

EXERCISE 4: Weak Constraint Dual Formulation 4D-Var

Introduction

A considerable advantage of the dual form of 4D-Var over the primal form is that the dimension of the minimization problem is *unaffected* by the size of the control vector. Therefore, dual 4D-Var is the obvious choice for weak constraint data assimilation where the control vector is augmented with corrections for model error at every grid-point and potentially every model time step. This exercise is a repeat of Exercise 3 except now you will run RBL4D-Var with RPCG subject to the weak constraint.

Broquet et al. (2011) found that ROMS CCS tends to underestimate SST near the California coast during the peak of the upwelling season. Evidence suggests that this is associated with errors in the model rather than errors in the surface forcing. Using strong constraint 4D-Var, Broquet et al. (2011) found that data assimilation tends to compensate for model error by adjusting the surface forcing. These findings were corroborated in a more recent study by Crawford et al. (2016). This is clearly undesirable, and a more appropriate approach to the problem is to account for model error during the data assimilation procedure.

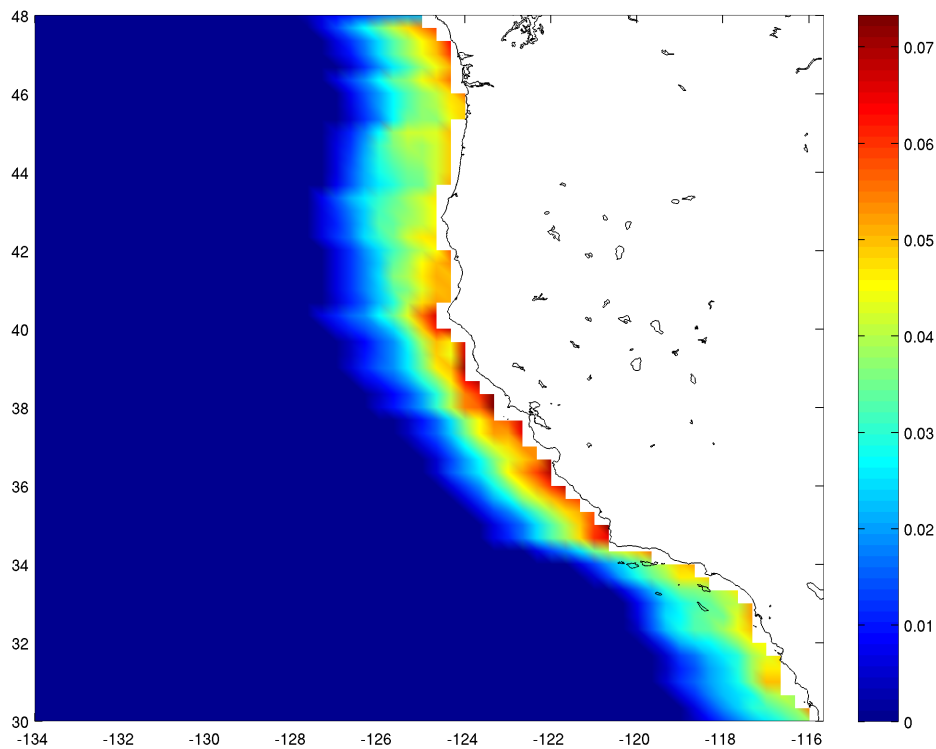


Figure 1: Assumed standard deviation for model error in the upper level of ROMS CCS.

Based on the findings of Broquet et al. (2011), a crude attempt is made here to account for errors in the model by assuming a *prior* model error covariance of the form

$\mathbf{Q} = \mathbf{K}_b \mathbf{W} \Sigma_x \mathbf{C} \Sigma_x^T \mathbf{W}^T \mathbf{K}_b^T$, where Σ_x is the *prior* initial condition error standard deviation, and \mathbf{W} is a diagonal matrix with all elements zero except those that correspond to temperature grid points within 300 km of the coast. The non-zero elements w of \mathbf{W} decrease linearly from a value 0.05 at the coast to zero 300 km offshore. During weak constraint 4D-Var, a correction η is applied to the model temperature at every grid point and time step, so this choice of \mathbf{Q} corresponds to the situation where the standard deviation (std) of η is 5% of the std of errors in the temperature initial condition at the coast, decreasing linearly to zero offshore. The structure of $\mathbf{W} \Sigma_x$ corresponding to the surface level temperature is illustrated in Fig. 1. A more sophisticated representation of \mathbf{Q} in the coastal upwelling zone is discussed in Crawford et al. (2016).

When running 4D-Var subject to the weak constraint, it is important to impose a temporal correlation on the model errors otherwise the model error corrections η may anticipate the availability of the observations in time, and impose large corrections on the model errors at the observation times to better fit the observations. This undesirable behavior can be eliminated by regularizing the model error corrections η in time by assuming that the η are correlated in time. This is controlled using the cpp option `TIME_CONV` and by setting the parameter `TdecayM` in `s4dvar.in` to a suitable decorrelation time.

Running weak constraint RBL4D-Var

First edit the file `build_roms.csh` or `build_roms.sh` and choose the following cpp options

```
setenv MY_CPP_FLAGS "${MY_CPP_FLAGS} -DRBL4DVAR"
#setenv MY_CPP_FLAGS "${MY_CPP_FLAGS} -DMINRES"
setenv MY_CPP_FLAGS "${MY_CPP_FLAGS} -DRPCG"
setenv MY_CPP_FLAGS "${MY_CPP_FLAGS} -DTIME_CONV"
```

Recompile the model and then run it.

Follow the instructions in the `Readme` file. Before running the model, you will need to edit `roms_wc13_2hours.in` to indicate that the weak constraint will be used. This is controlled by the parameter `NADJ`. If `NADJ < NTIMES`, then the weak constraint is assumed, otherwise dual 4D-Var is run using the strong constraint. In this case `NADJ` controls the frequency at which the model error corrections η are applied. The resulting corrections are linearly interpolated in time to yield a correction at every time step, a procedure that also introduces an artificial temporal correlation in the model error corrections. Choosing `NADJ=1` removes the artificial temporal correlation but is not advisable due to the large amount of I/O that will be involved. In this exercise, the choose `NADJ=48` (i.e. daily). The explicit decorrelation time `TdecayM` is chosen to be 1 day for the errors in all state variables. You should feel free to play with these parameters.

Create a new subdirectory `EX4_WEAK`, and save the solution in it for analysis and plotting to avoid overwriting solutions when playing with different CPP options and rerunning and recompiling:

```
mkdir EX4_WEAK
mv Build_roms rbl4dvar.in *.nc log EX4_WEAK
cp -p romsM roms_wc13_2hours.in EX4_WEAK
```

where log is the ROMS standard output specified.

Plotting your results

Plot the 4D-Var cost function J using [plot_rbl4dvar_cost_compare_weak.m](#) which can be found in [WC13/plotting](#).

Plot next a selection of the increments using or [plot_rbl4dvar_increments.m](#).

NOTE: Before moving on to Exercise 5, please move all of the *.nc files and the log file into a new directory.

References

Broquet, G., A.M. Moore, H.G. Arango, and C.A. Edwards, 2011: Corrections to ocean surface forcing in the California Current System using 4D-variational data assimilation. *Ocean Modelling*, **36**, 116-132.

Crawford, W.J., P.J. Smith, R.F. Milliff, J. Fiechter, C. Wikle, C.A. Edwards and A.M. Moore, 2016: Weak constraint 4-dimensional variational data assimilation in a model of the California Current system. *Advances in Statistical Climatology, Meteorology and Oceanography*, **2**, 171-192.