Summary of 4DDA Algorithm for CRPC Talk, Geoffry C.Fox

Gregor von Laszewski

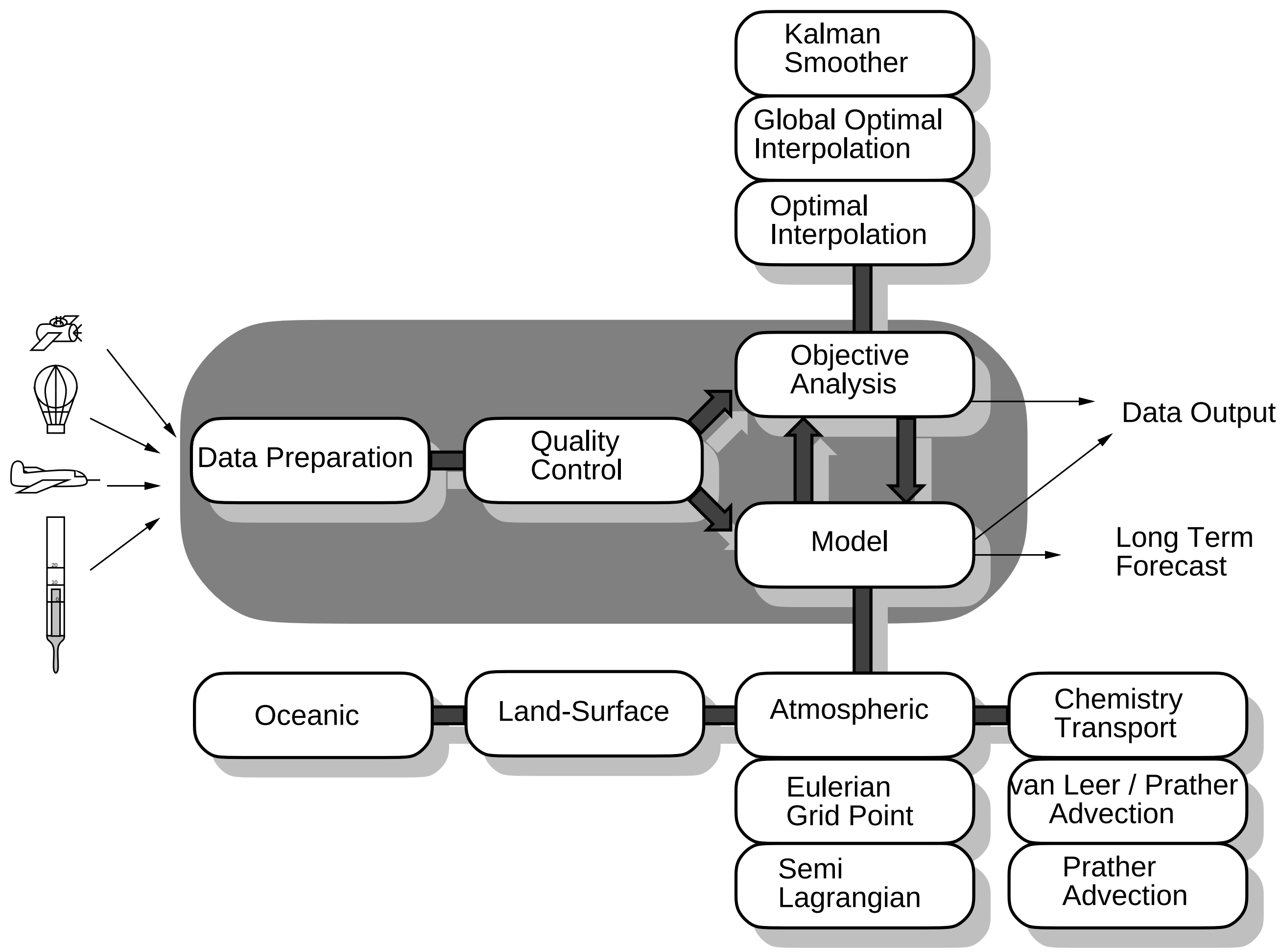
May 9, 1996

1 Four Dimensional Data Assimilation

- Data Assimilation is motivated by climate modeling
- In climate modeling calculations are based on observations gathered from diverse instruments.
- Climate models often use regular grids as “domain”
- Problem: there is not an instrument (observation) available on every grid point
- Solution: Use of an interpolation algorithm which determines statistically the missing gridpoints, while using a set of neighboring points.
- This data analysis is repeated every 6h.

- Where is it used: Recalculation of atmospheric state for the last decade in order to test accuracy of climate models.

2 Overview



Kalman Smoother
Global Optimal Interpolation
Optimal Interpolation
Objective Analysis
Data Output
Data Preparation
Quality Control
Model
Long Term Forecast
Oceanic
Land-Surface
Atmospheric
Chemistry Transport
Eulerian Grid Point
van Leer / Prather Advection
Semi Lagrangian
Prather Advection

3 The Algorithm

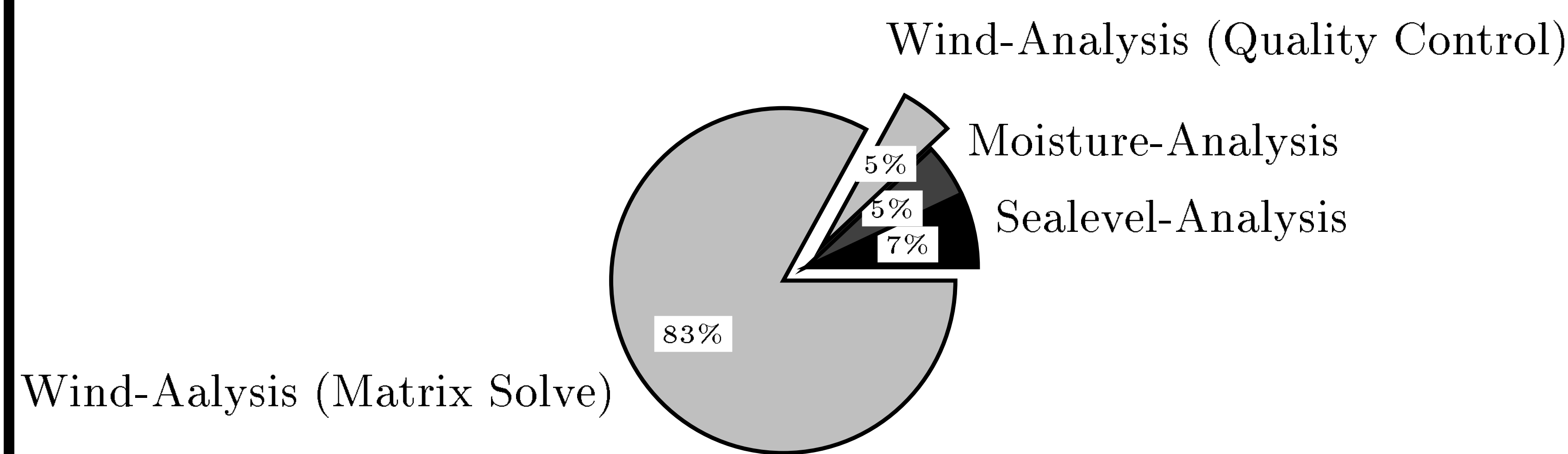
There are three essentially equal parts, corresponding to

1. the sea-level analysis,
2. the moisture analysis,
3. and the wind analysis.

All three analysis parts follow the general scheme:

1. Quality Control – foreach observation
   (a) Eliminate obviously wrong observations
   (b) Conduct an averaging process with neighboring observations, e.g. to preserve storms
2. Interpolation Step – foreach gridpoint
   (a) Determine the neighbors
   (b) Set up a system of linear equations. The solution minimizes the error between the observed values and the analyzed value. The size of the matrix depends on the number of observations in the neighborhood of the gridpoint, thus introducing load imbalance.
   (c) Solve the equations.

4 Sequential Computational Analysis



5 Parallelization Strategy

1. Improve overall performance of the sequential program before parallelization. (From 6000 CPU seconds to 600 CPU seconds on an RS6000)

2. In the quality control the calculation is conducted forall observation independently.

3. In the matrix solve the calculation is conducted forall gridpoints independently.

5.1 Remark

- Each observation/gridpoint can be calculated independently after all observations in the region of influence are gathered. This is done before the calculation is started.
- At the end of the calculation the result has to be forwarded to the Processors which contain neighboring observations.
- Different distributions of the observations are possible (stripped, block, bisection, cyclic). The best load balance can be achieved with the cyclic distribution.