SPEEDY EXPERIMENTS.

About the SPEEDY model (from Miyoshi PhD Thesis):

The SPEEDY model (Molteni 2003) is a recently developed atmospheric general circulation model (AGCM) with a spectral primitive-equation dynamic core and a set of simplified physical parameterization schemes (SPEEDY stands for Simplified Parameterizations, primitivE-Equation DYnamics). The goal of this model is to achieve computational efficiency while maintaining characteristics similar to the state-of-the-art AGCMs with complex physics. The resolution of the model is T30L7 (horizontal spectral truncation of 30 wave numbers and 7 vertical levels), the computational cost is one order of magnitude less than that of state-of-the-art AGCMs at similar horizontal resolution. According to Molteni (2003), the SPEEDY model simulates the general structure of global atmospheric circulation fairly well, and some aspects of the systematic errors are similar to many AGCMs, though the error amplitude is larger than state-of-the-art models. The SPEEDY model includes basic components of physical parameterizations used in more complex GCMs, such as convection (a simplified mass-flux scheme), large-scale condensation, clouds, short-wave radiation (two spectral bands), long wave radiation (four spectral bands), surface fluxes of momentum and energy (bulk aerodynamic formula), and vertical diffusion. Details of the simplified physical parameterization schemes of the SPEEDY model can be found in Molteni (2003), especially in its Appendix which is available on the website: "http://www.ictp.trieste.it/~moltenif/speedy-doc.html". The boundary conditions of the SPEEDY model includes topographic height and land-sea mask, which are constant, and sea surface temperature (SST), sea ice fraction, surface temperature in the top soil layer, moisture in the top soil layer and the root-zone layer, and snow depth, all of which are specified by monthly means, and bare-surface albedo and fraction of land-surface covered by vegetation, which are specified by annual-mean fields. The lower boundary conditions such as SST are obtained by ECMWF's reanalysis in the period 1981-90. The incoming solar radiation flux and the boundary conditions (SST etc.), except bare-surface albedo and vegetation fraction, are updated daily.

Software requirements:

1. Unix operating system
2. little-endian machine
3. Fortran 90 compiler (ifort, gfortran or Portland Group).
4. GrADS to visualize the results.

How to install:

Download the file SPEEDY_DA.tar from ftp://ftp.cima.fcen.uba.ar/pub/juan into your home directory. Then "untar" the file:

- tar –xvf SPEEDY_DA.tar

  If you save and untar the file in another directory you will have to change all the paths accordingly.

Before start running the scripts check the compiler option (value of the "compiler" variable) inside each script.

Paths

The required scripts and files are located in your home directory (**/home/$user/SPEEDY_DA**). (Where $user is the name of your own user).

/home/$user/SPEEDY_DA/tdvar: The scripts and fortran code for the 3DVAR experiments are stored in this folder. The main files in this folder are:

- tdvar.f90 : main 3DVAR fortran code.
- tdvar_tools.f90 : most of the subroutines used by tdvar.f90. Uses the minimizer in the file minimizelib.f90
- tdvar.sh : shell script to run the assimilation cycle.
- tdvar_response.sh : shell script to test the assimilation scheme.
- rmse_zm.sh : shell script to compute the zonally averaged RMSE.
- ex_obs.f90 : fortran program that sets the location of the observation points.
- dat_stat : folder which contains the files with the information about the error statistics in the SPEEDY model.
- Initial : folder which contains the initial guess for the analysis cycle and the ctl files to open the files in grads.

/home/user/SPEEDY_DA/letkf: The scripts and fortran code for the LETKF experiments are stored in this folder. The main files in this folder are:

- letkf.f90 : main 3DVAR fortran code.
- letkf_tools.f90 : Interface between model and LETKF assimilation.
- common_letkf.f90 : The LETKF core.
- letkf.sh : shell script to run an assimilation cycle.
- letkf_response.sh : shell script to test the assimilation scheme.
- rmse_zm.sh : shell script to compute the zonally averaged RMSE.
- ex_obs.f90 : fortran program that sets the location of the observation points.
- Initial : folder which contains the initial ensemble for the analysis cycle and the ctl files to open the files in grads.
- ens_mean_spread.sh : computes the ensemble mean and spread (can be used for both the analysis ensemble and the first guess ensemble)

/home/user/SPEEDY_DA/model: This folder contains the SPEEDY model source code and static data (i.e. topography, albedo, etc).

/home/user/SPEEDY_DA/truth: A two month simulation performed with the SPEEDY model is stored in this folder. This simulation will be treated as the true evolution of the atmosphere in the different experiments. This evolution was obtained with the shell script run_cycle.sh located in the folder /home/user/model/run.

/home/user/SPEEDY_DA/obs: A randomly perturbed version of the true evolution of the atmosphere is stored in this folder. Observations will be generated from these files. The observation error (i.e. the amplitude of the random perturbation added to these files) can be set in the file /home/user/letkf/common_speedy.f90. In order to generate a new observation set with different errors the shell script /home/user/SPEEDY_DA/letkf/create_obs.sh must be executed.

/home/user/SPEEDY_DA/DAS_result: The folder where the results of the DA experiments will be stored. It contains two main folders:

- 3dvar : To store the 3dvar experiments.
- letkf : To store the LETKF experiments.

Under each of these folders new folder will be created with the name of the different experiments. For example is the name of a new 3dvar experiment is "FULL" then the folder **/home/user/SPEEDY_DA/DAS_result/FULL** is created during the experiment execution. Inside **FULL** three more folder are created: anal (which contain the analysis files), analf (which contains the SPEEDY filtered analysis files) and gues (which contain the 6 hour SPEEDY forecast which are used as first guesses in the assimilation cycle). For the 3dvar case, the analysis, filtered analysis and first guess will be stored in fortran unformatted files (big endian, direct acces) and one time per file. The name of the file is the same for the analysis, the filtered analysis and the first guess (YYYYMMDDHH.grd for the sigma level files and YYYYMMDDHH_p.grd for the p_level files). Two ctl files will be present in each folder to allow the visualization of the results with GrADS.

In the case of the LETKF experiment an ensemble is used as the first guess and as the initialization for the next assimilation cycle. Inside the anal, analf and gues folders N folders are created and named after the number of the corresponding ensemble member. Two more folders are included in the anal and gues folders, the mean and spread folders which contain the average and standard deviation of the ensemble respectively.

How to run:

3dvar (Developed by Takemasa Miyoshi and Junjie Liu):

Single observation experiment: This experiment shows the impact of a single observation in the analyzed fields and can be used to test the sensitivity of the scheme to different parameters.

To perform a single observation experiment, do the following:

- Go to the /home/user/SPEEDY_DA/tdvar folder.
- Edit the fortran program ex_obs.f90 and set the value of msw_test to .TRUE. and the value of msw_real to .FALSE.
- In the same file, in the line 23 choose the observation location. For example ex_t(76,35,4) means that the observation will be located in the grid point X=76, Y=35, Z=4. SPEEDY has 96x48 grid points starting at the South Pole in Y and at Greenwich in X. It also has 7 sigma vertical levels.
- The variable can also be changed: for example in line 23, ex_t(76,35,4) means that the temperature (t) will be observed. Observations for other variables won't be available. If the line is changed to ex_u(76,35,4) then the location of the observation is going to be the same as before, but in this case the zonal wind component is going to be observed and assimilated. The available variables are t, u,v, ps and q.
- Edit the fortran program tdvar.f90 and change the value of the parameter msw_test to .TRUE. Note that in this mode the value of the observational increment is fixed to 1m/s for u and v, 1 K for t, 0.1 g/kg for q and 1 hPa for ps. These values can be changed in lines 59-63 of the same file.

- Edit the tdvar_response.sh shell script and check that all the paths are properly defined. (note that you can select any date to perform the experiment but an analysis corresponding to this date must be present in the "initial" folder).
- From the tdvar folder type ./tdvar_response.sh experiment_name
- A new folder /home/user/SPEEDY_DA/DAS_result/3dvar/experiment_name will be created and 2 data files (.grd) will be stored in it: analysis.grd and guess.grd. Those files are the analyzed field (with only one available observation) and the first guess. The corresponding ctl files will also be stored in this folder.

Multiple observation experiment: Here we explain how to perform a multiple observation assimilation experiment using 3dvar.

- Go to the /home/user/SPEEDY_DA/tdvar folder.
- Edit the fortran program ex_obs.f90 and set the value of msw_test to .FALSE. To perform an experiment with a realistic rawindsonde observation network change the value of the parameter msw_real to .TRUE. . An alternative to this experiment is to use a regular spaced dense observational network. To do this set the value of the parameter msw_real to .FALSE. and set the value of msw_dnsobs to .TRUE. to get a dense observational network (i.e. one observation every 2 grid points). If the parameter msw_dnsobs is .FALSE. then the observations will be located every 4 grid points. The observational network is specified in the ex_obs.f90 file and can be modified to allow other configurations.
- In the case of a realistic observational network the location of the stations is specified in the file obsmark.gs which is also a GrADS scripts that plots the location of the observations.
- The file ex_obs.f90 has different blocks where the location of the observations in the different network configuration is set. At each of this blocks the observed variables can be chosen. For example between lines 36 and 40 the location of the observations for the realistic network is defined, each line corresponds to one variable. To avoid assimilating a particular variable just comment the line corresponding to this variable. (E.g. if the line ex_u(ilon,ilat,1:nlev) is commented, then the u observations won't be available during the 3dvar assimilation process).
- Edit the file tdvar.f90 changing the value of the parameter msw_test to .FALSE. after this change the 3dvar assimilation system will use the observations located in the /home/user/SPEEDY_DA/obs folder.

Analysis cycle: Here we explain how to perform the simulation of an analysis cycle using the randomly generated observations and the 3dvar scheme developed for the SPEEDY model.

- Go to the /home/user/SPEEDY_DA/tdvar folder.
- As in the previous experiments set the value of the parameters corresponding to the desired observational network configuration.
- Edit the tdvar.sh shell script and check that all the paths are properly defined. Set the start and end dates for the analysis cycle simulation by changing the value of the variables IYYYY (start year), IMM (start month), IDD (start day), IHH (start hour), EYYYY (end year), EMM (end month), EDD (end day) and EHH (end hour). Remember that to start in a particular day; an analysis file

corresponding to that day must be located in the /home/user/SPEEDY_DA/tdvar/initial/anal/ folder.

- From the tdvar folder type ./tdvar.sh experiment_name
- A new folder /home/user/SPEEDY_DA/DAS_result/3dvar/experiment_name will be created with 3 folders in it (anal, analf, gues and rmse), see the PATH description to get more information about the files stored in these folders.
- The tdvar.sh shell script has a variable named "store" which can reduce the amount of output files. In the case of the 3dvar experiments, if store is set to 0 only the analysis and model output in pressure coordinates are stored.
- The folder rmse in the experiment output directory contains two files, one called rmse_zm.grd with the zonally averaged rmse for each day, level and latitude and rmse_energy.grd with the globally averaged energy rmse for each day. The corresponding ctl files to visualize the results in GrADS are also stored in the same folder.

3dvar Error Statistics (Developed by Takemasa Miyoshi and Junjie Liu):

In this section we will see how to run the scripts to obtain background error statistics using the NMC method (Parrish and Derber, 1992).

The scripts to compute the background error statistics are available under /home/user/SPEEDY_DA/tdvar_stat .

The NMC method estimates background error statistics from differences between forecast started at different times that verify at the same time (for example forecast at 18 and 24 lead time). The differences between these forecast is taken as an estimation of the background error. Error statistics are computed from a relatively big sample of these forecast differences.

To generate the required forecast we need a set of analysis. To generate the analysis when no error statistics are available we can use 3DVAR using some prescribed values for the error parameters. For example we can use a fixed length scale for horizontal error correlation and assume no inter-variable correlation.

The first analysis set (performed with inaccurate estimations of the parameters) and the forecast needed to use the NMC method can be computed using the **tdvar_nmc.sh** script.

After running this script a set of 24 hour forecast will be available. Then the script nmc_stat.sh can be used to compute the background error statistics using the NMC method. The results will be stored in the dat_stat directory under tdvar_stat.

The results can be viewed with GrADS. Note that this files won´t be used by the 3DVAR experiments that are started from /home/user/SPEEDY_DA/tdvar/ unless you copy the content of the folder /home/user/SPEEDY_DA/var_stat to /home/user/SPEEDY_DA/tdvar/var_stat .

LETKF (Developed by Takemasa Miyoshi and Junjie Liu):

Single observation experiment: This experiment shows the impact of a single observation in the analyzed fields and can be used to test the sensitivity of the scheme to different parameters.

To perform a single observation experiment do the following:

- Go to the /home/user/SPEEDY_DA/letkf folder.
- Edit the fortran program ex_obs.f90 as in the 3dvar example (i.e. set the value of the parameters in order to assimilate only one observation and chose the observed variable and the location of the observation.
- Edit the fortran program letkf_tools.f90 and change the value of the parameter test to .TRUE. Note that in this mode the value of the observational increment is fixed to 1m/s for u and v, 1 K for t, 0.1 g/kg for q and 1 hPa for ps. This values can be changed in lines 386-390 of the same file.
- Edit the letkf_response.sh shell script and check that all the paths are properly defined. (Note that you can select any date to perform the experiment but an initial ensemble corresponding to this date must be present in the "initial" folder).
- From the letkf folder type ./letkf_response.sh experiment_name
- A new folder /home/user/SPEEDY_DA/DAS_result/letkf/experiment_name will be created and 3 data files (.grd) will be stored in it: analysis.grd, gues.grd and gues_spread.grd. Those files are the analyzed field (with only one available observation) and the first guess. The corresponding ctl files will also be stored in this folder. The guess ensemble spread is also available as a .grd file.

Multiple observation experiment: Here we explain how to perform a multiple observation assimilation experiment using letkf.

- Go to the /home/user/SPEEDY_DA/letkf folder.
- Edit the fortran program ex_obs.f90 as in the 3dvar experiment with multiple observations.
- Edit the file letkf_tools.f90 changing the value of the parameter test to .FALSE. after this change the LETKF assimilation system will use the observations located in the /home/user/SPEEDY_DA/obs folder.

Analysis cycle: Here we explain how to perform the simulation of an analysis cycle using the randomly generated observations and the LETKF scheme developed for the SPEEDY model.

- Go to the /home/user/SPEEDY_DA/letkf folder.
- As in the previous experiments set the value of the parameters corresponding to the desired observational network configuration.
- Edit the letkf.sh shell script and check that all the paths are properly defined. Set the start and end dates for the analysis cycle simulation by changing the value of the variables IYYYY (start year), IMM (start month), IDD (start day), IHH (start hour), EYYYY (end year), EMM (end month), EDD (end day) and EHH (end hour). Remember that to start in a particular day; an analysis ensemble corresponding to that day must be located in the /home/user/SPEEDY_DA/letkf/initial/anal/ folder.
- From the letkf folder type ./letkf.sh experiment_name
- A new folder /home/user/SPEEDY_DA/DAS_result/letkf/experiment_name will be created with 3 folders in it (anal, analf, gues and rmse), see the PATH description to get more information about the files stored in these folders.

- The folder rmse in the experiment output directory contains two files, one called rmse_zm.grd with the zonally averaged rmse for each day, level and latitude and rmse_energy.grd with the globally averaged energy rmse for each day. The corresponding ctl files to visualize the results in GrADS are also stored in the same folder.
- The letkf.sh shell script has a variable named "store" which can reduce the amount of output files. If store is set to 0 only the ensemble mean and spread of the analysis and the first guess in pressure levels are stored. If store is set to 1 then the full ensemble for the analysis and the first guess in pressure and sigma levels are stored in the experiment output folder. (note that if the full ensemble is stored then 1 month of simulated assimilation cycle needs approximately 10 Gb of free space).

Other parameters of the LETKF escheme: Here we list some parameters involved in the LETKF scheme. All of these parameters are specified in the letkf_tool.f90 file.

As the current implementation of the LETKF uses an analysis localization to avoid spurious correlations the size of the local patch is an important part of the method.

nlocal: this parameter specifies the horizontal size of the local patch. If the value is 4 then the local patch centered at each grid point and has an extension of 4 grid points in each direction (W-E-N-S). The shape of the patch is squared. In the edges of the domains, cyclic boundary conditions are used to define the local patch.

nlocalv: this parameter specifies the vertical size of the local patch. As the previous one if the value is set to 2 then the local patch is composed by the current level plus the 2 previous levels and the 2 following levels. If the level is near the top or the bottom then the vertical extent of the local patch is reduced accordingly. The default value is 0 which means that the analysis is performed at each vertical level independently.

nlocal2: this parameter specifies how many neighboring local patches are averaged over one particular location to get the analysis value. The default value is 0 which means that only the patch centered at each grid point is used as the analysis for this particular grid point.

nskip_h: As the local patches overlap with each other it is possible not to perform the analysis in every grid point. This parameter allows the analysis to be performed every 2 or more grid points. The default value is 1 corresponding to the case where the analysis is performed at every grid point. For short and test experiments the optimal value is 2 to reduce the computational cost of the experiments.

parm_inflation : This parameter sets the value of the multiplicative inflation for the covariance. The default value is 0.05.

sigma_obs: This parameter controls the observation localization. The default value is 2.5.