- Support for sequential chlorophyll data assimilation (Hemmings, J.C.P, Barciela, R.M. & Bell, J.M., 2008, *J. Mar. Res.* 66, 87-126)

- Simulation input tables (Each item is optional by default - MarMOT provides default values for missing data where appropriate and checks that configuration-specific data requirements are met)
 - **Model parameter set items** - one table for each ecosystem model
 - **Data assimilation parameter set items** - one table for each ecosystem model with chlorophyll data assimilation
 - **Option parameter set item**: ecosystem model selection (initialized variables are treated as passive tracers if no model), photosynthesis options & controls, advection scheme selection, diffusion and mixing controls, perturbation options & controls, tracer relaxation controls, time-stepping parameters, temporal interpolation modes for forcing data, default tracer relaxation rates & initial conditions
 - **Time axis parameter set item**: base year, periodic forcing switch for annual forcing, year length
 - **Time period parameter set item**: simulation start and finish times
 - **Environment parameter set item**: maximum depth, latitude, default values for forcing variables below
 - **Perturbation seed parameter set item**: seed values for autoregressive random perturbations
 - **Misfit control parameter set item**: variable-specific log or sqrt transformations for model-data misfit calculation
 - **Vertical grid item** (on level grid): depths for up to 500 levels
 - **Initial profile item** (defined on level grid): initial profiles of primary tracers and/or composition ratios for deriving secondary tracers (default to values in option parameter set item)
 - **Scalar forcing item/items** (defined on an arbitrary time grid): solar irradiance, mixed layer depth, surface salinity, atmospheric pCO₂, dust deposition, chlorophyll observations for data assimilation, depth-invariant full-depth forcing variables (default to constant forcing from environment parameter set item or option item); option for dividing variables between multiple items on different time grids
 - **Full-depth profile forcing item/items** (defined on level and arbitrary time grid): vertical diffusion coefficient, vertical velocity, temperature, means and standard deviations for tracer perturbations, relaxation rates and reference profiles for tracer relaxation (default to depth-invariant forcing from scalar forcing item); option for dividing variables between multiple items on different time grids
 - **Observation set item** (defined at arbitrary times and depths): 'observed' values of any output variable, observation weights; can use previous output as synthetic observations

- Simulation output tables (Data are selectable for output at table or variable level)
 - **Options table:** options applicable to simulation including active defaults
 - **Surface and scalar variables tables** (end of time step, daily mean, monthly mean & annual mean versions): top-level values of full-depth variables below plus surface forcing and scalar diagnostics
 - **Full-depth profile tables** (end of time step, daily mean, monthly mean & annual mean versions): full-depth forcing, primary tracers, composition ratios, derived tracers and full-depth diagnostics
 - **Surface and scalar data assimilation variables table** (output at analysis times): surface/scalar data assimilation diagnostics
 - **Full-depth data assimilation profiles table** (output at analysis times): full-depth data assimilation diagnostics
 - **Misfit table:** simulated and observed values, model-data differences, weights and weighted misfits (squared differences) at each observation point
 - **Cost table:** mean weighted misfit over all observations
- Output cost function value given by mean weighted misfit over all observations for all specified simulation cases

GFAAn Data Manager with Multi-case Support

- Supports multi-site and ensemble runs via an input *case table*; each record, identified by alphanumeric *case variables* 'site' and 'member', determines input for one case
- Each simulation input table contains a single instance of a data item or multiple instances identified by one or both case variables or by an item-specific key
- Cross-referencer selects instances of input items for each case by context, using case variable(s), or explicitly, using item-specific keys
- Solutions for different cases are identified by their case variables in the simulation output tables

GFAAn Optimizer

- Seeks minimum of cost function (over all cases specified)
- Parameters for each optimization experiment are selectable by name from any one input parameter set item via a *free parameter table*
- Micro-genetic algorithm (Krishnakumar, K., 1989, Proc. SPIE: Intelligent Control and Adaptive Systems 1196, Philadelphia, PA, 289-296), based on an

implementation by Carroll (Carroll, D. L., 1996, in *Developments in Theoretical and Applied Mechanics*, Vol. XVIII, eds. Wilson, H. B, Batra, R. C., Bert, C. W. Davis, A. M. J., Schapery, R. A., Stewart, D. S. & Swinson, F. F., School of Engineering, The University of Alabama, pp.411-424)

- User-definable bounds for each parameter or log-transformed parameter
- Initialization from input table of parameter vectors or random
- Local non-gradient direction set algorithm (Powell, M.J.D., 1964, *Computer J.* 7, 155-162) using Brent's line minimization algorithm (Brent R.P., 1973, in: *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, N. J., 61-80)
 - User-definable bounds or no bound option for each parameter or log-transformed parameter
 - Repeat searches with different initial parameter vectors from input table or genetic algorithm output
- Table of final parameter vectors and costs; optional output of all cost function values and corresponding parameter vectors to *evaluations table(s)*
- Detailed optimizer report
- Simulation output tables produced for final optimal parameter vector

GFAAn Experiment Control Interface

- Handles multiple experiments, with and without optimization, in batch runs
- Works with ASCII format tables with records in rows and variables in columns (variable names in row 1) or accepts a transposed format (variable names in column 1) for single record tables; alternative double precision binary format (Table Toolbox Binary)
- Supports extraction from one or more NetCDF data sets onto a user-defined grid as an alternative to ASCII/binary input for gridded items; option to save extracted data as ASCII/binary tables for later use
- Assigns experiment-specific file names to each input and output table
- Avoids expensive re-loading of resident data where input file names are unchanged between experiments
- Comprehensive log output includes view of all input data or selected tables only
- Generates templates for input tables on request

Ecosystem Models Supported by MarMOT 1.1

- NPZD model of Oschlies and Garçon - Ref: Oschlies, A. & Garçon, V., 1999, An eddy-permitting coupled physical-biological model of the North Atlantic. 1. Sensitivity to advection numerics and mixed layer physics, *Global Biogeochem. Cy.* 13, 135-160
- HadOCC (UK Met Office code; restricted availability) - Ref: Palmer, J.R. & Totterdell, I.J., 2001, Production and export in a global ocean ecosystem model, *Deep-sea Research I* 48, 1169-1198; current version includes a variable C:Chl ratio (Geider, R.J., MacIntyre, H.L. & Kana, T.M., 1997, *Mar. Ecol. Prog. Ser.*, 148, 187-200) and modified pathways for products of grazing and mortality; chlorophyll data assimilation is implemented
- MEDUSA 1.0 - Ref: Yool, A., Popova, E. E., Anderson, T. R., 2011. MEDUSA-1.0: a new intermediate complexity plankton ecosystem model for the global domain. *Geosci. Model Dev.* 4, 381-417

Supporting Software Tools: Table Toolbox

- Suite of command line tools for creating, viewing, processing, joining and splitting GFAN compatible data files in ASCII or Table Toolbox Binary formats
- Programs support standard input and output streams, allowing complex operations combining functions of different tools to be performed by concatenating commands via UNIX/LINUX pipes