

Organism Knowledge Base

Using and adding to the OKB with carbo*

Chris Jenkins, Don Potts, Peter Burgess
(INSTAAR, UCSC, RHUL) 19 Aug 2013



Strategy for the OKB

- The OKB is a data/knowledge resource on the properties of organisms for use in numerical models. The immediate application is with the carbo* suite of population-ecology models for seafloor carbonate-production
- The OKB table is human readable – for efficient use and maintenance; it is also computer readable to allow for complete, accurate transfer of information to the models
- An intermediate computing step (by carboOKB.py) transfers properties and values from the human-readable, somewhat variable codes to standardized model variables; it also performs completeness and QAQC tests
- Vocabularies on the word-based data and units of measurement are narrow, but not controlled. They are extensible. Unexpected cases raise a prompt to the operator, asking for the program code to be updated to deal with them.
- We envision the OKB eventually becoming a public, community data-resource

Why a 'Knowledge Base' ?

- A spreadsheet or data table is not a knowledge base (KB). A KB needs to involve machine readability, logical structure, deductive reasoning.
- The carbo* involves elements of these, achieved by the combination of human-interface with metadata, deduction powers for example to convert and in-fill data, machine readability for model pickup of the data.
- The carbo* OKB package holds:
 - _carboOKB_7.xlsx: human readable form, to which all editing and metadata should be confined
 - _carboOKB_7.txt: exported text version of that
 - carboOKB_2.py: the Python interpreter, QAQC checker, producer of machine-suitable text
 - _carboOKB_7.okb: the machine-suitable stage
 - carboOKB_Docn_2.pdf: documentation (this file)

Notes: Version numbers, 7,2,1 as of Aug 2013. Underscore on a file indicates it is needed by or produced by the carbo system.*

What is in the OKB ?

- A select group of organisms to represent the functions of carbonate production, competition and symbioses; actual species are used to represent functional groups; those organisms are chosen to be widespread, well-documented (e.g. in reproductive strategy), have a stable taxonomy
- The organisms are set to survive, grow and compete in what we call a 'numerical aquarium' – a population ecology model
- Example functional groups are: massive coral, branched coral, bryozoan, halimeda green algae, encrusting calcareous algae, borers, grazing fish, larger forams, blue-green algae, boring echinoids, grazing echinoids
- Information on organism's: sizes, growth and reproduction, skeletal forms and compositions, susceptibility to external events; habitat preferences, etc
- As the organisms survive, grow and compete they leave a carbonate record which is accumulated in the models reefs, banks, sediments, etc

One species in the OKB

#parameterName	parameterValue	parameterUnit	dataDetails	parameterDetails
Organism=	Halimeda_tuna	5	Green alga, calcified	
.AlternativeTaxa		\$		other names/organisms included
.Size	mx=10	cm		linear size
.Occupancy	mx=224; av=50	#/m2		Observed areal coverages
.Tissue.AreaFraction	10	%		Areal fraction (vertical view) of living tissue
.Skeleton.AreaFraction	10	%		Areal proportion of skeletal frame (view vertically)
.Skeleton.GrowthRate	100	%/yr	complete replacement	linear extension
.Skeleton.Morphology	attached(holdfast); articulated frondose	\$		Descriptive
.Skeleton.Zonation	[0;0;100;0]	%	% presences of creature's materials	[%Presence] Set
.Skeleton.Disintegrated.Fraction	5	%		Fraction volume it will disintegrate to after death/transport/burial
.Skeleton.Disintegrated.Grainsize	1	cm		Clast/grain size after death/transport/burial
.Enviro.Threats		\$		Limitations on growth/fecundity/survival
.Enviro.Prefs	coral reef lagoons; mid-water banks; nutrient flux	\$		Preferences for growth/fecundity/survival
.Enviro.Range.irrad	[0,0;100,50;800,100;1600,66]	umol_photons/m2/s		[EnviroParam, Fitness] LineSet
.Enviro.Range.salin	[15,0;36,100;40,0]	[psu,%]		[EnviroParam, Fitness] LineSet
.Enviro.Range.nutr		ppm		[EnviroParam, Fitness] LineSet
.Enviro.Range.temp	25>>32	degC		[EnviroParam, Fitness] LineSet
.Enviro.Range.wd	[0,0; 8,1; 20,10; 55,19; 76,15; 121,0]	m		[EnviroParam, Fitness] LineSet
.Enviro.Range.oxygen		ml/l		Dissolved Oxygen
.Mortality.Longevity	1	yr		Maximum, usual lifespan
.Mortality.Rate	mx=100;mn=30	%/yr	actually /284d; orig as #/yr - changed by cjj to %/yr	Rate of natural, unexplained mortality
.Mortality.Type	seasonal dieback	\$		Causes of mortality even in optimum conditions (e.g predation, dieback)
.Repro.Clone.ProductionRate	2.3+1.7	#/adult	individual; instantaneous	Clones produced per adult per year under normal conditions
.Repro.Clone.SuccessRate	1	%		Successful formation of colonies from clones (e.g., unattached colonies)
.Repro.Clone.Type		\$	with_dieback	
.Repro.Spawn.MaturitySize		m		Organism/colony size to commence spawning
.Repro.Spawn.ProductionRate	600	%/yr	+1x each 2 mo	Spawn (eggs,larvae) produced per adult per year
.Repro.Spawn.SuccessRate	10	%/yr		Propagules surviving to grow; /m2 instantaneously
.Repro.Spawn.Type	gametes; holocarpic; summer to early autumn	\$		The process of reproduction; seasonality

- Properties are arranged into super-properties such as ‘.Enviro’ for environmental preferences
- Each of these has specific properties and even sub-properties such as ‘.Enviro.Range.temp’
- The values for the .Properties may be numeric or word-based

The Syntaxes of the OKB

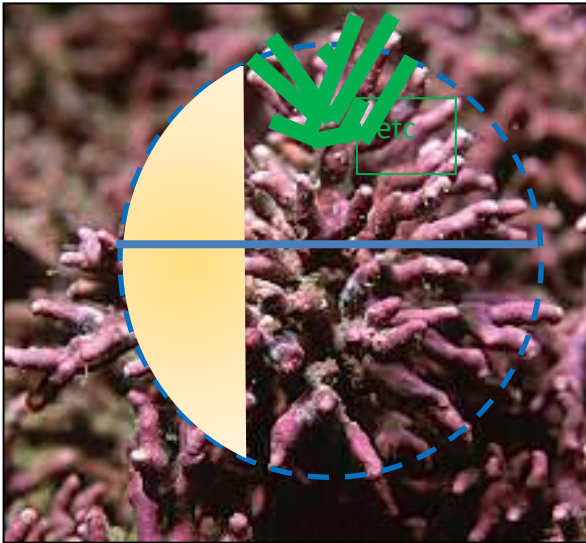
- The layout is a familiar one – an Excel table with columns for: *parameterName*, *parameterValue*, *parameterUnit*, *dataDetails*, *parameterDetails* under each organism
- The parameters are grouped into super- parameters such as ‘.Enviro’ for environmental preferences; and they may have sub- parameters such as ‘.Enviro.Range.temp’
- The parameter values may be numeric or word-based. They are encoded, with 3 formatting options:
 - “av=2.0; mn=1.8; mx=2.4; md=2.0” setting max, min ranges or average values; or “ty=100” giving a typical value
 - “aggregated; globular” giving word-values for a parameter; the vocabulary is narrow but not controlled; unexpected terms lead to a prompt for the operator
 - “[200,0;300,50;1300,50]” giving a linearly-segmented range like a fuzzy membership; here the membership ramps from 0% at 200 to 50% at 300, flat-lining to 1300, then it is assumed that it returns to 0% thereafter.
 - Only one type may be used in a parameter line
- The units of measurement are a narrow set. They are transformed to a single unified set by the program that prepares the OKB for the model.

Standardizations

- Obviously, there may be several different ways of measuring a parameter. The OKB documentation sets out the concept of measurement method as it interprets data.
- It is the responsibility of OKB preparers to approach that standard as closely as possible. If that can't be done, then some commentary should be entered on the right of the OKB table entry.
- Basically, enterers are encouraged to enter data as it appears in the original sources – for instance with units. Then the translating program will work on those to produce the unified, standardized values.

Some Definitions

Dimensions and Stocks

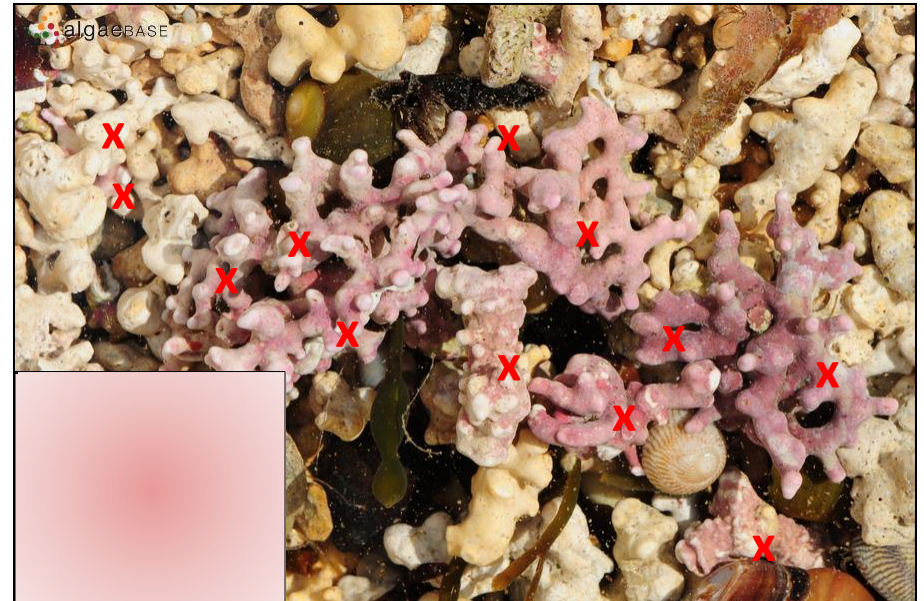


colonySize (diameter; m)

frameFraction (seen from vertical; /1;
obviously quite small with branching)

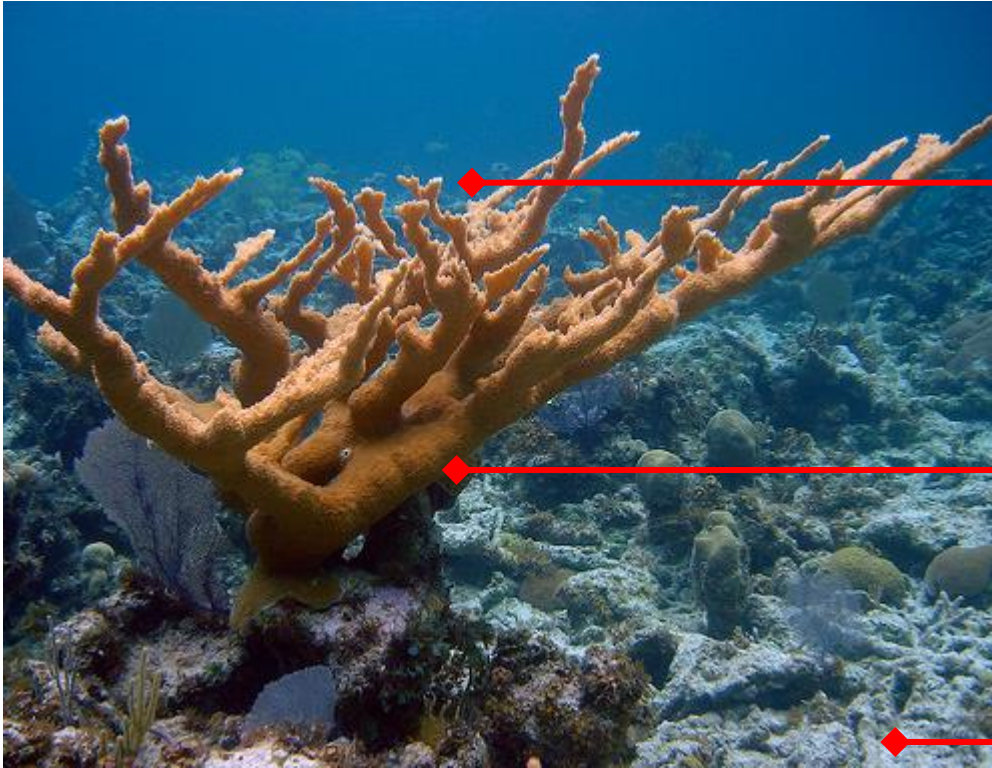
colonyArea (colonySize*frameFraction; m²;
here only a small fraction of the
encirclement)

creatureStock (colonyArea*numbers;
live, attached or not; /m²)



Vocabulary Example

Structural Bin Levels



Superstructure – collapses,
more water movement

Framework – remains in place,
inside benthic boundary layer

Bed – accreting, sometimes
mobile, shaded

Underworld – solid,
enclosed, has diagenesis



CORE OF A DECOMMITTED REEF FACIES OF MAINLY CORAL SKELETONS

Fig. 1

(Not) The Last Word...

- An acknowledged problem with population-ecology models is the need for copious detailed data about the organisms being modeled. The OKB is a method to deal with and mitigate that problem.
- We are exploring the way forward ...