V. Test of Kalman Filter Method in Schematic Case

V.1. Introduction

The objective of this application is to consider the one-dimensional (1D) linear advection equation to model the water level perturbed by a wave in a 1D circular domain composed of several cells. This application follows the example presented by Evensen (1992): starting from a state where all the cells verify a null water level, one aims to use a Kalman filter to make the system approach the analytic solution of the advection equation, using as measurements the values of the analytic solution of the advection equation for some cells. The implementation is made in MATLAB software.

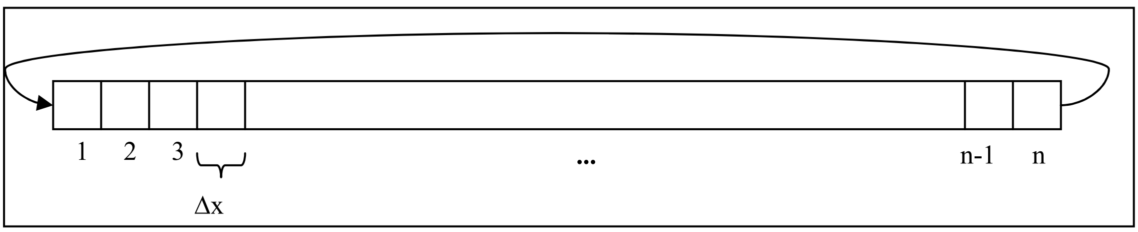
This application allows a first contact with the Kalman filter method in its simpler form, providing knowledge about the way it works and about practical problems that arise while implementing it in particular and in data assimilation in general. This first application is intended also as a way of showing the specific advantages of data assimilation.

V.2. System description

V.2.1. Grid configuration

The configuration of the grid considered for the system is as presented in the following picture.

Figure V-1: Grid configuration for schematic case.



The wave propagates from left to right and, as seen, the first cell is driven by the last cell (n), and thus the boundary condition is cyclic.

Evensen (1992) considers a total of n=30 cells.

V.2.2. Dynamic model

The dynamics considered for this application consists in the 1D linear advection equation:

$$\frac{\partial \eta}{\partial t} + c\frac{\partial \eta}{\partial x} = 0 \qquad \text{(V.1)}$$

This equation describes the advection of a wave of quantity η (in this application the water level), at constant velocity c and without distorting its shape or amplitude. t is the time dimension and x is the space dimension.

Using the upwind method to the numerical discretization of the equation one finds:

$$\eta_j^f\left(t_i\right) = \frac{\Delta t.c}{\Delta x}\eta_{j-1}^f\left(t_{i-1}\right) + \left(1 - \frac{\Delta t.c}{\Delta x}\right)\eta_j^f\left(t_{i-1}\right) \quad \text{(V.2)}$$

where Δt and Δx are the time and spatial intervals of discretization; j is the index of cell; f in superscript means forecast.

Evensen (1992) seems to assume the following values:

- Δt = 1 time step;
- Δx = 1 space step identified as a cell's length.

No information is given on the chosen value for c, except that it is positive (wave propagating from left to right). The ratio (Δt.c)/(Δx) is the Courant number (C_r). If C_r is greater than 1 then reportedly numerical instability occurs. Because of the values assumed for Δt and Δx this means that c cannot be greater than 1. But one also knows that if C_r is less than 1 then the wave is degraded through the domain, approaching progressively its average level, and will not allow the approximation of a periodic analytic solution. From this reasoning, a value of c of 1 is considered in the present application.

V.2.3. Measurements

The exact solution of the advection equation is:

$$\eta_j\left(t_i\right)=\overline{\eta}+a\sin\left(\left(\frac{t_i}{T}-\frac{x}{cT}\right)2\pi\right) \qquad \text{(V.3)}$$

Where $\overline{\eta}$ is the average water level, a is the amplitude of the wave, T is the period of the wave and x is the distance to the west (left) boundary of the domain. Considering that the first cell is at zero distance from the west boundary, then x is given by $x = j-1$ where j is the cell number.

About the wave parameters Evensen (1992) assumes (c is valued as previously indicated):

- $\overline{\eta}$ = 0;
- a = 1.

T is derived from the dimension of the domain considering that the wavelength (λ) is half the dimension of the domain. Assuming from the previous section that c equals 1 and the following equation, T can be found as 15s.

$$T=\frac{\lambda}{c} \qquad \text{(V.4)}$$

From Evensen (1992), the measurements values consist in the exact solution (without an error term) for two cells: 5 and 20. These measurements first become available after 5 time steps and afterwards from 5 to 5 time steps.

To achieve the phase presented in the results of Evensen (1992) the equation (V.3) is slightly changed to the following equation, which does not change the qualitative response of the system.

$$\eta_i\left(t_i\right)=\overline{\eta}+a\sin\left(\left(\frac{t_i+3}{T}-\frac{x}{cT}\right)2\pi\right) \qquad \text{(V.5)}$$

V.2.4. State vector

In this application the state vector has to be chosen taking into account not only the variables whose value one aims to know (the water levels in each cell) but also the way the dynamics and measurements are specified. This specification drives a state vector consisting on the levels in all the cells, because:

- the water level in each cell depends not only on the water level in the same cell in the previous time step but also on the one in the previous cell in the previous time step;
- measurements are taken only at some (two) cells and their assimilation should have consequences also in the other cells.

V.2.5. Covariance matrices

It is assumed that initially no information is known about the solution of the system. The initial conditions considered by Evensen (1992) are:

- initial water level: null in all cells (null amplitude of wave);
- initial covariance matrix of estimation, P_0:
    - variances = 1;
    - correlation of error = given by the function f(r) = $e^{-\left(\frac{r}{r_c}\right)^2}$, where r is the distance between cells and r_c is the e-folding scale of the correlation function (gaussian correlation model).

This correlation function ensures smoothing in all state vector of the information provided by measurements and also ensures that the spatial relationship between cells is accounted in the process of assimilation.

Then, P_0 is an nxn matrix with the configuration given in Figure V-2.

Figure V-2: Configuration of P_0 for schematic case.

$$P(t_0) = \begin{pmatrix} 1 & f(r) & \cdots & f(r) \\ f(r) & 1 & \cdots & f(r) \\ \vdots & \vdots & \ddots & \vdots \\ f(r) & f(r) & \cdots & 1 \end{pmatrix}$$

Evensen (1992) does not refer what r_c he uses in its paper.

No information is given by Evensen (1992) for the values of the covariance matrices of the model (Q) and the measurements (R) errors, except that the measurement error at cell 20 has a variance that is double of the one at cell 5 and the variance of model error seems to be considered not null. The correlation between errors is not provided but it is assumed in this report to be null since errors are considered to be independent.

Because two measurements are supposed then the R is as indicated in Figure V-3, assuming var_{me5} the variance of the measurement at x = 5.

Figure V-3: Configuration of R for schematic case.

$$R = \begin{pmatrix} \mathrm{var}_{me5} & 0 \\ 0 & 2.\mathrm{var}_{me5} \end{pmatrix}$$

Considering that the variance of the model error is the same in every cell ($\mathrm{var}_{\mathrm{mod}}$)[111], Q is, likewise R, an nxn diagonal matrix (Figure V-4).

Figure V-4: Configuration of Q for schematic case.

$$Q = \begin{pmatrix} \mathrm{var}_{\mathrm{mod}} & 0 & \cdots & 0 & 0 \\ 0 & \mathrm{var}_{\mathrm{mod}} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathrm{var}_{\mathrm{mod}} & 0 \\ 0 & 0 & \cdots & 0 & \mathrm{var}_{\mathrm{mod}} \end{pmatrix}$$

[111] This is a reasonable assumption due to the cyclic characteristic of the domain considered here and the configuration of the vector of state (all variables consist in the water level of a cell).

Both R and Q are considered constant in time because no changes are assumed in the measurement and dynamics processes.

V.2.6. Kalman filter equations

As no forcing term is considered the system evolution is according to the equation:

$$\eta^t(t_i) = M\eta^t(t_{i-1}) + \nu_i \qquad \text{(V.6)}$$

where t in superscript means the true value and ν is the nx1 vector with unbiased model system error[112]. *M* is the dynamic or model matrix of dimension nxn, defined according to the equation (V.2) and the cyclic boundary condition. η is the state vector of dimension nx1.

Considering:

- $a_1 = \dfrac{\Delta t.c}{\Delta x}$;
- $a_2 = 1 - \dfrac{\Delta t.c}{\Delta x}$.

The matrix *M* has a_2 on the principal diagonal and a_1 on the lower second diagonal, as presented in Figure V-5.

Figure V-5: Model matrix configuration for schematic case.

$$M = \begin{pmatrix} a_2 & 0 & \cdots & 0 & 0 & a_1 \\ a_1 & a_2 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \ddots & a_2 & 0 & 0 \\ 0 & 0 & \cdots & a_1 & a_2 & 0 \\ 0 & 0 & \cdots & 0 & a_1 & a_2 \end{pmatrix}$$

112 Error average is null.

Measurements relate with the true state of the system through the measurement equation:

$$y^o\left(t_i\right) = H\eta^t\left(t_i\right) + \omega_t \qquad \text{(V.7)}$$

where y^o is the measurement vector (dimension px1, with p the number of variables measured), *H* is the measurement operator (dimension pxn, one line for each variable) and ω is the px1 vector with the unbiased measurement error. In this application measurements are made for variables considered explicitly in the state vector (cells 5 and 20), so the content of *H* is 1 for these variables and 0 elsewhere (see Figure V-6). *H* is the same for all update times, since the measurement process is assumed regular.

Figure V-6: Measurements operator configuration (p=2) for schematic case.

$$H = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{array}{l} \text{line for meas.1} \\ \text{line for meas. 2} \end{array}$$

x=5 x=20

As in every Kalman filter application, one finds that equations can be grouped in two sets: the corrector equations and the predictor equations. With the conditions given previously these sets for this application are as described in the following table.

Table V-1: Kalman filter equations for the schematic case.

Corrector equations	Gain: $K_i = P_i^f H^T \left[H P_i^f H^T + R \right]^{-1}$ (V.8) Update of state vector: $\eta^a\left(t_i\right) = \eta^f\left(t_i\right) + K_i\left(y_i^o - H\eta^f\left(t_i\right)\right)$ (V.9) Update of estimate's covariance matrix: $P_i^a = P_i^f - K_i H P_i^f$ (V.10)
Predictor equations	Prediction of state vector: $\eta^f\left(t_i\right) = M\eta^a\left(t_{i-1}\right)$ (V.11) Prediction of estimate's covariance matrix: $P_i^f = M P_{i-1}^a M^T + Q$ (V.12)

The order in which the two sets of equations are presented in Table V-1 is the order in which they are used in this application. However, measurements are not available for every time step, so between the update instants only prediction is made. This seems to be the procedure of Evensen (1992). Another approach to this problem would be to consider a Δt equal to the time interval between updates.

It is important to recognize that predicting P without updating afterwards and considering a null (or very small comparatively to P) Q may cause in some cases a reduction of the content of P, thus reducing the state estimate error. In the present application because in the dynamic matrix every non zero element has value one this does not occur. While the prediction of the state vector between updates seems useful if possible, because the discretisation error is reduced, by the literature review made it is not conclusive whether the prediction of the P is harmless and correct[113]. It seems however reasonable to assume that if the state vector is predicted then P should be predicted also and consistently. This question should be explored in subsequent research.

As there is an equation for the propagation in time of the water level for each cell in the domain, there is also an equation for the propagation in time of the covariance of the state estimate error for each element of P. This equation was deduced for this work and is thought to be useful for the interpretation of the results of the sensibility analysis in the next sections. Considering the nomenclature provided before, but with j equal to line and k equal to column of P, the equation is as following for j=k and j≠k, respectively.

$$P_{i,j,j} = a_2{}^2 P_{i-1,j,j} + a_1{}^2 P_{i-1,j-1,j-1} + 2a_1 a_2 P_{i-1,j,j-1} + Q_{j,j} \qquad \text{(V.13)}^{114}$$

$$P_{i,j,k} = a_2{}^2 P_{i-1,j,k} + a_1{}^2 P_{i-1,j-1,k-1} + a_1 a_2 P_{i-1,j-1,k} + a_1 a_2 P_{i,j,k-1} \qquad \text{(V.14)}$$

Because of the cyclic boundary condition it should be considered that before line 1 is the last line of P and before column 1 is the last column of P[115].

For the conditions assumed for this application a_1 equals 0 and a_2 equals 1 so that the equations resume to:

[113] Mónica (2004) is of the opinion that the sequential prediction of P without its update with measurements is incorrect. He states that the prediction of the state vector and P after the update by Kalman filter should be done for the time of the next update. He suggests that if needed the prediction of the state vector and of P for times in the intermediate time interval should be done with a different model matrix, consistent with the new time interval. Brown and Hwang (1997) while discussing round off errors refer the possibility of propagating the P matrix several times between assimilations.

[114] In this case $P_{i,j-1,j}=P_{i,j,j-1}$, because as any covariance matrix P must be symmetric.

[115] This is analogous to the condition that the first cell in domain is driven by the last cell in the domain.

$$P_{i,j,k} = a_1^{\,2} P_{i-1,j-1,k-1} + Q_{j,j} \qquad \text{(V.15)}^{116}$$

$$P_{i,j,k} = a_1^{\,2} P_{i-1,j-1,k-1} \qquad \text{(V.16)}$$

The effect of the model (not considering Q) is equivalent to the elements of P to propagate in the respective diagonal (Figure V-7).

Figure V-7: Effect of the model on P (not accounting Q) in schematic case. All elements in each diagonal are equal.

$$P = \begin{pmatrix} 1 & f(r) & \cdots & f(r) \\ f(r) & 1 & \cdots & f(r) \\ \vdots & \vdots & \ddots & \vdots \\ f(r) & f(r) & \cdots & 1 \end{pmatrix}$$

[116] In this case $P_{i,j-1,j}=P_{i,j,j-1}$, because as any covariance matrix P must be symmetric.

V.3. Methodology of assessment of the method

Several analyses are made of the performance of the assimilation system as a way to increase the knowledge of the method and also to test its robustness.

These analyses can be grouped in the following sets:

1. Sensibility analysis of the influence of the following parameters: R (variance of the measurement at cell 5), r_c and Q, considering the measurements in each update exactly equal to the analytical solution;
2. Explicit inclusion of a random error term in the measurements;
3. Sensibility analysis of the specification of the initial P matrix;
4. Consideration of measurements and model errors that are a function of the water level;
5. Model matrix with C_r=0,5 and measurements obtained with C_r=1;
6. Sensibility analysis of the number of measurements in the domain and of the frequency of measurements in time;
7. Analysis of the response of the filter to a change in amplitude in measurements.

In these analyses the evolution of four aspects is studied, following the approach of Evensen (1992):

- the state vector value and its comparison with the analytical solution and the measurements;
- the Kalman gain for each cell, for a measure of the correction that is being applied to each cell; the gain value in each cell is obtained through the sum of the lines in matrix K;
- the squared error (SE) for each cell, that is, the values at the diagonal of P;
- the mean squared error (MSE), obtained by the ratio between the trace of P and the dimension of the domain.

The study of these four aspects allows the assessment of the performance of the filter in terms of achieving the correct state vector and of the quality of the analysis. When considering the quality of the analysis is important also to account the values of the following parameters:

- the time to stabilization;
- the value of the MSE at stabilization.

V.4. Results and discussion

V.4.1. Sensibility analysis of the influence of R, r_c and Q

As base of this analysis the system with the values of parameters presented in Table V-2 is considered. These values were chosen because are the values presented by Evensen (1992) or their results approach reasonably well the results presented by this author.

Table V-2: Values of system's parameters for the reference run.

Discretisation parameters				Wave parameters			Kalman filter parameters				
domain's dimension (no. cells)	Δt	Δx	c	a	average level	wavelength	no. meas.	P(j,j)	Q(j,j)	R(1,1)	r_c
30*	1	1	1	1*	0*	15[a]*	2*	1*	0	0.15[b]	4

Notes: [a] (total dimension)/2; [b] R(2,2) = 2R(1,1), from Evensen (1992); * values from Evensen (1992).

The results for this reference run at the time of 1, 4 and 11 updates with Kalman filter are presented in Figure V-8, Figure V-9 and Figure V-10. From these figures it can be seen that for the water level exists a bigger approximation of the solution to the analytic solution at cell 5; this is visible also by the higher gain in this cell and in the smaller SE in the first update. With the increase of updates one verifies that the solution approaches the analytic solution. Because of the correlation of errors specified in matrix P the change of the levels is visible not only in cells 5 and 20 (where assimilation occurs) but also in the other cells. The updates cause a decrease in the Kalman gain and in the error, both for each cell (SE) and for the overall process (MSE), that seems to stabilise after 3 updates at the value of 0.15. As Evensen (1992) states, the reduction of the gain is due to the fact that the P decreases with the updates (note that for this run no model error is assumed as Q is a null matrix as presented in Table V-2) and so the filter is giving more weight to the background relative to the measurements.

After 1 update SE presents minimums at cells 5 and 20, which is reasonable because it is in these cells that the assimilation occurs. It should be noted that the SE at cell 5 has approximately the same value as R(1,1). Because the measure at cell 5 has lower variance than the measure at cell 20 the value of the minimum at cell 5 is smaller (approximately half). These minimums propagate, through the dynamic, to the others cells so at the time of 4 updates several minimums appear. The interval between minimums is 5 cells because both Δx and C_r have value 1. Thus, the difference of SE between cells reduces with the updates. When stabilisation of MSE is achieved the SE is regular in the domain. This seems to be linked with the circular boundary condition that is assumed. This isotropy is not however verified in Evensen (1992)'s results: there is noted the occurrence of anisotropy with the error in cells downwind of the assimilation points being lower than in cells upwind. This discrepancy of the results may indicate that Evensen is considering a different boundary condition.

Figure V-8: Solution of the advection equation after 1 update with Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) for the reference run.

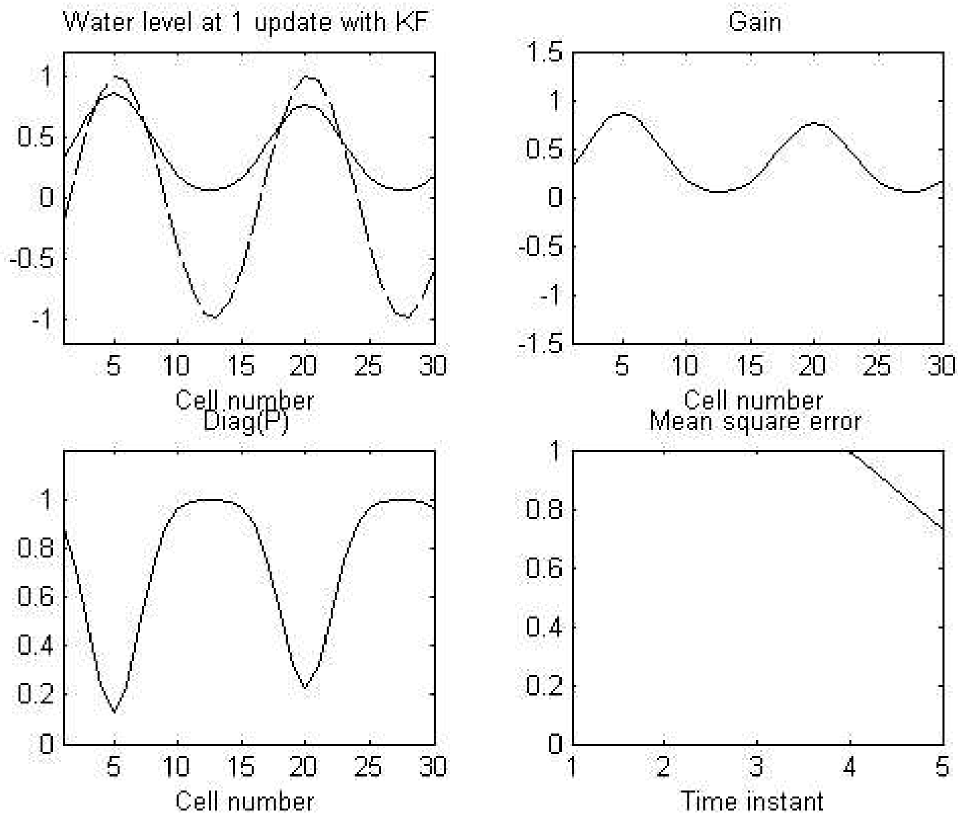


Figure V-9: Solution of the advection equation after 4 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) for the reference run.

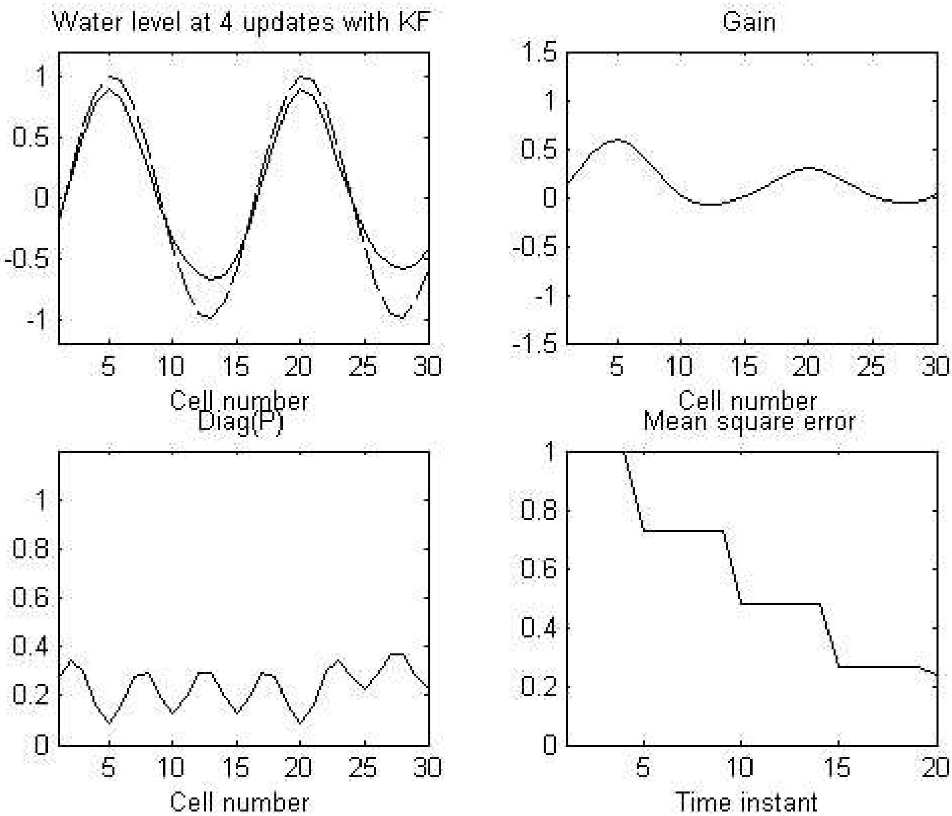
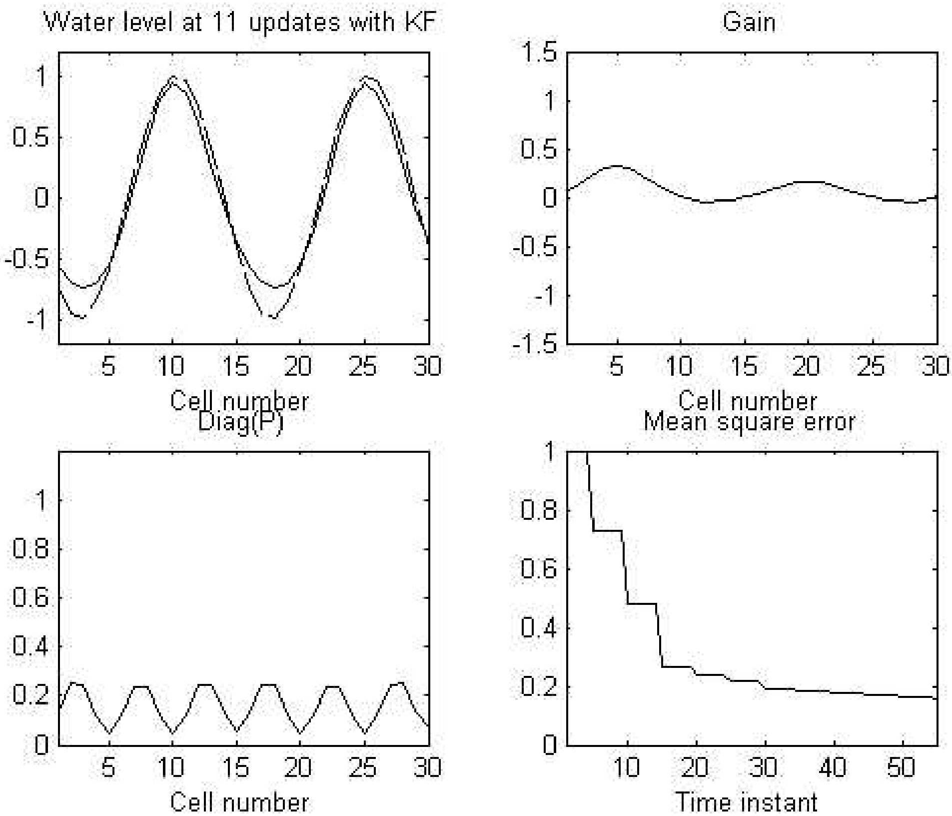


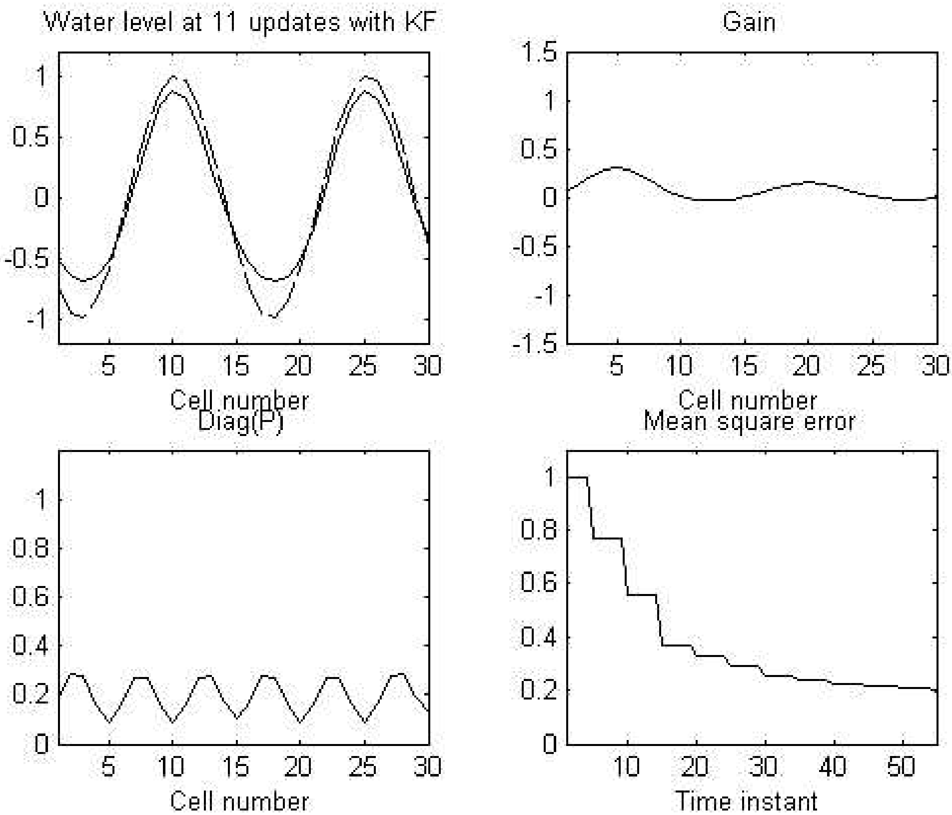
Figure V-10: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) for the reference run.



For the analysis of the influence of the R matrix two runs were made with different values for R(1,1): 0.3 (double of the reference value) and 0.015 (90% lower than the reference value). The relation between R(1,1) and R(2,2) and the value of all other parameters remains the same. The results after 11 updates are presented in Figure V-11.

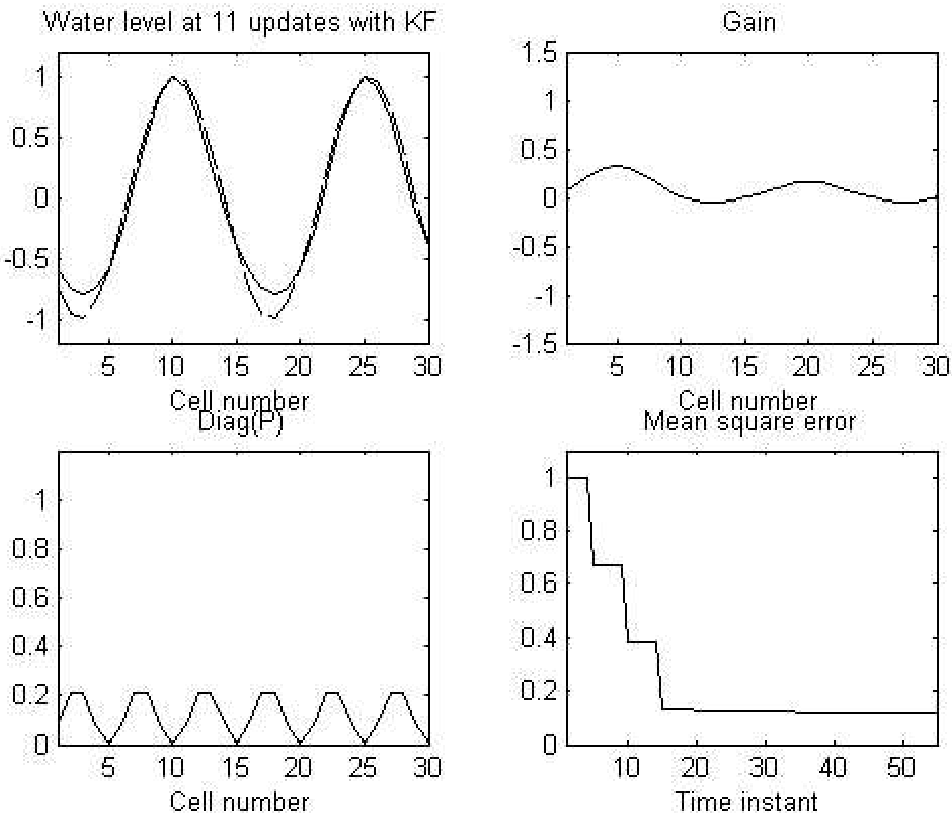
For R(1,1)=0.3 the results show, as expected (because of a smaller importance is given to the measurements), a smaller approximation to the analytic solution after 11 updates. The results for the gain and error show that the filter is not giving so much credit to the measurements as in the reference run: gain values are smaller in update 1 and error (SE and MSE) is larger. Gain behaviour after update 1 (achieving about the same values that occur in the reference run) can be explained with the smoothing of the gain due to error correlations in the first estimate of P and associated convergence of the filter. The qualitative evolution is similar to the one of the reference run but the value of stabilization of the MSE is now bigger: about 0.2. The increase in error is not proportional to the increase of R(1,1) (and R(2,2)): the average value of the SE seems to increase just about 25% after update 1 and just like the stabilization value of the MSE. But the increase of the SE in cells 5 and 20 seems to be proportional. The increase of R does not affect the stabilization time that remains about 3 assimilations.

Figure V-11: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with R(1,1)=0.3 (increase of 100%).



Decreasing R 90% (R(1,1)=0.015) one finds that the approximation to the analytic solution is larger than previously (Figure V-12), but the degree of approximation seems not proportional to the variation of R. Measurements are now more valuable when considering the background estimate and so the Kalman gain is larger, but as in the case of R(1,1)=0.3 the difference with the reference run is only visible in update 1, possibly because of the convergence of the filter. Because the difference between the variances of the first and second measurements (R(1,1) and R(2,2) is small in update 1 the gain for cells 5 and 20 is very similar and the same happens with the SE on the same cells. For updates 4 and 11 the gain shows clearly that the measure at cell 20 is given less value. The MSE is lower than previously with stabilization value of about 0.125 (17% lower, not proportional to the change of) which agrees with the fact that measurements are now regarded as more precise. Stabilization time does not change.

Figure V-12: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with R(1,1)=0.015 (decrease of 90%).

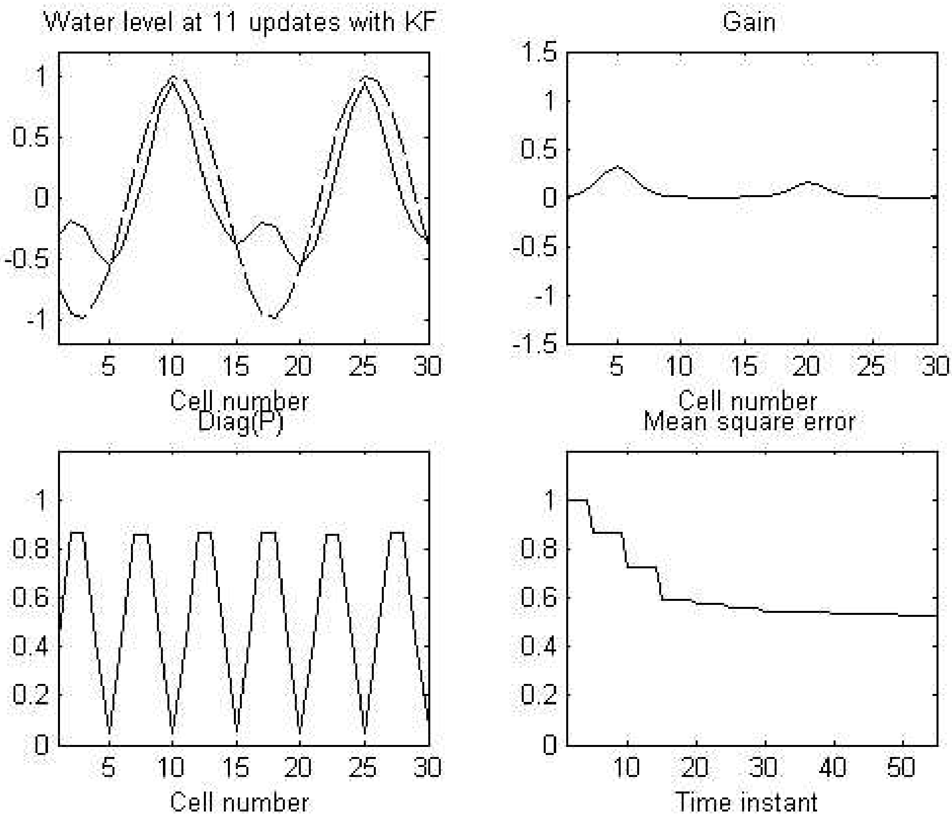


The results for the runs with the reference parameters and r_c equal to 2, 6 and 8 after 11 updates are presented in Figure V-13, Figure V-14 and Figure V-15, respectively.

For r_c = 2 after 1 update one finds a less smooth water level distribution than the ones found in previous runs: the information assimilated at cells 5 and 20 seems only to affect the 3 adjacent cells contrarily to what was observed before where all cells were affected by the first assimilation. This clearly shows that in this run there is a lower dispersion of the information assimilated (lower r_c). The solution after 11 updates shows less concordance with the analytic solution than the solution of the reference run. The point that strikes the most is that the solution does not present minimums agreeing with the minimums of the analytic solution. Instead one finds local maximums, a feature one can understand as the maximums in previous times that have not been dispersed in the other cells.

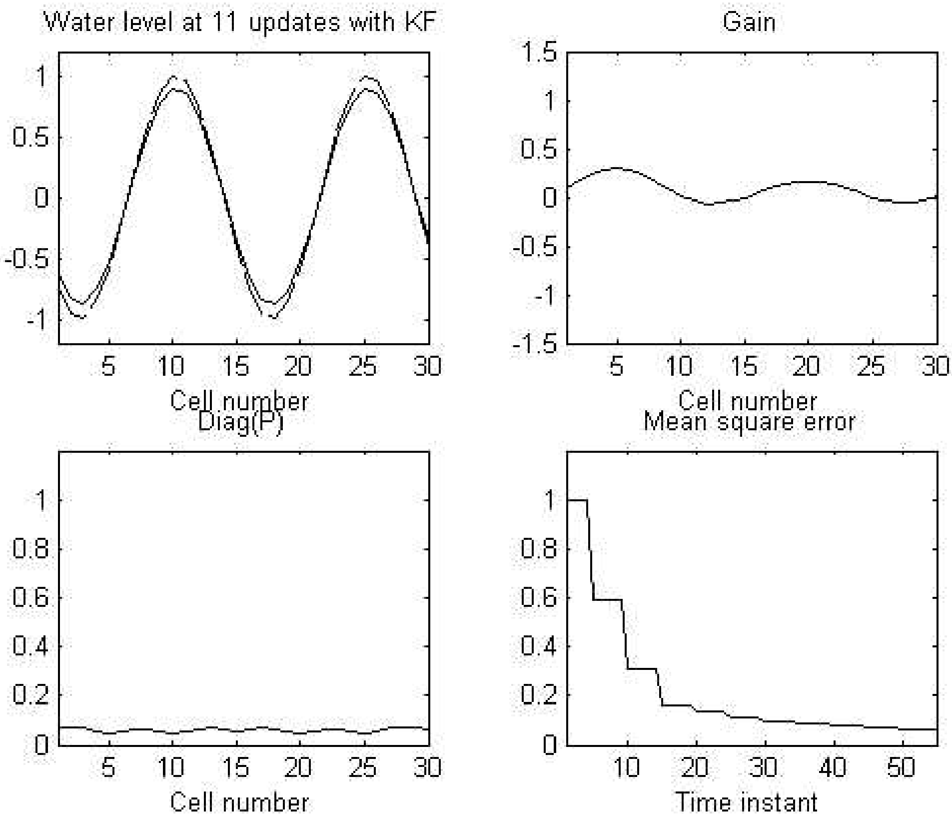
The Kalman gain shows less smooth peaks but the maximum values are equal to the ones of the reference run, showing that the filter is considering that observations have the same weight in both situations. Likewise, the SE verifies a lower diffusion around cells 5 and 20 at update 1 and a higher difference between maximum and minimum SE in updates 4 and 11, consequently increasing the MSE. Although stabilization is achieved about the same time as the reference run the value of stabilization is higher, about 0.53, which means an increase of 253%. Thus, the Kalman filter interprets the reduction of correlation of error in estimates as a decrease in precision of the solution. This behaviour can be understood if one considers the autoregressive characteristics of the dynamic matrix.

Figure V-13: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with r_c=2 (decrease of 50%).



The situation for r_c equal to 6 shows a greater approximation to the analytic solution after 11 updates when comparing with the reference run. It is also noted that the approximation at the maximum and minimum values is about the same contrarily to the reference run where the approximation to the maximum is larger than in minimum, possibly due to the effect of the first assimilation (measures are at maximums) that has not dispersed. The larger approximation is clearly a consequence of a higher diffusion through the domain of the information assimilated. This diffusion also appears translated in the Kalman gain and SE distribution: after 1 update no cell still presents the initial estimate's error of 1. However the minimum values of SE after update 1 are the same as the one in the reference run. The MSE of stabilization is consequently lower than in the reference run, about 0.07 (reduction of 54%). The time for stabilization is not altered.

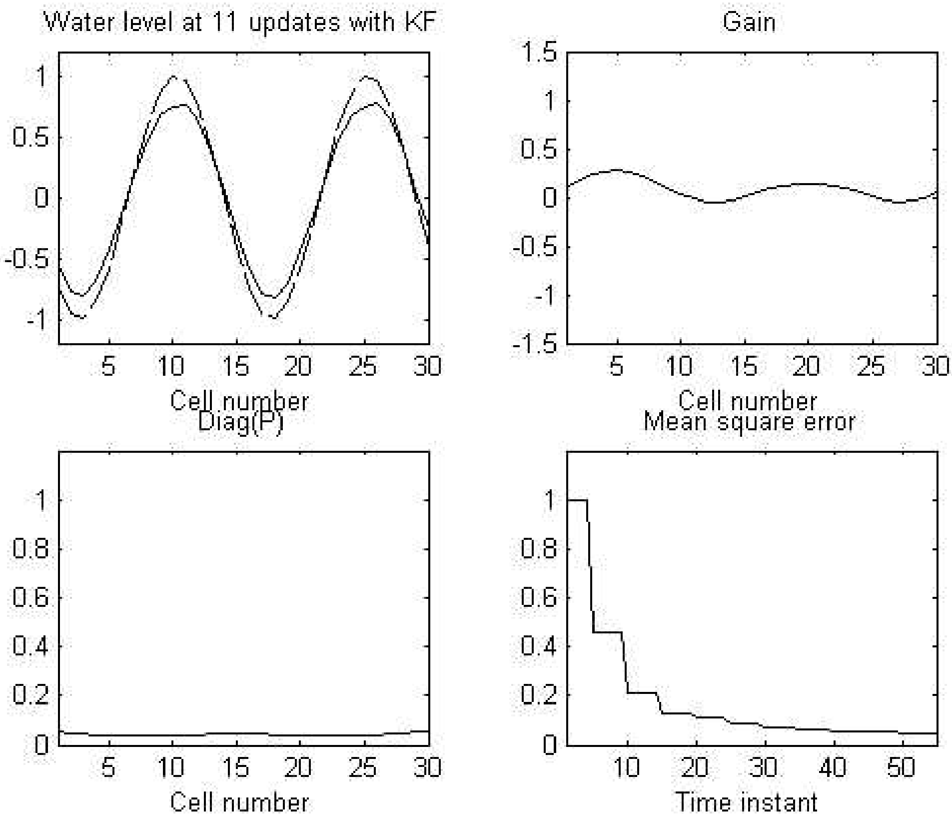
Figure V-14: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with r_c=6 (increase of 50%).



Contrarily to what could be expected the further increase of r_c to 8 does not result in a higher concordance between the solution of the Kalman filter after stabilization and the analytic solution. In this case the diffusion reduces the individuality of the level in each cell resulting in a solution with lower amplitude than the analytic solution. This diffusion is visible in the distribution of the Kalman gain trough the domain after 1 update and in the distribution of the SE. The maximums and minimum values visible in the SE's distribution after 4 and 11 updates are now practically inexistent. As before the variances of the estimate's error after 1 update are equal to the ones of the reference run. The MSE of stabilization is about 60% lower than the one of the reference run, but stabilization is achieved at the same time.

To analyse of the influence of the value of Q, three runs are made: one with value of Q equal to R(1,1), other with value of Q equal to the double of R(1,1) (0.3) and other with value of Q lower than R(1,1) – 0.00015. Q is considered a diagonal matrix with the variances equal in all cells. The results for update 11 are presented in Figure V-16, Figure V-17 and Figure V-18, respectively.

Figure V-15: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with r_c=8 (increase of 100%).

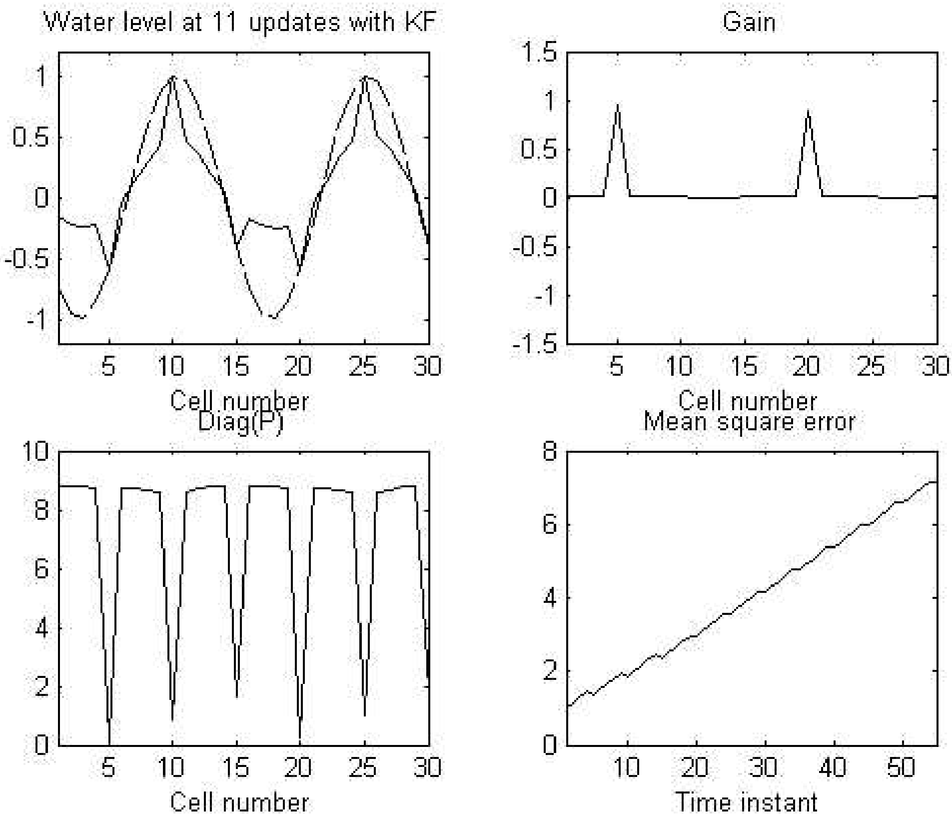


The results for var_{mod}=0.15 lead (at update 11) to a solution that differs not only quantitatively but also qualitatively from the analytic solution of the advection equation. In fact, at the cells where measures are available the water level shows discontinuities. Considering the Kalman filter equations, the inclusion of a not null Q in the system affects P (through equation (V.12)) and K (through equation (V.8)).

In terms of P, and because Q is a diagonal matrix, it is verified that a non null Q results in an increase of the difference between the diagonal and the non diagonal elements of P. This can be seen as equivalent to a reduction of covariance and thus the cells appear comparatively less correlated, an effect somewhat similar with the one caused by the reduction of r_c. Thus, the SE presents a very small smoothing around the minimums when compared with the reference run just as is observed with a low r_c (r_c=2). With the increase of the diagonal of P the value of K at the cells where measurement exist also increases and thus this explain why K exhibits a discontinuity at cells 5 and 20 and why the discontinuity increase with the updates. So, although diffusion of the assimilated information occurs it is as if only a part of the information is diffused and as if just a part of the assimilated information affects only the cells directly associated with measurements. When considering the reasoning behind the Kalman filter method, this is the result of the fact that the filter is disregarding dynamics that are less reliable than in the reference run and thus the shape of the wave (a feature of the dynamic) appears attenuated. Probably discontinuities appear in the cells with measurements because in these locations the effect of the dynamic is less important.

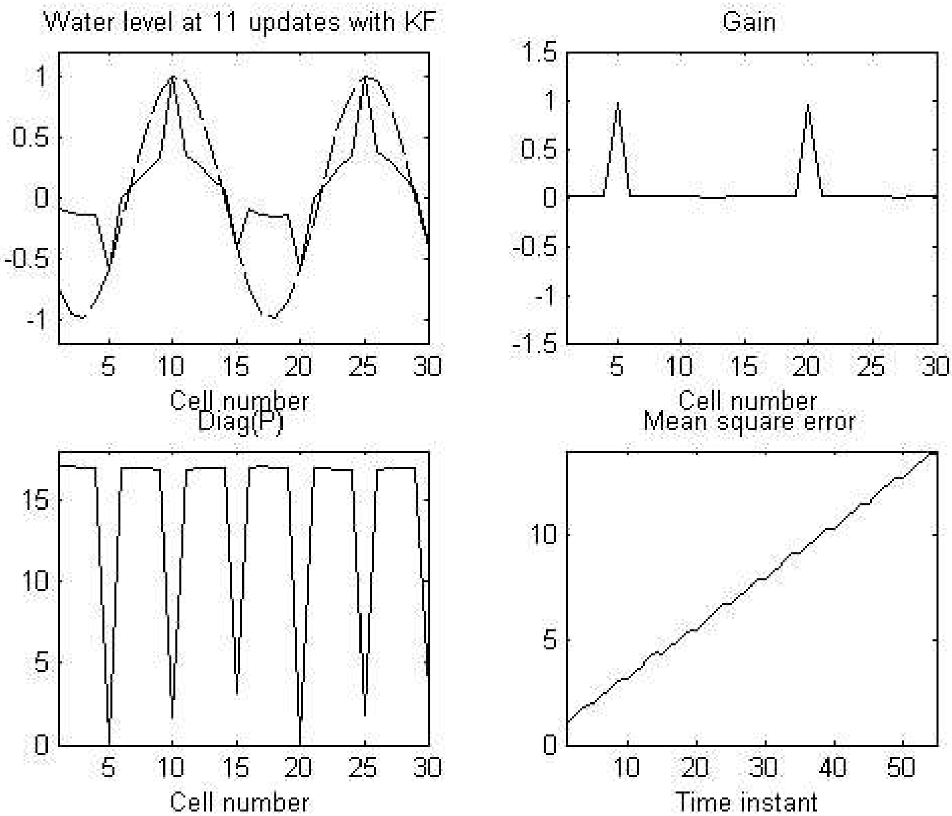
However, one of the most important findings of this run is the fact that MSE increases since the beginning and that stabilization is not obtained at the time of 11 updates. This makes one wonder if stabilization ever occurs. In fact, after running for 1000 time instants stabilization still does not occur. Considering the results again it is visible that the Kalman gain distribution is approximately the same after update 4 and update 11, suggesting that it is stabilizing. Since the measurements are periodic then the solution should also stabilize. The problem is that the Kalman gain is approximately null in the domain outside cell 5 and 20, meaning that the variance of the error of the analysis at these cells is not reduced by the filter and thus it is allowed to increase very much (also with the prediction) pushing up the MSE.

Figure V-16: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.15.



It is expected that with Q equal to 0.3 the situation is similar, besides an increase of the error of the estimate, what is in fact verified in Figure V-17.

Figure V-17: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.3.



Lowering Q could mean that the behaviour of the system approaches the one of the reference run (Q=0). In fact, with Q=0.00015 one finds basically the same results as the ones of the reference run, both quantitatively and qualitatively. However, what happens when K reaches the stabilization level? The diagonal of P increases further between the updates but at the time of the update, because C_r=1, the cells where measurements are available have associated a lower error variance because of the progression through the dynamic of estimate's error variances at cells where measurements where associated previously and also because of the circular boundary condition. The filter will consider then these small errors in the background estimate and will therefore not increase the Kalman gain as much as needed outside the measurement points. In this way, the filter will be «blind» to some elements of the diagonal of P and the MSE will be allowed to increase forever. So it is expected that even in this case of Q very small the MSE increases after reaching a minimum. This actually happens as can be seen in Figure V-19, with the system allowed to run for 5000 time instants. Evensen (1992) seems to assume a non null Q but it is unknown if in his conditions the instability of the MSE is also verified.

Taking into account this discussion, the MSE increase is though bound to happen has long as the time needed to propagate in the domain and reach the same position is a multiple of the update time, being update time not unitary and the boundary condition circular.

Figure V-18: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.00015.

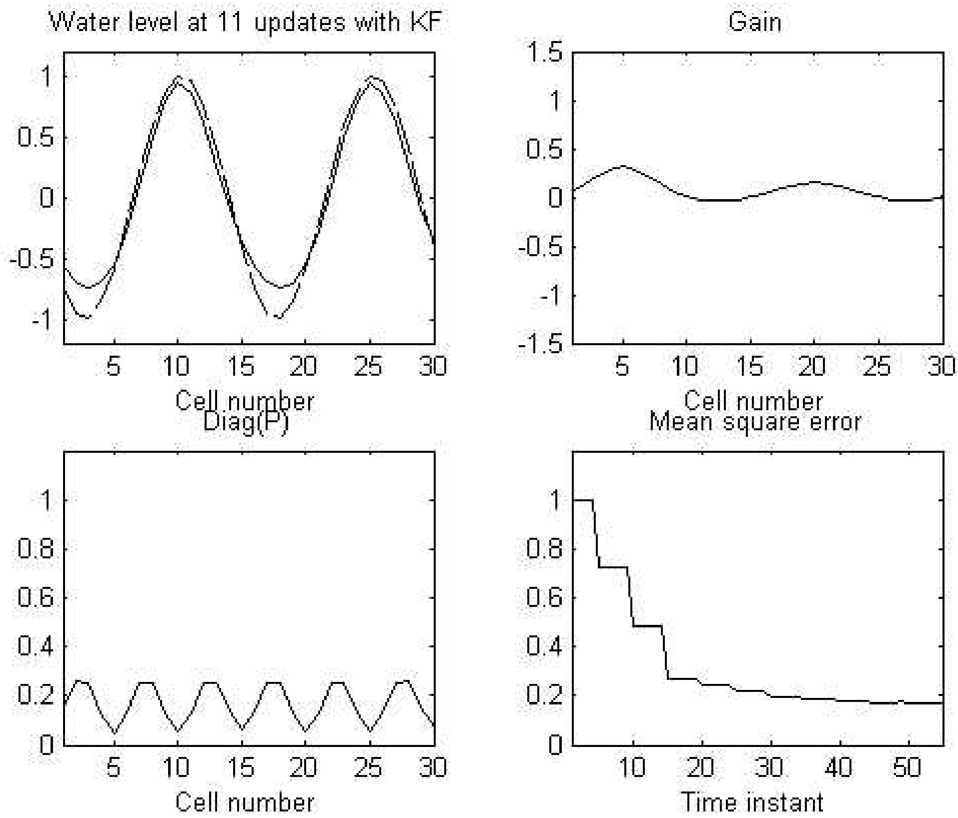
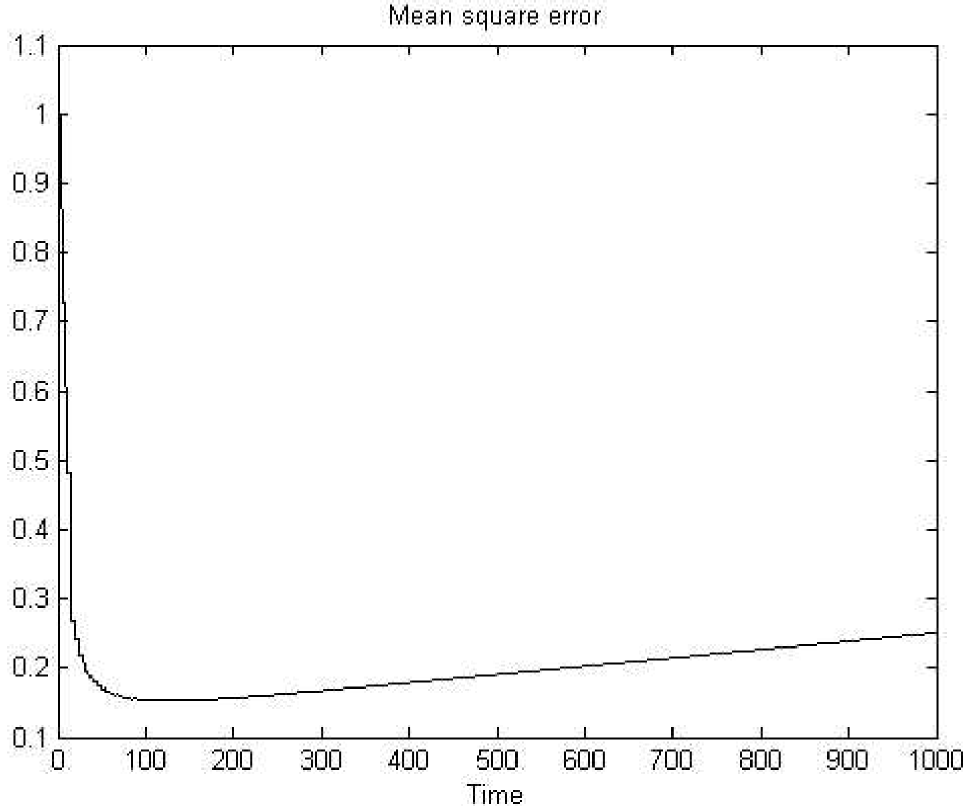


Figure V-19: Evolution of the MSE with var_{mod}=0.00015 in schematic case.



To assess this explanation of the unstable behaviour of the MSE some runs are made with the same conditions of the reference run, var_{mod} equal to 0.00015 and considering:

- run 1: domain length equal to 29 cells and update time equal to 5 time steps[117];

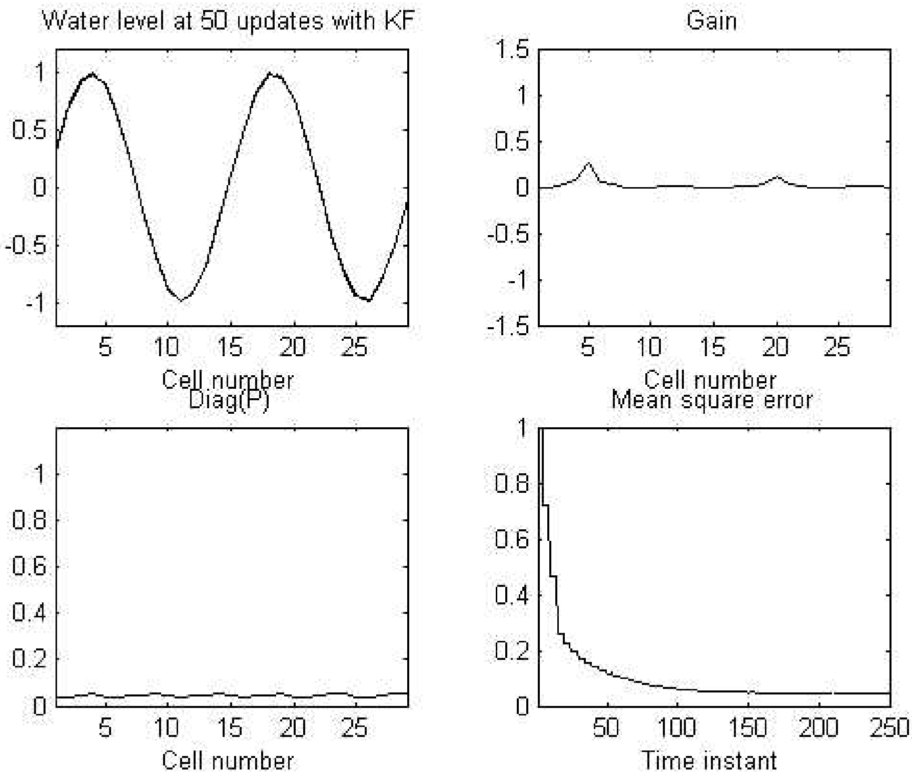
- run 2: domain length equal to 20 cells and update time equal to 5 time steps.

The results for these runs after 50 updates (about the time where the instability of the MSE could be detected previously) are presented in Figure V-20 and Figure V-21, respectively. These results show that stabilization of MSE is achieved for run 1, at a lower level than the minimum verified with domain length 30 cells. For run 2 MSE reduces to a minimum but begins to rise again. Thus, for this run one verifies the same problem found before and so the instability seems to be indeed linked with the relation between the time needed to propagate in the entire domain and the update time. The effect

[117] The number of 29 cells was chosen to approximate the domain length to the one considered in Evensen (1992) and because, contrarily to a number of cells of 31, it does not introduce *per se* improvements in the dynamic due to the discretization of the domain (more cells in the domain implies lower discretization error).

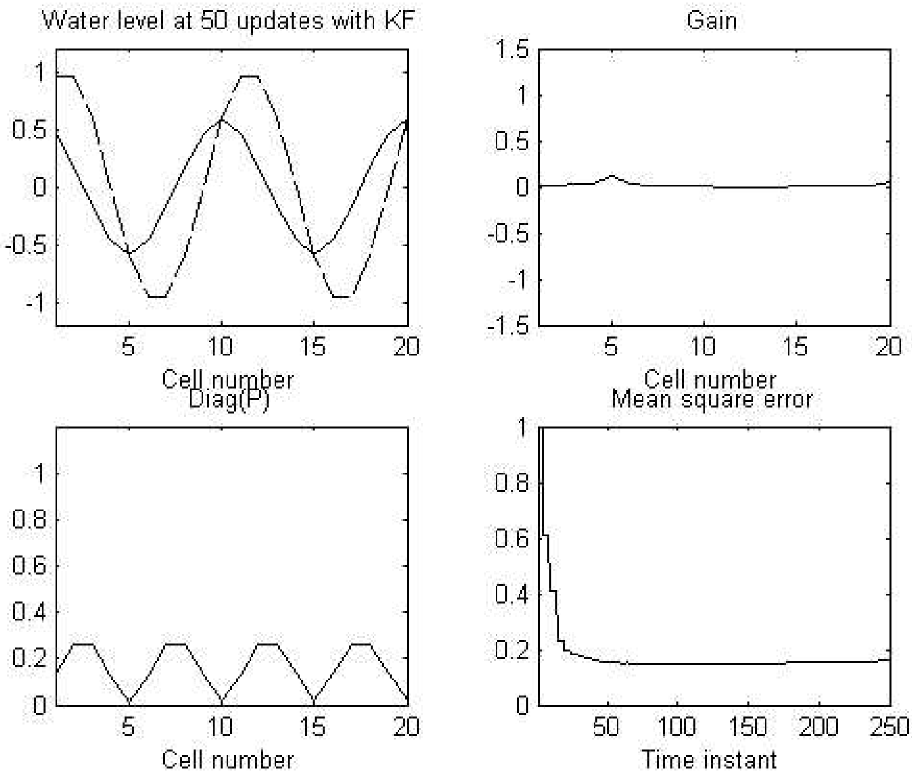
of this problem is like the existence of a resonance. In Figure V-22 it is tried to exemplify this phenomenon.

Figure V-20: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.00015 and domain length of 29 cells.



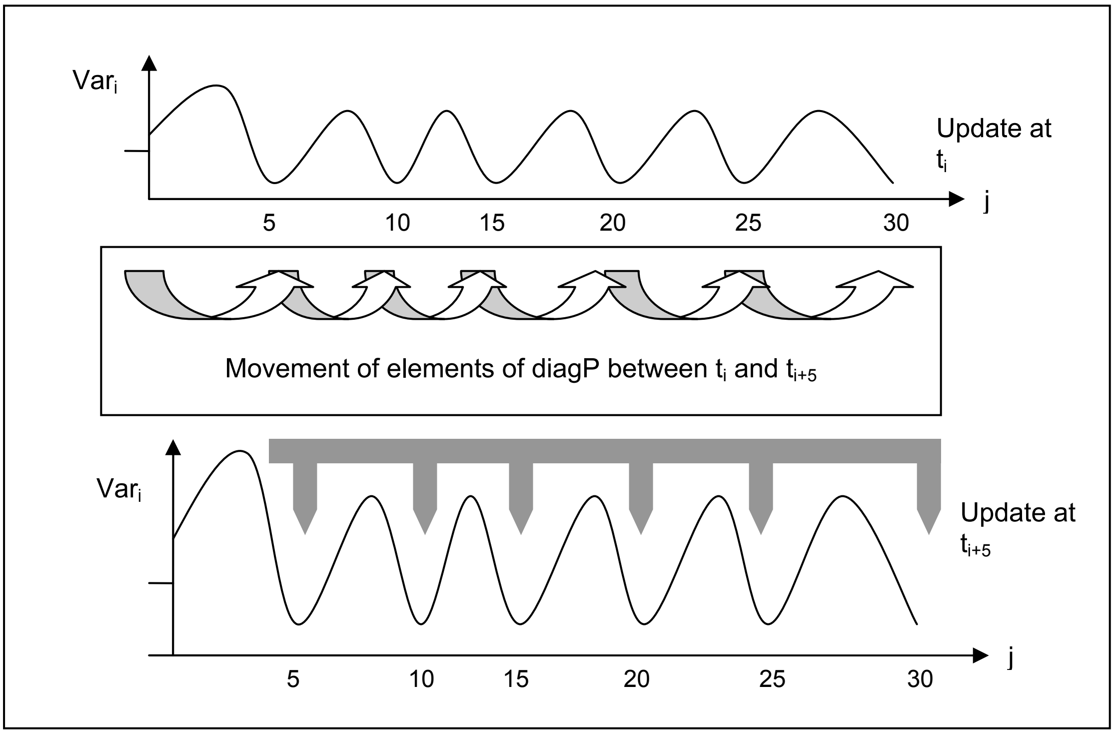
The results of the MSE running for 2000 time instants confirm these inferences. In run 2 it is also visible that the solution of water level that is reached after 50 updates is a poorer approximation of the exact solution than is found in previously presented results. This could be the result of considering a lower number of cells in the discretization of the domain.

Figure V-21: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.00015 and domain length of 20 cells.



Taking into account these results and to correct this problem of the cyclic condition, the reference run in the sensibility analysis from now on is the one with the reference parameters and a domain composed of 29 cells. This number was chosen to approximate the number of cells considered in Evensen (1992) while not verifying the problem of resonance. Q is as before 0.00015. The results for this **reference run 2** are presented in Figure V-23, Figure V-24 and Figure V-25, for 1, 4 and 11 updates. These results show an evolution of Kalman gain and error of the estimate similar to the one observed in reference run 1. The MSE increases slightly between updates due to the dynamic error that is being considered.

Figure V-22: Resonance caused by the circular boundary condition, domain length and update time interval.



In these new conditions, an analysis of the effect of the value of Q is performed again. Tenfold increases of var_{mod} where considered from 0.00015 (the reference run 2) till 0.15 (var_{mod}=R(1,1)). The results are presented in Figure V-26 and Figure V-27.

For var_{mod} = 0.0015 the results are very similar to the ones obtained for reference run 2 but one can observe a lower approximation to the exact solution and, accordingly, an increase of the error of the estimate, about 100% in average. The time for stabilization of the MSE is about the same but the stabilization value is higher (0.175 m). In the Kalman gain it is observed that its maximum values increase, because the filter now has to account for a less reliable dynamic (so the estimate obtained from prediction has more uncertainty associated) and that the distribution is less smooth.

With a tenfold increase of var_{mod} (0.015) the approximation to the exact solution is still smaller and one finds that the MSE decreases till the 4th update and then increases. The Kalman gain is now higher and with a lesser smooth distribution than before. In Figure V-28 are presented the results at the 50th update: it can be seen that MSE ends up stabilizing, but now at a slightly higher level on average (1.04) than the assumed first value of the error of the estimate. The time of stabilization is now about 200 time steps as opposed to the about 120 time steps observed with lower var_{mod}. Why this increase of MSE before stabilization? One explanation is that the transient phase before stabilization is the effect of the initial specification of P. In fact, if the initial estimate of P is too small it can cause the filter to «trust» too much in the estimate. The Kalman gain evolves accordingly with this estimate and it can happen that at a certain level where the gain should be stabilizing the increase of P results larger than the correction by the gain. In this case the gain must adjust till it reaches a level where it will counteract the increase of P due to the dynamics. With a less reliable dynamic it is reasonable to assume that the estimates of the state vector should also be less reliable so the choice of the initial value of P could not be considered innocent. It is advocated in the literature and by Mónica (2004) not to consider, if no information is available about the accuracy of the initial estimate, a very low value of P.

The results for the run considering var_{mod} equal to R(1,1) are presented in Figure V-29. In this run MSE increases in average (although decreases after each update relative to level in the previous time step) till stabilization is reached at about 9 m^2. This increase could be probably explained by an inadequate initial specification of P. Nevertheless the MSE evolution, the solution reached after 50 assimilations is very good compared with the known exact solution.

Figure V-23: Solution of the advection equation after 1 update with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.00015 and domain length of 29 cells – reference run 2.

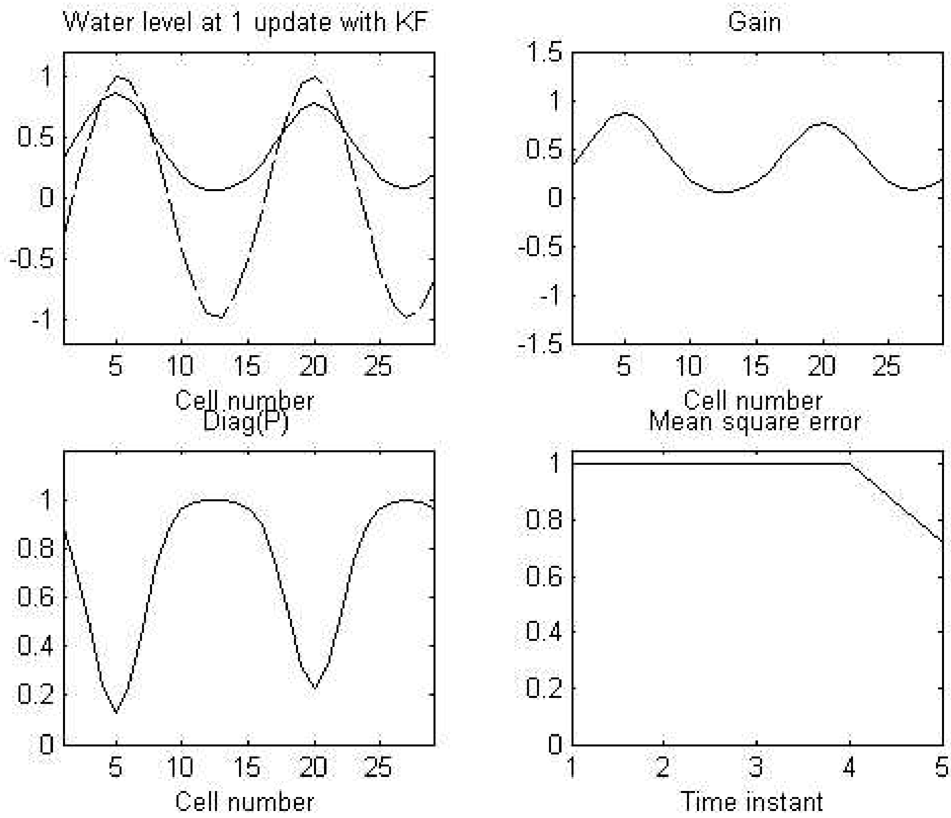


Figure V-24: Solution of the advection equation after 4 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.00015 and domain length of 29 cells – reference run 2.

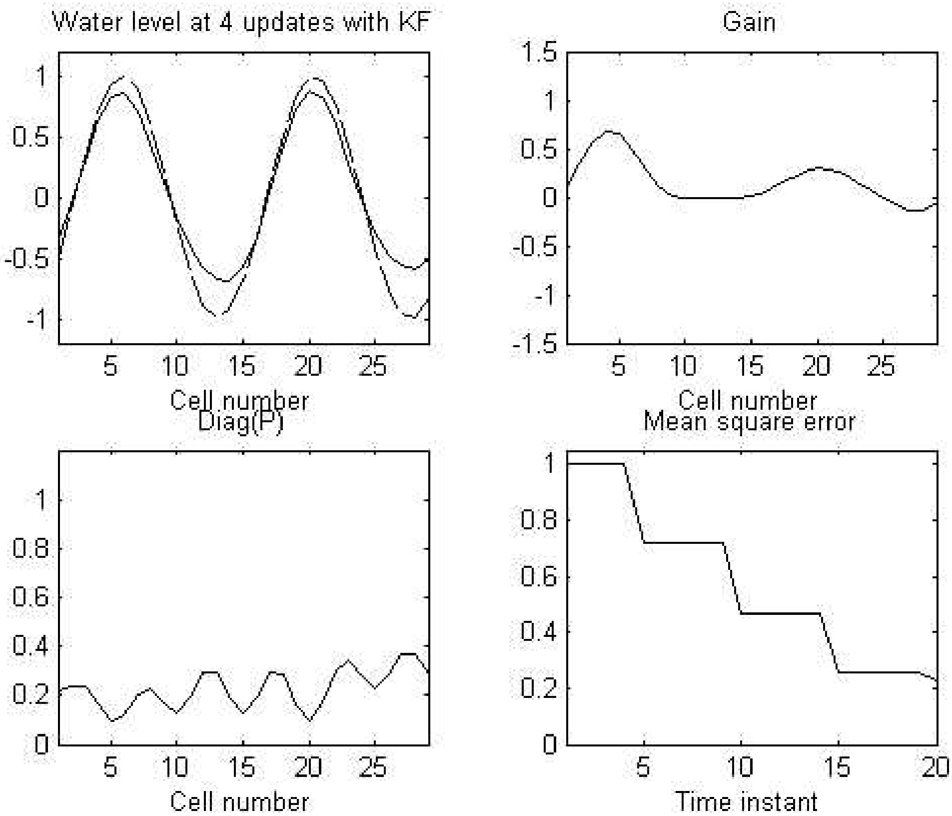


Figure V-25: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.00015 and domain length of 29 cells – reference run 2.

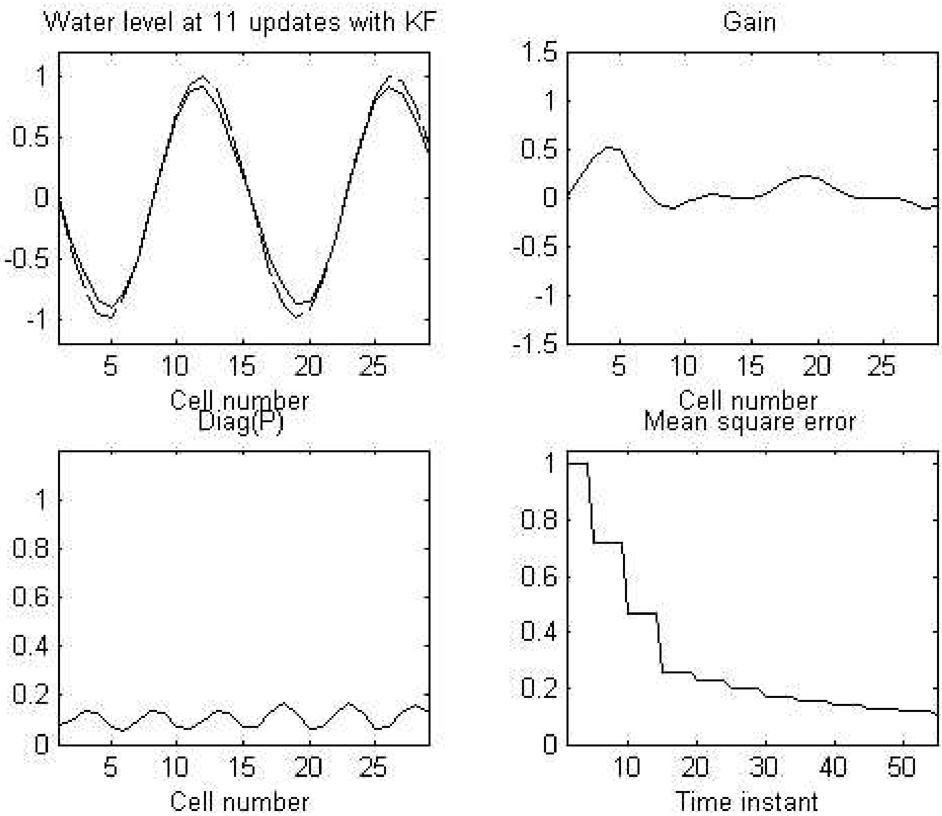


Figure V-26: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.0015 and domain length of 29 cells.

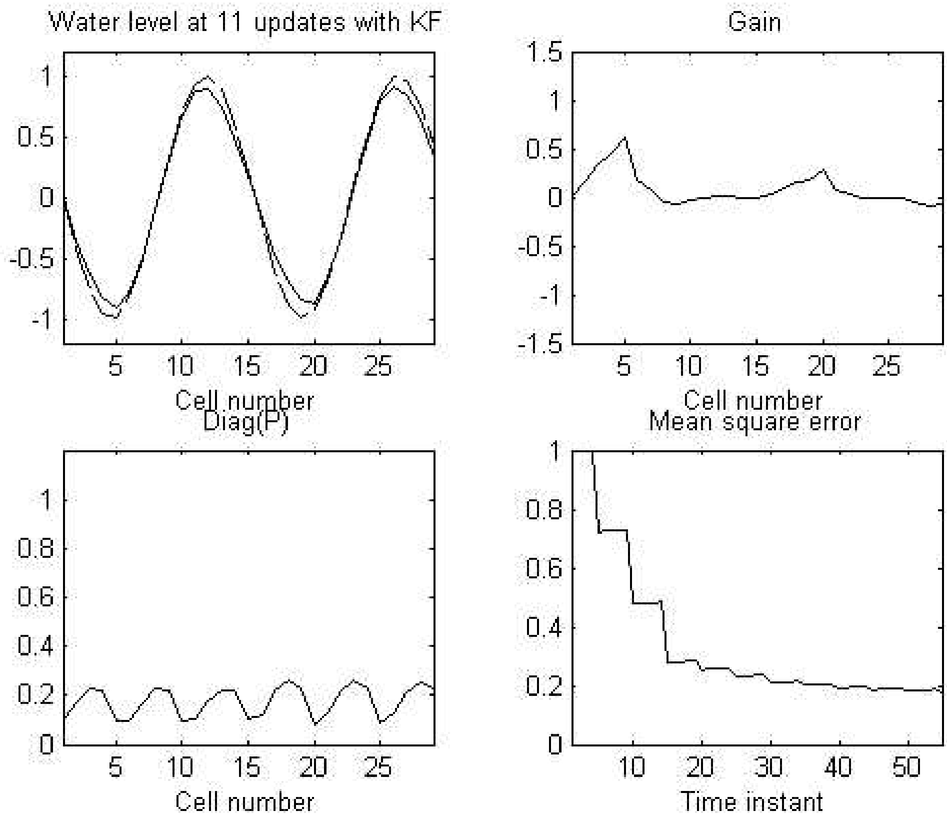


Figure V-27: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.015 and domain length of 29 cells.

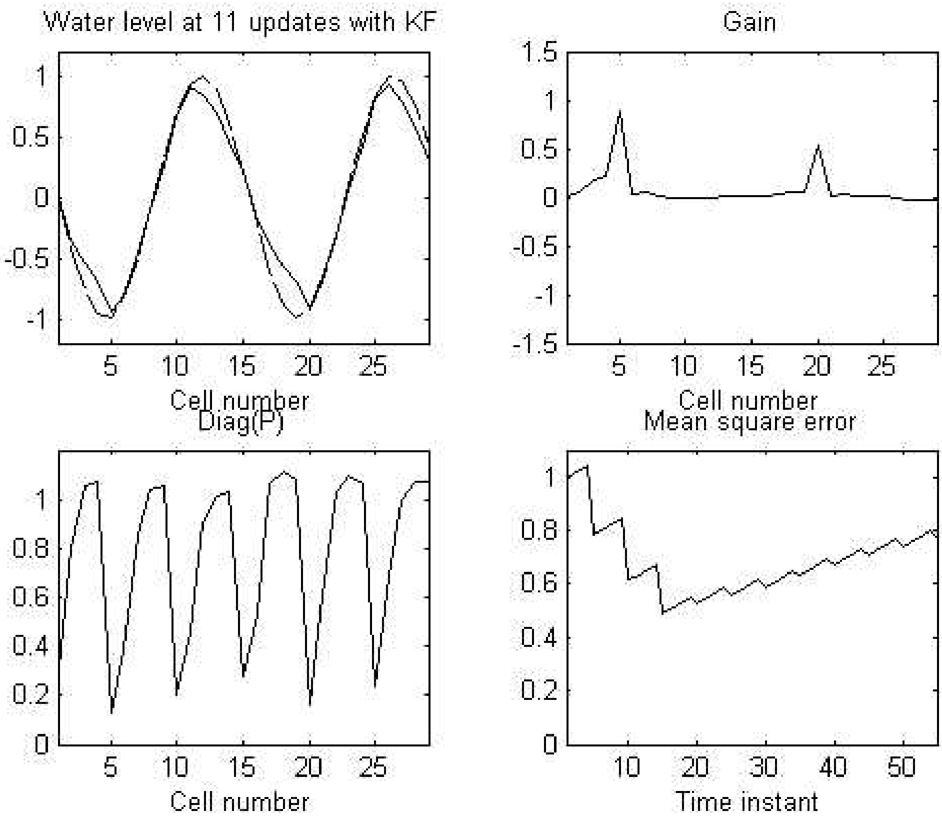


Figure V-28: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.015 and domain length of 29 cells.

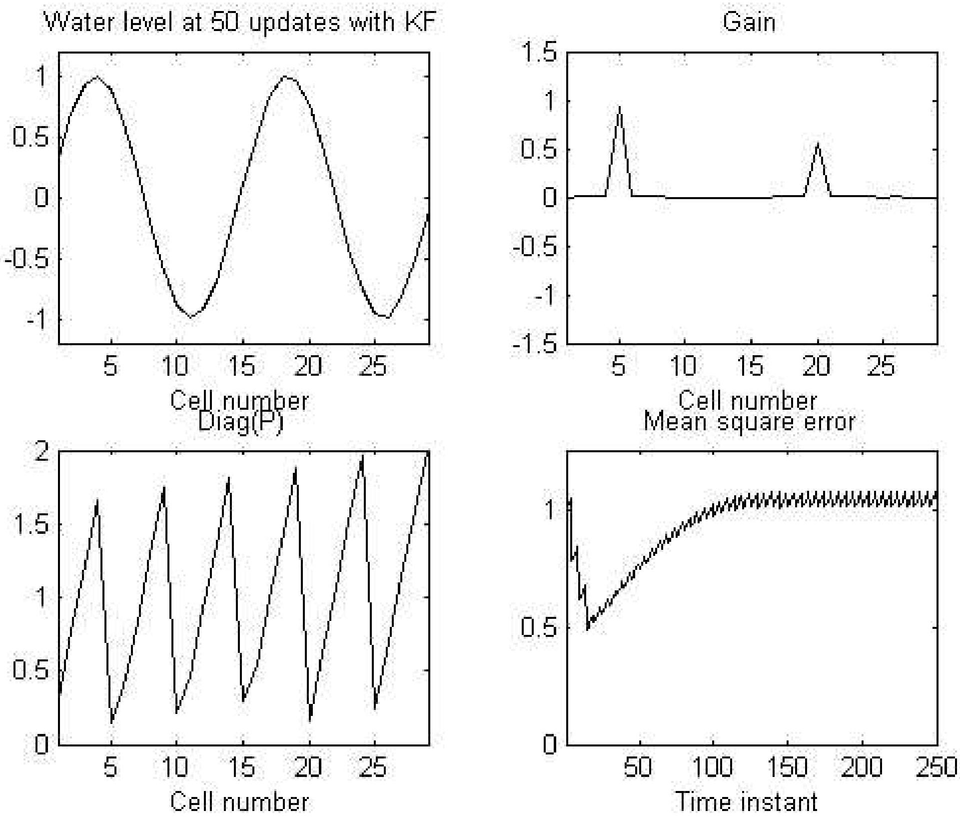
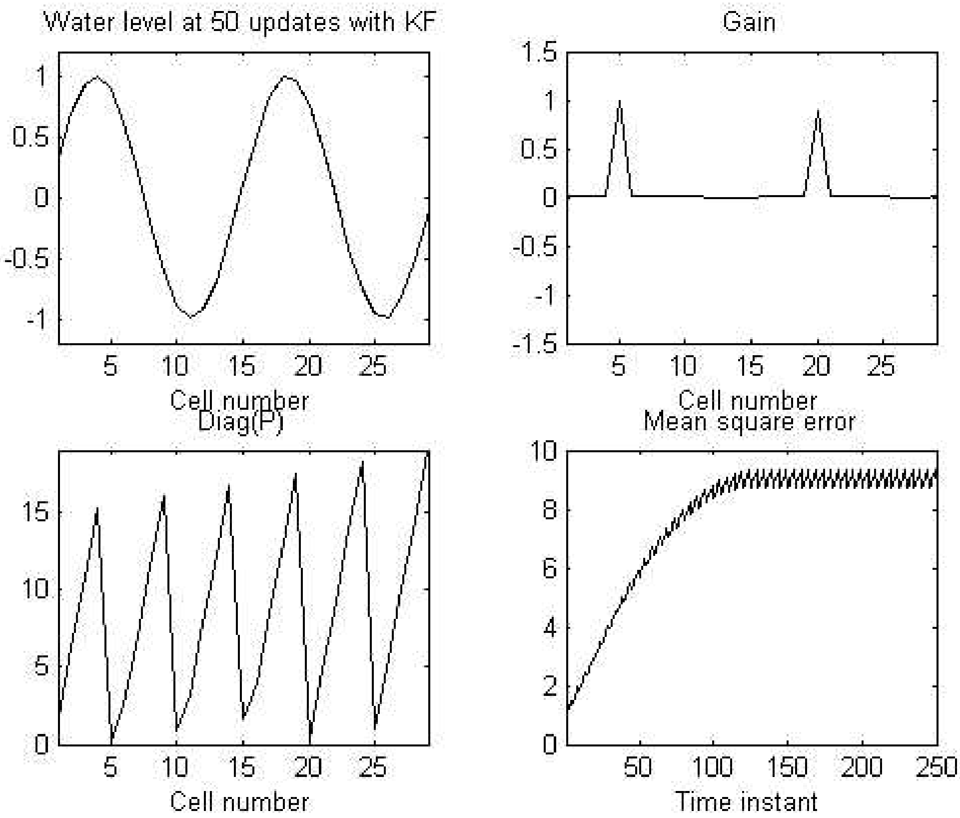


Figure V-29: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with var_{mod}=0.15 and domain length of 29 cells.



Based on the sensibility analysis presented in this section can be pointed the following conclusions relative to the effects of R, r_c and Q change:

- R influences the approximation of the state vector to the measurements, with larger values of variance implying a less weight given to measurements and a larger final background error estimate;

- r_c influences the dispersion of the assimilated information in the domain: dispersion (smoothness in Kalman gain distribution) is directly related to the value of r_c (large r_c causes a large dispersion); a r_c too large can reveal to be counterproductive because the system can be caused to disregard the dynamic provided distribution of levels; if dispersion is very low the MSE of stabilization increase;

- Q affects the approximation to measurements: a larger Q implies a larger approximation to measurements and a higher final level for MSE.

The Kalman filter reveals to be robust enough to result in similar values of the state vector given a variety of values attributed to parameters in R, r_c and Q.

V.4.2. Sensibility analysis of the influence of the initial specification of P

As seen in the sensibility analysis of the influence of Q's specification, the initial specification of P can be important, determining the occurrence or not of transient phases of filter adaptation before final stabilization. In this section the results of some runs where initial variances in P are different from those specified by Evensen (1992) are considered.

To specify different initial P it is assumed that the relation between the principal diagonal's elements of P (variances) and the non principal diagonal elements of P (covariances) is the same. This is a reasonable requirement due to the definition of variances and covariances:

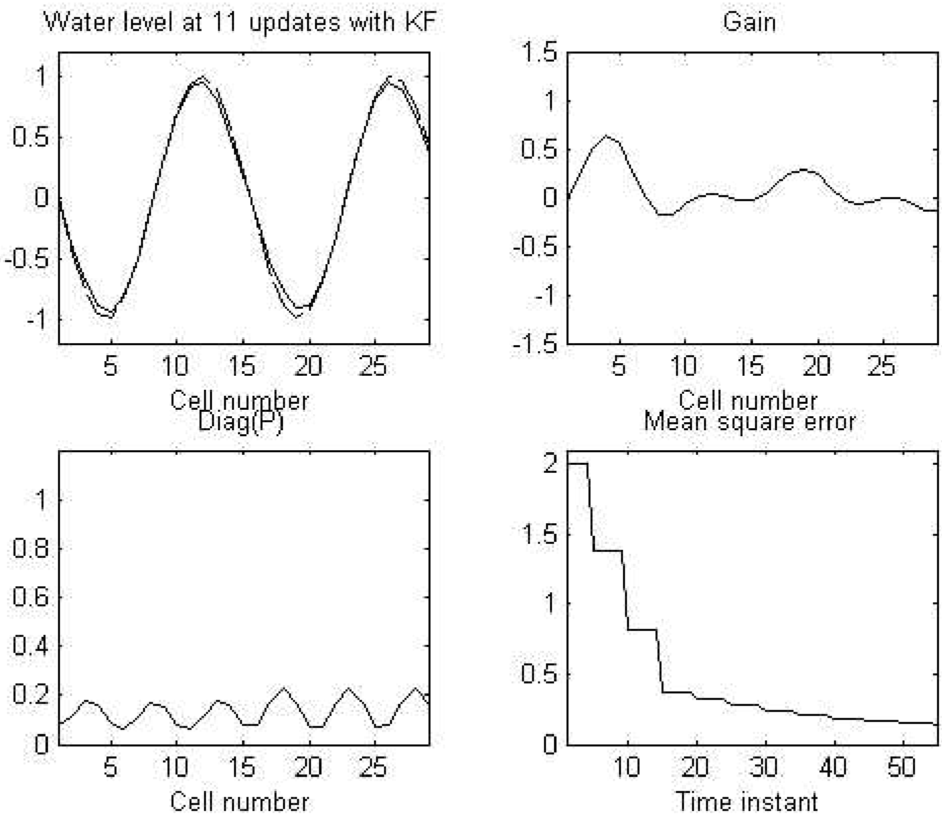
Variance of error of variable j = σ_j^2

Covariance between variables j and k errors = $\rho\sigma_j\sigma_k$

In this way, if an n-fold increase of $\sigma_j^2 = \sigma_k^2$ and an n-fold increase of $\rho\sigma_j\sigma_k$ happen then the correlation between the errors of these variables is maintained, allowing for the results to translate only an uncertainty of the initial variance of the errors. The measurements are considered equal to the exact solution to focus on the potential changes induced by the different initial P.

First is considered the case with double variance of the error of the initial estimate. In this case it is expected that the filter will give more weight to the measurements and so the exact solution will be approached sooner, although in the end the MSE of the estimate of the state vector should be the same as before. The results for a run with double variance of initial estimate are presented in Figure V-30 and Figure V-31 after 11 and 50 updates, respectively.

Figure V-30: Solution of the advection equation after 11 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with reference run 2's conditions and $P_{j,j}$=2.

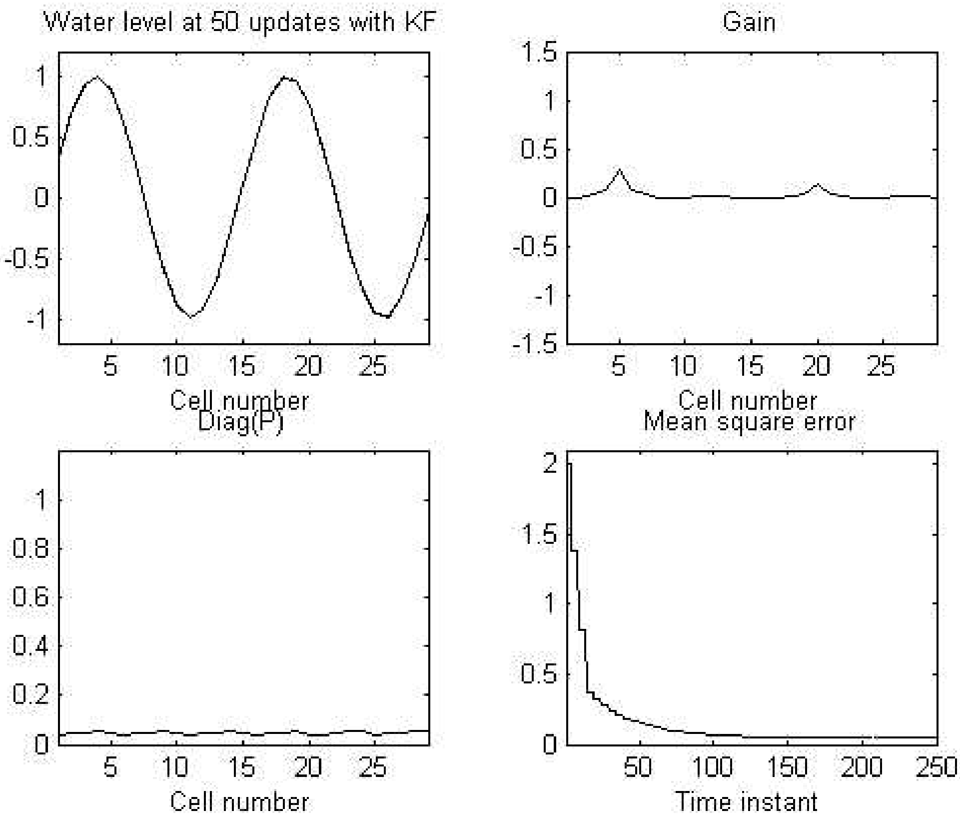


After 11 updates it is noticeable that the run with larger initial P obtains better results for the water level distribution (compare Figure V-30 with Figure V-25). Consistently the Kalman gain is larger with the larger initial P, because the correction of the background state is perceived as more important by the filter, and the predicted background error is larger too, still the effect of the initial specification of P. Thus, the expectations about this run are fulfilled.

The results of 50 updates, at a time instant when filter is well into stabilization, are remarkably similar with the results from the run with Evensen (1992)'s initial specification of P, in all aspects analysed: water level, Kalman gain and background error. This shows that the filter converge for the same solution regardless of the initial specification of P. This is a particular useful property of the Kalman filter method since in real applications it is not uncommon to be difficult to estimate the assumed initial state vector's error.

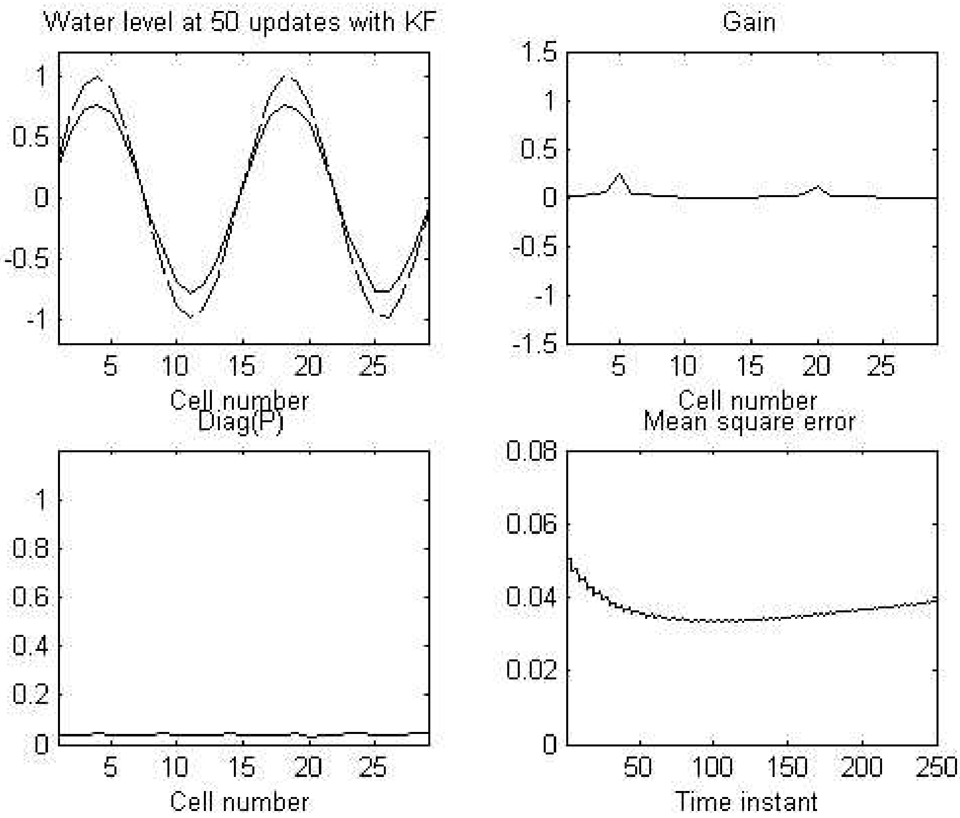
Following is presented the case with a very small variance of the error of the initial estimate of P: $P_{j,j}$=0.05. In this situation the initial estimate of the state vector is considered much more accurate and so it is expected that the convergence to the exact solution is much slower.

Figure V-31: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with reference run 2's conditions and Pj,j=2.



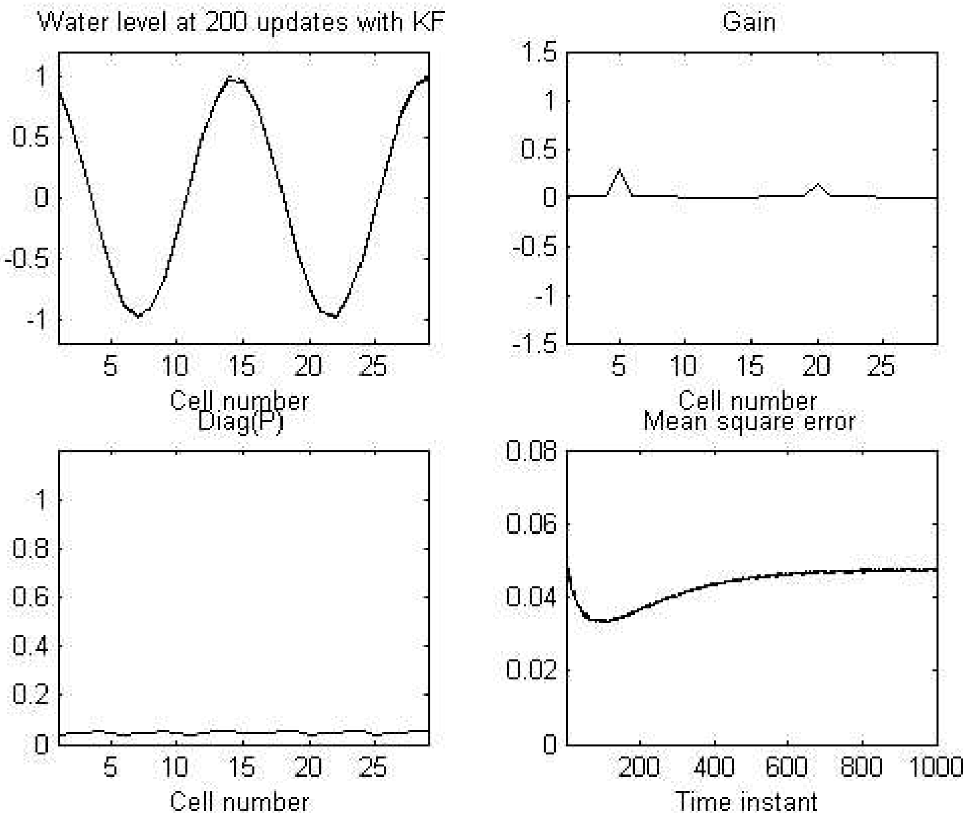
In Figure V-32 are presented the results of this run. In fact although the Kalman gain and SE distributions are very similar to the ones obtained with $P_{j,j}$ equal to 1 and 2, the estimate of the water level is much poorer than the ones obtained before. After 50 updates the MSE shows that the filter is still adapting from the initial value.

Figure V-32: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with reference run 2's conditions and $P_{j,j}$=0.05.



The filter manages to achieve the exact solution but only after more time. The results of this run after 200 updates are presented in Figure V-33. The water level, gain and SE are similar to the ones obtained after 50 updates for $P_{j,j}$ equal to 1 and 2. The final value of MSE is about the same too, but the evolution in MSE shows the transient phase that was earlier detected in the case of Q=0.015 (Figure V-28): if then it was the initial P that was too low for the dynamic's error considered, now it is the initial P that is too low for the initial state vector considered. Both are situations where the initial P is perceived by the filter as being wrong.

Figure V-33: Solution of the advection equation after 200 updates with the Kalman filter (in water level, filter solution in continuous line and analytic solution in dashed line) with reference run 2's conditions and $P_{j,j}$=0.05.



V.4.3. Model with C_r<1 and measurements with C_r=1

Besides the special case of water level dependent model error considered in the previous section, in the analyses performed so far it was usually assumed that the model/dynamic considered in this application is not exact and associated an error to it. This procedure is not totally correct because it is known that the model considered here provides the exact solution of the advection equation given the adequate initial condition and C_r=1. In this section one aims to assess the correcting performance of the Kalman filter method given a manifestly wrong model, a situation that is prone very likely to occur in real world applications of this method.

An easy way to specify a wrong model is to consider a C_r < 1[118]: given the correct initial condition the model will lead to the attenuation of the wave propagated through the domain till achieving the constant average level. The use of data assimilation in such a model is then fundamental to maintain meaning estimates of the state vector. The adequate measurements can be obtained from the exact solution with C_r=1. The error associated with this model can be easily calculated by comparing the

[118] A C_r higher than 1 will cause numeric problems as stated in section V.2.2.

results provided by this model with the ones provided by the exact solution with C_r=1, that will then be assumed as producing the true state vector. With the error variance estimated the filter could then be adequately specified.

In this application one considers a wrong model with C_r=0.5 and a true model with C_r=1. Taking into account what was assumed before, to obtain a C_r=0.5 one can change either the spatial or the time resolution of the discretization. c is a physical parameter that should not be altered because both models are describing the same physical phenomenon. Since it is important to obtain the water level for each cell, the spatial discretization cannot be changed. This leads to the realisation that the only way to achieve C_r=0.5 is to consider in the wrong model a Δt equal to 0.5.

To assess the error of this model it is considered a running time for both models equal to 5 times the assumed update interval (meaning 25 time instants). This running time was chosen by hint considering the time the filter takes to correct the solution in previous experiments. But the wrong model has a time step of half the time step of the true model and only at the true model time step can be obtained the error. The variance of the error thus obtained is an approximation of the needed variance of the model error: this would be the variance of the error when propagating one time step, thus a time step of 0.5. The error that is possible to be obtained in this application is the error of the model propagation for two consecutive time steps of 0.5. Although it is logical to think that the solution of the «true» model could be interpolated to achieve solutions at the time step of the wrong model and thus obtain the error of propagation through the model for one time step of 0.5, it is considered that this would involve interpolation errors that could be potentially larger than the error that is involved by assuming the variance of the model error obtained with a two time step interval as the correct one[119].

Taking into account this discussion the variance of the model error was calculated by obtaining the error at the true model time step (Δt equal to 1) and computing the variance of this error for each variable.

In a first run it was considered the measurements given by the exact solution of the advection equation and the conditions of reference run 2, except for the specification of Q.

The mean and variance of the model error for each variable are presented in Table V-3.

[119] Having the error after two consecutive time steps of 0.5 it is not possible to know the error after one time step of 0.5. Due to the dissipative nature of the dynamic of the wrong model it is reasonable to assume that the error will not remain constant in the time interval of 1.

Table V-3: Characteristics of the error associated with the model with C_r=0.5 comparing to the model associated with C_r=1, for schematic case.

Cell	Mean	Variance
1	-0.0735	0.1033
2	-0.0624	0.1115
3	-0.0398	0.1127
4	-0.0098	0.1062
5	0.0220	0.0964
6	0.0497	0.0904
7	0.0683	0.0923
8	0.0742	0.1008
9	0.0664	0.1100
10	0.0464	0.1132
11	0.0177	0.1084
12	-0.0142	0.0988
13	-0.0435	0.0912
14	-0.0647	0.0910
15	-0.0740	0.0984
16	-0.0696	0.1080
17	-0.0523	0.1132
18	-0.0254	0.1102
19	0.0062	0.1012
20	0.0367	0.0925
21	0.0604	0.0903
22	0.0729	0.0961
23	0.0720	0.1058
24	0.0577	0.1126
25	0.0328	0.1117
26	0.0018	0.1037
27	-0.0295	0.0943
28	-0.0554	0.0901
29	-0.0710	0.0940

Slight differences are verified between cells in the mean and variance, but it can be said that mean is approximately null and variance is about 0.1 for each cell. A null mean of the model error is concurrent with the requirement that model error to be white noise.

Having determined the accuracy of the model with C_r=0.5 these variances are used for the construction of Q. The results after 1 and 50 updates are presented in Figure V-34 and Figure V-35. They show that stabilization of the filter is achieved after about 3 updates but the solution that it converges into has lower wave amplitude than the exact solution. It is possible to see that the wrong model not submitted to assimilation converge to the mean water level.

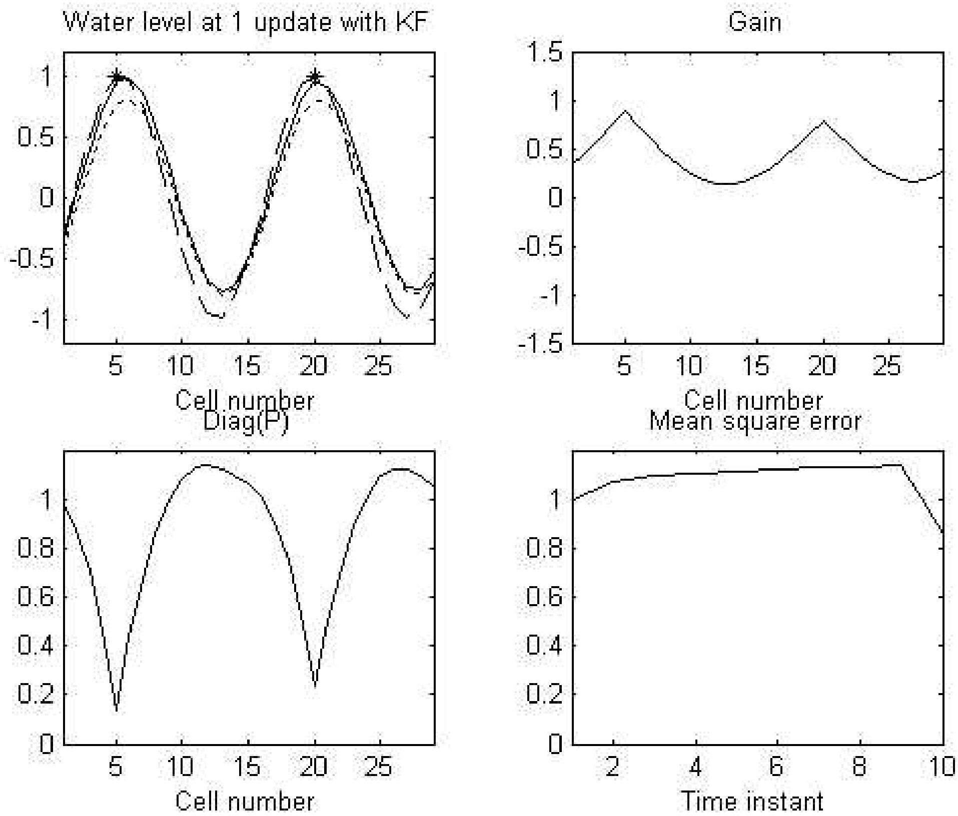
Important features of this run are the evolution of the Kalman gain and evolution of the squared error. About the Kalman gain it is verified little change from the first update to the 50th update: contrarily to previous results the gain does not approach a near zero value in the whole domain after this time instant. This clearly shows that the solution of the model is not stable, because otherwise it would maintain the effect of correction from the assimilation and gain would decrease. It is the existence of an attenuating effect that requires an almost constant gain. The distribution of the gain shows also a concave distribution as opposed to a traditionally convex distribution.

When it comes to the SE one finds that, although it verifies a similar spatial distribution after one update, after 50 updates its distribution is much different from the ones found in previous experiments: not only the mean error is higher than most but one does not find a set of SE minimums scattered through the domain. Pronounced minimums are found at the cells where measurements are located but other minimums are not very visible: it seems to be the effect of the attenuation of the dynamics that fails to propagate completely the error through the domain.

An important question is why the solution degrades after the first assimilation. Between assimilations the solution degrades because of the dissipative dynamic; when the next assimilation time is reached the solution needs a stronger correction but because of the reduction of P after the last assimilation the Kalman gain cannot increase to satisfy this need for a stronger correction, instead it decreases slightly. So the solution after several assimilations is less approximated to the exact solution than after the first one. Then why does stabilization occurs, that is, why the water level does not attenuates till reaching mean level? P decreases after each assimilation, till at a certain time the level of P is so low that its decrease is compensated by the increase by constant Q. K starts to be constant. Then the solution is corrected equally in each assimilation time from this point on and in this way the solution achieves stabilization: it degrades always in the same way after each one of the assimilations and then it is corrected always by the same amount.

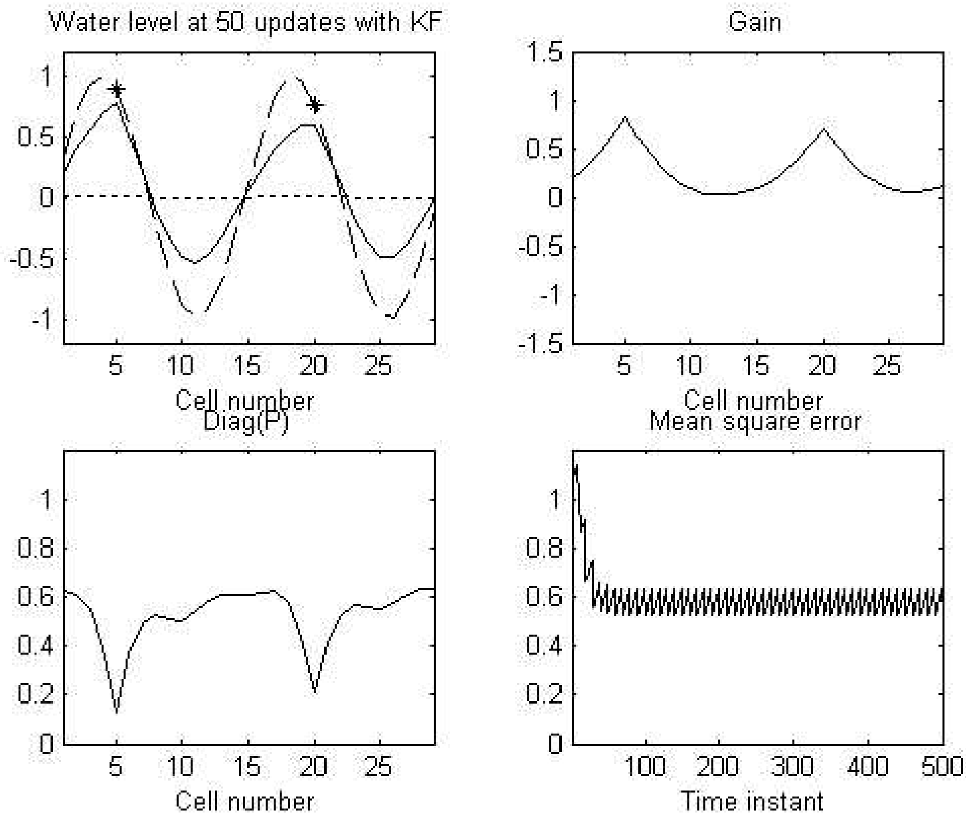
After stabilization the post-assimilation water level distribution is not rigorously the same. This happens again because of the attenuation in the dynamic. At each one of the assimilation times measurements are available at different cells; now, because the diffusion of the information assimilated through the domain is small the solution is very sensible to assimilation in these cells, producing some discrepancies in the solution between assimilation times. This can be also used to explain the «bias» that is observed with the average of the solution's water level not being equal to the analytic solution's average water level (assimilated levels are above or beneath the mean water level at each particular assimilation time).

Figure V-34: Solution of the advection equation after 1 update with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with C_r=0.5, Q specified comparing with model with C_r=1, other conditions of reference run 2 and measurements obtained from model with C_r=1.



To find if the variance attributed to measurement error is important in the stabilized water level configuration a run was performed with an almost null variance of measurement error (thus measurements are taken to be approximately exact): R(1,1)=0.00015. The results after 50 updates are presented in Figure V-36. They show that improvement in water level is small, although the SE and stabilization ME are somewhat lower. SE distribution is remarkably similar. This seems to indicate that what is limiting the convergence to the exact solution consists mostly in the anomalous dynamic and the attenuation induced by it.

Figure V-35: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with C_r=0.5, Q specified comparing with model with C_r=1, other conditions of reference run 2 and measurements obtained from model with C_r=1.



To study the effect of the model, where performed two runs considering the C_r of the wrong model respectively 0.25 and 0.1[120]. In

Figure V-37 and Figure V-38 are presented the results after 50 assimilations, when the system is well stabilized. The model characteristics are presented in Table V-4.

On water level one finds that the results is amazingly similar to the one found with C_r=0.5, although it is shown that the reduction of C_r implies a larger approximation to the analytic solution. This happens because the variance of the associated error increases with decreasing C_r: if the model is less reliable from the view point of the filter then measurements have larger weight in the statistical interpolation. Accordingly, Kalman gain is larger with lower C_r. SE distribution is similar, although the maximum is higher, because model error is higher. At stabilization MSE is higher in lower C_r models and the range of variation of the MSE is also higher. As before the models show high sensitivity to measurements at the associated cells.

[120] When experimenting with C_r in the context of the wrong model one has to consider that both the wrong and the true model have to have common time instants for the wrong model error to be calculated.

The configuration of water level at stabilization is almost independent of C_r: it is thought that this occurs because the change in C_r, and thus the change in the attenuation effect, is counterbalanced by the change in the variance of the model error, and consequently the model is regarded as less accurate. In this way, if a model has smaller C_r then it will degrade more in the interval between assimilations but when assimilation time is reached it is also corrected more heavily. It seems that the change in attenuation obtained with the range of change in C_r used here is not very important because the lower variance of model error factor is showing on results (higher approximation to the exact solution with lower C_r).

In the next run it is considered explicitly a random error in measurements with variance equal to the one specified in reference run 2. The results after 50 updates are presented in Figure V-39. The water level distribution shows the great sensibility of the model to measurements but in cells where no measurements are available the difference is small.

Figure V-36: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with C_r=0.5, Q specified comparing with model with C_r=1, other conditions of reference run 2, measurements obtained from model with C_r=1 and R(1,1)=0.00015.

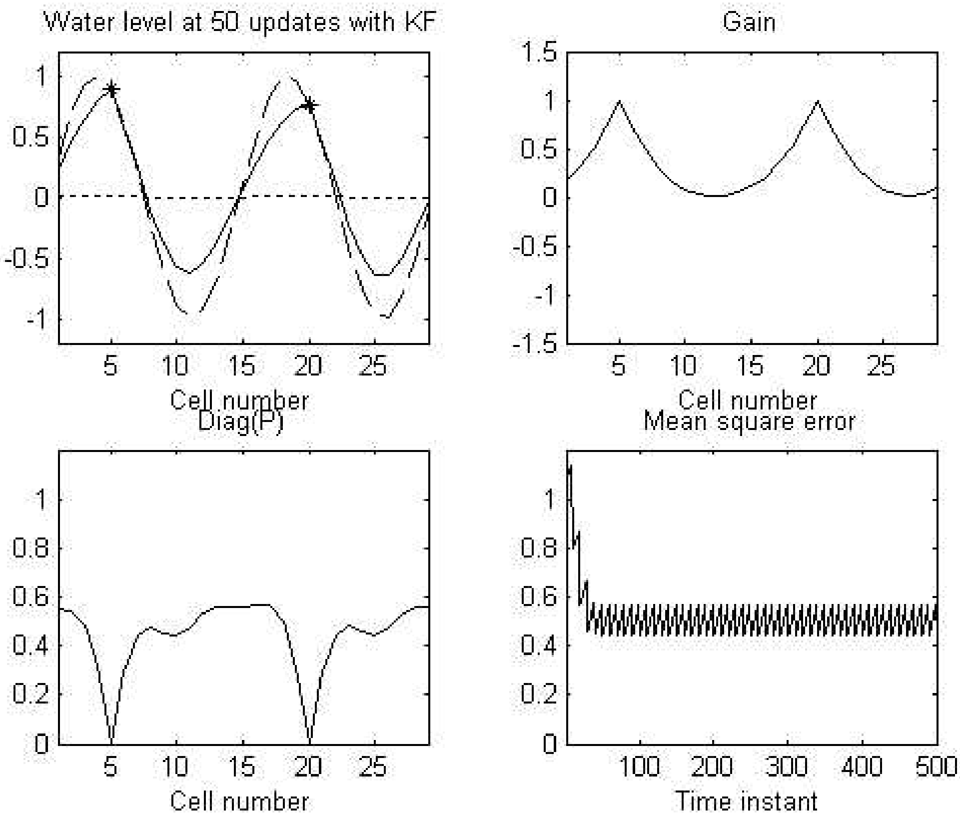


Figure V-37: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with C_r=0.25, Q specified comparing with model with C_r=1, other conditions of reference run 2 and measurements obtained from model with C_r=1.

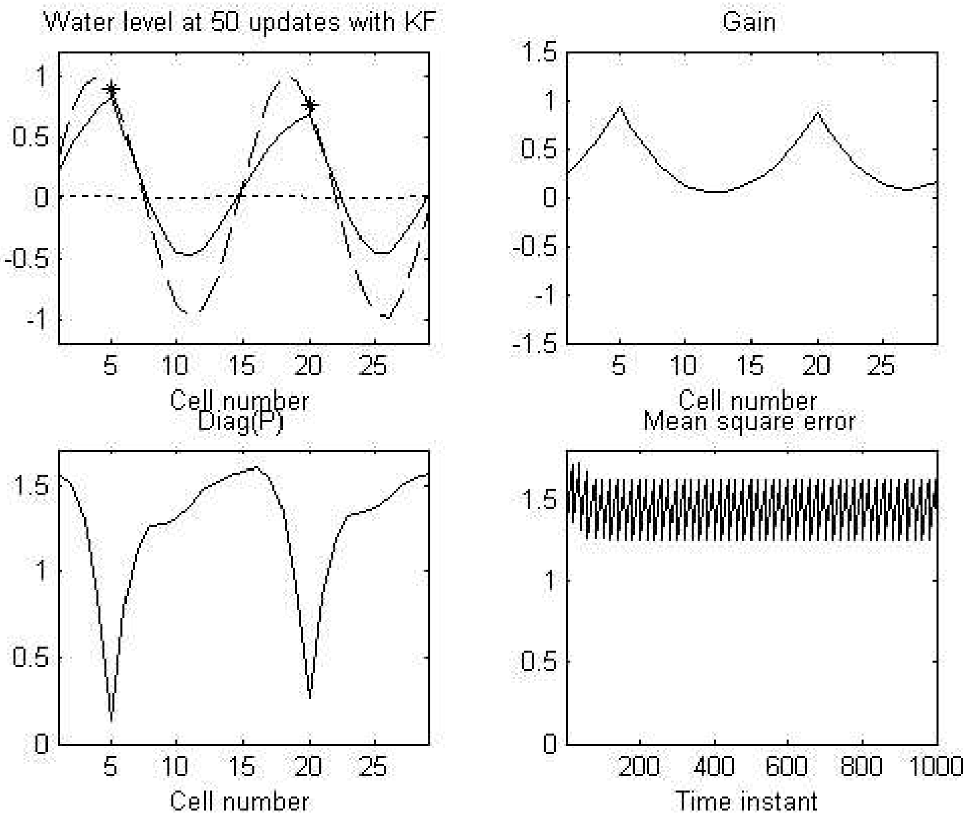


Figure V-38: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with C_r=0.1, Q specified comparing with model with C_r=1, other conditions of reference run 2 and measurements obtained from model with C_r=1.

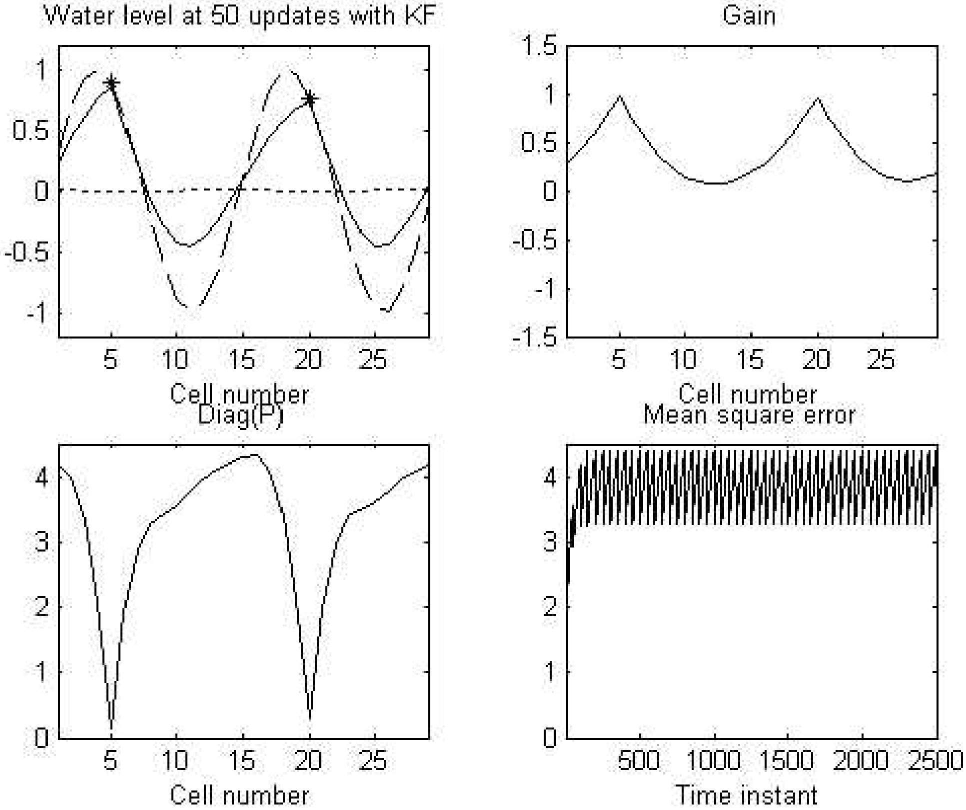
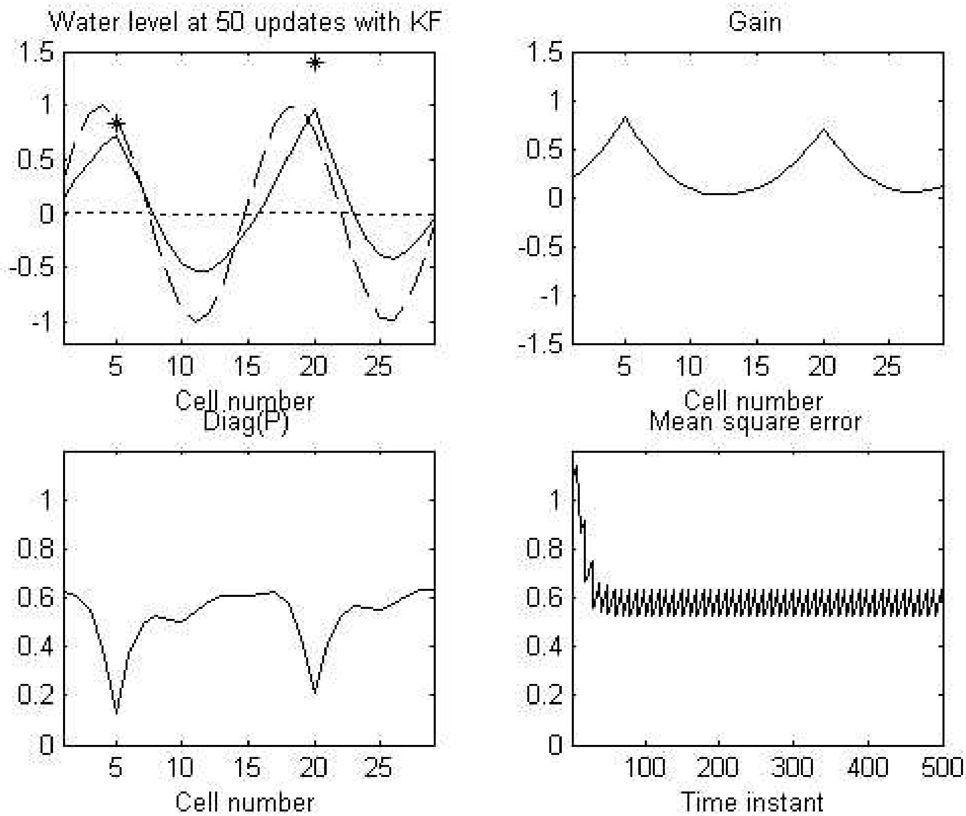


Table V-4: Characteristics of the error associated with the model comparing to the model associated with C_r=1, for schematic case.

Cell	Model with C_r = 0.25		Model with C_r = 0.1	
	Mean	Variance	Mean	Variance
1	-0.0740	0.1833	-0.0730	0.2294
2	-0.0574	0.1946	-0.0535	0.2420
3	-0.0302	0.1942	-0.0241	0.2405
4	0.0026	0.1824	0.0098	0.2258
5	0.0349	0.1675	0.0419	0.2084
6	0.0607	0.1601	0.0662	0.2006
7	0.0754	0.1653	0.0783	0.2078
8	0.0761	0.1795	0.0759	0.2250
9	0.0627	0.1927	0.0595	0.2400
10	0.0377	0.1956	0.0321	0.2423
11	0.0058	0.1861	-0.0012	0.2302
12	-0.0272	0.1709	-0.0343	0.2123
13	-0.0552	0.1608	-0.0611	0.2012
14	-0.0729	0.1628	-0.0766	0.2046
15	-0.0772	0.1756	-0.0779	0.2203
16	-0.0673	0.1901	-0.0648	0.2371
17	-0.0448	0.1961	-0.0397	0.2432
18	-0.0141	0.1894	-0.0073	0.2343
19	0.0192	0.1747	0.0264	0.2166
20	0.0490	0.1623	0.0553	0.2027
21	0.0697	0.1611	0.0740	0.2023
22	0.0775	0.1718	0.0789	0.2158
23	0.0710	0.1869	0.0693	0.2336
24	0.0514	0.1958	0.0469	0.2431
25	0.0223	0.1921	0.0158	0.2377
26	-0.0109	0.1785	-0.0182	0.2212
27	-0.0422	0.1646	-0.0489	0.2052
28	-0.0656	0.1602	-0.0705	0.2010
29	-0.0769	0.1683	-0.0791	0.2115

Figure V-39: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with C_r=0.5, Q specified comparing with model with C_r=1, other conditions of reference run 2, measurements obtained from model with C_r=1 and explicit random error in measurements.



The results presented in this section show that when facing a model producing a transient solution and measurements associated with a stable solution the Kalman filter method is capable of producing a stable solution, albeit attenuated relatively to measurements, which demonstrates the potential and usefulness of the method.

In the next section it is attempted to find the influence of the spatial and time frequencies of the measurements, through a sensibility analysis.

V.4.4. Sensibility analysis of the space and time measurements' frequencies

In this section one aims to study the influence of the frequency in time and space of the measurements on the filter performance. This issue is extremely relevant in real world applications of the Kalman filter method because then most likely one aims to process the minimum amount of measurements to achieve good filter results, due to economical reasons or availability of measurement and computational resources. Knowing the influence of the frequency of the measurements helps designing the measurement network requirements.

In the first part of this section it is considered the time frequency of the measurements.

When studying the effect of the time frequency of the measurements it has to be determined which is the maximum frequency according to the Nyquist rule (Emery and Thomson, 2001). Because the period of the signal provided by the model in this application is 14.5 time instants (half of the domain length) then frequency is 0.069 (time instants)$^{-1}$. For the frequency of measurements to be double then the period of the time grid of measurements has to be lower than 7.3 time instants. According to this, three runs are considered with time interval for update of 1, 2 and 7 time instants, respectively.

The state vector is initialized from the mean water level and a random error is explicitly included in the measurements following the procedure presented before. The results after 1 and 50 updates are presented in figures from Figure V-40 to Figure V-45.

Figure V-40: Solution of the advection equation after 1 update with the Kalman filter in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2, explicit random error in measurements and update time interval of 1 time step.

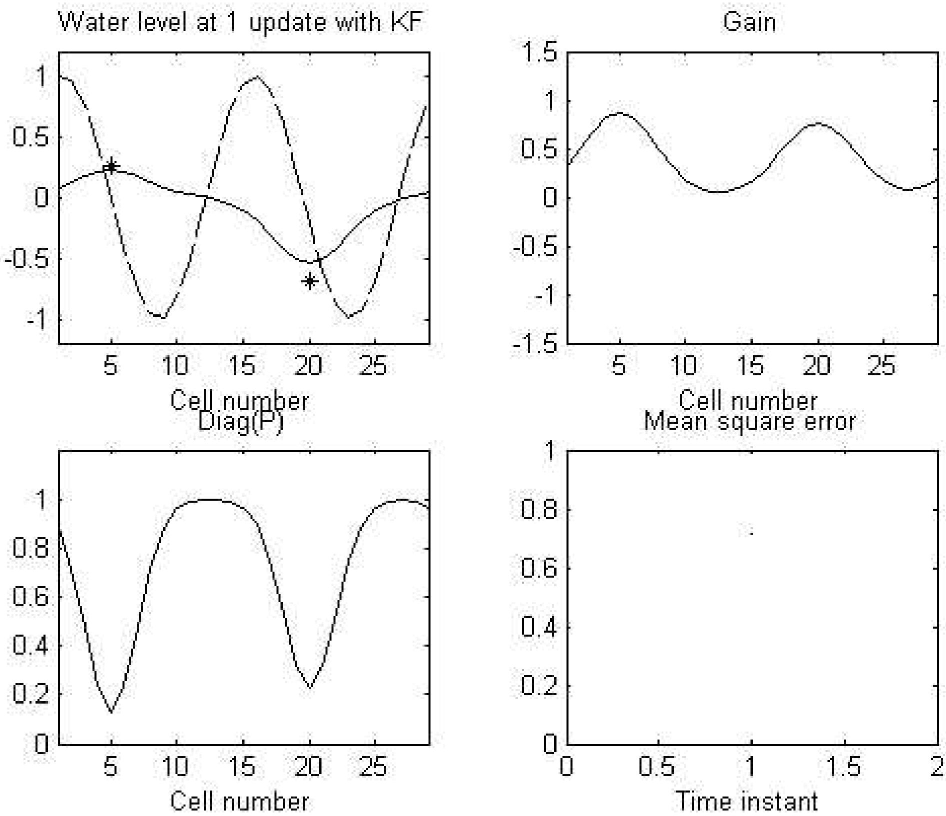
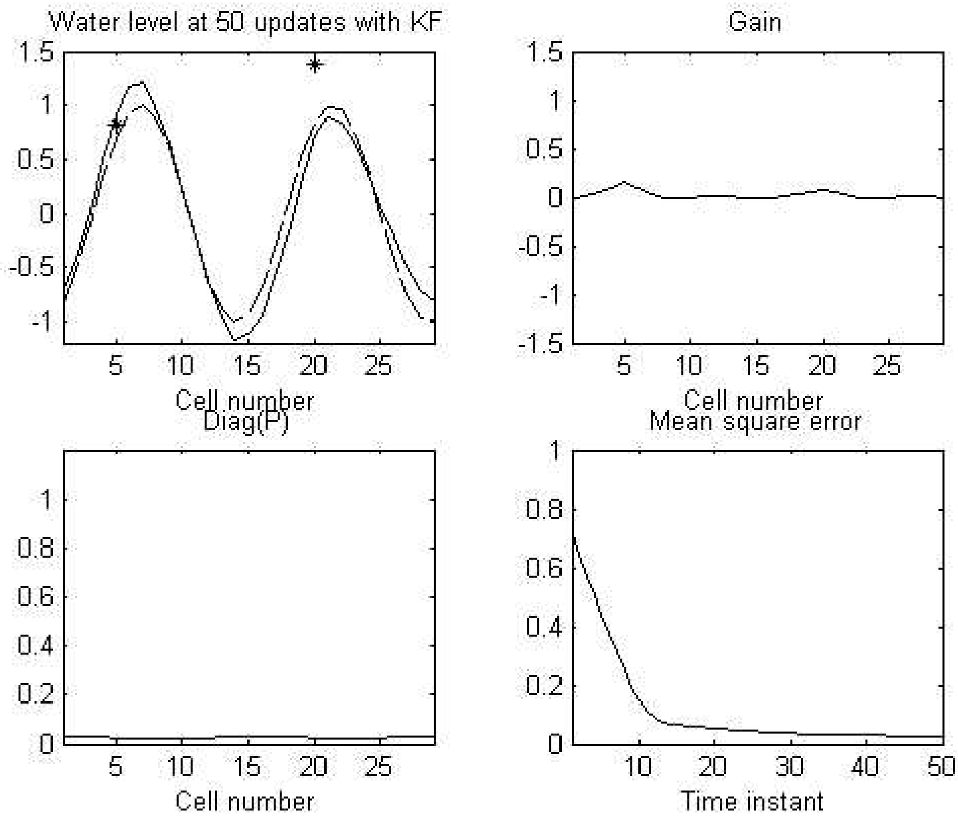
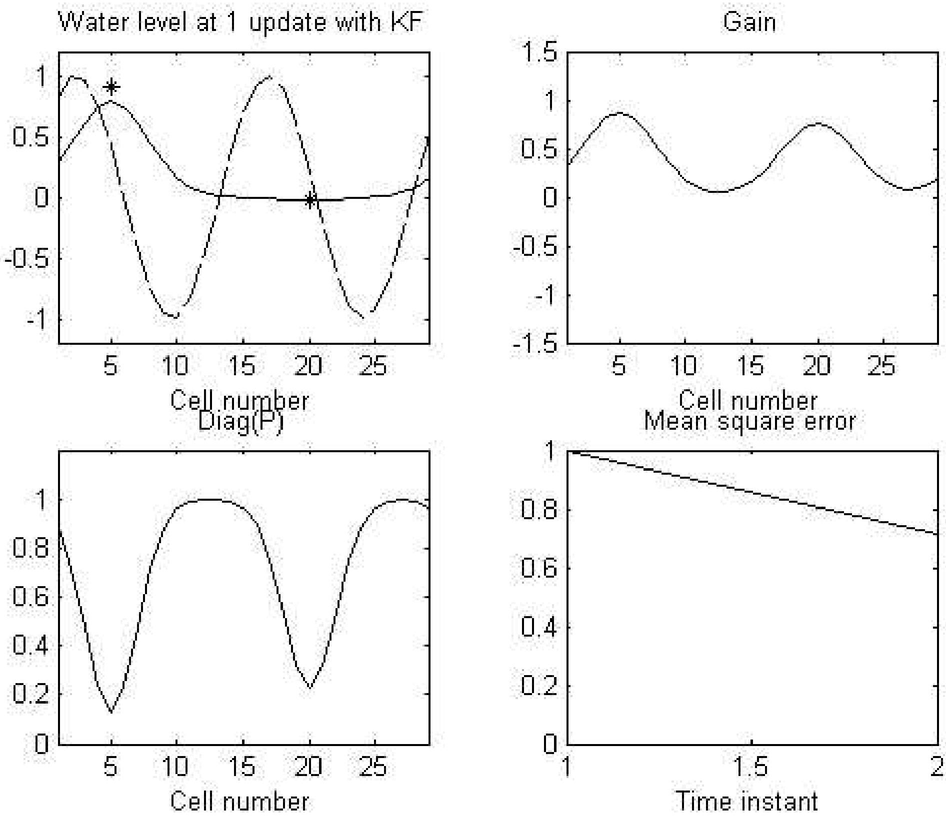


Figure V-41: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2, explicit random error in measurements and update time interval of 1 time step.



The situation after the first update for the water level is necessarily different for each run, because the wave phase is different and measurements have different values, even considering only the non random part of the measurements. But the spatial distributions of the Kalman gain and the SE are, as far as can be seen, equal to the ones verified for the update time interval of 5. It is thought that this happens because of the small value of the variance of model error that is being assumed (Q = 0.00015). From the equation of the propagation of the error of the estimate of the state vector it can be found that for the particular case of this application, where a_1=0 and a_2=1 (refer to the nomenclature provide before), the effect of the dynamic only on the elements of P is null because it causes each element to «walk» in the diagonal, all diagonals having isotropy. Thus P remains basically the same till the first update by assimilation. Because P is the same in the runs with different update interval so K is also the same.

Figure V-42: Solution of the advection equation after 1 update with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2, explicit random error in measurements and update time interval of 2 time steps.



After 50 updates the situation is very dissimilar between runs. It can be seen that higher approximation to the exact solution is achieved with lower update time intervals and solution after stabilization has also a lower SE and MSE. Gain is consistently larger with larger update time interval. When considering absolute time, stabilization is reached sooner with a higher update frequency. This happens because the error of the estimate in high frequency updates is not allowed to grow so much as with lower frequency of update. It should be noted that MSE evolution in the run with update time interval of one does not have the characteristic stair pattern because of this effect.

This analysis shows that a larger update time frequency ensures that stabilization of the solution provided by the filter is achieved sooner. Although the solution has a higher quality with hlarger time frequency of update, the system in this application is not very sensible to frequency change in terms of quality: with a seven fold change of update frequency the quality in terms of MSE shows only a two fold change. Thus the main advantage of increasing the time frequency of update seems to be, in this application, a lower stabilization time. This means that with a good quality model and a stable pattern provided by measurements an initial phase of running with higher frequency of updates can be considered to speed up the solution and after stabilization is reached the update frequency could be decreased.

Figure V-43: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2, explicit random error in measurements and update time interval of 2 time steps.

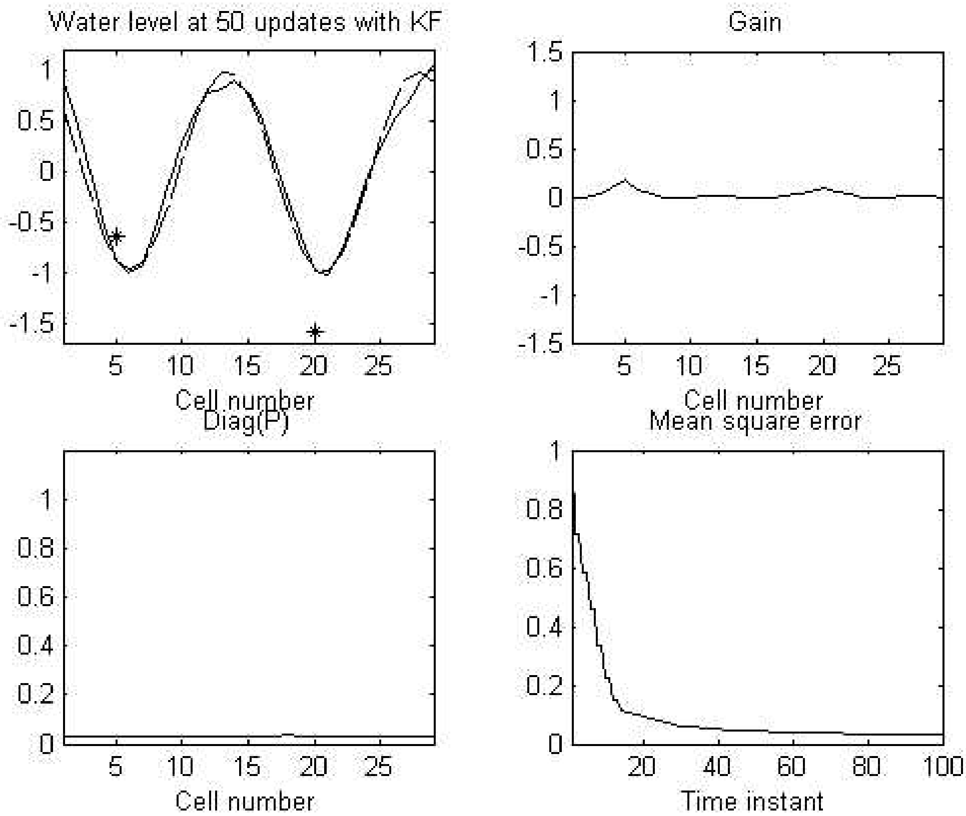


Figure V-44: Solution of the advection equation after 1 update with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2, explicit random error in measurements and update time interval of 7 time steps.

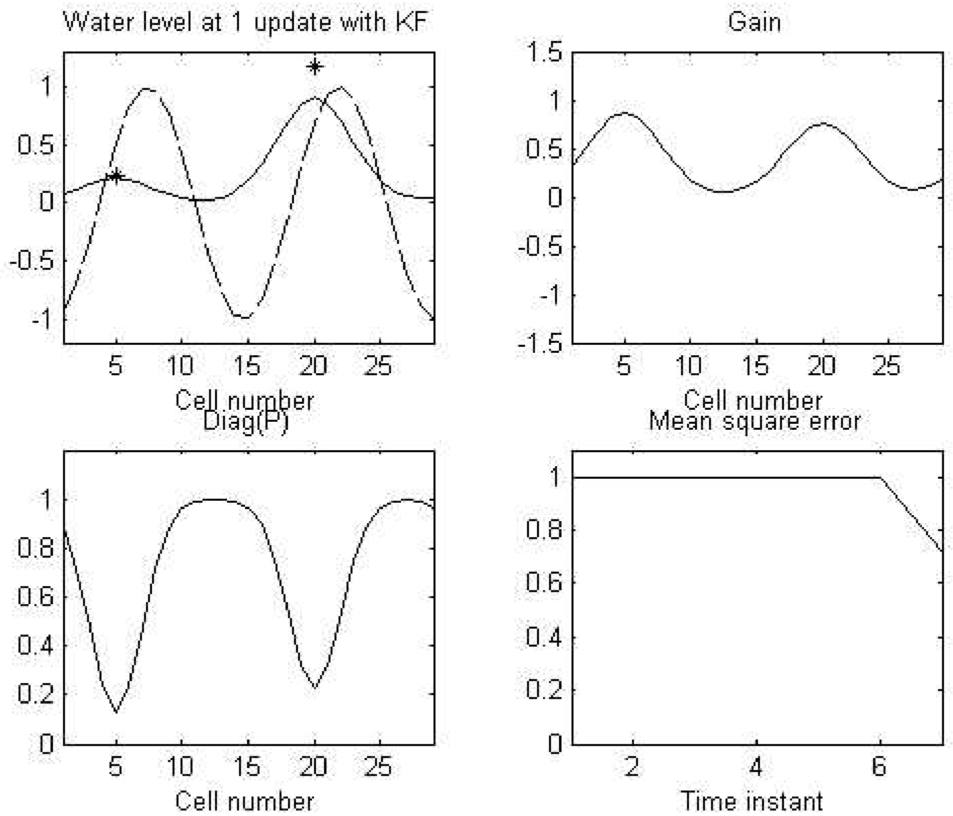
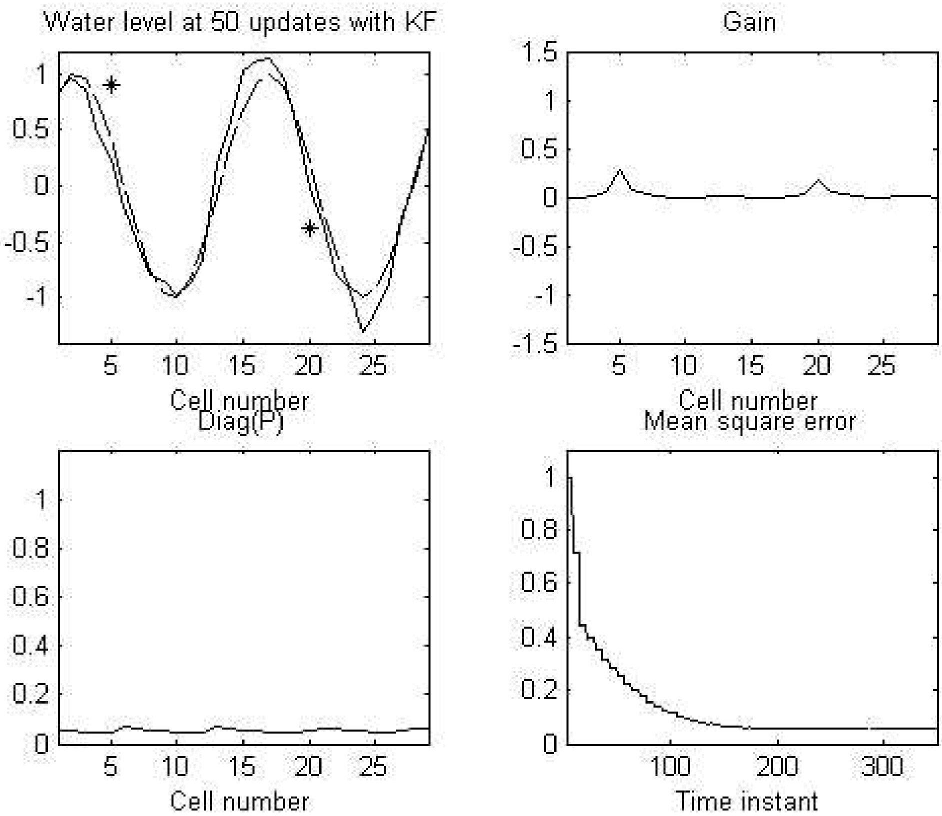


Figure V-45: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2, explicit random error in measurements and update time interval of 7 time steps.



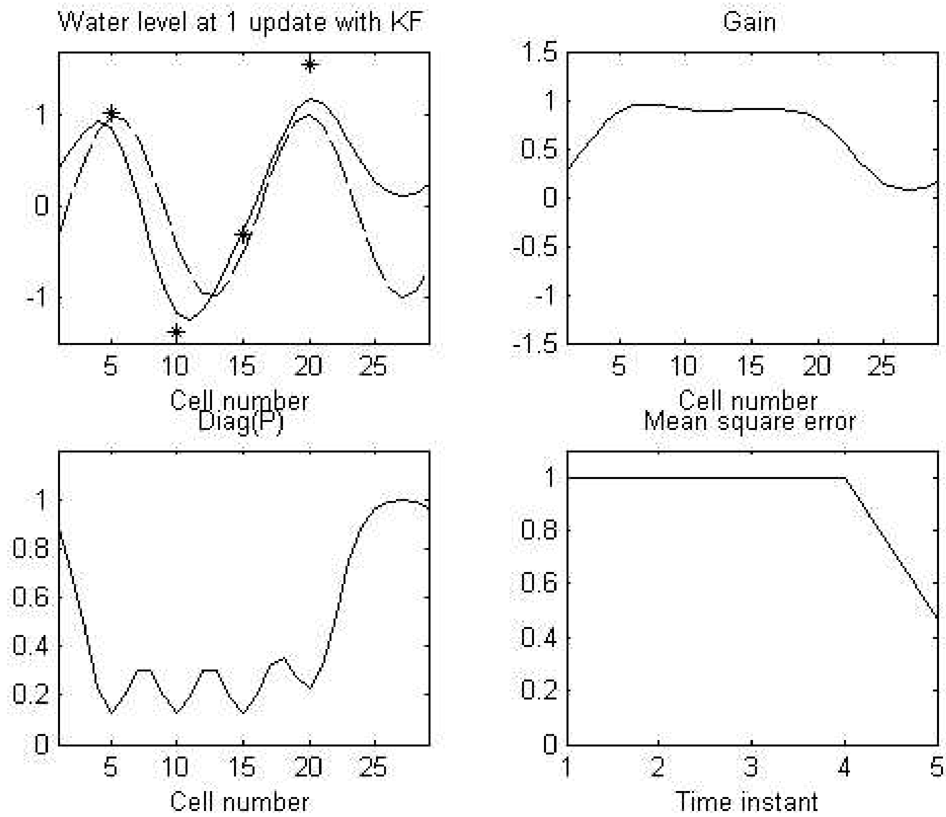
In the rest of the section it is studied the influence of the spatial frequency of the measurements. In this analysis it is assumed that measurements are taken in a spatial grid that can be chosen to be more or less thick. The runs performed are:

- thicker spatial grid: 4 measurements taken at cells 5, 10, 15 and 20;
- coarser spatial grid: 1 measurement only, taken at cell 15 (chosen to be approximately half of the domain).

To maintain comparability with the previous runs it is assumed that variance of the measurement error in cell 5 is half the one of cell 20 and that measurements in cells 10 and 15 have error variance equal to the one of cell 5. The general conditions are the ones of reference run 2 and random error is included explicitly in the measurements.

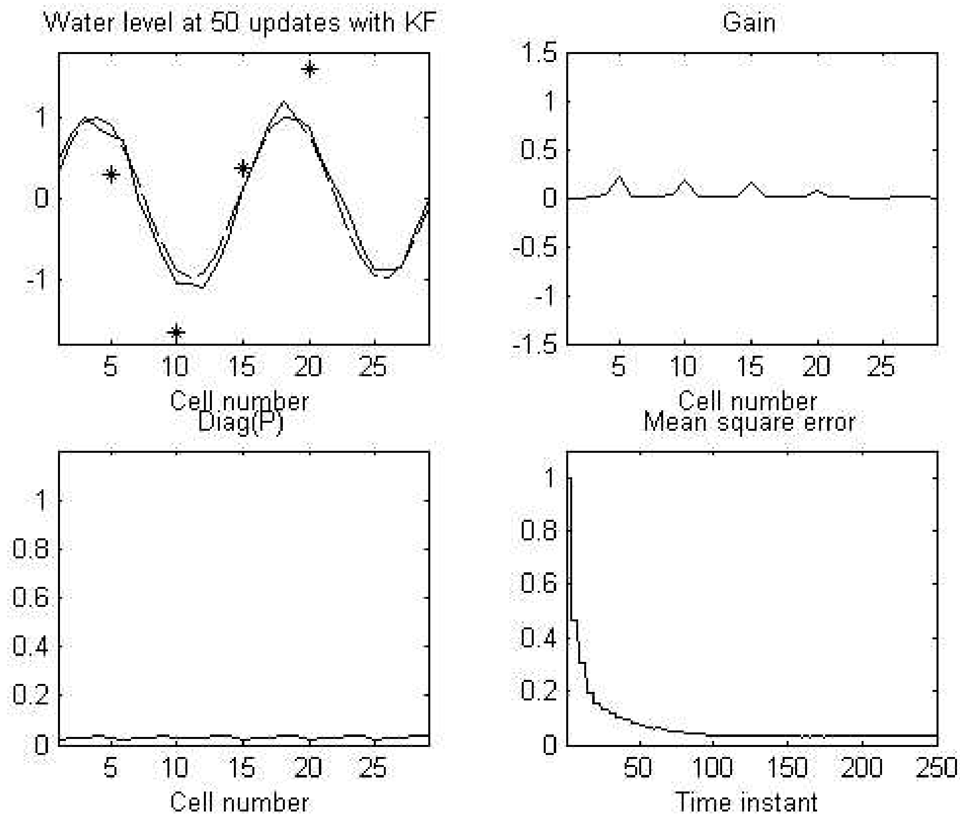
The results for the thicker spatial grid after 1 and 50 assimilations are provided in Figure V-46 and Figure V-47. After 1 update the solution is more approximated to the exact solution than the observed for a grid only with two measurements. For cells located between the cells where measurements are available the level is more corrected, as is also shown by the more smooth Kalman gain distribution. SE evidences the higher accuracy of the solution in these cells. MSE is consequently smaller after 1 update in this run. All these results are concurrent with the fact that in this run more information is available to correct the model estimated solution.

Figure V-46: Solution of the advection equation after 1 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2, four measurements and explicit random error in each measurement.



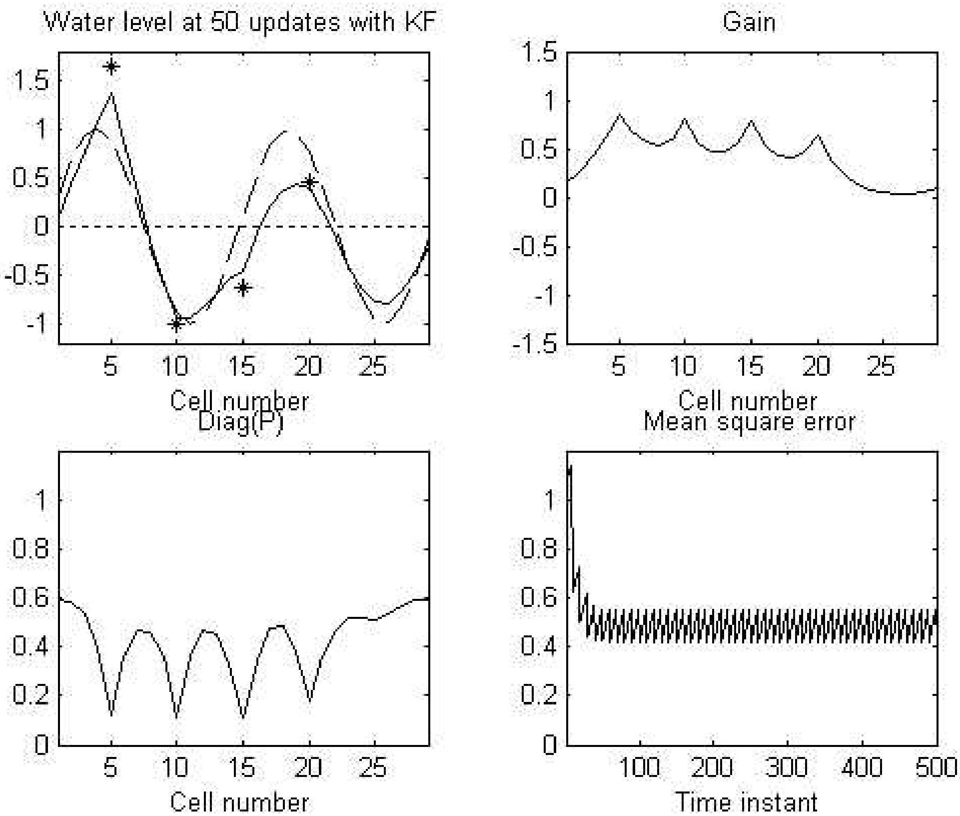
After 50 updates the solution is well stabilized: the time to stabilization does not change, contrarily to what could be expected *a priori*. From the MSE evolution in time it can be seen that corrections become smaller after the big correction verified in the first update, contrarily to the two measurements picture where the reduction of the error was more uniform. The MSE stabilization value is very similar in the cases of four and two measurements. The approximation to the exact solution is also similar in both runs. Gain necessarily shows a different distribution in the domain, with each peak signalizing the cells where measurements are available. Since in this run the statistical interpolation that forms the basis of the data assimilation is performed with more points the distribution of the gain by the points is more smooth: the cells 5, 10 and 15 verify a higher gain than cell 20 but nevertheless their value is smaller than the gain observed for cell 5 with only two measurements. This likeness in results can be explained by the fact that the model is very efficient in transmitting in the domain the information assimilated: the information from past measurements is maintained efficiently in the system and measurements refer to a constant signal so the fact that a thicker grid is used is not relevant in this application. With a less qualified model this could prove to be more important to approach an adequate approximation to the exact solution. A wrong model run was made with the wrong model specified with C_r equal to 0.5 and the true model/measurements with C_r equal to 1. The results after 50 updates are found in Figure V-48. The apparent bias relative to the mean level seems to have been partially removed from the solution and the MSE in stabilization has decreased, revealing that the estimate is perceived as more accurate.

Figure V-47: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2, four measurements and explicit random error in each measurement.



The results for the run with only one measurement after 1 and 50 updates are provided by Figure V-49 and Figure V-50. After 1 update the water level in each cell is necessarily less approximate to the exact solution than the equivalent solution obtained assimilating two measurements. The value of the Kalman gain is equal in this run and in the runs with two and four measurements, because the state vector estimate starts with the same mean water level. SE shows a distribution consistent with the ones observed in the runs with two and four measurements assimilation. Because the information assimilated in this run is lower, the reduction of the MSE obtained after one update is lower than previously, confirming the pattern of more measurements implying a larger reduction of the MSE after the first update.

Figure V-48: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with C_r=0.5, Q specified comparing with model with C_r=1, other conditions of reference run 2, measurements obtained from model with C_r=1, four measurements and explicit random error in measurements.



The situation after 50 updates confirms the findings of the runs with two and four measurements, in terms of the approximation of the water level to the exact solution (rather similar) and gain (larger in each measurement with coarser measurement grid). However when it comes to the error of the estimate the situation is different: it is perceived an increase of the MSE level after stabilization about one half of the MSE verified with four measurement assimilation. Stabilization of MSE is achieved later, which is necessarily linked with a lower initial correction of the estimate of the state vector: since less information is assimilated each update time the system needs to rely more on the information propagated from past assimilations and so stabilization takes longer. The results for these runs show the importance of past assimilations information propagation in the system. In longer domains, where the propagation of past assimilations information is slower, the coarseness of the measurement grid may be found quite relevant in determining the quality of the analysis.

Figure V-49: Solution of the advection equation after 1 update with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and one measurement with explicit random error.

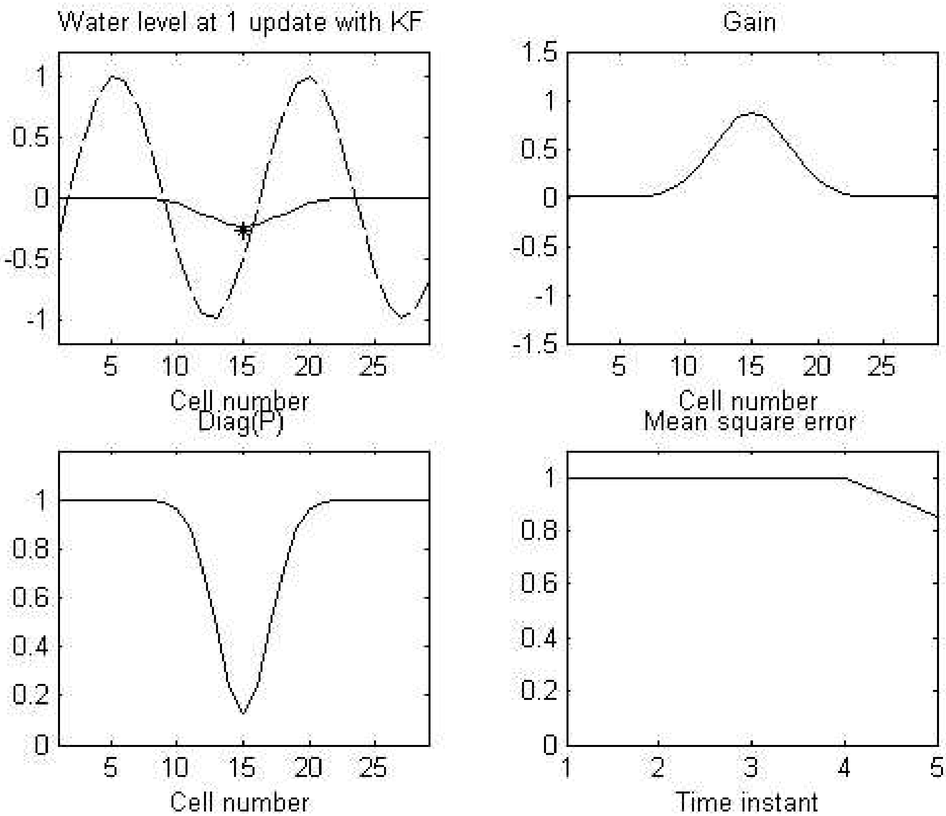
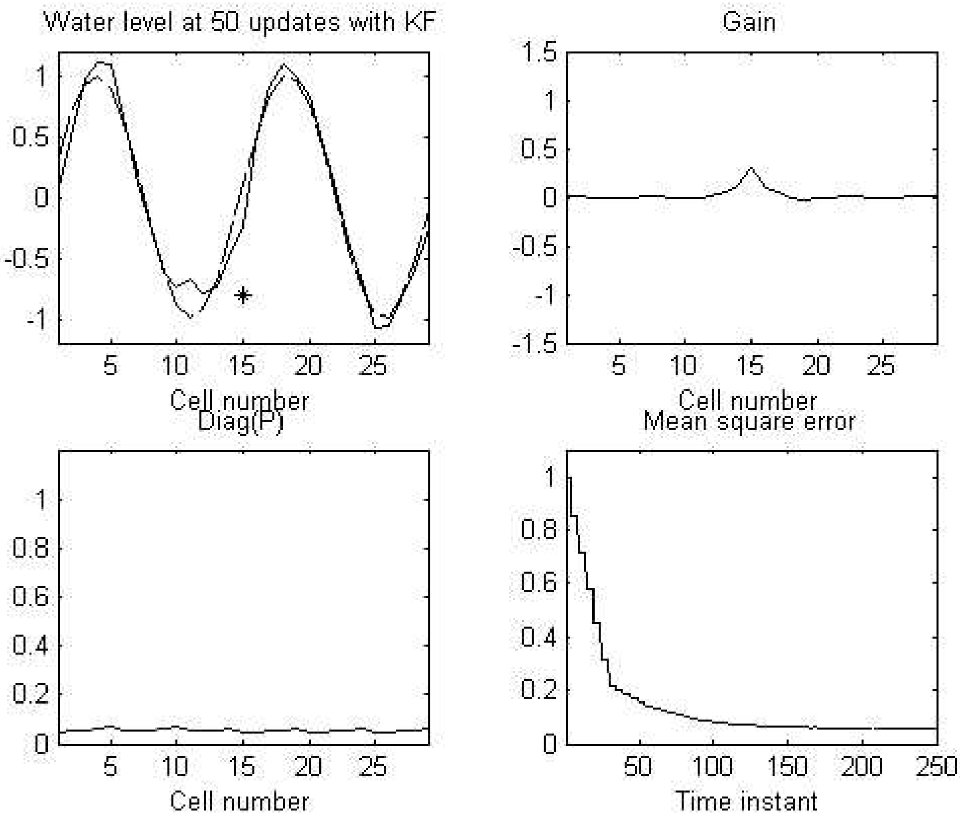


Figure V-50: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and one measurement with explicit random error.



V.4.5. Analysis of behavior with sudden dynamic change

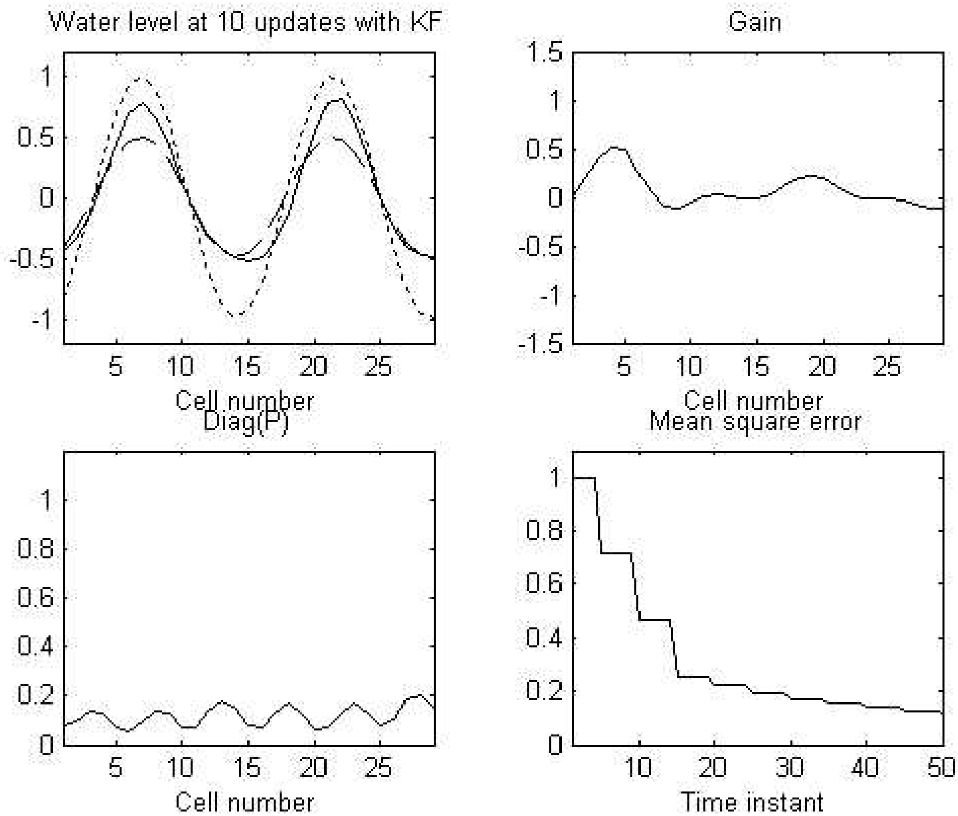
In this section it is analysed the effect on the filter of a sudden change in the dynamic. This change should be visible in a sudden change of the signal provided by the measurements[121]. These analyses are very important because in real world applications sudden changes of the parameters measured can happen. To be useful, the Kalman filter should prove robust enough to account for these changes in dynamic and correct the model in the right way. A sudden change in the signal could be theoretically conceived as the change in any of the following parameters of the wave equation: amplitude, mean water level, period and celerity. But in real world applications is not normally expected a sudden change in period or celerity of a wave of tidal characteristics. A change in amplitude or mean water level can be the effect of a change of atmospheric pressure or of a tectonic vertical movement of the solid bottom of the water mass, for instance. Another possible situation is the appearance of another wave: this wave could be originated from atmospheric processes and thus add to a pre-existent tidal wave.

[121] A sudden change could also be considered in the model but since this is used defined (and then controlled by the user) it is not foreseen that the knowledge gained with this analysis is relevant to real world applications of the Kalman filter.

A first set of runs are made with a sudden change of the amplitude of the wave to the value of 0.5, introduced before and after stabilization. Since the conditions of reference run 2 are considered, the sudden change is introduced at t = 25 in one run and t = 150 in the other run. No random error is explicitly included in measurements in these runs for assessing with more detail potential effects of the sudden change in the filter. It is though *a priori* that the response of the filter should be slower after stabilization because the error of the estimate of the state vector and the Kalman gain are very much reduced. To assess the speed of response the run behaviour is assessed at the time of 5 updates post the change in amplitude, thus after 10 updates for run with change of amplitude at t = 25 and after 35 updates for run with change of amplitude at t = 150.

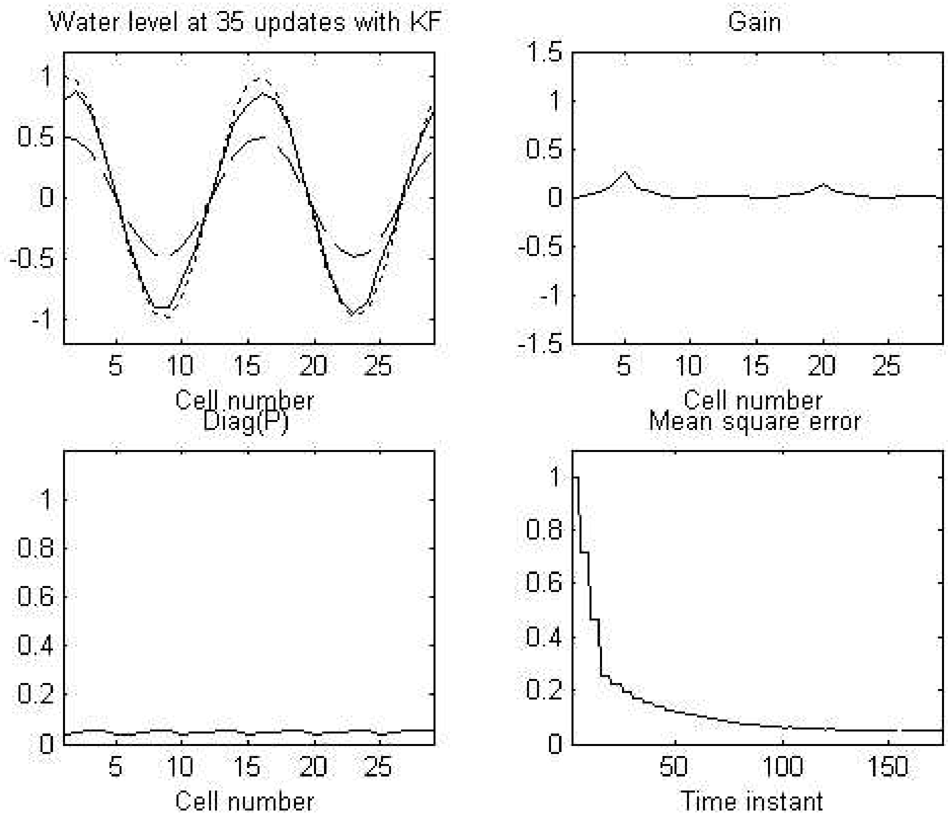
The results are presented in Figure V-51 and Figure V-52, respectively for sudden changes of amplitude introduced at t = 25 and t = 150.

Figure V-51: Solution of the advection equation after 10 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and amplitude of the wave measured suddenly changed to 0.5 at t=25.



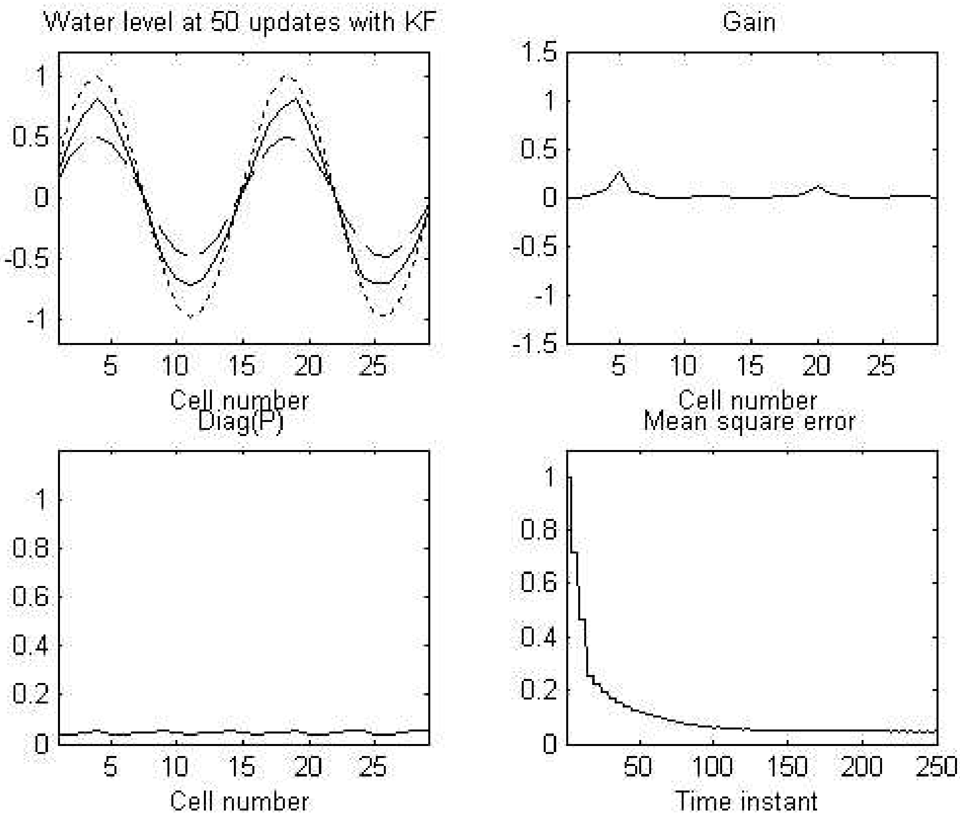
The results show that in either run after 5 updates past the sudden change of amplitude the exact solution is still not reached. Nevertheless, the situation of the change introduced in t=25 shows a better approximation, which is concurrent with the expectations: the reaction of the filter is indeed slower if the change is introduced when stabilization of the solution is installed, because the Kalman gain is considerably smaller. The distributions of Kalman gain and error of the estimated state vector are not very different from the ones obtained before. This highlights that the evolution of K and P depends only on the initial specification of P, R and Q and thus an externally produced change in the measurements or the estimate of the state vector is not traduced in the quality of the estimate. This is a feature of the Kalman filter method that should be always kept on mind. After 50 updates (Figure V-53) in the run with change at t=150 the solution is still somewhat apart from the exact solution. It is thought that if the measurement error has associated a smaller variance then the convergence of the filter to the exact solution would be faster.

Figure V-52: Solution of the advection equation after 35 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and amplitude of the wave measured suddenly changed to 0.5 at t=150.



In the runs presented next it was the mean level that changed to 0.5 m. This change is introduced as before at t=25 in one run and t=150 in another. The results after 10 updates and after 35 updates are presented in Figure V-54 and Figure V-55.

Figure V-53: Solution of the advection equation after 50 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and amplitude of the wave measured suddenly changed to 0.5 at t=150.



As verified with the amplitude change, the filter responds more quickly when the change of mean level is introduced before stabilization. The distributions of the Kalman gain and of the error of the estimate of the state vector do not change from the situation without dynamic variation.

Figure V-54: Solution of the advection equation after 10 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and mean level of the wave measured suddenly changed to 0.5 at t=25.

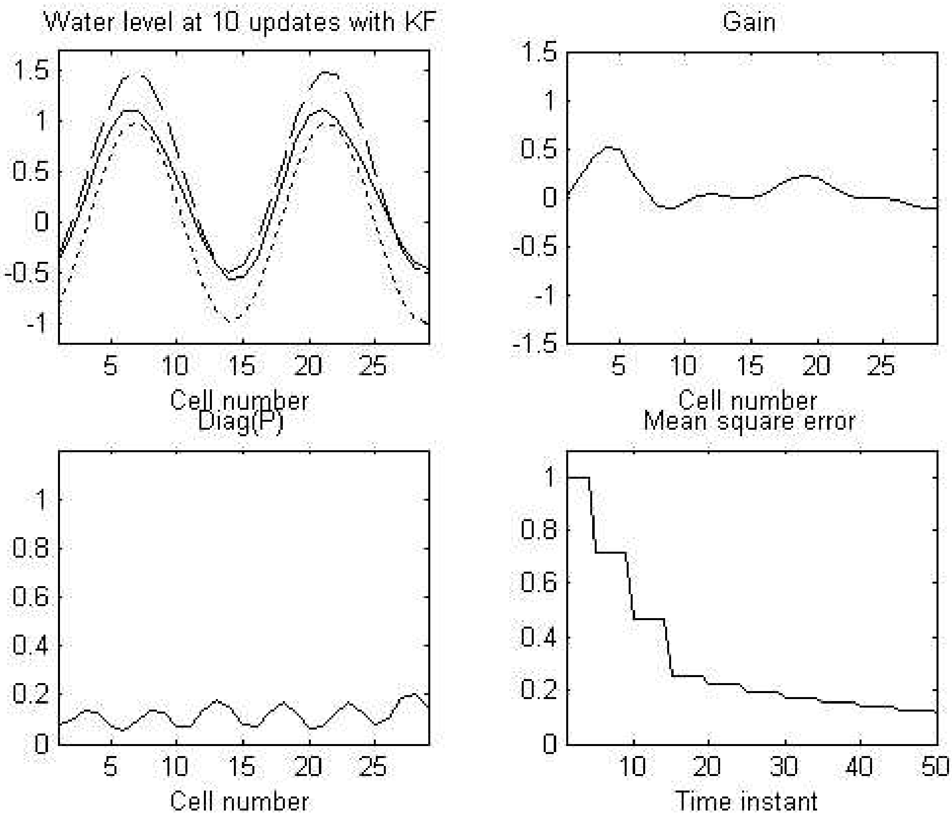
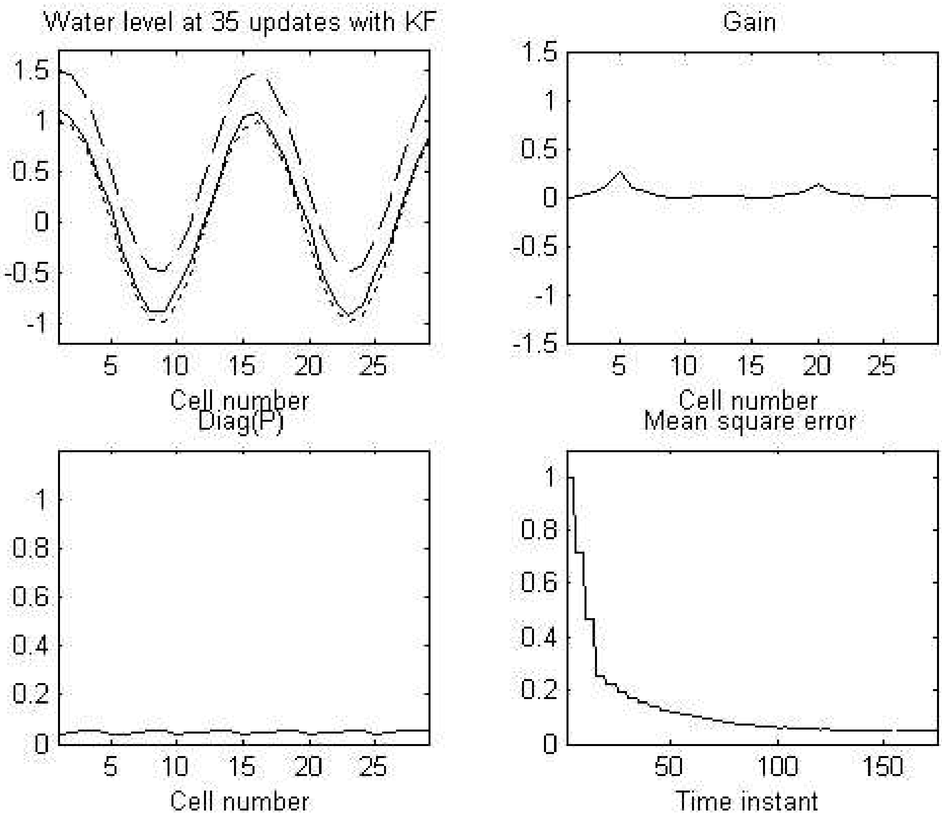


Figure V-55: Solution of the advection equation after 35 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and mean level of the wave measured suddenly changed to 0.5 at t=150.



Some runs were also made with the introduction of a wave with smaller amplitude and larger period. The following characteristics were considered for the new wave relatively to the wave already considered:

- period: 10 times larger;
- amplitude: 3 times smaller;
- mean level: equal.

As before the new wave was introduced before stabilization in one run and after stabilization in another run, being considered for that purpose the same time instants. The results are presented in Figure V-56 and in Figure V-57, respectively for the run with the change occurring at t = 25 and t = 150.

The effect of the introduction of the new wave consists in the vertical translation of the water level distribution in the domain. In these runs it is verified that the reaction of the filter to the change in signal is slower than before, possibly because the new dynamic is more complex than the already specified in the filter and the model considered is not as adequate as before at propagating this dynamic. After 600 updates (Figure V-58) in the run with the new wave introduced at t = 25 the estimated solution is still very much apart from the exact solution, although theoretically the measurements contain all the information needed to reconstruct the signal according to Nyquist rule. As expected the distributions of the Kalman gain and the error of estimate of state vector are equal to the ones observed before.

Figure V-56: Solution of the advection equation after 10 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and introduction of another wave with ten times the period and one-third of the amplitude at t=25.

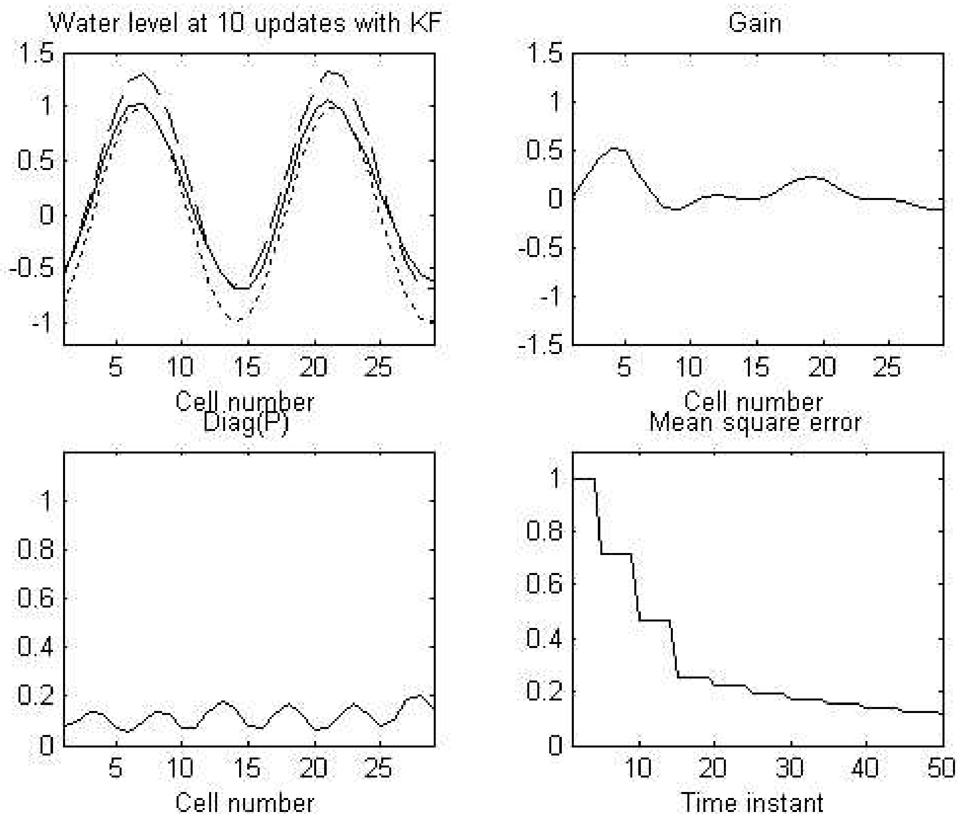


Figure V-57: Solution of the advection equation after 35 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and introduction of another wave with ten times the period and one-third of the amplitude at t=150.

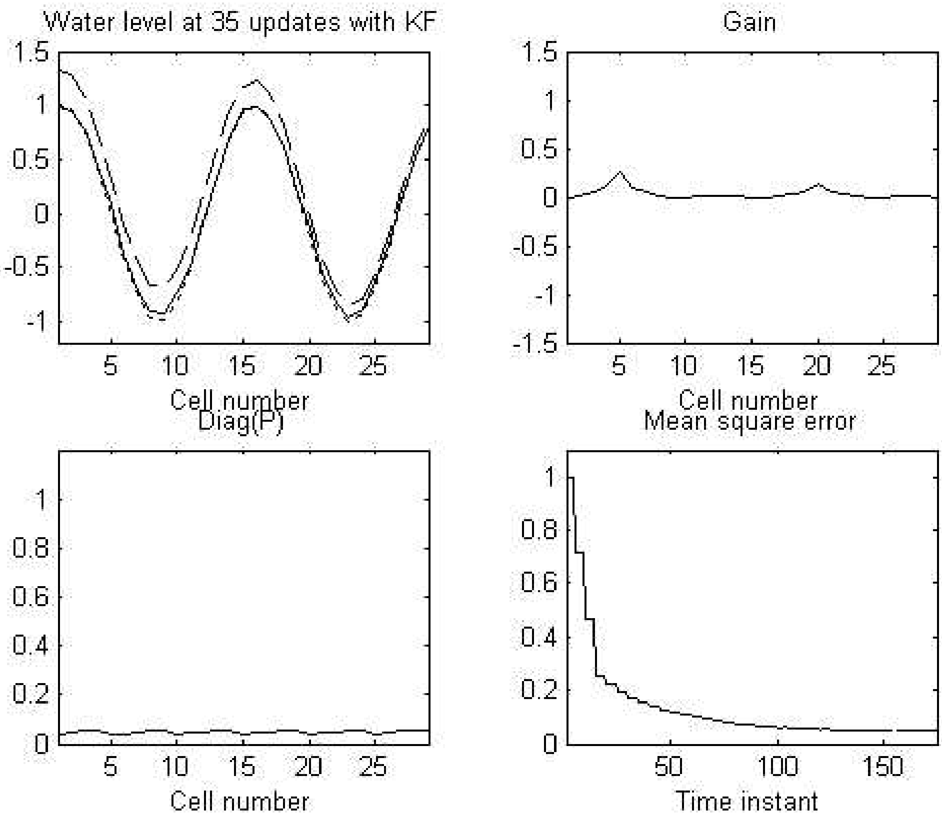
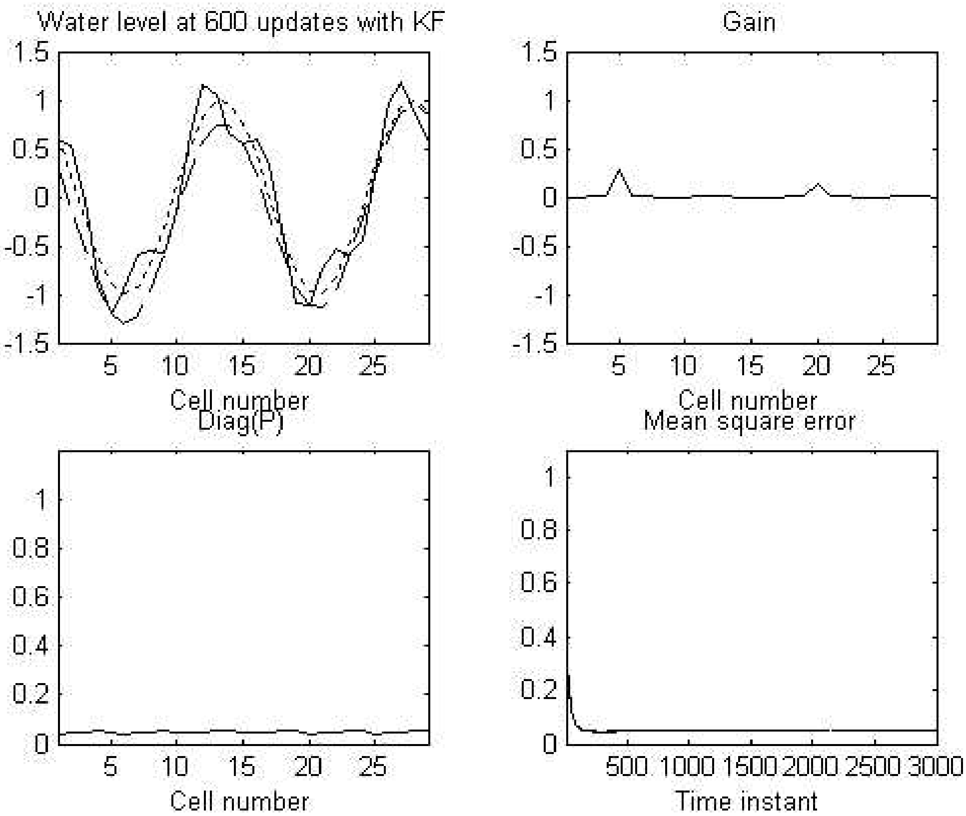


Figure V-58: Solution of the advection equation after 600 updates with the Kalman filter (in water level, filter solution in continuous line, analytic solution in dashed line, solution without assimilation in small dashed line and measurements with star mark) with conditions of reference run 2 and introduction of another wave with ten times the period and one-third of the amplitude at t=25.



V.5. Conclusive remarks

In this section is presented a summary of the influence of different modelling parameters and situations in the key parameters that are interesting to the user of the modelling tool because they allow the assessing of the quality of the modelling. Table V-5 presents those findings.

A major finding of this study is the possibility of occurrence of a resonance problem when one has a cyclic boundary condition. The specification of Q as proved to be a key factor for a good performance of the filter, which agrees to what is stated in the literature.

Some useful features of the Kalman filter method were detected in this study, namely the robustness of the method's solution given a variety of specifications of the initial P. Nevertheless some faults were also present: the small sensibility of the filter relatively to large differences between the model produced state vector and the measurements (this is the problem usually called filter's divergence).

The study of this application of the Kalman filter has allowed the acquisition of a considerable amount of knowledge about the practical use of the Kalman filter method, its potentials and faults, providing a good basis for future research.

Table V-5: Influence of several modelling situations in key parameters of the system of schematic case.

Modelling situation	**Approximation to true solution**	**Kalman gain**	**Variance of error of state vector's estimate (after stabilization)**	**Time to stabilization of the solution**
R ↑	↓ (before stabilization)	↓	↑ (not proportional)	-
r_c ↑	↑ but may ↓ (if too high tends to uniformity)	more distributed	↓ (not proportional)	-
Q ↑	↓ (before stabilization)	↑ less distributed	↑	-
Explict random error in measurements	↓	-	-	-
Initial P ↑	↑ (before stabilization)	↑	-	-
R dependent of water level measured	↓	-	↓	-
Q dependent of water level from model	dampening of extremes very sensitive to random numbers specified	↓	↓	-
Wrong model (lower C_r)	lower amplitude than the exact solution	↑	↑	- (not very sensible)
Time frequency of measurements ↑	↑	↑	↓ (sensibility is small)	↓
Spatial frequency of measurements ↑	↑	more distributed	↓	↓ (not very sensible after a frequency of 2)
Sudden change of amplitude to 0.5 m	response is slower if the change is introduced after stabilization	-	-	n.a.
Sudden change of mean level to 0.5 m	response is slower if the change is introduced after stabilization	-	-	n.a.
Sudden appearance of another wave	response is slower if the change is introduced after stabilization	-	-	n.a.
Increase of P by the difference between model and measurements	↑ in cells with measurements	↑	n.a.	n.a.

Note: ↑ = increase; ↓ = decrease; - = maintains; n.a. = not applicable; + = plus.