Thinning techniques for remote sensing observations in support of ocean data assimilation

UC SANTA CRUZ

Juliana Matranga[1] - Andrew Moore[2] - Christopher Edwards[2]

1: Dept of Applied Mathematics, UCSC, U.S.A.; 2: Dept of Ocean Sciences, UCSC, U.S.A.

Motivation: 4D-Var

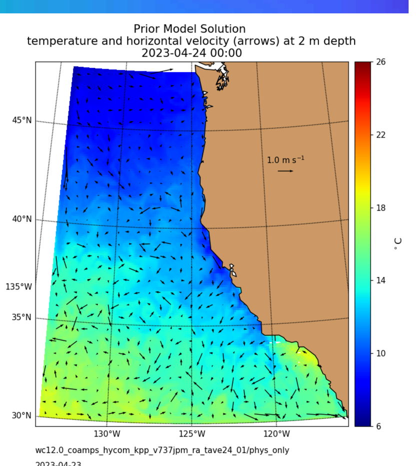
- The goal of 4D-Variational data assimilation is to **calculate the best estimate of the ocean circulation** (analysis, $\boldsymbol{x_a}$, a.k.a. best linear unbiased estimate) using a numerical model solution (background, $\boldsymbol{x_b}$) and observations ($\boldsymbol{y}$) of the state of the ocean.
- This problem can be solved by minimizing a cost function that provides the best estimate of the ocean given our set of observations. The solution takes the form of an increment ($\boldsymbol{\delta x_a}$) that corrects the background state. The analysis then can be expressed as: $\boldsymbol{x_a} = \boldsymbol{x_b} + \boldsymbol{\delta x_a}$
- To calculate the increment, we need to invert the stabilized represented matrix $(\boldsymbol{HBH^T} + \boldsymbol{R})$, which represents the total error covariance matrix.
- Observations located very close to each other, and which have correlated errors, introduce practical challenges to invert the stabilized represented matrix.
- **By thinning observations, we seek to reduce the number of redundant observations, and get closer to the 4d-var assumption of uncorrelated observation errors.**

Findings

Our analysis of the ocean circulation was not significantly degraded after reducing the amount of assimilated observations.

Future work

- Apply thinning of observations to improve initial conditions for a near real time ocean forecasting system of the US West Coast.



- Combine thinning of observations with other methods to optimize 4D-var performance, such as the use of mixed resolution or the implementation of saddle point algorithm

Experiments

We tested two different thinning techniques for two observation sets:

- A simple technique creating superobservations for high frequency radar (HFR) radial velocity observations.
- An intelligent data thinning algorithm based on R2005 applied to gridded satellite data of sea surface temperature (SST). The goal is to retain observations in regions of high variance and subsample regions of low variance.

After an initial set of single cycle assimilation tests, we assimilated data for a 1-year period divided in 4-day cycles for three different data sets:

1) Using all the observations
2) Thinned HFR radial velocity observations
3) Thinned SST observations

Configuration

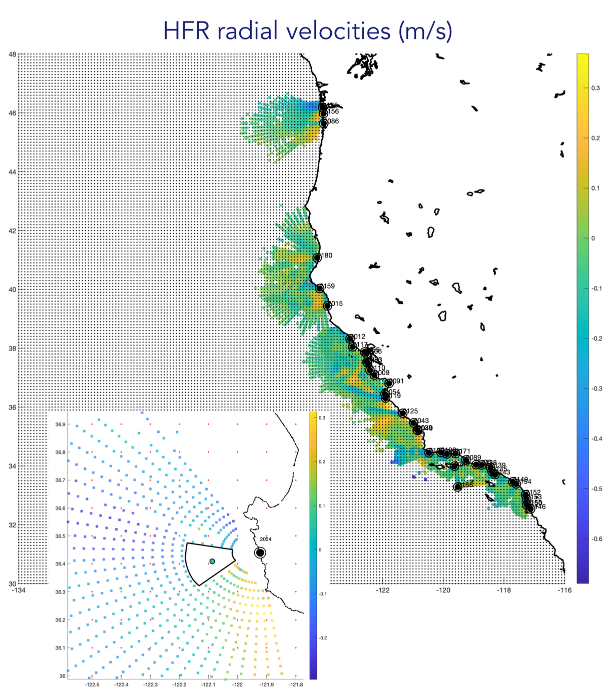
- Model: Regional Oceanic Modeling System (ROMS): 3D, free surface, terrain following coordinate numerical model
- Domain: California Current System (CCS)
- Grid resolution ~10km – 42 vertical levels
- Atmospheric Conditions: atmospheric forecasting system
- Boundary Conditions: global scale ocean circulation model
- Initial Conditions: Assimilating data from the previous 4 days
- Observations: AVISO Sea Level Anomalies, OSTIA SST, AQUARIUS Sea Surface Salinity, HF Radar data, Argo gliders, buoys.

Notation

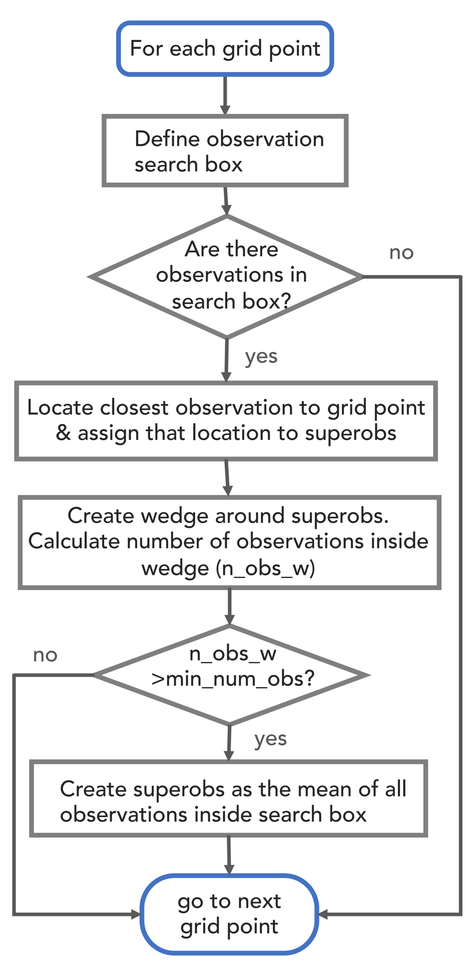
Increment: $\boldsymbol{\delta x_a}$
Innovation: $\boldsymbol{d_b} = \boldsymbol{y} - H(\boldsymbol{x_b})$
Residual: $\boldsymbol{d_a} = \boldsymbol{y} - H(\boldsymbol{x_b} + \boldsymbol{\delta x_a})$
Background error covariance matrix: $\boldsymbol{B}$
Observation error covariance matrix: $\boldsymbol{R}$
Tangent linearization of the observation operator: $\boldsymbol{H}$
Total error covariance matrix (stabilized representer matrix): $\boldsymbol{HBH^T} + \boldsymbol{R}$

Observations and thinning algorithms

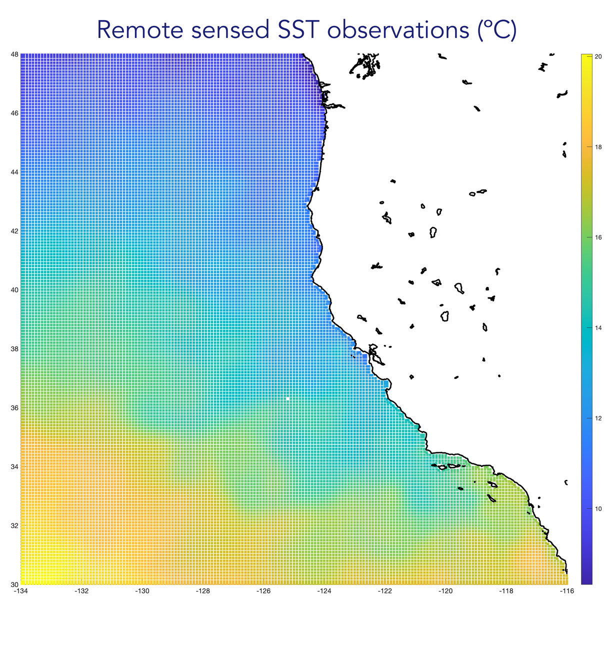
Superobing radial velocity observations from a network of HF radars operated by NOAA



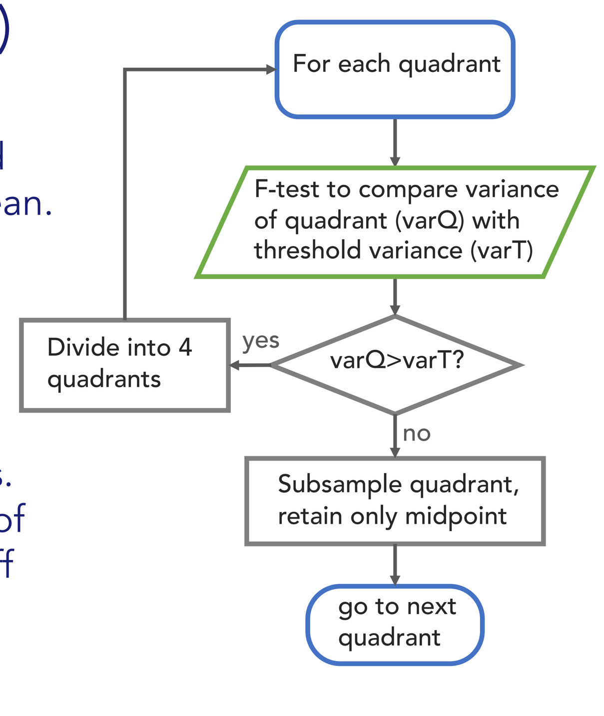
- Each HF radar site is treated individually.
- For each grid point locate closest observation .
- Define pie-wedge area that depends on grid and distance to antenna.
- Average observations in that area to create superobservation.



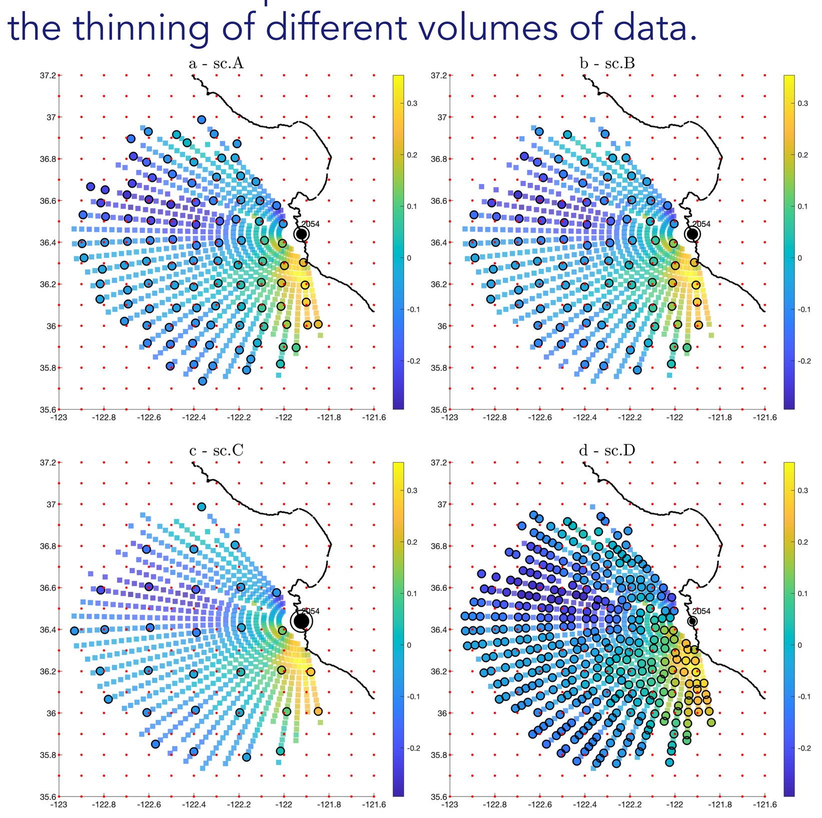
Subsampling regions of low variance for satellite data from operational SST (OSTIA)

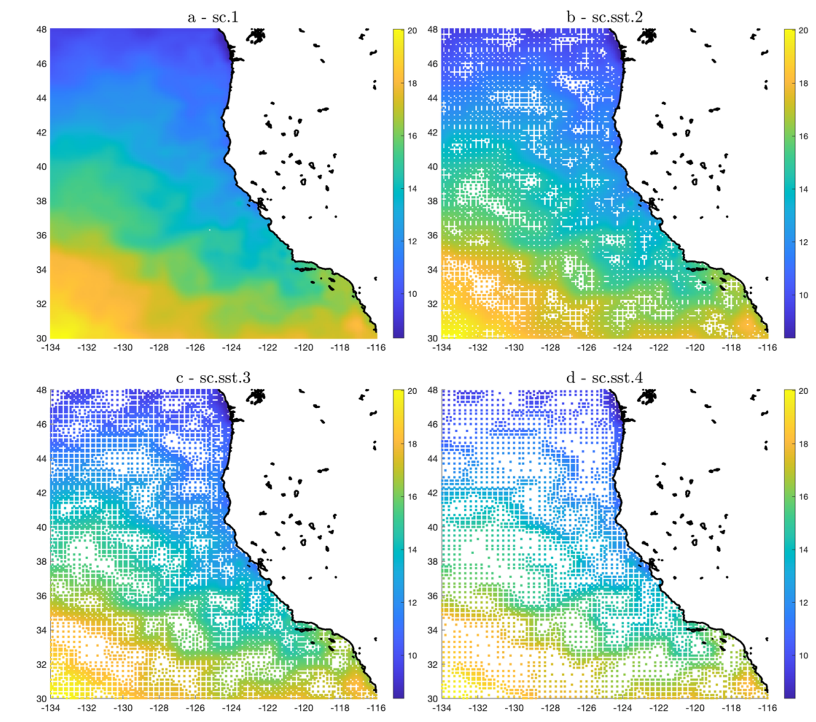


- Normalize data and calculate global mean.
- Define cutoff threshold.
- Calculate cutoff threshold variance.
- Recursively divide data into quadrants.
- Compare variance of quadrant with cutoff threshold variance.



We performed a series of single cycle data assimilation experiments to test and visualize the thinning of different volumes of data.





The analysis of errors before and after the assimilation (by means of the standard deviation of the increment and the innovation) showed that thinning in general increased the errors, but the calculated uncertainties were within a tolerable margin for all cases.

A numerical optimization problem

- ROMS 4D-Var uses an iterative Gauss-Newton method based on an incremental approach to minimize the cost function.
- The solution consists of estimating an increment, which can be expressed as: $\boldsymbol{\delta x_a} = \boldsymbol{BH^T}(\boldsymbol{HBH^T} + \boldsymbol{R})^{-1}(\boldsymbol{y} - \mathrm{H}(\boldsymbol{x_b}))$
- The inversion of the stabilized representer matrix is usually solved using the equivalent matrix: $(\boldsymbol{R^{-1}HBH^T} + \boldsymbol{I})^{-1}$
- This inversion involves some practical challenges, such as:
  - For remote sensing observations closely separated in space and time, $\boldsymbol{HBH^T}$ can have columns very similar to each other, resulting in an ill-conditioned matrix.
  - The evaluation of $\boldsymbol{R^{-1}}$ is trivial if $\boldsymbol{R}$ is diagonal, but this implies the unrealistic assumption of uncorrelated observation errors.
- **Data thinning can ameliorate the challenges outlined above.**

Analysis

- **Innovation statistics:** all experiments showed a good agreement with the assumption of normally distributed unbiased innovations, except during the upwelling season, where the model is known to be biased.
- **Background error and observation error consistency:** comparing an analysis of the increments with the a priori defined background and observations errors we conclude that the thinning did not significantly degrade the analysis.
- **Properties of the total error covariance matrix:** the superobing of HFR velocities slightly improved the conditioning of the matrix. Detection of regions of high variance for the SST observations had the undesired effect of increasing the condition number of the matrix to invert.
- **Observation impacts:** the thinning of radial velocities drastically reduced the impact of these observations in the analysis. On the other hand, thinning of SST had a good effect on evening the impact of different types of observations on the analysis.

References:

- Bick T et. al, 2016 "Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale"
- Desroziers, G. "Diagnosis of Observation, Background and Analysis-error Statistics in Observation Space."
- Moore, Andrew M., et al. "Regional Ocean Modeling System (ROMS) 4-Dimensional Variational Data Assimilation Systems Part I – System Overview and Formulation."
- Ramachandran, R., Li, X., Movva, S., Graves, S., Greco, S., Emmitt, D., Terry, J. and Atlas, R. 'Intelligent Data Thinning Algorithm for Earth System Numerical Model Research and Application'. (R2005)