**C-system observations in the Southern California Bight**

**Task 1: dataset compilation**

**Purpose**: compile all available observations of C-system parameters in the Southern California Bight.

**Deliverable**: A matlab (and ASCII) structure containing observations of C-system from the region, with ancillary data, from reliable data sources (NOAA cruises and buoys, CalCOFI cruises, SMBO observation, and other published datasets).

**Details**: Fayçal will provide observations.

To be included in the database, they need to include:

* At least 1 C-system parameter observation, i.e. pCO2, pH, dissolved inorganic carbon (DIC), alkalinity (Alk)
* Basic physical variables: T, S
* Longitude, Latitude, Time (year, month, day and possibly hour of the day). Time should be processed as a time vector.

Other variables that, if present, should be included:

* Nutrients: phosphate (PO4), silicate (SIO2), nitrate (NO3) …
* Chlorophyll or fluorescence
* Oxygen (O2)
* ...

Information on the data source (cruise, buy, project) should be included in the database, for each observation, as an unique identifier (e.g. a string of characters). The database should be organized as a matlab structure with multiple fields, e.g.:

database.data\_source

database.lon

database.lat

database.time

database.temperature

etc.

The data could also contain flags for measurements (good, bad, questionable), but if we include only reliable measurements, probably that’s not needed in the first iteration. Also, we should keep track of the data sources (e.g. a small write up on each of the sources, linked to the unique identifier), with info on the error associated to each measurement if present.

We may want to add an additional layer to the data\_source, for example profile number for in situ profiles, or cruise number for zodiac data – this could be discussed. The purpose would be retrieving individual “meaningful units” from the dataset.

**Task 2: development of C-system algorithms**

**Purpose:** adding all C-system parameters to the Southern California Bight database, and potentially other datasets.

**Deliverables:**

* An updated database that includes the main variables (pCO2, pH, DIC, Alk, OmegaA, OmegaC, …?) either observed or derived from existing observations.
* A set of algorithms that take as inputs observations and return C-system parameters.

**Details:** (to be fleshed out)

When C-system variables are missing, they can be derived by a combination of algorithms, e.g. C-system function from DOE, or empirical relationships with measured variables (e.g. Leinweber’s; Alin’s, etc.). We will identify and develop 4 tiers of algorithm combinations:

* **Tier 1:** DOE C-system algorithm only, where all inputs are observations. For example, one can determine pCO2 exactly with the DOE algorithm, if 2 C-system parameters are known, in combination with PO4 and SiO2.
* **Tier 2:** DOE C-system algorithm using 2 observed C-system parameters, and PO4 and SiO2 (and any other required inputs) derived from empirical relationships from the Southern California Bight dataset.
* **Tier 3:** DOE C-system algorithm using 1 observed C-system parameter, and 1 C-system parameter empirically derived from the Southern California Bight dataset. Other inputs can be either observed or derived. This will be the case of algorithms applied to the zodiac measurements, since the only measured variables are T, S, and pCO2.
* **Tier 4 (to be determined):** C-system parameters are all derived from other variables, using empirical relationships.

This task will require developing empirical relationships for PO4, SiO2, Alk, etc., based on the observation dataset. The algorithms should be tested against available observations, documented, ranked, and the error associated to the algorithm should be quantified. Starting points could be existing algorithms from the literature, against which the new algorithms should be compared.

The algorithms range of applicability could also be refined, for example splitting algorithms that use only surface data (useful for the zodiac measurements) and algorithms that use full water column data.

**Task 3: analysis of zodiac pCO2 data**

**Purpose:** Analyzing the zodiac data, and complementing zodiac pCO2 observations with derived variables, i.e. our “best guesses” of C-system parameters (and nutrients if needed) from the algorithms in “tier 3” of the previous task.

**Deliverables:** A set of stand-alone functions that process zodiac data by cleaning it up (e.g. removing outliers), add derived variables (e.g. pH, DIC, Alk), save it as self-contained matlab structures, plot it as maps, etc. This function should be usable by anyone interesting in the zodiac data.

**Details:** Part of this task (Adding derived C-system parameters) will require using the algorithms developed in Task 2 – in particular the “tier 3”. The default should be the algorithm identified as best under task 2, but other options could be allowed.