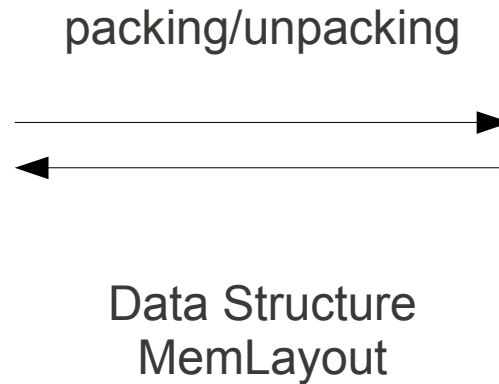# Data structures in Ocean Assimilation Kit

Alexander Barth, François Laenen,
Jean-Marie Beckers

Sangoma, video conference,
8th May 2012

State vector

Temperature

Salinity

Velocity

...

packing/unpacking

Data Structure
MemLayout

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

List of variables (including
possibly land points)
Fortran 1-, 2- or
3-dimensional arrays

Fortran vector
only with sea-points
of size `x(n)`

n: size of the state-vector

Type MemLayout

- **Name** of the individual variables

- Land-sea **mask**

- For each variable: **start-** and **end-index**

- Optional:

  - **Permutation** (useful for local assimilation)

  - **Distributed** across nodes

- Example:

  - `packVector(ML,x,temp,salt,uvel,vvel,..)`

  - `loadVector(path,filenames,ML,x)`
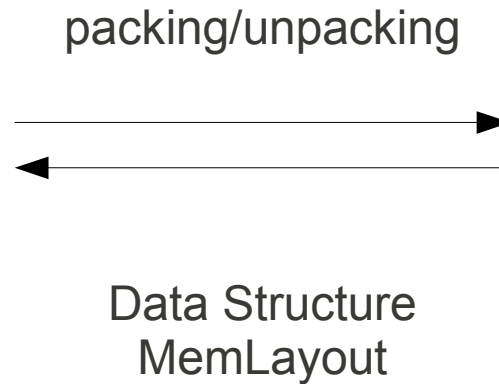
Arguments defined as optional
in the subroutine interface

Model grid

- Array of model grids (type(grid))
- Grid type contains:
  - n: dimension of the grid
  - mask: land-sea mask
  - coordinates for every grid point
- The dimension can be arbitrarily high for module `ndgrid`
- Only structured grids are supported

Observation vector

SST

Profile

Glider data

...

packing/unpacking

Data Structure
MemLayout

$$\mathbf{y}^o = \begin{pmatrix} y_1^o \\ y_2^o \\ \vdots \\ y_m^o \end{pmatrix}$$

List of variables (including
possibly land points)
Fortran 1-, 2- or
3-dimensional arrays

Fortran vector
only with sea-points
yo(m)

m: size of the observation vector

Observation operator

- New data type `SparseMatrix`
  - Only contains non-zero elements
  - Implements operators for multiplication with matrix and vector
- For every element of the vector yo(m), we have the vectors `x(m), y(m), z(m)` for the coordinates
- Bilinear interpolation coefficients are based on `x,y,z` and model grid or provided by the user
- Observation operator can also be distributed across nodes (as function call)
- Non-linear observation operator: $\rightarrow$ state vector augmentation

Use of data types

- Derived types (`MemLayout`, `grid`, `SparseMatrix`) are only used in high-level routines

  - Loading/Saving of the state vector, ensemble, observations,...

  - Assimilation diagnostics per variable

- Low-level computational routines use only Fortran arrays:

  - Vectors: `x(n)`, `Hx(m)`, `yo(m)`

  - Matrices: `S(n,r)`, `HS(m,r)` for error modes (ens. member – ens. mean) and its observed part (r: number of ens. members)

Local assimilation

- Partition vector: `part(n)` of type integer

- Every element in `part` with the same number belong to the same zone

  - For example forecast `xf` for zone `i` would be `pack(xf,part==i)` or `xf(part == i)`in matlab

- Often partition works per vertical water column, but it can be something else

- Permutation vector is applied such that all variables belonging to the same zone are contiguous in memory