DANISH METEOROLOGICAL INSTITUTE

TECHNICAL REPORT

98-15

Kalman Filtering of DACFOS for Ozone Forecast

Jérôme Chenevez

DMI

COPENHAGEN 1998

ISSN-Nr. 0906-897X

SUMMARY

The DMI's air pollution group has developed an operational forecast system for ozone concentrations over Europe ; this system is called DACFOS [1]. A statistical after-treatment of the DACFOS results with the Kalman filter provides a better description of the local variations of ozone concentration.

The ozone concentration depends on the meteorological conditions. Given ozone concentration measurements, the Kalman filter combines DACFOS' ozone forecast with HIRLAM's meteorological data to adjust the ozone forecast. The variations of parameters such as temperature, wind speed, surface heat flux and Atmospheric Boundary Layer- height are compared with DACFOS' ozone forecast and measured ozone concentrations during four days. It is then possible to filter DACFOS' ozone forecast to predict the following two days ozone concentration. This statistical after-treatment is performed for stations where real time ozone observations are available.

Several tests of different combinations of meteorological parameters have been carried out using ozone observations during the Summer 1997 in two stations in North Zealand and three stations in England. A statistical analysis comparing observations and 48-hour forecast presents better results than DACFOS forecast for a few Kalman filters using specific combinations of the meteorological parameters. The correlation coefficients for observed concentrations and Kalman filters forecasts are between 0.50 and 0.70. The Kalman filters have generally smaller biases and rms values than DACFOS, and the common slight underestimation of DACFOS is generally well compensated by the Kalman filter.

Since May 1998 the Kalman filter is operational for the DMI's station in Jægersborg with four different combinations of meteorological data, among which the surface heat flux, which was not available in the Summer 1997. The same set of Kalman filters runs in a semi-operational setup for two other stations in Denmark and the three English stations.

The forecasts obtained by the different Kalman filters are not alike and the best forecast is not always given by the same filter. The performances of the Kalman filters are very dependant on the accuracy of the meteorological data. The four different Kalman filters are then kept in the operational system in order to give the best possible forecast according to the meteorological conditions.

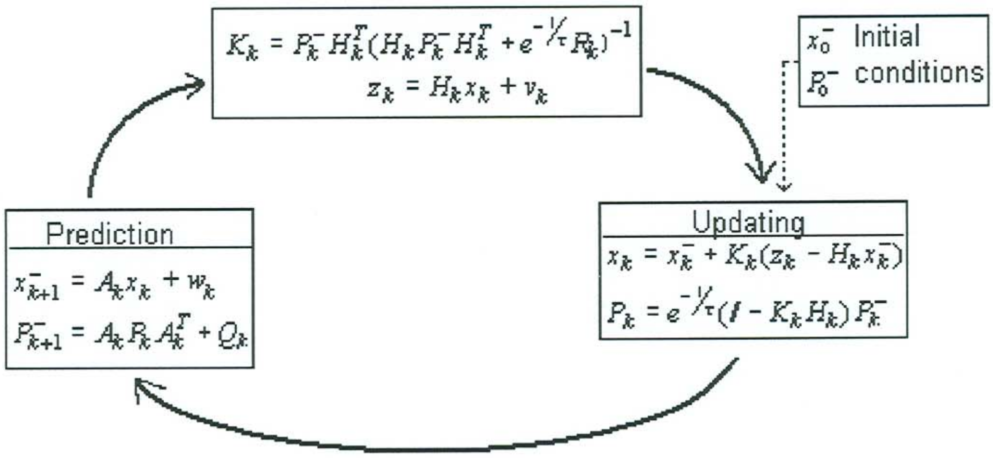
1. Introduction

The present note continues the two preceding reports by Christian Ødum Jensen (CJ) "Kalman filtrering af DACFOS modellen" (Annexe 1) and "Rapport over arbejde med Kalmanfiltrering af DACFOS" by Joachim Jacobsen (JoJ) included in Annexe 2. As a theoretical background one can refer to literature cited in the references list.

After a brief overview of the theory of the Kalman filter, the section 2 resumes the objectives concerning the development of the Kalman filter of DACFOS for ozone forecast. The tests and results of the new version of the Kalman filter for the DMI's station in Jægersborg during the Summer 1997 are presented in section 3. In section 4 the statistics obtained during the same period for the DMU's station in Lille Valby and three other stations in England are summarised. The present operational setup for ozone forecast is exposed in section 5. Finally some conclusions are given in section 6.

1.1 Algorithm of the Kalman filter

Generally speaking the Kalman filter is a statistic tool to predict the behaviour of a signal given observations ; it is a filter because it reduces observations noise. The discrete linear Kalman filter algorithm can be represented by a loop as follows:



Let n denote the dimension of the system and m the dimension of the measurement (number of "observables"), the vector of state $\mathbf{x}_k \in \Re^n$ and the observation $\mathbf{z}_k \in \Re^m$. The dimensions of the different matrices used in the algorithm are the following : $\mathbf{A}_k$(n,n) is the state transition matrix, $\mathbf{P}_k$(n,n) is an error covariance matrix, $\mathbf{K}_k$(n,m) is called the "Kalman Gain" and $\mathbf{H}_k$(m,n) is the measurement matrix ; at last the process error covariance $\mathbf{Q}_k$(n,n) = var($\mathbf{w}_k$) and the measurement error covariance $\mathbf{R}_k$(m,m) = var ($\mathbf{v}_k$) describe the system noise ($\mathbf{w}_k$) and the observation noise ($\mathbf{v}_k$) respectively. The index k is here a discrete time and τ is the memory (a time constant) of the filter. On the left-hand side of the loop is the prediction step of the algorithm, where $\mathbf{x}^-$ and $\mathbf{P}^-$ are the a priori (before observation) estimations of the system state and of the error covariance respectively. On the right-hand side is the measurement updating step, where $\mathbf{x}$ and $\mathbf{P}$ are then the a posteriori (after observation) estimations of state and error covariance respectively. Between these two steps, before updating, the Kalman gain $\mathbf{K}_k$ is calculated and a new measure $\mathbf{z}$ (or observation vector) is introduced. The point is that the observation vector is written as a linear combination of the matrix **H** via the vector of state: at an instant k, given n independent variables in the matrix **H** and m observations that are dependent variables of the system, the algorithm seeks for the best coefficients x_i (components of the vector of state) between the independent variables and the observations. The n+m variables form chronological series that are used by the filter to learn the behaviour of the system. When the coefficients are calculated, the filter can predict the next observations from the independent variables and then forecasts the future behaviour of the system.

1.2 Kalman filter for ozone forecast

This work concerns the DMI's project of ozone forecast and is combined with DACFOS [1]. The Kalman filter is run here to filter DACFOS ozone forecast with parameters such as temperature and wind variations using ozone observations at DMI's station in Jægersborg. The number of these filtering parameters is the number n of independent variables ; we will see later some examples of filters using different variables. In our case m=1, for the only measurement is the ozone concentration in Jægersborg ; the state transition matrix **A=1** because the state does not change from step to step, thus the coefficients x_i are constant. The algorithm finds these constants by comparing the variations of the different variables with ozone concentration variations in a learning period before the forecast onset, during which the coefficients values vary (see Annexe 2). We are in the conditions where $\mathbf{Q}_k = \mathbf{Q}$ and $\mathbf{R}_k = \mathbf{R}$ are constant, allowing both the estimation error covariance $\mathbf{P}_k$ and the Kalman gain $\mathbf{K}_k$ to stabilise quickly and then remain constant [2]. Thus, we can rewrite the preceding equations for our specific system as follows:

$$\text{Prediction :}\quad \mathbf{x}_{k+1}^{-} = \mathbf{x}_k \qquad\qquad \text{Updating :}\quad \mathbf{x}_k = \mathbf{x}_k^{-} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^{-})$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k \quad\Rightarrow\quad [\mathrm{O}_3](t) = \sum_{i=1}^{n} h_i(t)x_i \qquad (1)$$

During the learning period the algorithm optimises the coefficients values by using the observations and the extern independent variables in the updating equation ; the difference $(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^{-})$ is called the measurement innovation. During the forecast the coefficients are set in the prediction equation and the ozone concentration is forecast using the linear combination (1).
A Kalman filter is also determined by its intrinsic parameters Q, R and τ. On the one hand the covariance matrix **Q** includes the error in the assumption that the coefficients x_i are constant, on the other the covariance matrix **R** includes the error in the measurement and in the assumption that the observation is written as a linear combination of the h_i ; these matrices are diagonal and may be defined by their diagonal values Q and R respectively. The parameter τ is called the memory of the Kalman filter because it gives recent measurements a larger influence on the filter estimates and as measurements get older their influence decreases [3]. This last parameter is however not as critical as the two former.
After the work done by CJ and JoJ, the matrix **H** contained up to six variables (see Annexe 1) and the best values for the parameters were found to be : $Q=10^{-4}$, R=1000, τ=200 (see Annexe 2). The variables h_i are DACFOS ozone concentrations in Jægersborg (5 levels or mean), HIRLAM's surface temperature and 10 meter wind in Jægersborg, and eventually a sinusoidal variation (possibly with a given constant) to simulate the diurnal cycle. The best filter was found statistically to use DACFOS' mean with temperature and wind. The

Kalman filtered ozone concentration is then a linear combination of DACFOS and HIRLAM forecasts. The learning period of the Kalman filter is four days, which seems to be a good compromise between ozone forecast quality and a realistic necessary time before the filter becomes operative. The forecast range of DACFOS is two days and so is the forecast range of the Kalman filter.

2. Objectives

a) Improve performances of the Kalman filter, if necessary in adding new extern variables and look for the best options concerning combinations of parameters and variables.
b) Simplify utilisation and organisation of the Kalman filter system for user **disperse**.
c) Make Kalman filter forecast operational for the next "ozone season" starting in spring 1998, eventually in expect of its exploitation on DMI's web page.
d) Run Kalman filter for other stations involved in the DACFOS set (for example DMU's observation stations in Lille Valby and Keldsnor and three other stations in England).

3. Tests for the period Summer 1997

The technical aspects concerning the developments of a new version of the Kalman filter for DACFOS are described in Annexe 3. In brief, the main features of this new version are the use of two new options (-s 7 and -s 8) for HIRLAM's ABL-height and surface heat flux (sHf) respectively. In this section the results obtained with the new version for the period including June, July and August 1997 in Jægersborg are presented.

The routine carried out to perform the following tests is described at the end of Annexe 3.

3.1 Results

One of the new Kalman filters uses HIRLAM's ABL-height besides surface temperature and 10 meter wind. Due to the large variations of the ABL-height, between a few tens meters and a few kilometers, this Kalman filter does not behave the same way as the others. It turns out that the parameters values Q=0.0001, R=1000 and τ=200 are not satisfying for the filters running with the option -s 7 (HIRLAM's ABL-height). After a series of investigations, the best parameters for the filters using the option -s 7 are : Q=0.00001, R=10000 and τ=400.
The table 1, by combining the results of the four daily forecasts, summarises the statistics obtained in the file **sum.dat** (cf. Annexe 2 for explanations about this table) for a specific set of filters using different variables (other filters have been discarded because of relatively poor results). The filter using ABL-height that gives the best results is obtained with the combination -s 2 -s 3 -s 4 -s 7, but it has an higher rms value than the filters without the option -s 7. The best filter, in the sense that it has the lowest rms and the highest correlation, is still the one using the combination -s 1 -s 2 -s 3 -s 4 (DACFOS' mean, temperature and wind, denoted in the following by KF -s 1234) ; like the most of the other filters, it is even better than the persistence. In the table 2, the performances for forecast origin at 00H (UTC) and at 12H (UTC) are compared, showing that the forecasts are slightly better on the afternoon.

Summary of Performance of Dacfos and Kalman Filters for Ozone-Concentration Forecasts.
Reading 311 logfile(s) each with 6 filter(s)
Kalman Filter Parameter Values:
Q-value = 0.0001 R-value = 1000 Memory = 200
u1: Use Dacfos mean
u2: Use Dacfos 5 levels
u3: Use Hirlam Temperature
u4: Use Hirlam Wind
u5: Use Sinus
u6: Use Constant
u7: Use ABL (Q-value = 0.00001 R-value = 10000 Memory = 400)
****

JEAGERSBORG (forecast origin at 00H, 06H, 12H, 18H)

Statistics for the Observations

Mean	St.Dev.	<Peak>	Peak-stdev.
28.62	13.37	44.64	10.57

Mean Errors and Root Mean Square Errors for Forecasts and Peak Forecasts — Best Performers (Best of the day):

	Total Average				Peak 1. 24h		Peak 2. 24h						
	<err>	<|err|>	rms	corr	<err>	rms	<err>	rms	<|err|>	Peak_1	Peak_2	high1	high2
Dacfos:	-3.77	12.18	15.14	0.28	-5.30	15.32	-5.56	15.12	7%	8%	12%	3/31	1/28
Persistence:	-0.19	10.36	13.46	0.49	-0.11	9.37	0.44	10.98	31%	31%	24%	13/31	6/28
Kalman Filter using:													
u1 u2 u3 u4 u5 u6 u7													
1 1 1 0 0 0 0	-1.13	9.69	12.09	0.50	-6.31	11.64	-6.24	13.45	5%	14%	11%	1/31	0/28
1 1 1 1 0 0 0	-1.16	9.60	11.96	0.51	-6.74	11.87	-6.52	13.48	15%	6%	7%	1/31	0/28
0 1 1 1 0 0 0	-1.09	10.29	12.99	0.51	-3.87	12.17	-3.51	14.82	7%	10%	10%	3/31	5/28
0 1 1 0 0 0 0	-1.15	10.41	13.13	0.50	-3.69	12.23	-3.40	14.96	2%	3%	1%	3/31	1/28
0 1 1 1 0 0 1	-0.62	10.69	13.50	0.58	1.87	14.48	3.03	17.29	12%	12%	10%	4/31	9/28
1 1 1 1 0 0 1	-0.02	11.74	15.65	0.55	5.33	20.63	7.57	23.26	14%	8%	13%	3/31	6/28

Table 1 : Global statistics obtained for the period Summer 1997 in Jægersborg.

JEAGERSBORG (forecast origin at 00H)

Statistics for the Observations

Mean	St.Dev.	<Peak>	Peak-stdev.
28.72	13.43	43.96	10.66

Mean Errors and Root Mean Square Errors for Forecasts and Peak Forecasts — Best Performers (Best of the day):

	Total Average				Peak 1. 24h		Peak 2. 24h						
	<err>	<|err|>	rms	corr	<err>	rms	<err>	rms	<|err|>	Peak_1	Peak_2	high1	high2
Dacfos:	-4.22	12.30	15.33	0.27	-5.17	16.18	-6.54	15.69	8%	9%	9%	1/7	0/7
Persistence:	-0.39	10.21	13.26	0.51	-0.21	9.87	0.03	11.32	36%	35%	26%	2/7	0/7
Kalman Filter using:													
u1 u2 u3 u4 u5 u6 u7													
1 1 1 0 0 0 0	-1.64	9.95	12.43	0.50	-6.52	12.60	-7.56	14.02	5%	9%	10%	0/7	0/7
1 1 1 1 0 0 0	-1.67	9.86	12.31	0.51	-6.91	12.77	-7.70	14.04	12%	4%	6%	0/7	0/7
0 1 1 1 0 0 0	-2.06	10.75	13.57	0.50	-4.63	14.37	-5.55	15.13	4%	13%	10%	2/7	3/7
0 1 1 0 0 0 0	-2.14	10.87	13.71	0.49	-4.47	14.45	-5.52	15.26	3%	3%	1%	1/7	0/7
0 1 1 1 0 0 1	-1.40	11.01	13.84	0.56	0.43	15.05	1.39	17.14	14%	6%	10%	0/7	3/7
1 1 1 1 0 0 1	-0.66	11.91	15.70	0.54	3.60	20.50	6.07	22.42	12%	12%	13%	1/7	1/7

****

JEAGERSBORG (forecast origin at 12H)

Statistics for the Observations

Mean	St.Dev.	<Peak>	Peak-stdev.
28.44	13.43	45.48	10.65

Mean Errors and Root Mean Square Errors for Forecasts and Peak Forecasts — Best Performers (Best of the day):

	Total Average				Peak 1. 24h		Peak 2. 24h						
	<err>	<|err|>	rms	corr	<err>	rms	<err>	rms	<|err|>	Peak_1	Peak_2	high1	high2
Dacfos:	-3.48	12.33	15.34	0.27	-4.82	14.94	-5.14	14.88	4%	4%	10%	0/8	0/7
Persistence:	-0.05	10.44	13.58	0.49	0.16	8.78	1.19	10.34	31%	26%	25%	4/8	3/7
Kalman Filter using:													
u1 u2 u3 u4 u5 u6 u7													
1 1 1 0 0 0 0	-0.79	9.47	11.89	0.51	-5.80	10.71	-5.38	13.20	6%	17%	8%	1/8	0/7
1 1 1 1 0 0 0	-0.83	9.39	11.75	0.53	-6.29	11.04	-5.76	13.24	21%	8%	8%	0/8	0/7
0 1 1 1 0 0 0	-0.27	9.93	12.70	0.53	-2.31	9.93	-1.75	14.79	6%	9%	13%	0/8	0/7
0 1 1 0 0 0 0	-0.32	10.05	12.85	0.52	-2.08	9.96	-1.59	14.96	3%	3%	1%	1/8	0/7
0 1 1 1 0 0 1	0.43	10.48	13.21	0.61	6.47	14.85	5.66	16.45	9%	16%	8%	2/8	1/7
1 1 1 1 0 0 1	0.97	11.54	15.29	0.59	11.46	22.28	11.35	23.38	12%	8%	13%	0/8	3/7

Table 2 : The same as table 1, but only for forecast origin at 00h and 12h (UTC).

It is also interesting to look at the quality of the different forecasts as a function of time after forecast origin. The three following figures show the behaviour of three indicative statistical values for each Kalman filter present in the table 1 (and given in the same order), as well as for DACFOS and the persistence. The correlation is showed in figure 1, the bias in figure 2 and the rms value in figure 3. Like in table 1 they are obtained by compiling the four daily forecasts and their variations are then modulated by the periodicity of the forecasts, that is each six hours. Accordingly to the results shown in table 1, the correlation factors and biases of the filters are obviously better than for DACFOS. In figure 1 the correlation coefficients of the filters slightly decrease with time after forecast. In figure 2 the filters lie between DACFOS, which underestimates the measurements, and the persistence, which bias is almost equal to zero. One can see that the filters using option -s 7 (especially the second one) overestimate the ozone concentration of the second forecast day. The figure 3 shows that the filters, but the last one, have better rms values than the persistence and DACFOS ; the best filter is again the one using the combination -s 1 2 3 4.

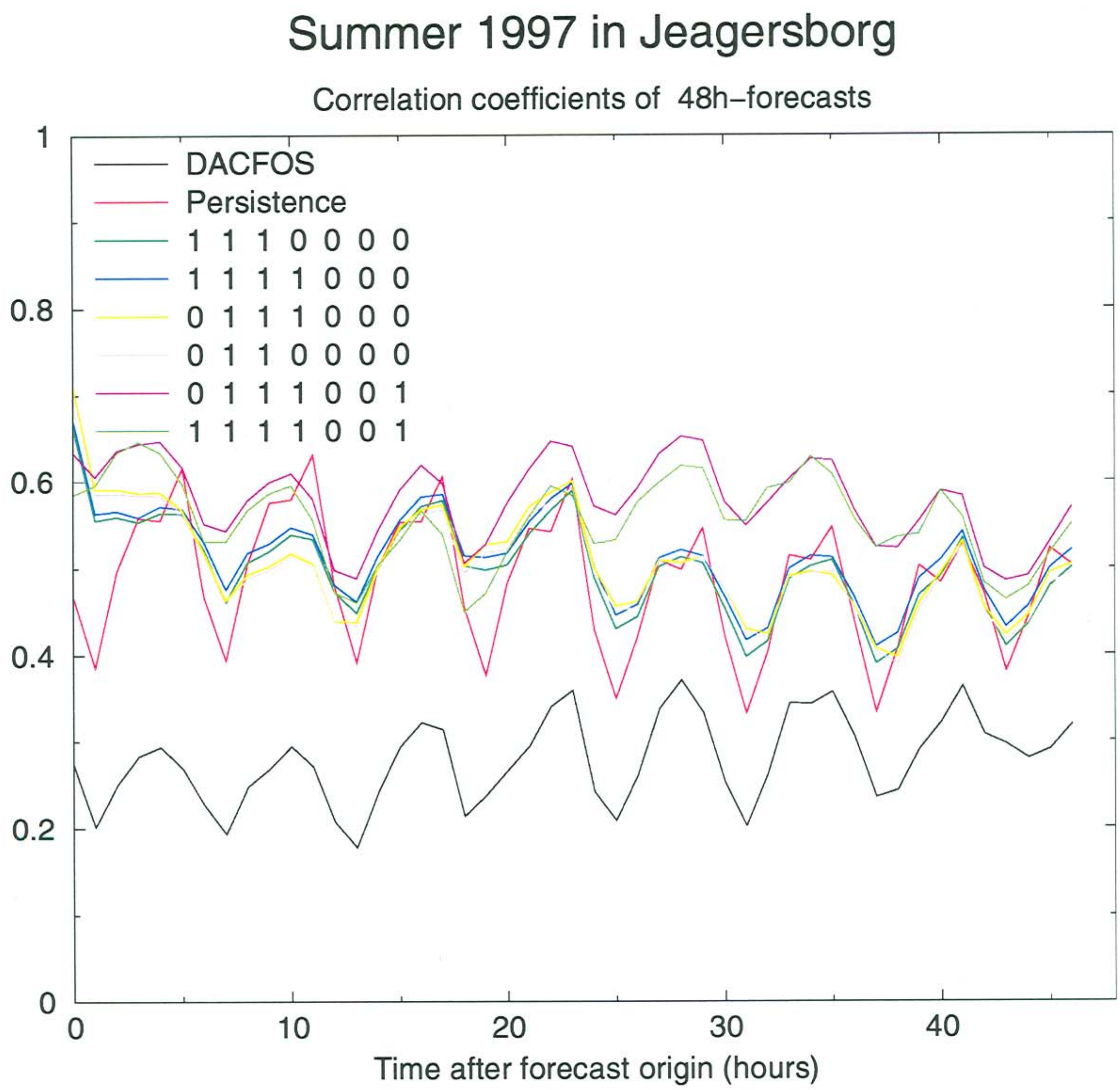


Figure 1 : Comparative evolution of the correlation coefficients for the Kalman filters, DACFOS and the persistence versus time after forecast origin.

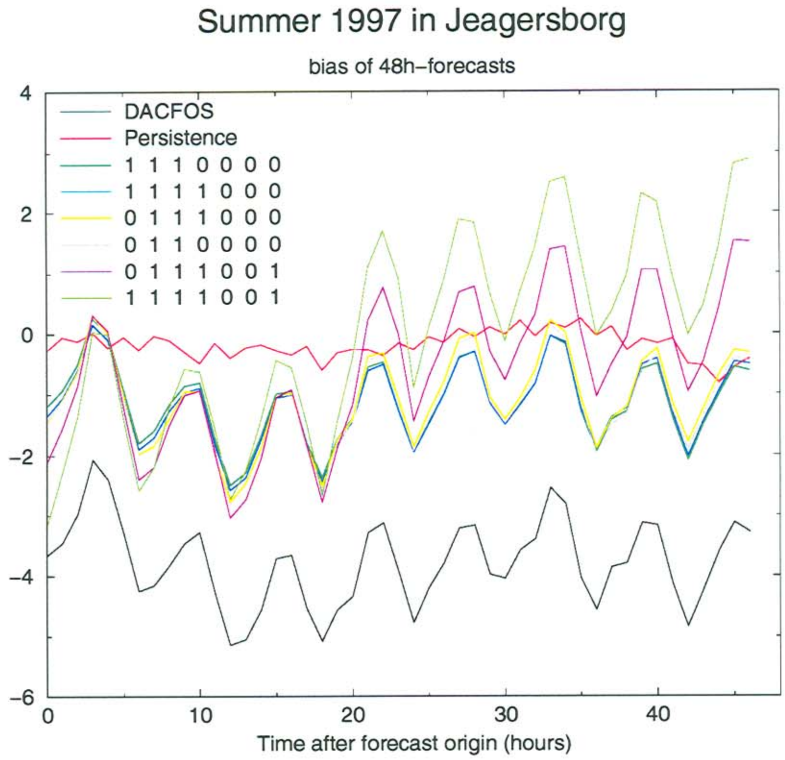


Figure 2 : Comparative evolution of the biases for the Kalman filters, DACFOS and the persistence versus time after forecast origin.

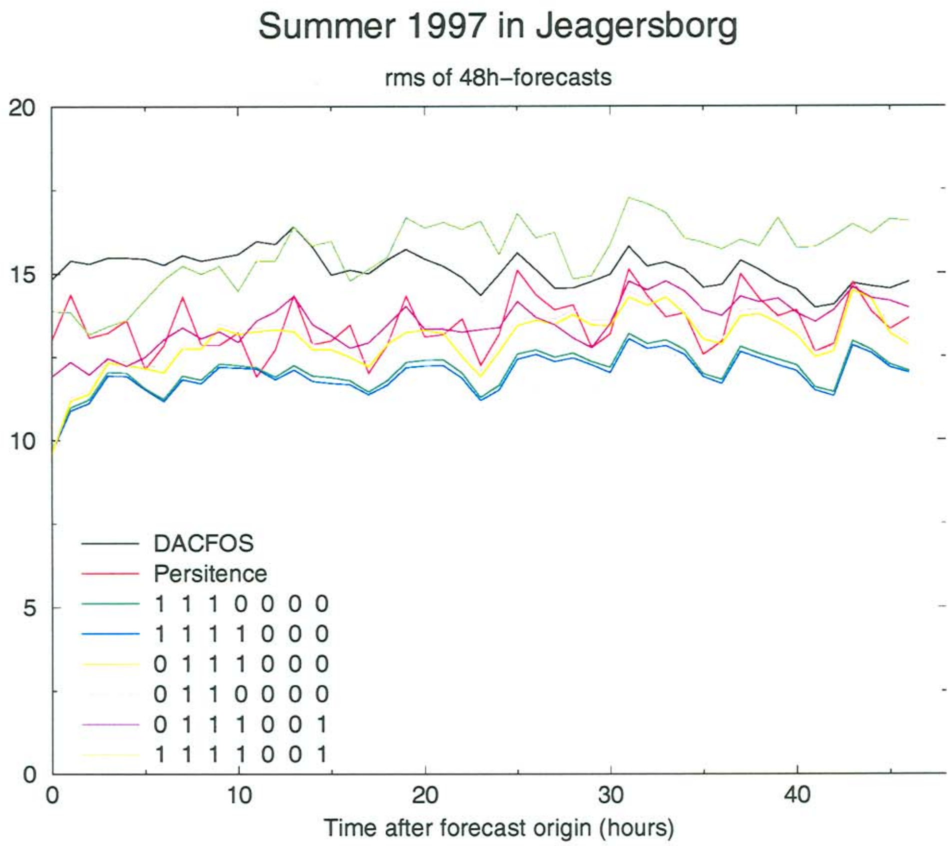


Figure 3 : Comparative evolution of the rms values for the Kalman filters, DACFOS and the persistence versus time after forecast origin.

Scatter plots of the Kalman filter (-s 1234) and DACFOS for each six hours after forecast origin during the 48-hour forecast period are shown in Annexe 4. From these scatter plots one can calculate the correlation factors between observed ozone concentration and Kalman filter and DACFOS ozone concentrations respectively. The evolution of these correlation factors versus forecast time is displayed in figure 4. The scatter plots and the corresponding correlation factors compile the four daily forecasts of DACFOS and the Kalman filter (-s 1234) during the whole period between 03/06/97 and 23/08/97.

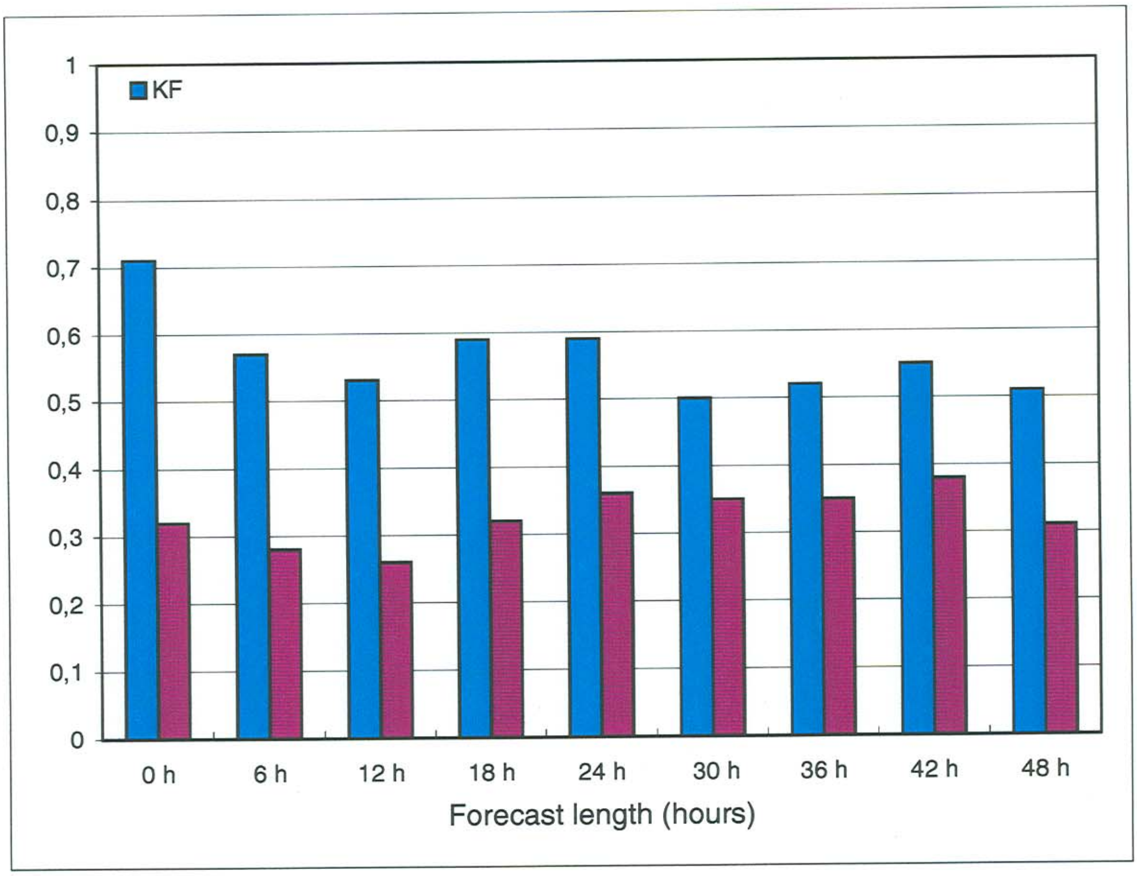


Figure 4 : Correlation coefficients for DACFOS and KF (-s 1234) in Jægersborg for the period between 03/06/97 and 23/08/97.

However, statistics are not always the best indicators to find the best filters. For instance, a filter using only HIRLAM's temperature and wind plus a constant, without DACFOS, has always the lowest rms ; but, looking at the daily forecast plots shows that the variations of this particular filter have not the same amplitude as the observations. It is then necessary to inspect some key days with specific conditions ; a few examples will now be given.

Despite its relatively weak statistical performances (comparatively to KF -s 1234), the filter using ABL-height in the combination -s 2 -s 3 -s 4 -s 7 is often the best to forecast the highest peaks in the period of summer 97 (this is also shown by the fairly good relative performances of this filter in tables 1 and 2, in the columns “Best performers”). To illustrate this, let us take the same example as JoJ (see Annexe 2), that is forecast at 97081400. The forecast statistics obtained at 97081400 for the filter KF -s 2347 are shown in table 3 (cf. Annexe 1 for explanations about this table).

forecast statistics	: ----- 47 points -----	
	KF	DACFOS
root mean sqr :	16.697	19.670
est. val. :	0.459	0.249
mean abs error :	13.194	15.633
mean bias :	5.510	-8.824
correlation :	0.826	0.638
significance :	0.000	0.000
Fishers z :	1.176	0.756

Table 3 : Example of statistics obtained for 48 h forecast at 97081400 with the KF -s 2347.

The figure 5 and 6 show that during the learning period, DACFOS always underestimates the actual ozone concentration. It is then relatively easy for the Kalman filter to find the best coefficient between DACFOS and observations. However the filter on figure 5, which does not use ABL-height, also underestimates the first day's peak, while the same filter, but using ABL-height (shown on figure 6) is very close to this peak. The second day the forecast peak is not so good because DACFOS overestimates it. The figure 7 showing the variations of observed ozone concentrations and meteorological variables helps to understand what is going on. It reveals that during the period around the 14/08/97, the ozone concentration was very correlated with the variations of the ABL-height (this is not always the case). That explains why the second filter is better than the first one.

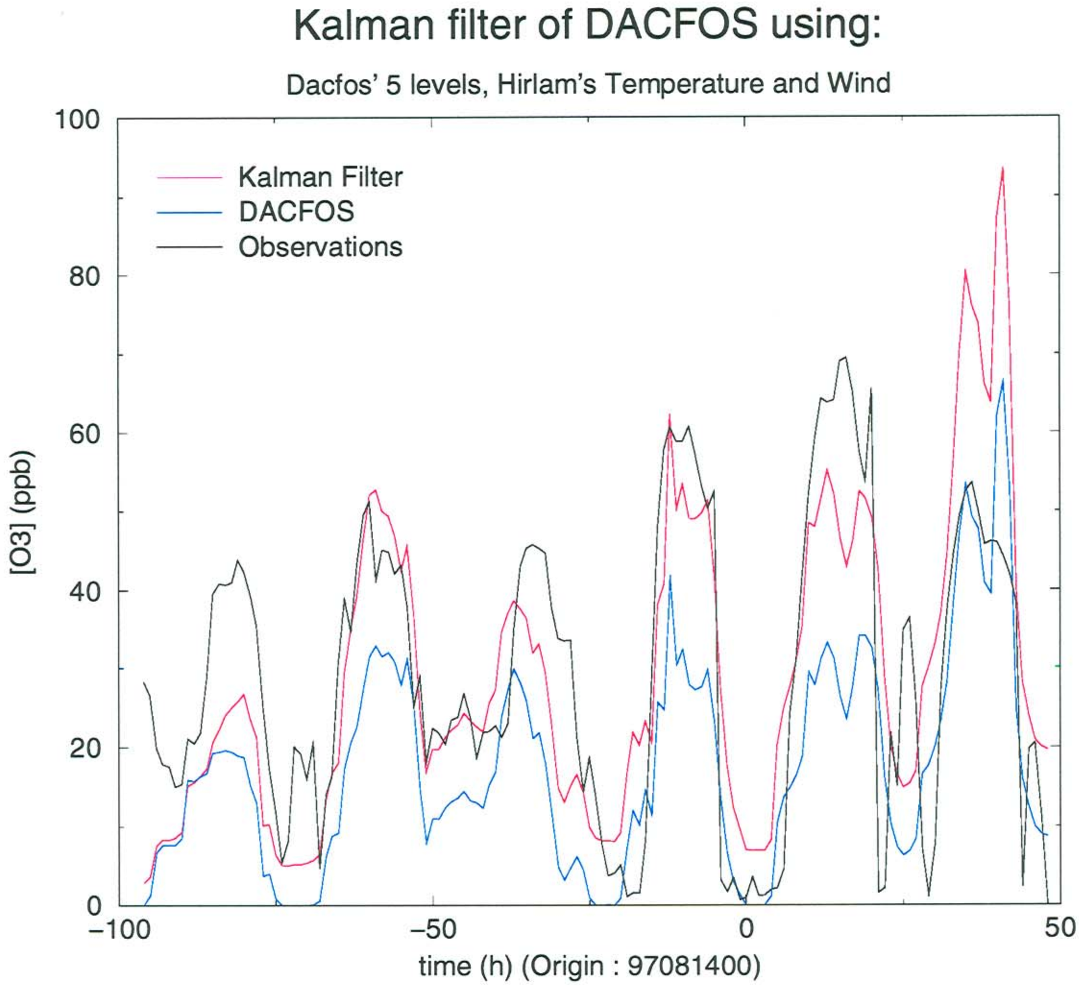


FIGURE 5 : Plot of observed [O3], DACFOS and KF (-s 234) including the learning period and 48-h forecast for the 14/08/97 in Jægersborg.

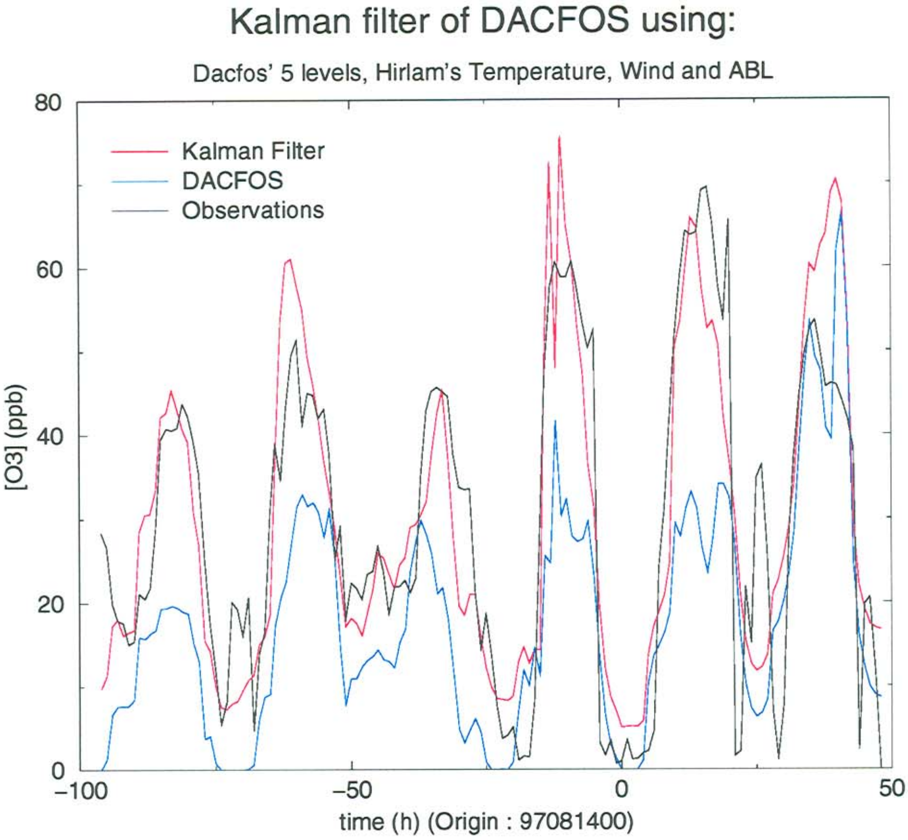


FIGURE 6 : Same as Fig. 5 with KF options combination -s 2347.

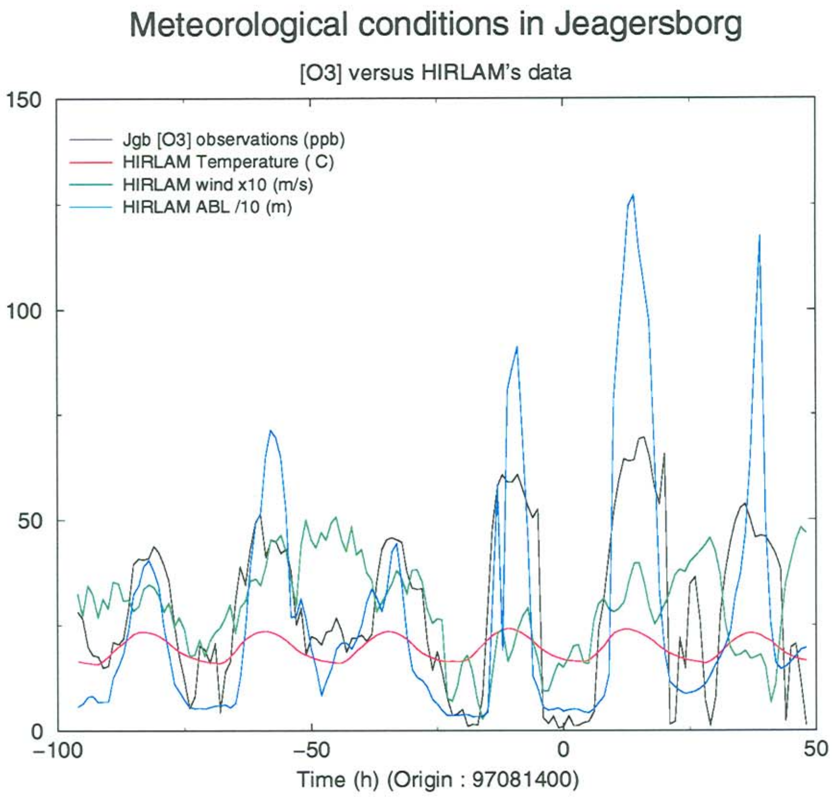


FIGURE 7 : Comparative plot of [O3] with meteorological parameters in Jægersborg during the same time interval as Fig. 5 and 6.

The situation where the ozone concentration is well correlated with ABL-height variations is relatively common in summer. A filter using ABL-height is thus better in summer than in winter. So one may trust a filter using ABL-height only in summer ; other filters are better outside the summer period. Since DMI's objective is to forecast the highest ozone concentration peaks especially in summer, a Kalman filter using the ABL-height is kept for operational forecast.

In Annexe 5, one can see a plot of the time series of observed and modelled ozone concentration in Jægersborg during the whole period considered, that is between 03/06/97 and 23/08/97. The Kalman filter represented in this plot is the one with the best statistic, using DACFOS' mean with HIRLAM's temperature and wind (-s 1234).

Figure 8 shows the daily maximum ozone concentrations in Jægersborg during the Summer 1997 for observations, DACFOS and the Kalman filter (-s 1234). The observed [O3] in Jægersborg exceeded the threshold of 60 ppb eight days in the Summer 1997. The correlation between daily maximum Kalman filter's [O3] and daily maximum observed [O3] is 0,63 while DACFOS' daily maximum values are slightly anti-correlated with the observations.

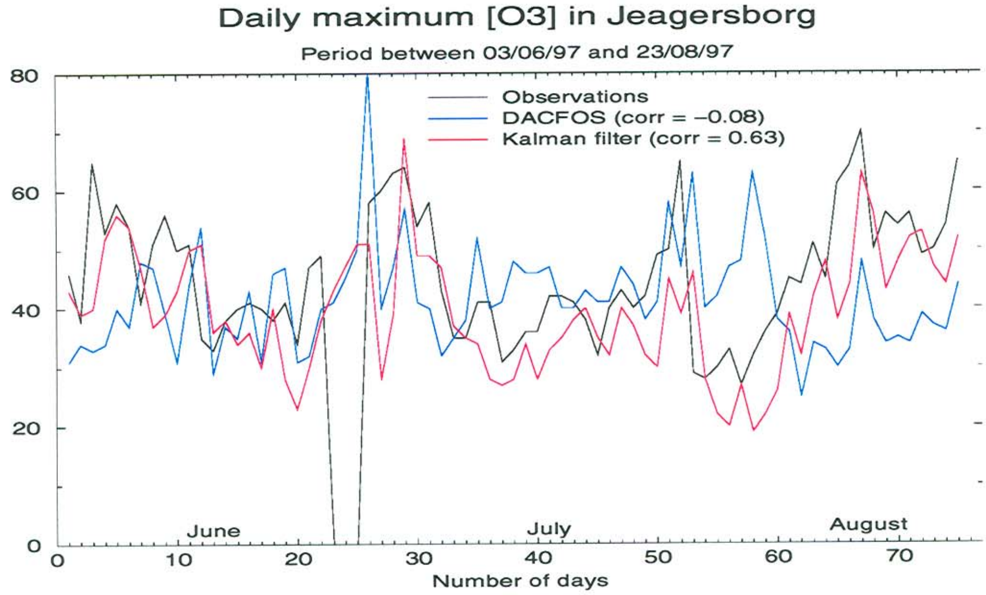


FIGURE 8 : Daily maximum [O3] for the Summer 1997 in Jægersborg ; KF (-s 1234).

As shown on figure 8 (see also the plot over all the period in Annexe 5), one of the highest peaks in the Summer 1997 happened on June 5th with a measured [O3]=65 ppb. The figure 9 shows the forecast of different Kalman filters for June 5th and 6th. On figure 9-a the peak is forecast by the Kalman filter using ABL-height (-s 2347) with a too high value of 95 ppb. On figure 9-b, the "classical" Kalman filter (-s 1234) underestimates the peak and is only slightly higher than DACFOS ; but, adding the use of the surface heat flux (figure 9-c) slightly increases the maximum value. At last, the Kalman filter using both ABL-height and surface heat flux gives a peak (59 ppb) close to the observations (figure 9-d). Unfortunately the surface heat flux was only available in the beginning of June 1997 (and since May 1998) and could not be used during the rest of the Summer 1997. However, it does not mean that the last configuration is always the best ; for instance, the first configuration (-s 2347) gives the best result for a forecast origin only 12 hours later.

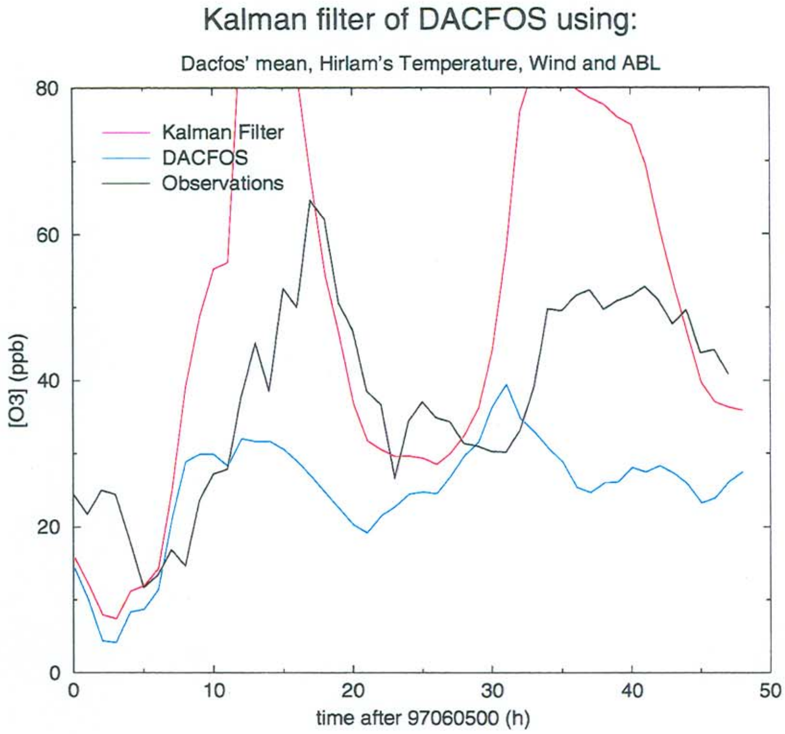


Figure 9-a : Plot of 48-h forecast for KF (-s 2347), DACFOS and observed [O3] in Jægersborg for the 05/06/97 (the KF reaches a peak at 95 ppb).

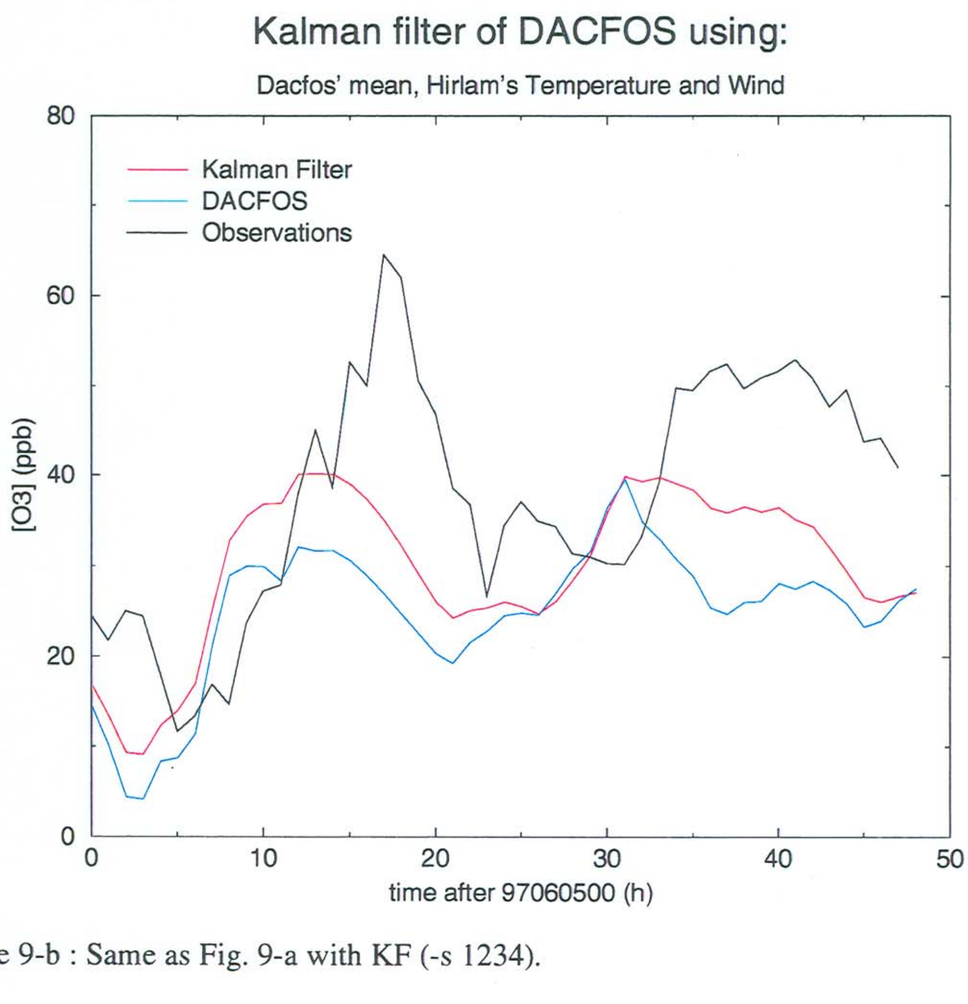


Figure 9-b : Same as Fig. 9-a with KF (-s 1234).

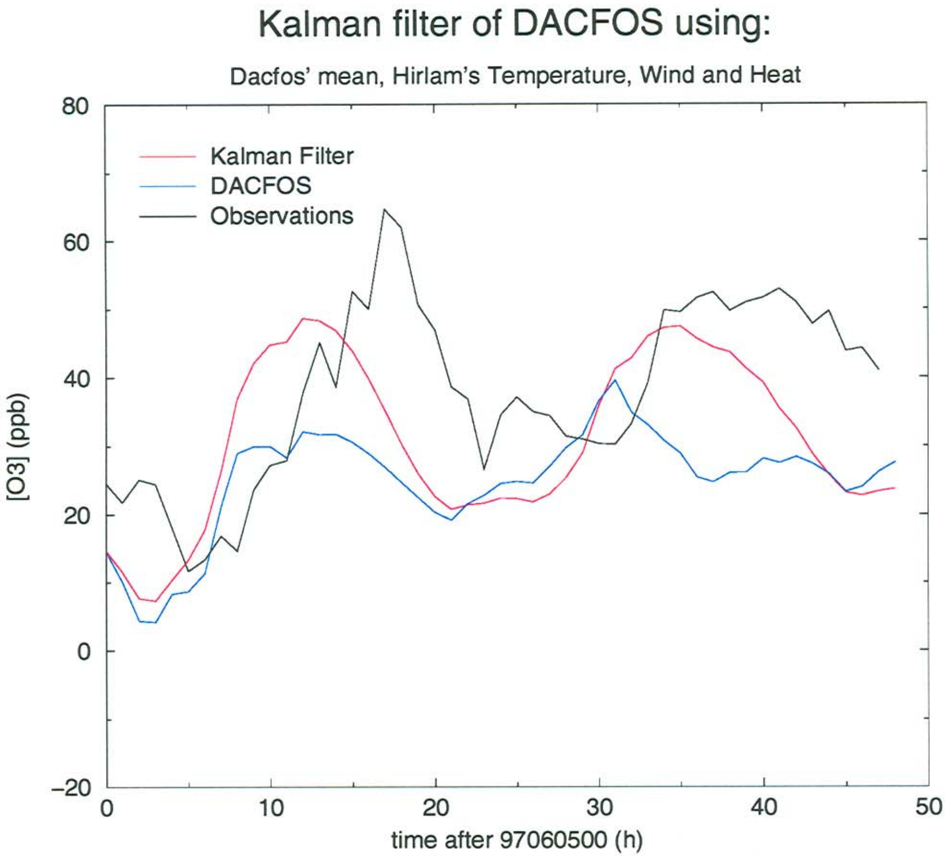


Figure 9-c : Same as Fig. 9-b with KF using surface heat flux.

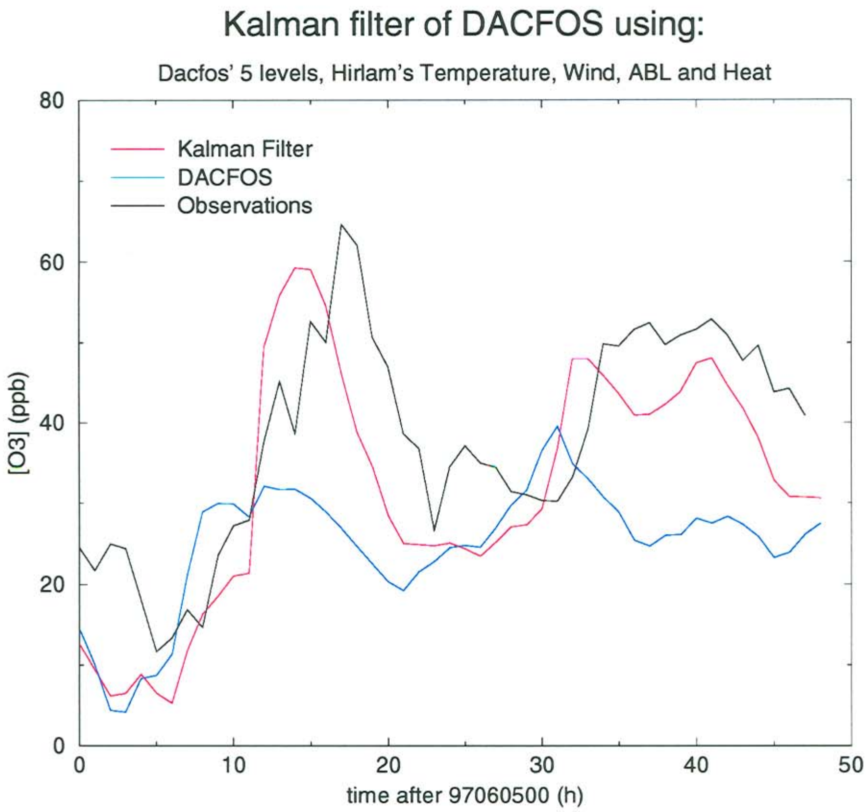


Figure 9-d : Same as Fig. 9-a with KF using surface heat flux.

3.2 Improving tests

The input file holding ozone concentration measurements for the Kalman filter contains observation every ten minutes at Jægersborg. It is easy to alter this file to keep a 1-hour running mean of the observations (see Annexe 1). As already mentioned by JoJ, it has been checked that smoothing the observations is not pertinent with the parameters values used in the filters. It gives only better statistics on the second digit after the comma. Therefore, it has been decided to keep putting 10-mn observations in the Kalman filter.
Besides, a 3-hour running mean on DACFOS ozone concentration has been investigated in order to prevent the Kalman filter from some artefacts in DACFOS' peaks. The statistics are only slightly better (less than 1%) and some forecast days are even worse. A test has also been tried using Jægersborg synoptic temperature and wind measurements instead of HIRLAM's. Of course this method cannot be used for operational forecast, but can only be applied when these observations are available after forecast origin. However, the statistics thus obtained are worse, probably because of the too frequent irregularities in actual temperature and wind.

3.3 Experimental version

There is now a late version including all the meteorological data contained in HIRLAM's trajectory file ; that is temperature, wind, ABL-height, relative humidity, precipitation intensity, total cloud cover, surface heat flux, surface moment flux and surface pressure. All the necessary modifications mentioned in Annexe 3 have been managed in the same way. Thus, the Kalman filter may use up to 13 extern variables and the number of valid combinations is very large ; only a few of them have been tested. However, it does not seem that using additional options yields better statistics than the previous filters ; the best one is a filter using DACFOS' 5 levels plus HIRLAM's temperature, surface moment flux and precipitation intensity, which yields statistics almost as good as the KF -s 1234. Moreover, these new variables are less relevant according to their influence on the ozone concentration ; it can hence be concluded that it is not advantageous to use further external variables.

4. Other stations

The measurements of ozone concentration during the Summer 1997 are also available for the DMU's ozone observation station in Lille Valby and three English stations in Harwell, Keldsnor and Strath Vaich. These four stations, like Jægersborg, are also DACFOS trajectory arrival points. We can then apply the previous procedure to Kalman-filter the DACFOS ozone forecast in these stations.
The program of the Kalman filter is of course able to use hourly observations instead of 10 min observations like in Jægersborg. This implies however to modify the characteristic parameters Q, R and τ of the Kalman filter. In these conditions the best parameters (in the sense that they give the best statistics) are found to be : Q=0.001, R=100 and τ=200 and Q=0.0001, R=10000 and τ=400 for the filters using the option -s 7. The same tests as for Jægersborg have been performed for the four stations. The statistical performances concerning the different filters agree with those obtained in Jægersborg and lead to the same conclusions. The results obtained for each of these stations are only briefly resumed hereafter by the statistics obtained in the respective files "sum.dat" concerning the period of the Summer1997 and for a forecast origin at 00 H (UTC).

4.1 Lille valby

Summary of Performance of Dacfos and Kalman Filters for Ozone-Concentration Forecasts.
Reading 77 logfile(s) each with 6 filter(s)
Kalman Filter Parameter Values:
Q-value = 0.001 R-value = 100 Memory = 200
u1: Use Dacfos mean
u2: Use Dacfos 5 levels
u3: Use Hirlam Temperature
u4: Use Hirlam Wind
u5: Use Sinus
u6: Use Constant
u7: Use ABL (Q-value = 0.0001 R-value = 10000 Memory = 400)

LILLE VALBY (forecast origin at 00H)

Statistics for the Observations

Mean	St.Dev.	<Peak>	Peak-stdev.
29.88	14.78	46.49	12.82

Mean Errors and Root Mean Square Errors for Forecasts and Peak Forecasts; Best Performers (Best of the day):

	Total Average				Peak 1. 24h		Peak 2. 24h		Best Performers				
	<err>	<|err|>	rms	corr	<err>	rms	<err>	rms	<|err|>	Peak_1	Peak_2	high1	high2
Dacfos:	-5.95	12.30	16.04	0.64	-8.25	18.41	-8.84	18.27	13%	7%	9%	0/6	0/5
Persistence:	-1.36	8.79	11.93	0.82	-2.09	9.63	-2.31	13.31	41%	28%	16%	2/6	2/5
Kalman Filter using:													
u1 u2 u3 u4 u5 u6 u7													
1 1 1 0 0 0 0	-1.25	9.86	12.52	0.70	-5.97	14.38	-9.07	15.38	5%	5%	11%	0/6	0/5
1 1 1 1 0 0 0	-1.26	9.73	12.34	0.71	-6.30	14.31	-9.05	15.27	5%	3%	1%	0/6	0/5
0 1 1 1 0 0 0	-1.56	10.19	13.23	0.67	-2.72	15.24	-6.59	15.08	8%	7%	9%	1/6	1/5
0 1 1 1 0 0 1	-1.37	10.31	13.55	0.59	1.61	16.53	-0.06	17.00	16%	16%	8%	3/6	2/5

4.2 Harwell

HARWELL (forecast origin at00H)

Statistics for the Observations

Mean	St.Dev.	<Peak>	Peak-stdev.
30.17	15.19	48.24	17.12

Mean Errors and Root Mean Square Errors for Forecasts and Peak Forecasts; Best Performers (Best of the day):

	Total Average				Peak 1. 24h		Peak 2. 24h		Best Performers				
	<err>	<|err|>	rms	corr	<err>	rms	<err>	rms	<|err|>	Peak_1	Peak_2	high1	high2
Dacfos:	-7.66	13.44	17.65	0.42	-4.06	20.00	-5.05	21.99	12%	19%	14%	2/16	4/14
Persistence:	0.37	11.67	16.05	0.49	0.44	14.78	2.21	18.18	34%	26%	30%	5/16	5/14
Kalman Filter using:													
u1 u2 u3 u4 u5 u6 u7													
1 1 1 0 0 0 0	-1.51	9.95	13.50	0.58	-7.82	16.49	-7.76	19.51	9%	16%	12%	2/16	0/14
1 1 1 1 0 0 0	-1.31	9.89	13.39	0.58	-7.96	16.32	-7.81	19.39	16%	8%	9%	1/16	0/14
0 1 1 1 0 0 0	-2.07	10.62	14.56	0.56	-5.35	18.22	-6.27	20.03	22%	12%	9%	3/16	0/14
0 1 1 1 0 0 1	-3.64	11.79	15.50	0.58	-1.07	17.50	-2.36	18.26	6%	13%	18%	3/16	5/14

4.3 Ladybower

LADYBOWER (forecast at 00H)

Statistics for the Observations

Mean	St.Dev.	<Peak>	Peak-stdev.
29.56	13.54	43.34	15.50

Mean Errors and Root Mean Square Errors for Forecasts and Peak Forecasts; Best Performers (Best of the day):

	Total Average				Peak 1. 24h		Peak 2. 24h		Best Performers				
	<err>	<|err|>	rms	corr	<err>	rms	<err>	rms	<|err|>	Peak_1	Peak_2	high1	high2
Dacfos:	-10.74	13.92	18.14	0.40	-4.90	19.50	-7.16	17.89	8%	9%	16%	1/11	3/9
Persistence:	0.09	11.61	15.58	0.39	0.79	16.11	0.84	18.12	31%	21%	21%	2/11	1/9
Kalman Filter using:													
u1 u2 u3 u4 u5 u6 u7													
1 1 1 0 0 0 0	-5.28	9.77	13.04	0.60	-4.83	12.48	-5.24	16.10	18%	17%	16%	0/11	1/9
1 1 1 1 0 0 0	-5.71	9.81	13.05	0.60	-6.06	12.77	-6.40	15.97	18%	13%	9%	2/11	1/9
0 1 1 1 0 0 0	-5.68	11.02	15.04	0.56	-2.12	14.58	-2.23	22.37	5%	17%	14%	3/11	3/9
0 1 1 1 0 0 1	-5.36	11.46	15.25	0.53	0.72	15.45	-1.64	14.41	17%	14%	12%	3/11	0/9

4.4 Strath Vaich

STRATH VAICH (forecast origin at 00H)													
Statistics for the Observations													
	Mean	St.Dev.		<Peak>	Peak-stdev.								
	29.73	10.23		37.49	11.12								
Mean Errors and Root Mean Square Errors for Forecasts and Peak Forecasts								Best Performers (Best of the day):					
	Total Average				Peak 1. 24h		Peak 2. 24h						
	<err>	<|err|>	rms	corr	<err>	rms	<err>	rms	<|err|>	Peak_1	Peak_2	high1	high2
Dacfos:	-7.15	11.16	13.98	0.21	-3.72	11.68	-4.75	11.37	5%	19%	26%	1/3	1/2
Persistence:	-0.44	7.92	10.78	0.50	-0.21	8.65	-0.03	11.75	35%	21%	14%	0/3	0/2
Kalman Filter using:													
u1 u2 u3 u4 u5 u6 u7													
1 1 1 0 0 0 0	-2.97	8.27	10.76	0.50	-1.50	10.27	-2.56	12.54	10%	10%	12%	0/3	0/2
1 1 1 1 0 0 0	-2.81	8.03	10.44	0.51	-2.03	9.82	-2.82	12.37	16%	8%	6%	0/3	0/2
0 1 1 1 0 0 0	-4.02	8.63	11.31	0.50	-2.64	10.71	-3.84	13.32	13%	14%	8%	0/3	0/2
0 1 1 1 0 0 1	-3.27	9.67	12.47	0.41	0.23	11.47	-0.51	13.91	6%	9%	13%	2/3	1/2

5. Operational setup

Four different Kalman filters have ultimately been retained for operational forecast. They correspond to the combinations -s 1 -s 2 -s 3 (DACFOS' mean plus HIRLAM's temperature), -s 1 -s 2 -s 3 -s 4 (the preceding plus HIRLAM's wind), -s 1 -s 2 -s 3 -s 4 -s 8 (the preceding plus surface heat flux) and -s 2 -s 3 -s 4 -s 7 (DACFOS' 5 levels plus HIRLAM's temperature, wind and ABL-height). The two first Kalman filters have given the best statistical performances in the Summer 1997. The two other, by taking account of other meteorological data, complete the two former and make wider the field of possibilities to forecast the ozone concentration in different meteorological conditions.

This setup is entirely operational for Jægersborg, which observations are automatically acquired every ten minutes, and runs four times a day at each six hours. Hourly observed ozone concentrations are published daily on the DMU's web-page for their stations in Lille Valby and Keldsnor (they are also published hourly since 1st May 1998). We have now observations and input data to run the Kalman filter for these stations since 01/01/98. The observations data for the three stations in England are also available hourly since 01/06/98. The Kalman filter for these five stations is ready to run operational but we need authorisations from the concerned organisations to use their data in a complete operational system.

As an example, the figure 10 shows 48-hour forecast results and observations for the 14/06/98 00H (UTC) at the DMU's station in Keldsnor. The four distinct Kalman filters do not give very different results and are all together in good accord with DACFOS result, which itself gives a relatively good forecast. This absence of divergence may be explained by the fact that the ozone concentration was not very high, although it was middle of June. Note that the observed ozone concentration almost did not vary during these two days. However, the models predict a usual decrease in the night, and therefore the forecast ozone concentrations are too low for both DACFOS and Kalman filter during the night. Nevertheless, we are rather interested by the diurnal ozone concentration, that is when it is more likely to reach high values. The maximum values are obtained late at the end of the afternoon (49 ppb and 48 ppb, at +18h and +40h respectively) and are actually well forecast by the Kalman filter using DACFOS' mean, HIRLAM's temperature, wind and surface heat flux.

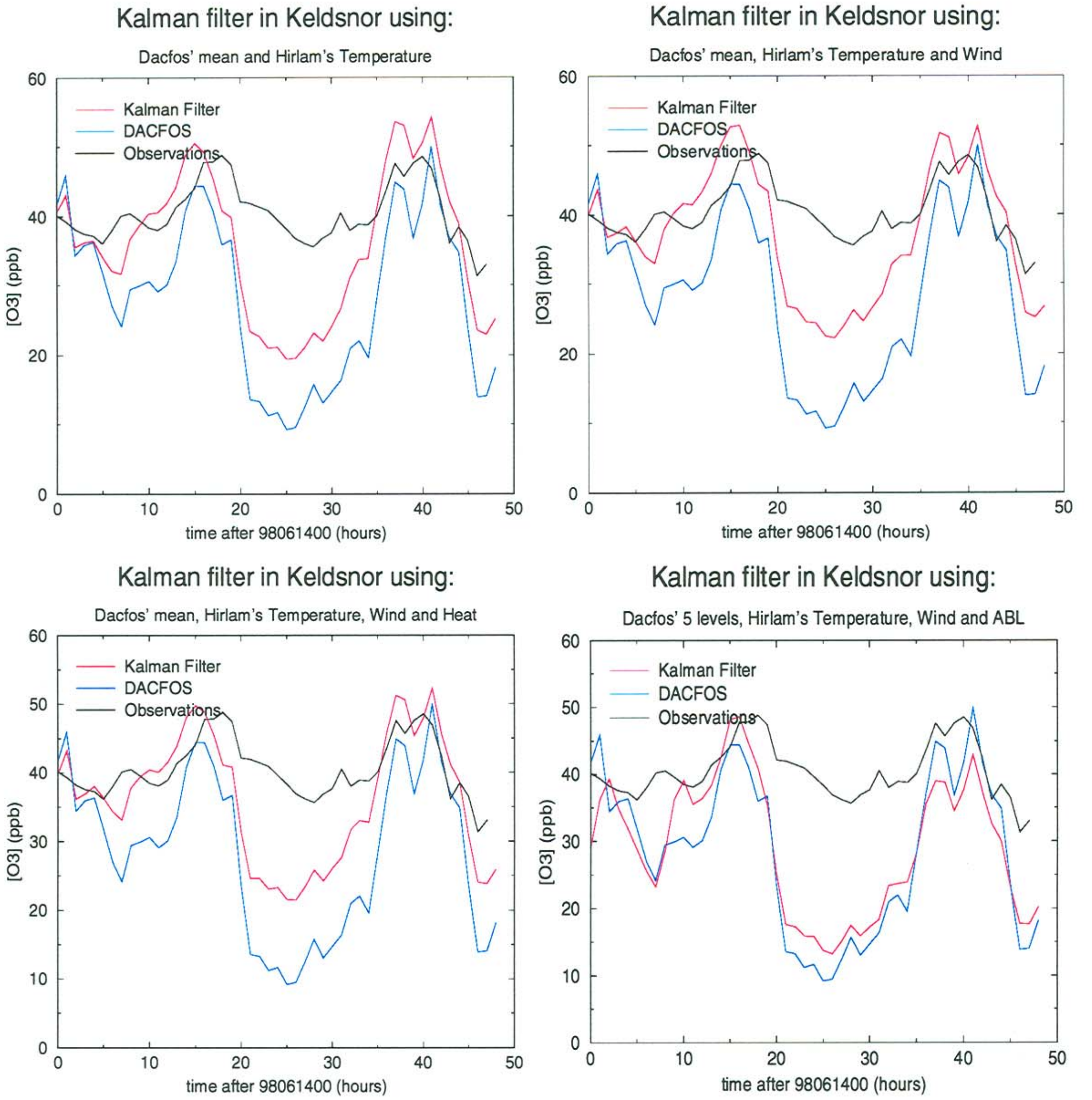


Figure 10 : Results of the four operational Kalman filters in Kelsdnor for a forecast origin on the 14/06/98 at 00H (UTC)

6. Conclusion

The Kalman filter of DACFOS for ozone forecast has been improved in order to use further meteorological variables, and it has been tested over the whole Summer 1997 for six ozone observation stations in Denmark and in England. It is found for all these six stations that the same combinations of DACFOS forecast with meteorological variables give better results than others. Generally, the Kalman filter results improve DACFOS results with correlation coefficients two times better than DACFOS'. Four Kalman filters have ultimately been retained to be used in the operational setup, which runs since May 1998 for the DMI's station in Jægersborg and in a semi-operational mode for the others. The quality of the forecast depends on the quality of the meteorological data, and the best forecast will often be given by one of these four Kalman filters according to the specific meteorological conditions.

Acknowledgement

I wish to thank Christian Ødum Jensen and Joachim Jacobsen for their contribution to this work and to have helped me in my first steps with the Kalman filter system. I am also grateful to Sissi Kilshom who got the observations data in the required format from the Internet.

References

[1] Jensen, M.H., Rasmussen, A., Svenmark, H., and Havskov Sørensen, J. *Danish Atmospheric Chemistry Forecasting System*, DMI, Technical Report 96-3

[2] Welch, G. and Bishop, G., "An Introduction to the Kalman Filter," University of North Carolina at Chapel Hill, Department of Computer Science, TR 95-041, 1995

[3] Handbook of Applicable Mathematics (Volume VI Statistics, part B), Chapter 20: "Kalman Filtering", Walter Ledermann Chief Ed.

[4] Grewal, M. S. and Andrews, A. P., "Kalman Filtering, Theory and Practice", Prentice Hall Information and System Sciences Series, 1993