CarboKB - A Knowledge Base Supporting Carbonate Modeling

Donald Potts, **Rachel Fabian, Helen O'Brien,** *University of California at Santa Cruz, California, USA*
Chris Jenkins, *University of Colorado, Boulder, USA* **Peter Burgess**, *Royal Holloway, University of London, UK*

INTRODUCTION

Models need data. For some modeling tasks quite a lot of detailed data is critical, as is the case with the population-ecology modeling of marine carbonates as produced by corals, bryozoa, molluscs, foraminifera, calcareous algae, and micro-organisms.

We show an example of a Knowledge Base (KB; i.e., a database with internal inferencing) that is supporting population ecology modeling of carbonates connected with the CSDMS facility.

A goal is that the same KB will be useful to other modeling efforts within the CSDMS and more widely. When that becomes the case, the KB should be an attractor of further community data inputs, perhaps even from programs unconnected modeling – from by data and analyses projects.

As shown in figure 1, carbonate production, accumulation to form sediment and rock is a complicated set of processes: PRODUCTION of by biological growth and mineral/ambient water processes; DESTRUCTION of biological agents (i.e. 'bioerosion' by fishes, urchins, polychaetes, molluscs, sponges, algae, micro-organisms) and physical processes (e.g. abrasion, waves, storms); SEDIMENTATION suspension, transport, deposition and cementation of biological chemical and physical agents.

Of course, where the agents are organisms, the process rates and styles of the products will be determined by the population abundances of those organisms. The KB provides the essential parameters by which to calculate the populations, and various competition and feedback mechanisms that can be expressed mathematically.

Important premises underlying the adoption and content of the KB include:

1. That despite the complexity and diversity of biological communities, a few KEY SPECIES (e.g., Figs 2,3) dominate the ecological processes of production and destruction in a particular system.
2. MODELS based on these key species will capture the essential features of carbonate structures on ecological to geological scales.
3. HYPOTHESES about processes and outcomes of carbonate accretion generated by such models can be tested against biological and/or geological field observations.

TECHNICAL GOALS

The immediate goal is to create a knowledge base (KB) which:

1. addresses the physiological and ecological properties of the KEY SPECIES,
2. contains reliable and complete empirical data;
3. expresses the spatio-temporal ecological variation which is inherent in biological processes;
4. allows for the efficient addition of new species and new data;
5. supplies information to numerical models.

In current carbo* suite models the concept of 'VIRTUAL AQUARIUM' is used. It is impossible to include all organisms in a model. Instead, a narrow set of the best-documented species is used to represent the primary ecologic processes to be modeled. Dealing with the organisms at species level is important, as opposed to dealing with hypothetical and/or blended reproductive, growth, niche characteristics of guilds, functional groups, genera or biofacies.

The KB structure is still evolving, but currently exists as a suite of Excel worksheets ('Organisms', one per KEY SPECIES; 'Citations'; 'Dictionary' for parameters and terms). These are aggregated, then processed by custom software to create a uniform output for models.

The KB outputs are useful for several kinds of population models: competition Lotka-Volterra, cellular stochastic models, and Leslie matrix models. An important technical aim is to widen the uptake of the KB to other existing and future models.

CONSTRAINTS ON DATA GATHERING

Composite data. Ideally, the output data for a KEY SPECIES will be from a single species. In practice, it may have to be a composite from several close species whose data is held separately in the KB. Users have the option of coalescing data into a "composite" species.

The data of interest may include, but not be limited to:

- Taxonomic precision and accuracy
- Dates, locations, habitats, frequency of observations
- Means, ranges, and variances for both biological properties and habitat characteristics
- Calcification / bioerosion rates
- Experimental vs. observational methods
- Degree of density dependence or other responses as limits are approached
- Known environmental limits or thresholds
- Variables affecting deposition/destruction of $CaCO_3$

Search expertise. Often, the required data were not primary purposes of the source studies. Locating them requires scanning a large body of literature across many branches of the sciences, and making judgments about data reliability and relevance.

Measurement methods and units. These vary greatly among studies. Data will be entered in their original units which the KB software converts to common (S.I.) units.

Distributed data sources. Ecological communities and their habitats change over time through predictable processes of ecological succession. During succession, habitats change, quantitative properties of species change, dominant species are often replaced, and rates of change tend to accelerate as environmental limits (e.g. sea level) are approached. Largely unpredictable external disturbances (e.g. storms, epidemics) commonly return communities to earlier successional states. Capturing such variation requires ecological, construction and erosion data from an array of habitats for each species.

INFORMATION DELIVERY TO MODELS

The syntax for the data in the KB is designed for efficiency of data entry – a key determinant of overall project success. The data are organized under species (as best can be identified). As individual data items of interest are discovered in the literature, they are entered in a succinct way (see Figure 5) but documented in a verbose way. Average, median, range, max/min or single values all have a dedicated notation. The units of measurement are also codified.

The KB data syntax is interrogated by software (of the carbo* model suite), which then deals with the options of syntax, values and units, etc. That software resolves the entries for each species-parameter combination to a single value with ±uncertainty and with tracking codes. Error traps detect incomplete and inconsistent cases.

In this way, the syntax is a compromise between human-editability and –readability, and machine-readability.

In the final step, the information is loaded into data arrays used in the models. The parameter set is shown in Figure 6.

Figure 1. Carbonate production processes.

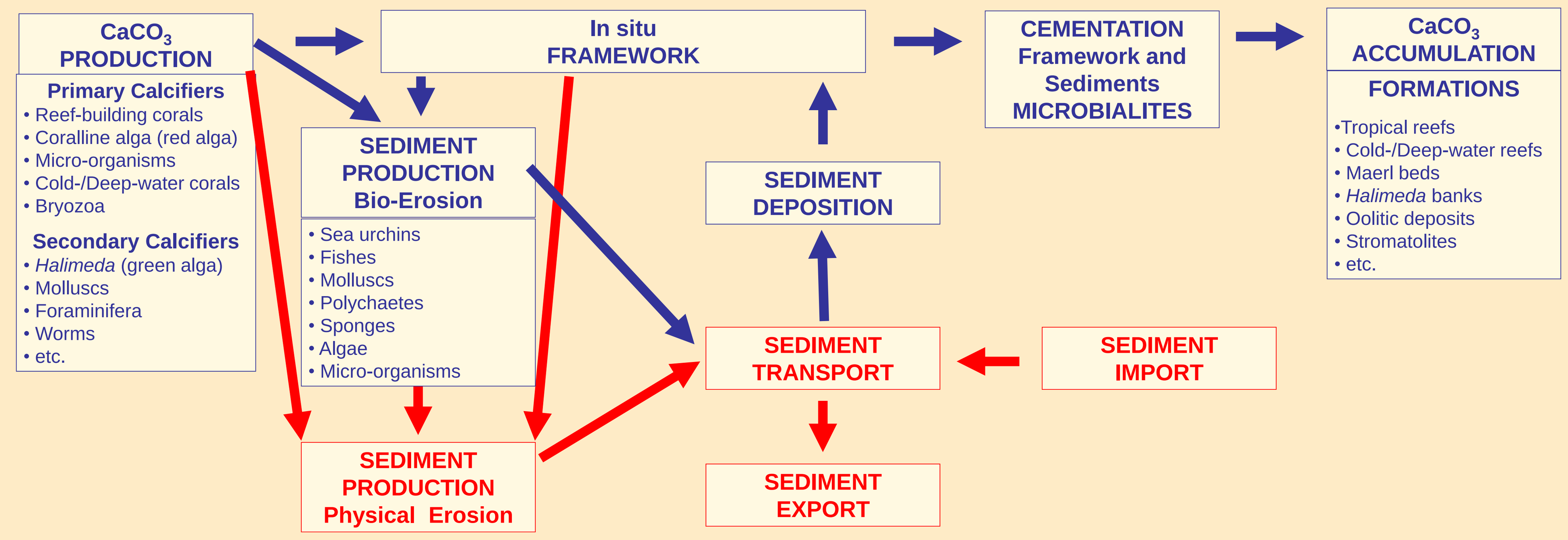
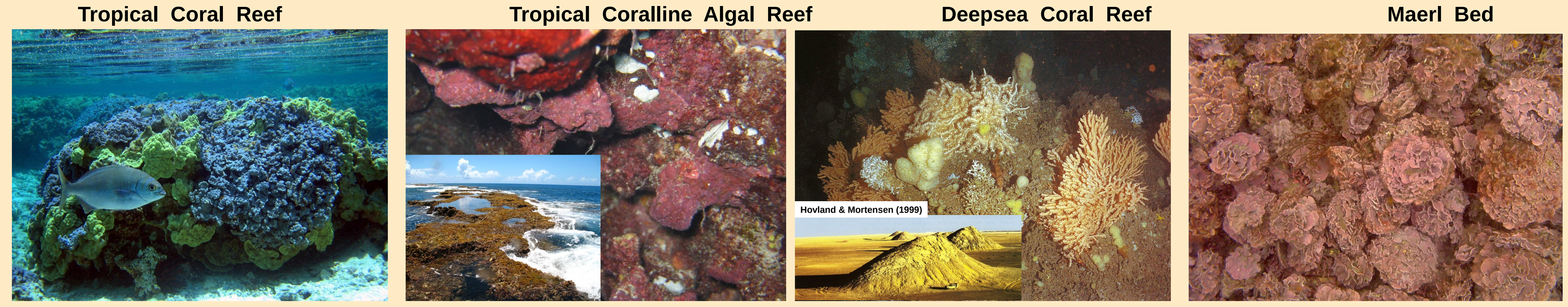


Figure 2. examples of major carbonate environments and key calcifying species

Tropical Coral Reef | **Tropical Coralline Algal Reef** | **Deepsea Coral Reef** | **Maerl Bed**



Modern coral framework – patch reef in sheltered lagoon, Midway Atoll

Modern encrusting coralline algae and accreting reef margin, Midway Atoll

Modern Lophelia pertusa reef, North Sea. Inset: Similar fossil structures, Algeria

Maerl (*Lithophyllum dentatum*) Kenmare River, south-west Ireland Hall-Spencer etal. (2008)

Figure 3. Examples of major key bioeroders and sediment producers

Sea Urchins *Heterocentrotus* *Echinometra* | **Parrot Fish** *Scarus* | **Rock-boring Molluscs** *Lithophaga* | **Boring Sponges** *Cliona*

Living coral (*Porites*) with urchin bioerosion Midway Atoll

Bioerosion by parrot fish feeding on dead coral Midway Atoll

Coral (*Porites*) skeleton with *Lithophaga* borings North Seymore Island, Galapagos

Limestone with *Cliona* borings New England

Figure 4. CarboKB structure and workflow

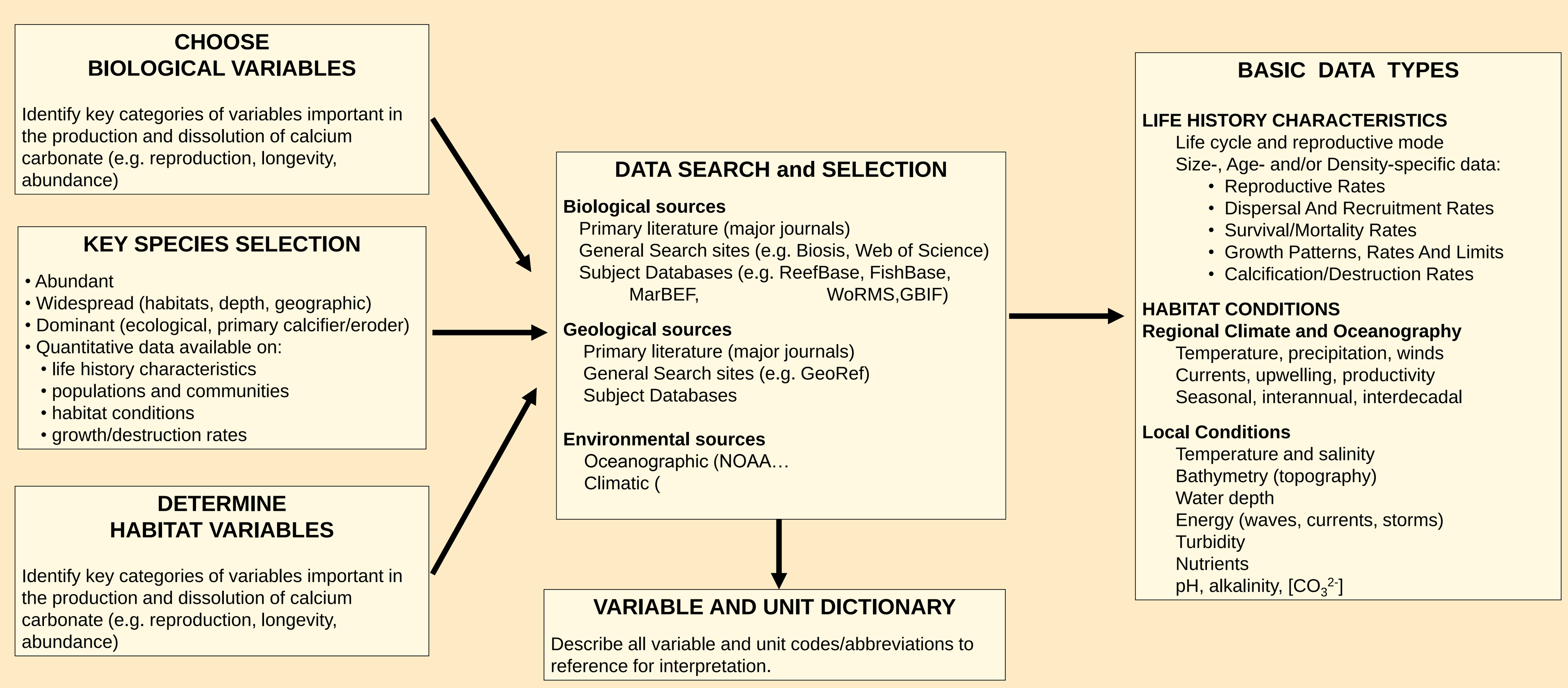


Figure 5. Example of KB data: Porites astreoides, a massive Caribbean reef-building coral

	A	B	C	D	E
1	Citation	Organism	Conditions	Parameter	Comments
2	Porites astreoides				
3	cite=Elizalde-Rendon, EM, Horta-Puga, GH, Gonzalez-Diaz, P, Carricart-Ganivet, JP 2010. Growth characteristics of the reef-building coral Porites astreoides under different environmental conditions in the Western Atlantic. Coral Reefs. 29, 607-614.; doi=10.1007/s00338-010-0604-7; url=http://www.springerlink.com/content/nh340w4617t671n6/	species=porites_astreoides ; morphology=massive	lat_min=19_03N; lat_max=19_12N; long_min=95_56W; long_max=96_04W; depth_min=2; depth_max=5; T_mean=26.53; PO43-_tot_mean=2.8{µM}; PO43-_tot_sd=0.84; N_tot_mean=14.65{µM}; N_tot_sd=0.96; light_atten_mean=0.5{m-1}	dens_mean=1.42{g_cm-3}; dens_sd=0.19; calc=0.52{g_cm-2_yr-1}; calc_sd=0.16; growth_mean=3.69{mm_yr-1}; growth_sd=1.08	location=Veracruz Reef System; habitat=reef; sed_rate=high; low density band at growth max, high density band at growth min; high turbidity, high nutrients from run-off
4			lat=22_09N; long=84_46W; T_mean=28; sed_rate_mean=422.2{mg_cm-2_mo}; sed_rate_sd=306.2; N_tot_mean=5.76; N_tot_sd=3.34; PO43-_tot_mean=1.11; PO43-_tot_sd=1.31; light_atten_mean=0.36	dens_mean=1.62; dens_sd=0.21; calc=0.71; calc_sd=0.23; growth_mean=4.33; growth_sd=1.22	location=Gulf of Guanahacabibes, northwest coast of Cuba; well-preserved reef; low turbidity, low nutrients, low sedimentation rates
5			lat=18_43N: long=87_41W; T_mean=27.73	dens_mean=1.71; dens_sd=0.16; calc=0.61; calc_sd=0.22; growth_mean=3.53; growth_sd=1.19	location=Mahahual Reef; low turbidity, low nutrients, low sedimentation rates

Figure 6. CarboKB output data: Porites lutea, a massive Pacific reef-building coral

#---

OrganismName Index |Taxon	Organism 1|porites_lutea		
PARAMETER	ArrayName	[value(recalculated from what)1, …	[label1,label2,… ,measurement unit]
Organism footprint	Covrg	[2.31, 0.45, 1.0]	['m:from m', 'fxn', '#/m2']
Organism growth rates	Grwth	[0.009]	['m/yr:from mm/yr']
Organism growth rates	Skltl	[1.0, 3.15, '-', 0.90]	['fxn', 'm:from m', '-', 'fxn']
Organism vertical zones	Zones	[0.0, 0.0, 100.0, 0.0]	['%', '%', '%', '%']
Water depth range	HabWd	['0,5; 12,8; 13,100; 20,0']	['m']
Temperature range	HabTemp	['12,0;14,10;27,66;29.6,10;30,0']	['degc']
Temperature irradiance	HabPar	['0.0,0;21.6,100;43.2,100;86.4,30']	['mol_photons/m2/d:from umol_photons/m2/s']
Normal mortality rate	MortR	[0.124, 500.0]	['fxn/yr:from %/yr', 'yr:from yr']
Reproductive rate	ReprR	[[0.050, 0.1480], [100.0, 0.029]]	[['clone_fxn/yr:from %/yr', ' spawn_fxn/yr:from %'], ['#/yr', 'fxn/#/m2']]