ADVANCED ENVIRONMENTAL DATA ANALYSIS
## HOMEWORK #3

Function fitting

1) Fit a plane to the Pacific SST, $SST(x,y) = \alpha x + \beta y + \gamma$, using the least square approach. The data is downloadable at http://ocean3d.org/eas-8803/hw/03-hw/Feb98˙SST.mat . What are the value of your model parameters? Make a plot of the SST derived using your linear model.

Dispersion Model

2) The figure below is an example of 2D view of the dispersion pattern of a pollutant over your favorite city 100 minutes after its initial release (T_{100}). The left panel shows the true dispersion pattern and the right panel shows the observed dispersion. In the observed dispersion pattern there is noise associated with measurement errors. The error has Gaussian distribution with zero mean and a signal to noise ratio of 1, corresponding to a standard deviation of 0.01.
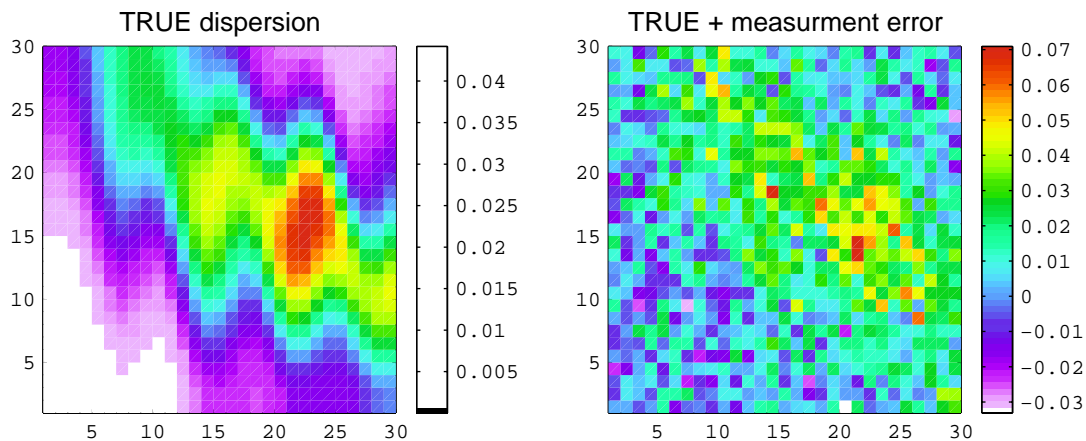
The dynamics controlling the dispersion of the pollutant T are perfectly known and controlled by an advection-diffusion equation

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = K\left[\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right]$$

where the diffusion coefficient K=0.04 m s^{-2}, and the exact velocity (u,v) field is given. An atmospheric scientist gives you a numerical model of the atmospheric circulation over this region (http://ocean3d.org/eas-8803/hw/03-hw/adv˙diff˙model.m). The model runs in MATLAB. The input parameters of the model are the initial condition at $t = 0$ (T_0), the velocity fields and the diffusion coefficient K. You are also told that the sources do not exceed a value of 5 and they always occupy a square area of 4 grid points.
Exercise: Download the MATLAB file http://ocean3d.org/eas-8803/hw/03-hw/run˙model.m and place it in the same directory where you have the circulation model above. This script contains information on how to run the atmospheric model and several hints useful to do this homework.

a) Load the true dispersion data at t=100 (http://ocean3d.org/eas-8803/hw/03-hw/pattern1.dat). Generate synthetic observations by adding random measurements error (the noise). The error must be normally distributed with zero mean and a signal to noise ratio of 5.

Use least squares and the synthetic observations to estimate the location of the point source/sources at $t = 0$. Make a plot of your estimate of the dispersion pattern at $t = 0$ and $t = 100$. Compare the pattern at t=100 with the true and the observed. Label the approximate location of the sources for t=0. Write down what cost function and weights you used for the least square estimate. Write on the title of the figure the value of the cost function and of the spatial correlation between your estimate of T_{100} and the truth. Compute the uncertainty associated with your estimate at $t = 0$. Explore and discuss how your results change with different definition of the weights in the cost function.

(*HINT*: from the picture in class you know that even though true sources may have amplitude of 5 the reconstructed one appear more like a smoothed version with typical size of 0.5, you may want to use this information when you pick your weights. I also suggest you repeat the fit at least 4 times with different realization of the random noise.)

b) Repeat (a) but change the signal to noise ratio to 1. Can you still retrieve the approximate location of the source/sources? If your cost function J followed a Chi-square distribution what would be its expected value? Can you confirm if the model is satisfactory and consistent with the data?

(*HINT*: To perform the Chi-square test you need to repeat the least square fit several times eg. 20 times. Each time add a different realization of the random noise.)

2) Load true dispersion data at t=100 for a different set of sources (http://ocean3d.org/eas-8803/hw/03-hw/pattern2.dat). Generate synthetic observations by adding random measurements error (the noise) with signal to noise ratio of 1 (as in the previous case). You are told that there is only one source and that there are no sources within 2 grid points along the boundary of the model domain. Estimate the location of the source and evaluate if the model is consistent with the data. What happens if you make the weights of your model-observation misfit artificially large? Does it make it easier to isolate the source/sources?

3) Load true dispersion data at t=100 for a different set of sources (http://ocean3d.org/eas-8803/hw/03-hw/pattern3.dat). Generate synthetic observations by adding random measurements error (the noise) with signal to noise ratio of 1. Now you are told that there are for sure two sources of different intensity located at point A (i=9:10 j=7:8) and source B (i=23:24 j=21:22). Again there are no sources along the boundary. Determine which source is likely to be bigger based on the correlation coefficient, check also the value of your cost function. Could you do this problem without knowing the two location before hand? Discuss your results.