ANNEXE 3

New technical scheme for the Kalman filter

A- 3.1 Environment

The files concerning the development and tests of the Kalman filter are located under the directory **/home/disperse/DACFOS/kalman_filter/jerome_test** assigned in the following to $tests. The operational forecasting system is found in **/home/disperse/OPER/SMOG/KALMAN** designated by $KALMAN. Under this directory, the source codes in C are placed in **$KALMAN/src**, while the binary files are placed in **$KALMAN/bin**, which also contains all the scripts necessary to handle Kalman filter inputs and outputs. The following environment variables are necessary to run the Kalman filter:
$DACFOSHOME = /home/disperse/OPER/SMOG/DACFOS/dat
$KALMANTRAJDATA = /net/chip/data/dispopr/SMOG/DACFOS/dat
$JGBHOME = /net/chip/data/dispopr/SMOG/OBS/dat/JGB

A- 3.2 New implementation

As already mentioned the Kalman filter of DACFOS uses independent variables like HIRLAM's temperature and wind in Jægersborg. The new operational version of the program now also includes HIRLAM's Atmospheric Boundary Layer (ABL) height and HIRLAM's surface heat flux (sHf). The modifications required by this new option in the filter are described below, as well as other changes made in the implementation of the system. In the following, the format YYMMDDHH will always designate the date corresponding to the origin of a forecast, where HH states for 00, 06, 12 or 18.

A- 3.2.1 Changes in the source program of the Kalman filter KForecast.c

The program of the Kalman filter itself is now called **KForecast** ; it is compiled by a **Makefile** with the command : **gmake all**.
Concerning the additional extern variables, the Kalman filter is now able to use up to eight options -s. The source code has been modified everywhere the algorithm operates with the matrix **H** by adding the new parameters in the diverse involved functions of the C program.
A new function "*printmatrix*" makes it possible to write (in "Standard Error") in a very comprehensible format all the components of all the matrices involved ; this may be useful for further development or tests of the program.
The output to the file "data/jgb" has been removed because it is not used ; so the program does not require to run a directory named "data" anymore.
When a data is missing, the symbolic value -00 (instead of the previous -99) is written in the output (Standard Error). The program now understands this as a missing data to be not used for statistics. This also avoids problems when plotting the results from the output files.
The program creates a new output file **kf.firstday** that contains the hourly results for the first forecast day together with observations when available. This file will be used to plot the ozone concentration day after day over a long period in order to appreciate the accuracy of the forecast (see below).

A- 3.2.2 Meteorological data

The extern variables issued from the transport model of HIRLAM are contained in the trajectory files **trajdata** in $KALMANTRAJDATA. The previous script get_kal_Tv.sh is

now changed to **get_kal_Tgv.sh**. It is run operational every six hours by DACFOS to get HIRLAM's surface temperature, 10 meter wind, ABL-height and sHf in Jægersborg for an interval including a 6-hour analyse period and a 48-hour forecast. The files issued from this script are named with the format **YYMMDDHH.hirlam.tvg** and are put in **$KALMAN/dat**.

A- 3.2.3 The program get_kalman_data

The new file issued from get_kal_Tgv.sh implies some minor modifications in the code get_kalman_data.c. This program orders the input data for the Kalman filter with a specific format. Though $KALMANTRAJDATA is set by default in the program, it also accepts other values for this environment variable to find the location of the *.hirlam.tvg files. The two options -b and -o are required to tell the program the dates between which the *.hirlam.tvg files must be read. Thus get_kalman_data reads HIRLAM's temperature, wind ABL-height and sHf over an interval of a few days corresponding to the learning period of the Kalman filter. These data are written in a file **kf.hirlam_YYMMDDHH** that also contains HIRLAM's forecast for the next 48 hours. Three other files with the same format are also created by get_kalman_data : **kf.dacfos_YYMMDDHH** receives the five levels DACFOS ozone forecast for Jægersborg, **kf.synop_YYMMDDHH** receives temperature and wind measured in Jægersborg and **kf.jgb_YYMMDDHH** receives Jægersborg's ozone observations each ten minutes.
Finally, the binary file get_kalman_data is compiled with the following command :
gmake -f makefile.get_kalman_data all.

A- 3.2.4 To run the Kalman filter

Each day at 00, 06, 12 and 18 (UTC) a crontab script runs a 2-day forecast for Jægersborg and renews the three preceding days' runs to allow comparison between forecast and observations. The script **run_kalman.scr** which is called by **crontab.scr**. actually handles the run of the Kalman filter.
This new script is almost the same than the previous run_filter_test.scr written by CJ and forecast_joj.scr by JoJ. To avoid any problems with the environment variables, they are defined in the beginning of the script. The program get_kalman_data is called to hunt the input data (observations and model variables) for four days, and the input files for the Kalman filter are then created. The file **kf.obs** is simply a copy of kf.jgb_YYMMDDHH, the file **kf.model** holds the components of the matrix **H** for each hours between -96h and +48h. A new file named **hirsyn** keeps together HIRLAM's forecast and synoptic observations. The three preceding files are systematically saved in **$KALMAN/input** with the suffix YYMMDDHH to allow future reused of their data. The output to the Standard Error is redirected to a "log-file" named **YYMMDDHH_log** and saved in **$KALMAN/output** together with the file **kf.firstday** renamed as **YYMMDDHH_day**.
Then the program KForecast is called with a specific combination of options -s (extern variables) and given values of the parameters Q, R, τ corresponding to a particular Kalman filter.
Previously, there were in run_kalman.scr ten distinct filters running with different extern variables but with the same parameters Q, R and τ. With all the combinations of the eight options there are several hundreds possible filters, but only a few of them are pertinent.

Finally, the temporaries files "update", "predict" and kf.* are now deleted to save disk space.

A- 3.2.5 New scripts for visualisation

In order to visualise the results, the graphics software **xmgr** has been reinstalled for the user disperse in **/home/disperse/lib/xmgr**.

We can use **plot_pre.sh** (see Annexe 2) to draw on the same plot KF and DACFOS forecasts (plus observations if available) two days after origin of forecast. Looking at the results over a longer period (a few weeks or a few months) is also advantageous.

The new script **merge_modelobs.scr** is based on the scripts **date_merge.scr** (see Annexe 1) and **plot_pre.sh** to plot on the same graph modelled and observed quantities. It reads the file **$KALMAN/input/hirsyn_YYMMDDHH** and writes together the observed ozone concentration in Jægersborg with HIRLAM's variables ; this is used to make a plot with **xmgr** in order to see how the ozone concentration is influenced by the meteorological conditions. Another plot is then obtained showing on the same graph HIRLAM's temperature and wind together with observations in Jægersborg. This script accepts an option -b to indicate the date one wants to see ; without any option one obtains a plot showing the present day data with four days analyse and two days forecast.

The late script **longplot.scr** reads a series of files **$KALMAN/ouput/YYMMDDHH_day** in a time-interval defined with the two options -b (for the beginnings' date) and -e (for the endings' date). Then it plots KF and DACFOS forecasts, beside observed ozone concentrations over the corresponding period. The software **xmgr** allows to zoom in the graph if necessary.

A- 3.3 Specific modifications for the other stations

The system of the Kalman filter for the other stations is found under the directory **$tests/**. Some necessary changes have been done in the script **get_kal_Tgv.sh** (in particular an additional option allows to choose the coordinates of the station) to get data from HIRLAM's trajectory files. The location of the observations implies to change the environment variable **JGBHOME** into the emplacements where the respective observations data are conserved. The coordinates of the station are put in at the beginning of the script **run_kalman.scr**. The rest of the system runs in the same way as for Jægersborg. The same modifications can now easily be made for any other stations involved in the DACFOS' stations set.

The parameters (Q, R, τ) of the Kalman filter must be changed in order to behave properly with only a one hour frequency for the measurements.

A- 3.4 Routine to test the Kalman filter for the period Summer 1997

Following the procedure of JoJ (see Annexe 2), the new version of the Kalman filter has been tested on the period including June, July and August 1997. The test and their results are found under **$tests/JGB**. The data files for this period (kf.obs_YYMMDDHH, kf.model_YYMMDDHH and hirsyn_YYMMDDHH) are saved in **$tests/JGB/Sum97input**. Thus hunting the input data with get_kalman_data is not necessary anymore, and the script run_kalman.scr is replaced by **start.scr** which, taking advantage of this, runs faster. The

execution of a test all over the period is run by **summer97.sh** and takes only a few minutes. However, there are some holes in the data during this period and only 77 of the 92 days are actually available.

To examine the quality of Kalman filters' forecasts, we can use the statistics done at the end of the program KForecast (see Annexe 1). JoJ has developed some scripts doing statistics to compare the Kalman filter with DACFOS and "persistence" forecasts. These scripts (located in **$KALMAN/stats**) are useful to test over a long period the results of filters using different combinations of variables. Some modifications have been done to take account of the additional variables in the new Kalman filter. A new script **scatterplot.scr** allows to draw scatter plot diagrams and calculate the corresponding correlation. A little script **clean_sum97.scr** discards the "log-files" containing errors, corresponding to the days data are missing. We can then type the command "**statistics**" which is an alias for :
readlog.sh *_log | readlogstr.