ministerie van verkeer en waterstaat

koninklijk nederlands meteorologisch instituut

The effect of wave data assimilation on the numerical simulation of wave energy advection

Miriam de Las Heras

and Gerrit Burgers

technical reports; TR - 121

technische rapporten; TR - 121

de bilt, 1990

KNMI

Abstract

Simulated observations are assimilated into a simple wave model, in which only propagation is considered. The approach used to assimilate them is called "model fitting". It looks for the solution of the model equations which fits best the observations to the space-time domain we are considering. It is also required that the deviation from our "a priori" knowledge (in our case the model initial state) remains as small as possible, in some well specified sense.

The method is based on the minimisation of a so called "cost function", which is a quadratic function of the differences between modelled data and observations on the hand, and between modelled and "a priori" data on the other hand.

The minimisation of the cost function is done by means of an iterative gradient-based descent method.

After assimilation, new model results, which fit better to the observations and prior knowledge considered, are produced. Therefore we conclude that the data assimilation process improves the wave field, but its effect is more clearly noticed around the observations than elsewhere, and it does not remain long after stopping assimilating.

1. Introduction

In general, modelled data are not equal to the measured observations available. Therefore the idea of introducing observed data into the modelling procedure is considered, in order to produce better analyses and predictions. This process is called data assimilation.
In this sense, one can say that data assimilation is the "correction" of a numerical simulation of a process, taking into account the observations available at the moment.
Various methods exist (Ghil, 1989). The simplest one consists of overwriting the modelled results with the measured data. Because this produces discontinuities and noise in the data, it is generally not preferred.
The method of "successive corrections" tries to reduce the noise introduced in overwriting by spreading out the observed information over the grid, according to the distance to the observation point.
An alternative method, "optimal interpolation", adds a fraction of the difference between modelled and observed values to the modelled ones. This difference is weighted according to the error covariance of the model. This method is the most advanced one which works operationally.
The "Kalman filter" is similar to optimal interpolation but the error covariances are computed in the numerical model dynamics. Therefore it is more accurate but it requires considerable computing power.
Finally, the procedure that we will consider here is called "model fitting" (Thacker, 1988; Long, 1989). This method has been considered for wave forecasting by (de Valk, 1989 and Schilperoort, 1989). It is an alternative to Kalman filtering which does not need as much computing effort.
It looks for a solution of the model equations, which is a compromise between the observations and the "a priori" knowledge available. This "a priori" knowledge could be, as in our case, the initial model state, but also some smoothness in the data, climatology, etc.
To evaluate if one solution is better than another, a "cost" is assigned to the differences between modelled and observed data, and between modelled and "a priori" knowlegde, and a so called cost function is calculated. This is a quadratic function of those differences. In this way, the solution which has minimal cost is considered to be the optimal one.

The process followed to find the minimum of the cost function is called the "adjoint method". It will be explained in detail in section 3.
The purpose of the present study is to get a better understanding of this data assimilation procedure in a relatively simple system: monochromatic advection on a "torus", i.e. a two dimensional grid with periodic boundary conditions.
On the one hand, this system is simple enough to analyse in full detail. On the other hand, the number of degrees of freedom is similar to that of meteorological models.
Interpreted in terms of wave physics, the advection equation gives the evolution of swell without any source terms. It is given by:

$$\frac{\partial}{\partial t} F(\vec{x},t;\vec{k}) + \vec{c}_g(\vec{k}).\frac{\partial}{\partial \vec{x}} F(\vec{x},t;\vec{k}) = 0 \qquad (1)$$

Since in this equation there are no derivatives with respect to the wave vector $\vec{k}$, only one wave vector is considered (monochromatic). The grid is chosen in such way that the stability condition is not violated (see section 2.1.).

Our model will be the first order upwinding discretized approximation of the continuous advection equation. Some observations will be simulated via ray back-tracking of the exact solution and then they will be compared with the "model counterparts", results of the interpolation of the four closest modelled data at the place of the observation (see section 2.4). After having compared them , it is noticed that a misfit exists between both. Therefore we assimilate the observations in the model, run the model again and compare the new modelled data after assimilation with the observations. The resulting misfit is now smaller.

The scheme of the process we follow is shown in figure 1.

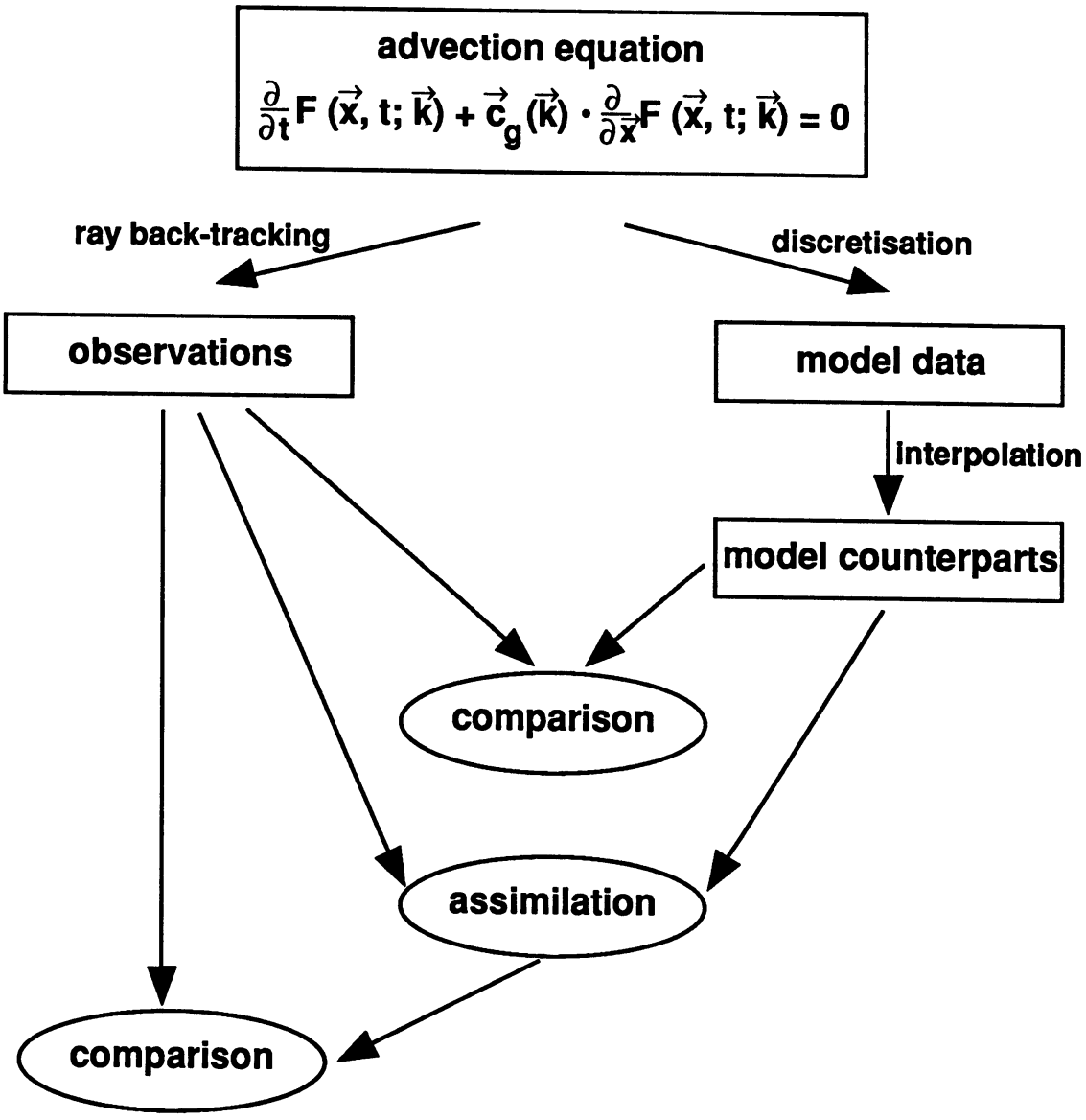

advection equation
$\frac{\partial}{\partial t}F(\vec{x}, t; \vec{k}) + \vec{c}_g(\vec{k}) \cdot \frac{\partial}{\partial \vec{x}}F(\vec{x}, t; \vec{k}) = 0$
ray back-tracking
discretisation
observations
model data
interpolation
model counterparts
comparison
assimilation
comparison

fig. 1

2. Generation of modelled and observed data.

2.1 Model

The grid we consider is a 20 x 20 points rectangular one, with periodic boundary conditions. Therefore the surface we are working with is a torus (see Fig. 2).

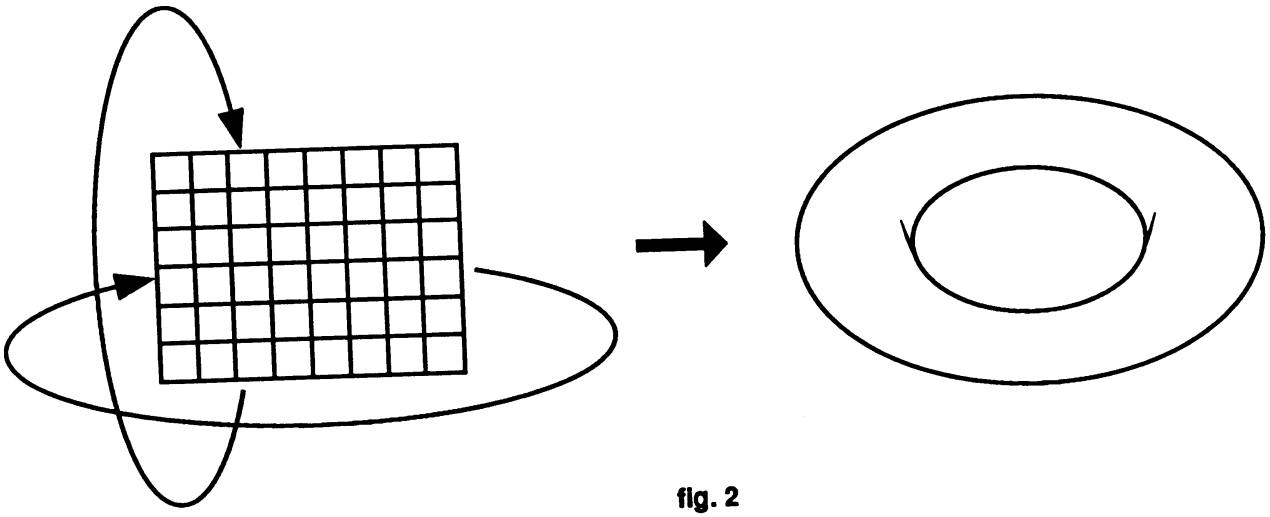
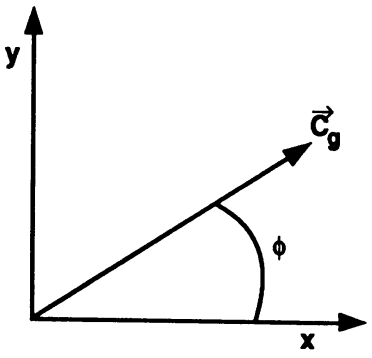


fig. 2

The discretisation scheme we use is a "first order upwinding" one:

$$F^{t+1}(i\Delta x, j\Delta y) = (1-\alpha_x-\alpha_y)\, F^t(i\Delta x, j\Delta t) + \alpha_x\, F^t[(i-\sigma_x)\Delta x, j\Delta y) + \alpha_y\, F^t(i\Delta x, (j-\sigma_y)\Delta y)]. \quad (2)$$



$$\vec{c}_g = (c_g\cos\phi,\ c_g\sin\phi)$$

$$\alpha_x = \frac{c_g\sigma_x\cos\phi\Delta t}{\Delta x},\quad \alpha_y = \frac{c_g\sigma_y\sin\phi\Delta t}{\Delta y}$$

From now on, the following notation will be used:

$$F^t_{ij} = F^t\,(i\Delta x, j\Delta y)$$

The above equation has a stable solution provided that α_x and α_y satisfy the following stability conditions:

$$\begin{aligned} &\alpha_x > 0 \\ &\alpha_y > 0 \\ &1 - \alpha_x - \alpha_y > 0 \end{aligned} \qquad (3)$$

These three conditions also guarantee that F remains positive if it was already so at the former time stage.

The first two conditions $\alpha_x > 0$ and $\alpha_y > 0$ imply that

$$\sigma_x = \cos\phi/|\cos\phi| \text{ and } \sigma_y = \sin\phi/|\sin\phi|$$

This means that the sigmas must be chosen in such a way that the scheme is "upwinding" i.e. that the waves at (x,y) are "coming from" $(x-\sigma_x,y)$ and $(x,y-\sigma_y)$.

The last condition is equivalent to

$$c_g \Delta t \left[\frac{|\cos\phi|}{\Delta x} + \frac{|\sin\phi|}{\Delta y}\right] < 1$$

and gives, for fixed Δ_x,Δ_y and Δ_t, an upper limit for the group velocity c_g. Waves are not allowed to propagate more than a grid unit per time step.

2.2 Simulation of observations

The exact solution of the advection equation has the following property:

$$F(\vec{x},t) = F(\vec{x}-\vec{c}_g t,0) \qquad (4)$$

Therefore, given an initial configuration:

$$F_{init}(\vec{x}) = F(\vec{x},0) \qquad (5)$$

the exact solution at any point in space and time can be generated in this way:

$$F(\vec{x},t) = F(\vec{x}-\vec{c}_g t,0) = F_{init}(\vec{x}-\vec{c}_g t)$$

the observations at point $\vec{x}$ and time t are taken to be the values of the exact solution at that point and time.
That means we go back in the opposite direction of the wave from $\vec{x}$ to $\vec{x}-\vec{c}_g t$ and then we take the initial value of the wave variance F at that point $\vec{x}-\vec{c}_g t$, as the observation value at point $\vec{x}$ and time t. (ray back-tracking).

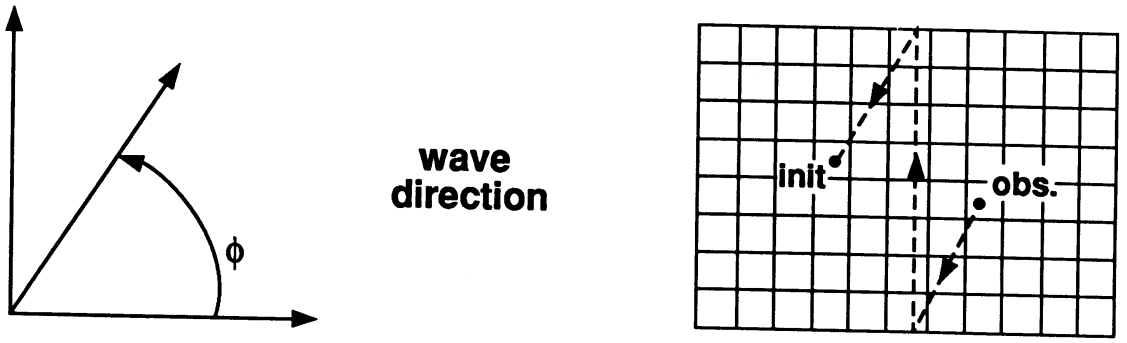


fig. 3

In this way, we generate 20 observations, at five arbitrary points and four different times. This means that we obtain "bogus" data from the exact solution at time t = 0.
Since our observation points are not necessarily grid points, later we will need to interpolate the model results to the observation points.
On the other hand, the observation times will always be taken at an integer number of model time steps and no interpolation in time will be needed.

2.3 Data for verification

Twenty additional values at five different points and four times are generated in the same way as the observations. But these data will not be assimilated. Our aim is to compare the effect of the assimilation at

points which are not observation ones, to see if the impact of the assimilation process is as important there as at the observation points. Therefore these five new points are called "verification points".

2.4 Model counterparts

Since the observation and verification points are not necessarily located at grid ones, an interpolation procedure is needed to compare observed and modelled data at the same place. Through this procedure, one obtains the "model counterparts" which can be considered as the result of the model at the observation or verification point.
Here we take as model counterpart the result of the bilinear interpolation of modelled data in the four closest grid points at the place of an observation or verification point.
In this way, if the κ^{th} observation at point (x_K, y_K) is taken into account, its model counterpart will be a weighted sum of the modelled data in the four closest grid points to it,
(i_K, j_K), (i_K+1, j_K), (i_K, j_K+1) and (i_K+1, j_K+1), (see also Fig. 3), as follows:

$$m_K^\tau = w_{1_K} F^\tau_{i_K j_K} + w_{2_K} F^\tau_{i_K+1\ j_K} + w_{3_K} F^\tau_{i_K\ j_K+1} + w_{4_K} F^\tau_{i_K+1\ j_K+1} \quad (6)$$

where the weights have the following expressions:

$$w_{1_K} = \frac{[(i_K+1)\Delta x - x_K]\ [(j_K+1)\Delta y - y_K]}{\Delta x \Delta y}$$

$$w_{2_K} = \frac{(x_K - i_K \Delta x)\ [(j_K+1)\ \Delta y - y_K]}{\Delta x \Delta y}$$

$$w_{3_K} = \frac{[(i_K+1)\Delta x - x_K]\ (y_K - j_K \Delta y)}{\Delta x \Delta y} \quad (7)$$

$$w_{4_K} = \frac{(x_K - i_K \Delta x)\ (y_K - j_K \Delta y)}{\Delta x \Delta y}$$

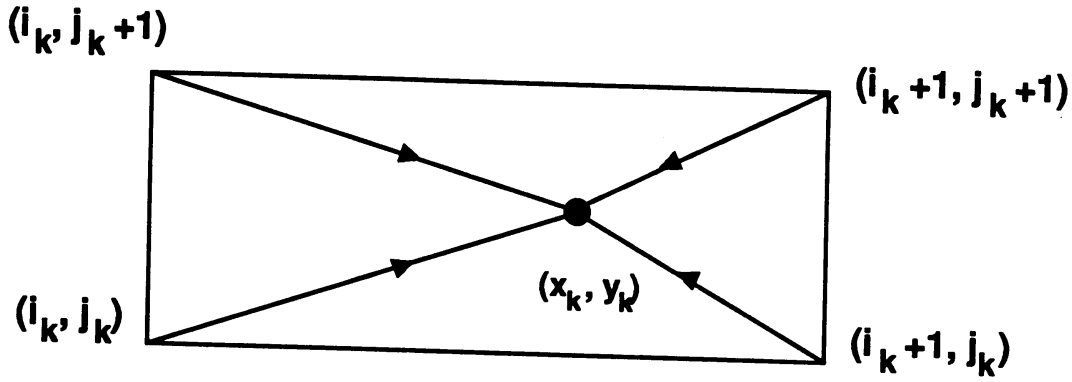


fig. 4

3. Data assimilation

The objective of the data assimilation is to find a solution to the model equations which is the best compromise between the data and the "a priori" knowledge. At an intermediate stage, this implies: that the misfit between model analyses and available observations should be as small as possible. But the final objective for forecasters is to minimize the difference between model predictions and future observations.

This is accomplished by minimizing the so called "cost function" J. This function has two parts:

$$J = J_1 + J_2 \tag{8}$$

The first one, J_1, is taken to be a quadratic function of the differences between modelled and observed data, which means that it is proportional to the sum of the squares of those differences, and it can be written so:

$$J_1 = \frac{1}{2} (m-d)^T A^{-1} (m-d) \tag{9}$$

where:

d is the vector of the observations
m is the vector of the model counterparts
A is the matrix of the error covariance of the observations.

$$A^{-1} = (\bar{a}_{\mu,\nu}) \tag{10}$$

where $\bar{a}_{\mu,\nu}$ is the covalidity between the μ^{th} and the ν^{th} observations. In our case, d and m have 20 components, and A is a 20 x 20 matrix. Assuming that all the observations are uncorrelated and of equal accuracy, we can write:

$$A^{-1} = \begin{Bmatrix} 1/\sigma^2 & & 0 \\ & 1/\sigma^2 & \\ 0 & & \ddots 1/\sigma^2 \end{Bmatrix}_{20x20} \tag{11}$$

where σ^2 is the variance of the measurement errors. Then the first component of the cost function can be written in the following way:

$$J_1 = \frac{1}{2\sigma^2} \sum_{k=1}^{5} \sum_{\tau=0,3,6,9} (m_k^\tau - d_k^\tau)^2 \qquad (12)$$

where k labels the position in space and τ denotes the time of the observations.

But noticing that we have:

- 5x4 = 20 "observed" data: d_k^τ,
- 20x20 = 400 independent variables: F_{ij}^0 (t=0),
- 20x20xN = 400 x N dependent variables: F_{ij}^t, $t \neq 0$, where N is the number of model time steps in the assimilation cycle,

the following problem arises: we have <u>less data than independent variables.</u>

In order to solve this, we take a guess for the independent variables:

$$G_{ij} \quad i = 1,\ldots, 20, \; j = 1,\ldots, 20, \; t = 0 \qquad (13)$$

which is the "best guess" before data assimilation, and we try to minimize the difference between G_{ij}^0 and F_{ij}^0 as well.

For that, the second part of J, J_2, is considered in the same way as the first one:

$$J_2 = \frac{1}{2\sigma_*^2} \sum_{i=1}^{20} \sum_{j=1}^{20} (F_{ij}^0 - G_{ij})^2 \qquad (14)$$

where, in this case, σ_*^2 is related to the model error. So J_2 is proportional to the sum of the squares of the differences between F_{ij}^0 and G_{ij}^0.

Now, we are interested in looking for the minimum of

$$J\left[F^o_{ij},\ F^t_{ij}\ (F^o_{ij})\right] = J_1 + J_2 \qquad \begin{matrix} i = 1,\ldots,\ 20 \\ j = 1,\ldots,\ 20 \\ t = 1,\ldots,\ N \end{matrix} \qquad (15)$$

We differentiate with respect to the independent variables, F^o_{ij}. But then, another complication has to be considered: the model counterparts are explicit functions of the dependent variables, but <u>implicit functions of the independent ones.</u>

$$m^\tau_k = m^\tau_k\ (F^t_{ij}) \qquad \begin{matrix} \tau = 0,3,6,9 \\ i = 1,\ldots,\ 20 \\ j = 1,\ldots,\ 20 \end{matrix} \qquad (16)$$

Therefore, we would get very complicated expressions if we would differentiate J with respect to the independent variables, F^o_{ij}, and for more complicated models (e.g. the North Sea wave forecasting model (Nedwam)), it is almost impossible to find analytical expressions for the model counterparts in terms of these initial variables. On the other hand, for most models (like Nedwam), finding analytical expressions for the model counterparts in terms of the dependent variables F^t_{ij} is not very difficult.

In order to solve this problem, the <u>"Lagrange multipliers method"</u> is used. We construct the so called "Lagrange function" adding a term to the cost function, as follows:

$$L(F^o_{ij},\ F^t_{ij},\ \lambda^t_{ij}) = J + \sum_{t=1}^{N}\ \sum_{i=1}^{20}\ \sum_{j=1}^{20} \lambda^t_{ij}\ E^t_{ij} \qquad (17)$$

The term we append is a sum for every point in space and time, of the product of an undetermined Lagrange multiplier, λ^t_{ij}, and the left hand side of the model equations.

$$E^t_{ij} = F^t_{ij} - (1-\alpha_x-\alpha_y)\ F^{t-1}_{ij} + \alpha_x\ F^{t-1}_{i-\sigma_x\ j} + \alpha_y\ F^{t-1}_{i\ j-\sigma_y} \qquad (18)$$

$E^t_{ij} = 0 \equiv$ model equations.

In this way, the complete explicit expression of the Lagrange function is the following:

$$L = \frac{1}{2\sigma^2} \sum_{k=1}^{5} \sum_{\tau=0,3,6,9} [m_k^\tau(F_{ij}^\tau) - d_k^\tau]^2 + \frac{1}{2\sigma_*^2} \sum_{i=1}^{20} \sum_{j=1}^{20} (F_{ij}^0 - G_{ij})^2 \qquad (19)$$

$$+ \sum_{t=1}^{N} \sum_{i=1}^{20} \sum_{j=1}^{20} \lambda_{ij}^t [F_{ij}^t - (1-\alpha_x-\alpha_y)F_{ij}^{t-1} + \alpha_x F_{i-\sigma_x\ j}^{t-1} + \alpha_y F_{i\ j-\sigma_y}^{t-1}]$$

Now, the minimum of J as a function of F_{ij}^0 can be found by looking for the saddle point of L, as the projection of this saddle point on the space of the independent variables (F_{ij}^0), is equal to the minimum of J on the manifold where the model equations are satisfied (see Appendix). This stationary point of L is found by differentiating L with respect to all the variables: λ_{ij}^t, (i = 1,..., 20, j = 1,..., 20, t = 1,..., N) and $F_{i,j}^t$ (i = 1,..., 20, j = 1,... 20, t = 0,... N), all of them treated as independent, and setting the derivatives equal to zero.

So, we get two sets of equations:

$$\frac{\partial L}{\partial \lambda_{ij}^t} = 0 = E_{ij}^t \qquad \text{model equations} \qquad (20)$$
$$i=1,...,20,\ j=1,...,20,\ t=1,...N$$

$$\frac{\partial L}{\partial F_{ij}^t} = 0 \qquad \text{adjoint equations} \qquad (21)$$
$$i=1,...,20,\ j=1,...,20,\ t=0,...N$$

The first set of equations reproduces again the model ones and the second constitutes the so called "adjoint equations". (Hence our method is called "adjoint method").

Now, the procedure to find the minimum of J is as follows:

1. First of all, we consider a first guess for the initial values of the wave variances:

$$F_{ij}^0 = H_{ij} \qquad i = 1,..., 20$$
$$j = 1,..., 20$$

Usually, we take H_{ij} equal to G_{ij}, the "best guess" before data assimilation.

2. Solving (20) forwards in time, the value of F^t_{ij} for every other $t \neq 0$, $t = 1,\ldots, N$ can be calculated.

$$\frac{\partial L}{\partial \lambda^t_{ij}} = E^t_{ij} = 0 \iff F^t_{ij} = (1-\alpha_x-\alpha_y)\, F^{t-1}_{ij} + \alpha_x F^{t-1}_{i-\sigma_x\, j} + \alpha_y F^{t-1}_{i\, j-\sigma_y}$$

$i = 1,\ldots, 20$

$j = 1,\ldots, 20$

$t = 1,\ldots, N$

3. Next, the adjoint equations (21) are solved backwards in time, for $i = 1,\ldots, 20$, $j = 1,\ldots, 20$ and $t = N,\ldots, 1$, and so the values of λ^t_{ij} ($i = 1,\ldots, 20$, $j = 1,\ldots, 20$, $t = N,\ldots, \underline{1}$) are calculated (there are no λ^o_{ij} because E^o_{ij} do not exist).

In order to do this, we need an explicit expression for $\frac{\partial L}{\partial F^t_{ij}}$. From the whole expression of L, (19), and of the model counterparts, (6) and (7), the derivative can be written as follows:

$$\frac{\partial L}{\partial F^t_{ij}} = \frac{1}{\sigma^2} \sum_{k=1}^{5} \sum_{\tau=0,3,6,9} (m^\tau_k - d^\tau_k) \frac{\partial m^\tau_k}{\partial F^t_{ij}} + \tag{22}$$

$$\lambda^t_{ij} - (1-\alpha_x-\alpha_y)\, \lambda^{t+1}_{ij} + \alpha_x\, \lambda^{t+1}_{i-\sigma_x\, j} + \alpha_y\, \lambda^{t+1}_{i\, j-\sigma_y}$$

where:

$$\frac{\partial m^\tau_k}{\partial F^t_{ij}} = [w_{1_k}\, \delta_{ii_k} \delta_{jj_k} + w_{2_k}\, \delta_{i\, i_k+1}\, \delta_{jj_k} + \tag{23}$$

$$w_{3_k}\, \delta_{ii_k}\, \delta_{j\, j_k+1} + w_{4_k}\, \delta_{i\, i_k+1}\, \delta_{j\, j_k+1}]\, \delta_{t\tau}$$

To simplify we call b^t_{ij} the whole first term of the expression of $\partial L/\partial F^t_{ij}$, which is known. So we can write:

$$\frac{\partial L}{\partial F^t_{ij}} = b^t_{ij} + \lambda^t_{ij} - (1-\alpha_x-\alpha_y)\,\lambda^{t+1}_{ij} + \alpha_x\,\lambda^{t+1}_{i-\sigma_x\ j} + \alpha_y\,\lambda^{t+1}_{i\ j-\sigma_y} = 0 \qquad (24)$$

$i = 1,\ldots, 20$

$j = 1,\ldots, 20$

$t = N,\ldots, 1$

As we know the values of F^t_{ij} $i = 1,\ldots, 20$, $j = 1,\ldots, 20$, $t = 0,\ldots,N$ from the solution of (20), we can calculate the values of m^τ_k, $\kappa = 1,\ldots, 5$, $\tau = 0,3,6,9$, and so the values of b^t_{ij} $i = 1,\ldots, 20$, $j = 1,\ldots, 20$, $t = 1,\ldots, N$.
Noticing that $\lambda^{N+1} = 0$ (it does not exist), we can solve the set of equations (24) backwards in time:

$$b^N_{ij} + \lambda^N_{ij} = 0$$

$$b^{N-1}_{ij} + \lambda^{N-1}_{ij} - (1-\alpha_x-\alpha_y)\,\lambda^N_{ij} + \alpha_x\,\lambda^N_{i-\sigma_x\ j} + \alpha_y\,\lambda^N_{i\ j-\sigma_y} = 0$$

$$\vdots \qquad (25)$$

$$b^1_{ij} + \lambda^1_{ij} - (1-\alpha_x-\alpha_y)\,\lambda^2_{ij} + \alpha_x\,\lambda^2_{i-\sigma_x\ j} + \alpha_y\,\lambda^2_{i\ j-\sigma_y} = 0$$

in this way we get the values for λ^t_{ij}, $i = 1,\ldots, 20$, $j = 1,\ldots, 20$, $t = N,\ldots, 1$:

$$\lambda^N_{ij} = -b^N_{ij}$$

$$\lambda^t_{ij} = (1-\alpha_x-\alpha_y)\,\lambda^{t+1}_{ij} - \alpha_x\,\lambda^{t+1}_{i-\sigma_x\ j} - \alpha_y\,\lambda^{t+1}_{i\ j-\sigma_y} - b^t_{ij} \qquad (26)$$

$i = 1,\ldots,20$, $j = 1,\ldots, 20$, $t = N-1,\ldots,1$

Note that also in the general case of a non-linear model, L (19) is linear in the λ's, hence the adjoint equations (21) are always linear in the λ's as well, and it is never a problem to solve for the λ's, even for a non-linear model.

4. Now, we have obtained values for all our variables:
λ^t_{ij} ($i = 1,\ldots, 20$, $j = 1,\ldots, 20$, $t = 1,\ldots, N$) and F^t_{ij} ($i = 1,\ldots, 20$, $j = 1,\ldots, 20$, $t = 0,\ldots, N$)
But we still have the subset of equations (21) concerning to $t = 0$, which are most probably not satified because of our guessing of the F^0_{ij} ($i = 1,\ldots, 20$, $j = 1,\ldots, 20$). But we can calculate the value of $\partial L/\partial F^0_{ij}$ ($i = 1,\ldots, 20$, $j = 1,\ldots, 20$) and it can be shown that this value equals the gradient of J, since both the adjoint and model equations are satisfied:

$$\operatorname{grad}_{F^0_{ij}} J = \frac{\partial J}{\partial F^0_{ij}} = \frac{\partial L}{\partial F^0_{ij}} \Bigg|_{\substack{E^t_{ij} = 0 \\ \frac{\partial L}{\partial F^t_{ij}} = 0}} \quad \begin{matrix} i = 1,\ldots, 20 \\ j = 1,\ldots, 20 \\ \\ i = 1,\ldots,20 \\ j = 1,\ldots,20 \\ t = 1,\ldots,N \end{matrix} \qquad (27)$$

and from (24) it follows that these derivatives have the following expression:

$$\operatorname{grad}_{F^0_{ij}} J = \frac{\partial L}{\partial F^0_{ij}} \Bigg|_{\substack{\text{mod.eq.}=0 \\ \text{adj.eq.}=0}} = b^0_{ij} - (1-\alpha_x-\alpha_y)\,\lambda^1_{ij} + \alpha_x\,\lambda^1_{i-\sigma_x\ j} + \alpha_y\,\lambda^1_{i\ j-\sigma_y} \qquad \begin{matrix} i = 1,\ldots, 20 \\ j = 1,\ldots, 20 \end{matrix} \qquad (28)$$

Note that λ^0_{ij} does not exist (see (19)) and that one has to add another term to b^0_{ij}, which comes from J_2 (14), and b^0_{ij} has now the following form:

$$b^0_{ij} = \frac{1}{\sigma^2} \sum_{k=1}^{5} (m^0_k - d^0_k) \frac{\partial m^0_k}{\partial F^0_{ij}} + \frac{1}{\sigma_*^2} (F^0_{ij} - G^0_{ij}) \qquad (29)$$

5. The next step is to evaluate if the gradient of J is small enough or not.

5.1 If it is, we consider our F^o_{ij} (i = 1,..., 20, j = 1,..., 20) as the "best guess" possible for the independent variables, and F^t_{ij} (i = 1,..., 20, j = 1,..., 20, t = 0,..., N) as the "best fit" to the observations, for the model values in the grid.

5.2 If it is not small enough, a "new guess" for F^o_{ij} has to be considered, which will make F^t_{ij} fit better to the observations.

To make this "new guess" we follow the "method of steepest descent", which consists on adding a term to the "old guess", in the direction opposite of the gradient of J, in the following way:

$$F^o_{ij_{new}} = F^o_{ij_{old}} - \Delta^o_{ij} \cdot \frac{\partial L}{\partial F^o_{ij}} \tag{30}$$

where Δ^o_{ij} is the increment we add in the opposite direction of the gradient. That moves our guess down towards the minimum of J (see Figure 2).

But we are interested in the best choice for Δ^o_{ij}, so let us look for it. The cost function is a paraboloid, since it is quadratic in the independent variables.

A section of this paraboloid with a plane of constant cost gives an ellipse of constant cost, which can be projected to the space of the independent variables. One point of this ellipse is our first guess F^o_{ij} and we know the gradient $\nabla_{F^o_{ij}} J$ at that point as well, which is a vector perpendicular to the ellipse at the point F^o_{ij}.

The section of the cost function by the plane

$$F^o_{ij} + x.\nabla_{F^o_{ij}} J \tag{31}$$

is a parabola, as we can see in Figure 5.

In this way,

$$J\,(F^o_{ij} + x.\nabla_{F^o_{ij}} J) = ax^2 + bx + c = J_{s1}(x) \tag{32}$$

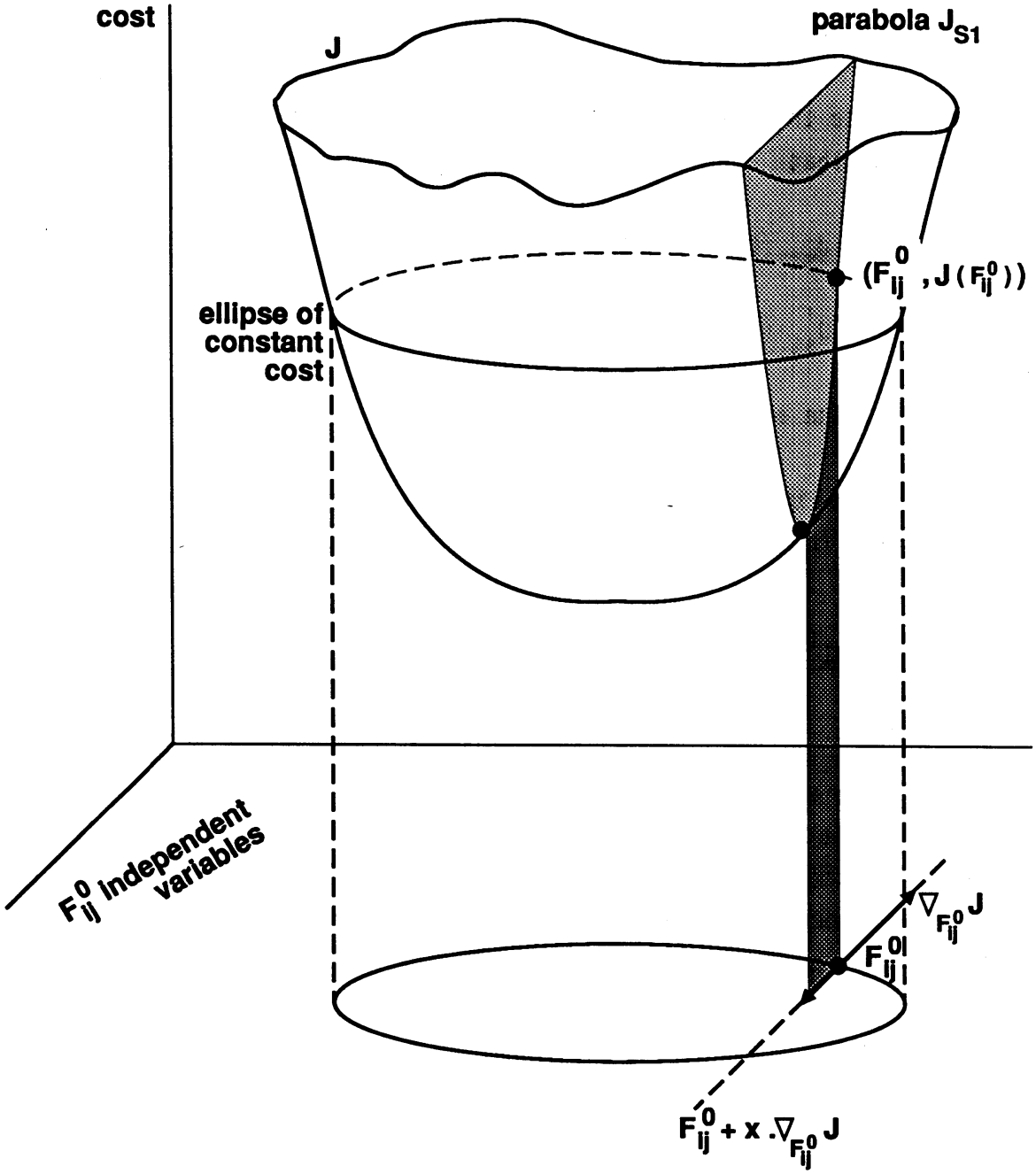

cost
J
parabola J_{S1}
$(F_{ij}^0, J(F_{ij}^0))$
ellipse of constant cost
F_{ij}^0 independent variables
$\nabla_{F_{ij}^0} J$
F_{ij}^0
$F_{ij}^0 + x \cdot \nabla_{F_{ij}^0} J$

fig. 5

Moving in the opposite direction of the gradient (in the space of the independent variables) implies going down in the parabola J_{s1} towards its minimum. Therefore the best would be to choose Δ^o_{ij} in a way that we would reach the projection of the vertex of J_{s1}.
J_{s1} has a minimum for $x = -b/2a$, therefore the best choice for Δ^o_{ij} is:

$$\Delta^o_{ij} = b/2a. \tag{33}$$

To calculate this value $\frac{b}{2a}$, we now proceed as follows:
as we know that

- $J_{s1}(0) = J(F^o_{ij}) = c$
- $J'_{s1}(0) = \nabla_{F^o_{ij}} J = b$

only the value of a is still unknown.

To calculate it, any other point x_1 can be used, and as we can always calculate $J(F^o_{ij} + x_1 . \nabla_{F^o_{ij}} J) = J_{s1}(x_1)$, the value of a is given by:

$$a = \frac{J_{s1}(x_1) - bx_1 - c}{x_1^2} \tag{34}$$

Now, that we are at the minimum of the parabola J_{s1}, we project it again in the space of the independent variables and we have then another ellipse of constant cost (of less cost than before) and so we can repeat the same procedure until we reach the absolute minimum of J. We will have another parabola J_{s2}, etc. (see Figure 6).

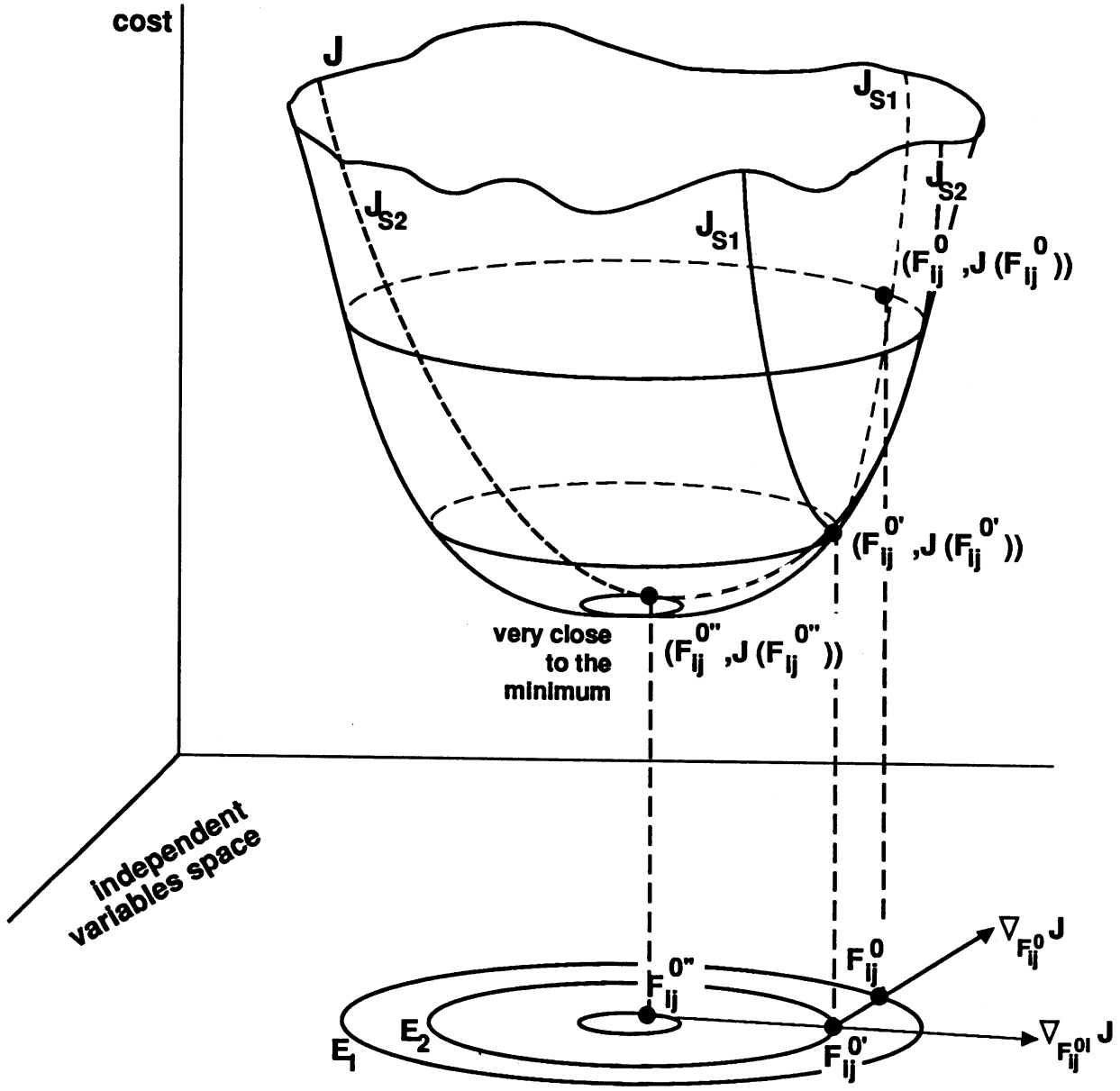

cost
J
J_{S1}
J_{S2}
J_{S2}
J_{S1}
$(F_{ij}^{0}, J(F_{ij}^{0}))$
$(F_{ij}^{0'}, J(F_{ij}^{0'}))$
very close to the minimum
$(F_{ij}^{0''}, J(F_{ij}^{0''}))$
independent variables space
$\nabla_{F_{ij}^{0}} J$
F_{ij}^{0}
$F_{ij}^{0''}$
E_2
E_1
$F_{ij}^{0'}$
$\nabla_{F_{ij}^{0'}} J$

fig. 6

Now, we go back again to step 2. and iterate the process in this way until convergence in the absolute minimum of J (see Figure 7).

This method of steepest descent is not optimal. There exist variations, which have a better convergence rate (de Valk, 1989). Note also that the trick of calculating b/2a is only valid for linear models, while most models are non-linear.
There can be more complications if the model is non-linear, due to the multiple minima that the cost function may have. Then if we choose the first guess too far away from the global minimum we are looking for, we could converge to a local one, in which we are not interested.
Trying sereval first guesses, H_{ij}, the probability of failing to find the absolute minimum is smaller.

We can summarize the procedure to find the minimum of the cost function J in the following scheme:

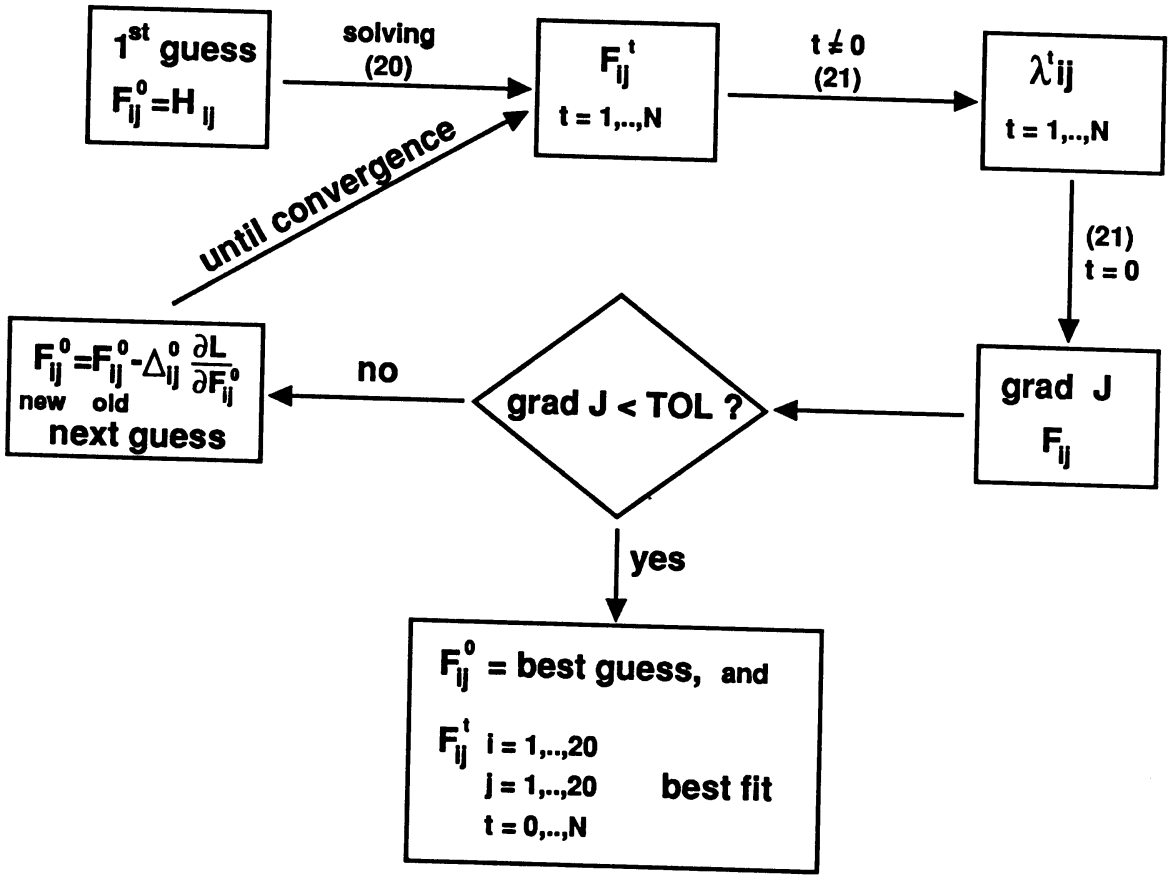


fig. 7

4. Assimilation cycles

What has been explained until now constitutes only one cycle of assimilation. But one can iterate the process so long as one likes. In that way, one would deal with more cycles of assimilation (see Fig. 4). If we choose a 9 hours assimilation window, in a model with a 9 hours analysis and a 18 hours forecast period, and if we have observations every 3 hours, we can assimilate the data available at 0h., 3h., 6h. and 9 h. With these data we can calculate a cost function, minimize it and produce a "best"guess for the independent variables which are the values of the wave variances at time 0h. From the improved field at 0h., improved analysed fields for 3h., 6h. and 9h., and improved forecasted fields for 12h., 15h. and 18h. are also obtained.
Next, as data for 12h. have arrived, we can assimilate the data at 3h., 6h., 9h. and 12 hours, using the "old" analysis for 3 hours as the "new" first guesses for 3h., and obtain a "best guess" for the wave variances at 3h.
In this way, model fitting could be used operationally.

TIME SCHEME

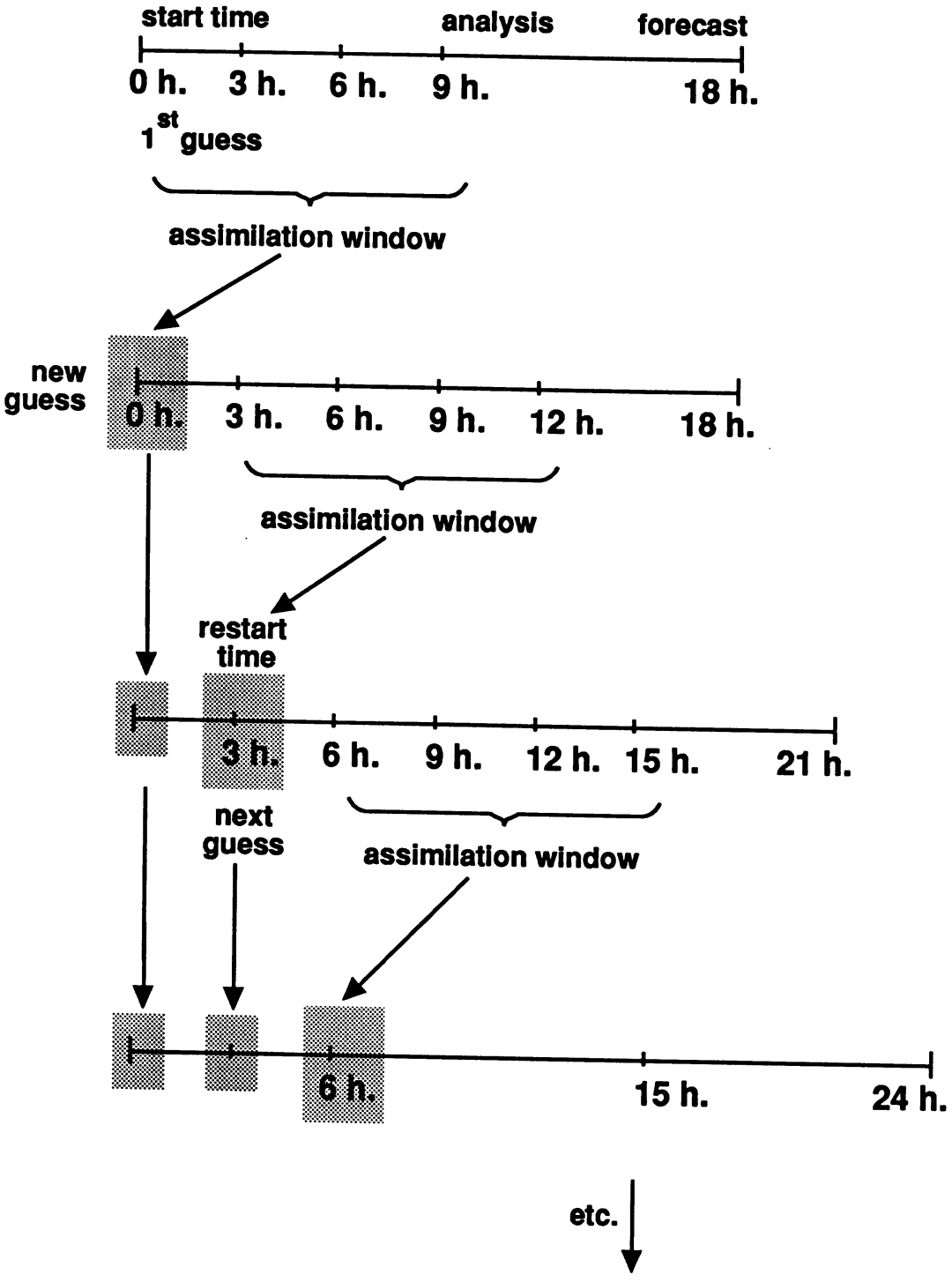


fig. 8

5. Results and conclusions

5.1 Graphs

In the Figures 8, 9, 10 and 11 (graphs) we can see a comparison of the results of the process before and after data assimilation.
The crosses are measured data, the squares are model results without data assimilation and the circles model result after assimilation.
This comparison is shown at some observation and verification points.

5.2 Table

The table shows the root mean square error between the measurements and the modelled data at the observation and verification points, before and after data assimilation, at every three hours.

TABLE 1.

	rms.error before assimilation		rms.error after assimilation	
time	obs.-mod.	ver.-mod.	obs.-mod.	ver.-mod.
0 h.	0.	0.	0.023	0.101
3 h.	0.444	0.125	0.076	0.068
6 h.	0.333	0.413	0.050	0.499
9 h.	0.346	0.140	0.099	0.221
12 h.	0.266	0.200	0.264	0.232
15 h.	0.691	0.619	0.609	0.655
18 h.	0.397	0.800	0.396	0.816

	before assimilation	after assimilation
mean error	0.341	0.293

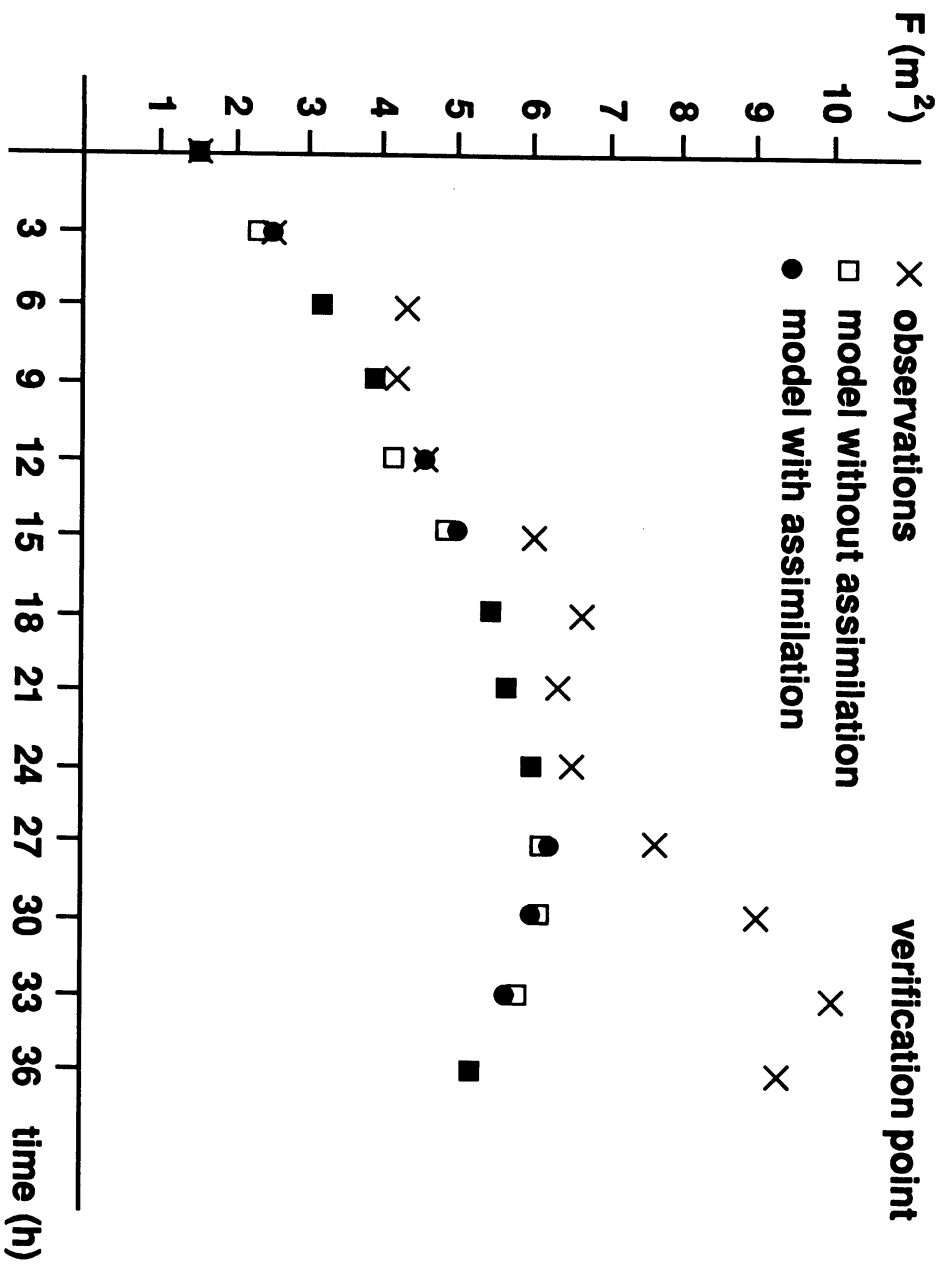

F (m²)
1
2
3
4
5
6
7
8
9
10
3
6
9
12
15
18
21
24
27
30
33
36
time (h)
× observations
□ model without assimilation
● model with assimilation
verification point

fig. 9

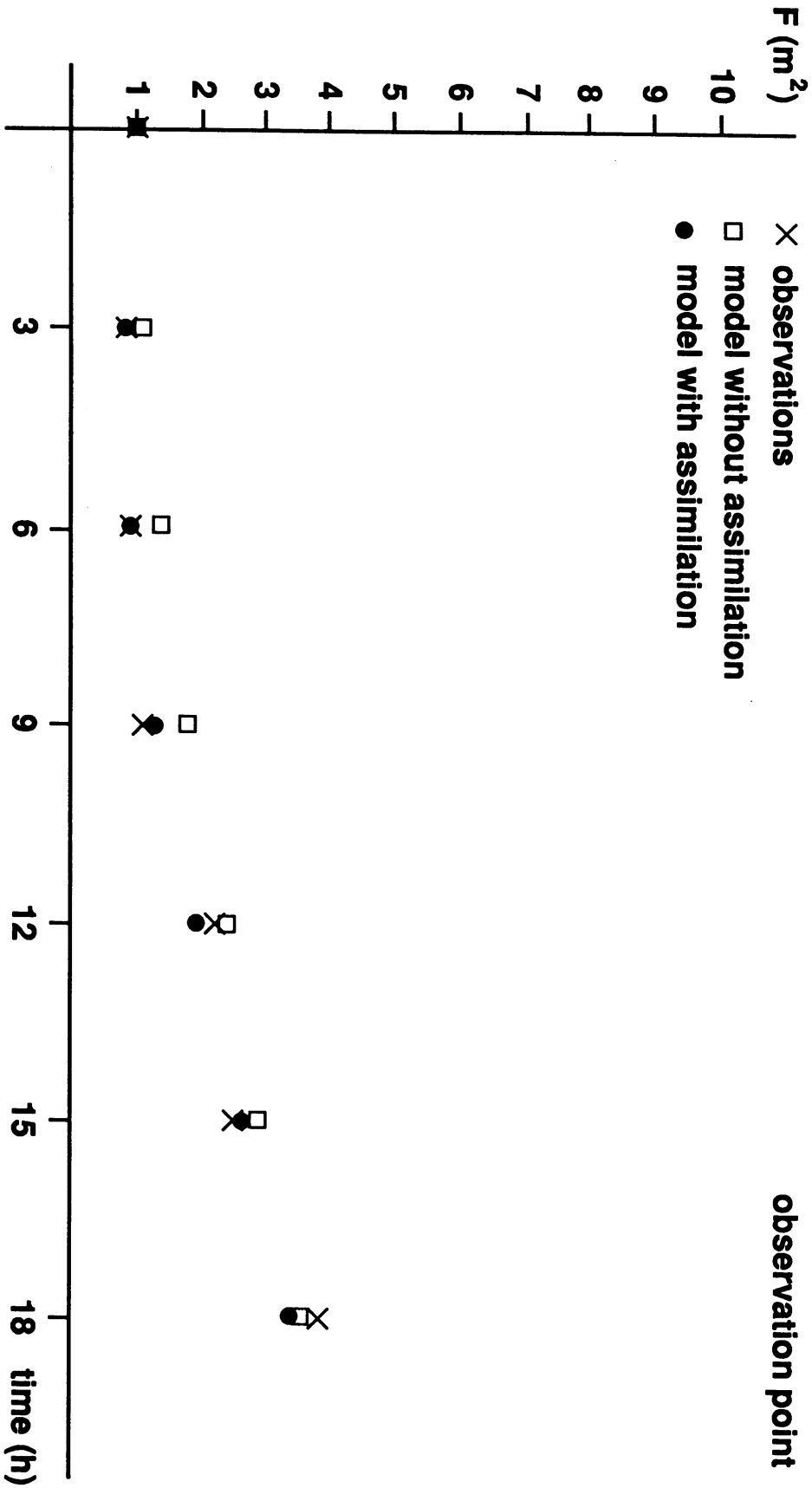

F (m²)
1
2
3
4
5
6
7
8
9
10
× observations
□ model without assimilation
● model with assimilation
observation point
3
6
9
12
15
18
time (h)

fig. 10

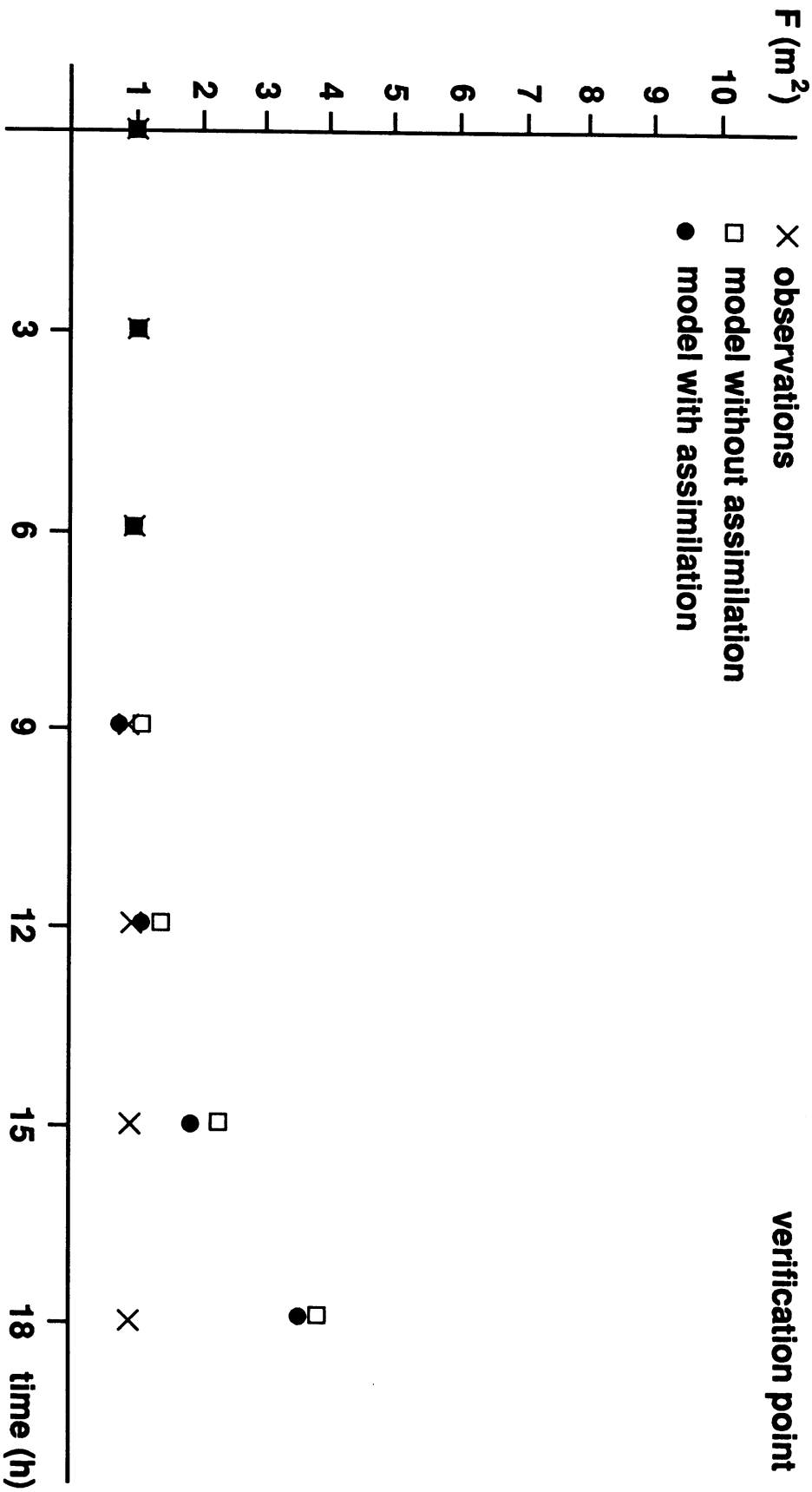

F (m²)
10
9
8
7
6
5
4
3
2
1
× observations
□ model without assimilation
● model with assimilation
verification point
3
6
9
12
15
18
time (h)

fig. 11

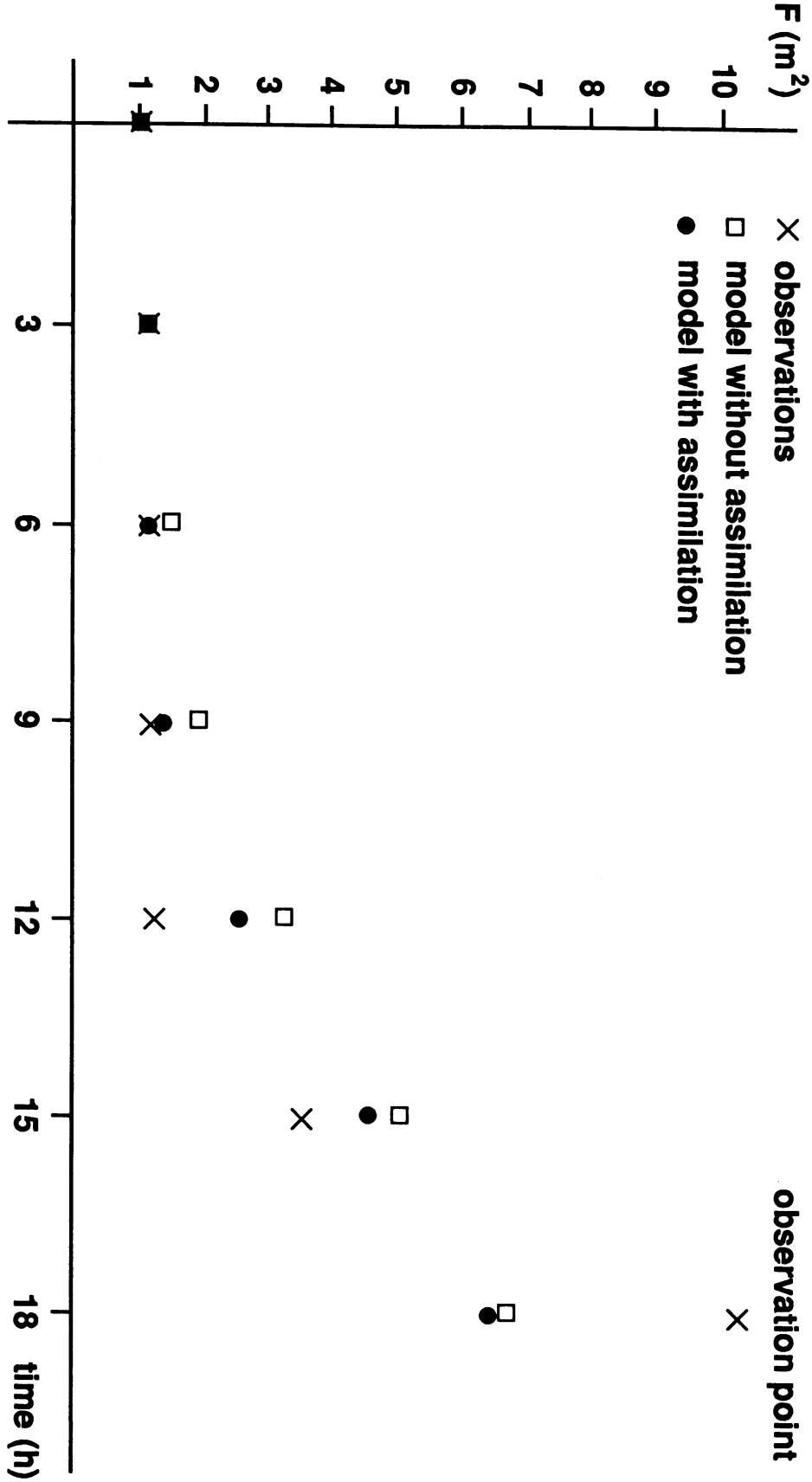

F (m²)
1
2
3
4
5
6
7
8
9
10
3
6
9
12
15
18
time (h)
observations
model without assimilation
model with assimilation
observation point

fig. 12

5.3 Conclusions

The analysis of the results, graphs and tables, leads to the following conclusions:

1. The "correction" due to the assimilation at observation points is larger than at verification ones.
   The effect of the assimilation is noticed more clearly around the observation points than elsewhere.

2. The average of the root mean squared errors between measurements and modelled data after assimilation is smaller than before, which implies that the data assimilation process improves the wave field, producing a new one which fits better the observations and the "a priori" knowledge we took into account.

3. The effect of the assimilation does not remain very long after stopping assimilating. This suggestes that some improvements in the system have to be made, if we want this effect to remain longer in time.

Acknowledgements.

One of the authors (M.H.) is financially supported by the Programa de Clima Maritimo (PCM), Madrid.
We thank Greet de Graaf for typing carefully the manuscript.

References

Thacker, W.C. 1987 Three lectures on fitting numerical models to observations. GKSS-Forschungszentrum Geesthacht GmbH, GKSS 87/E/65, Geesthacht.

Ghil, M. 1989 Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. Dyn.At.Oc. 13, (1989) p 171-218.

Long, R.B. 1989 Notes on assimilating observations into numerical models. Report Z274, Delft Hydraulics.

Schilperoort, T. 1989 Talk delivered at the VII WAM-meeting, May 25-28, 1989, Valencia.

Valk C.F. de 1989 Wave data assimilation in a third generation wave prediction model for the North Sea. An optimal control approach. Report H676, Delft Hydraulics.

Appendix 1.

Relation between the extrema of the cost and Lagrange functions.

The cost function J is a function of the wave variances at every time:

$$J = J\,(F^{o}_{ij},\ F^{t}_{ij}) \tag{35}$$

Since this cost function J is positive definite, it must have a minimum. On the other hand, L is a function of the wave variances and the Lagrange multipliers:

$$L = L\,(F^{o}_{ij},\ F^{t}_{ij},\ \lambda) \tag{36}$$

Both functions are related in the following way:

$$L\,(F^{o}_{ij},\ F^{t}_{ij},\ \lambda) = J\,(F^{o}_{ij},\ F^{t}_{ij}) + \lambda E\,(F^{o}_{ij},\ F^{t}_{ij}) \tag{37}$$

where $E\,(F^{o}_{ij},\ F^{t}_{ij}) = 0$ are the model equations.

Since L is odd in λ, if L has an extremum, it can not be a maximum or a minimum, it must be a saddle point.

Part 1.1.

Let us assume that L has an extremum at $(\bar{F}^{o}_{ij},\ \bar{F}^{t}_{ij},\ \bar{\lambda})$.

In this case, all the derivatives of L must vanish at that point. Thus,

$$\frac{\partial L}{\partial \lambda}\,(\bar{F}^{o}_{ij},\ \bar{F}^{t}_{ij},\ \bar{\lambda}) = 0 \tag{38}$$

and therefore

$$E\,(\bar{F}^{o}_{ij},\ \bar{F}^{t}_{ij}) = 0 \tag{39}$$

which implies that every extremum of L must lie on the manifold where the model equations are satisfied.
But for every point (F^o_{ij}, F^t_{ij}) on that manifold, (37) is reduced to:

$$L\ (F^o_{ij}, F^t_{ij}, \lambda) = J\ (F^o_{ij}, F^t_{ij}) \tag{40}$$

This, if L has an extremum at $(\bar{F}^o_{ij}, \bar{F}^t_{ij}, \bar{\lambda})$, J, as a function of F^o_{ij} and F^t_{ij}, where (F^o_{ij}, F^t_{ij}) must lie on the manifold $E(F^o_{ij}, F^t_{ij}) = 0$, has an extremum at $(\bar{F}^o_{ij}, \bar{F}^t_{ij})$.

This shows that if L has a saddle point at $(\bar{F}^o_{ij}, \bar{F}^t_{ij}, \bar{\lambda})$, then:

1.1.1 - $E(\bar{F}^o_{ij}, \bar{F}^t_{ij}) = 0$, and

1.1.2 - $J\ (\bar{F}^o_{ij}, \bar{F}^t_{ij})$ is a minimum of J on the manifold $E\ (F^o_{ij}, F^t_{ij}) = 0$

Part 1.2.

Let us now prove that if such an extremum of J exists, then L must have a saddle point (whose projection is that extremum (minimum) of J).

The vector space of the gradient of L with respect to the F^t_{ij} in a point on the manifold $E\ (F^o_{ij}, F^t_{ij}) = 0$, can be split in two parts: one of them parallel to that manifold, and another one normal to it, i.e.:

$$\vec{\nabla}_{F^t_{ij}} L = (\vec{\nabla}_{F^t_{ij}} L)_{par.} + (\vec{\nabla}_{F^t_{ij}} L)_{nor.} \tag{41}$$

In this way, the gradient of L consists of three parts:

$$(\vec{\nabla}_{F^t_{ij}} L)_{par.}\ ,\quad (\vec{\nabla}_{F^t_{ij}} L)_{nor.}\ \text{and}\quad (\vec{\nabla}_{\lambda} L). \tag{42}$$

Assuming that in the space $E\ (F^o_{ij}, F^t_{ij}) = o$, J attains its minimum

at $(\bar{F}^o_{ij}, \bar{F}^t_{ij})$, which satifies (39), and since

$$\vec{\nabla}_\lambda L = E\,(\bar{F}^o_{ij}, \bar{F}^t_{ij}) \tag{43}$$

we have

$$\vec{\nabla}_\lambda L = \vec{0}. \tag{44}$$

(37) also implies that:

$$(\vec{\nabla}_{F_{ij}} L)_{par.} = (\vec{\nabla}_{F_{ij}} J)_{par.} \tag{45}$$

because $\vec{\nabla}E$ is normal to E. Therefore, since $(\vec{\nabla}_{\bar{F}^t_{ij}} J)_{par.}$ vanish by assumption, also

$$(\vec{\nabla}_{\bar{F}^t_{ij}} L) = \vec{0}. \tag{46}$$

In this way, from (44) and (46), one has:

$$\begin{aligned} &(\vec{\nabla}_{F^t_{ij}} L)_{par.}(\bar{F}^o_{ij}, \bar{F}^t_{ij}, \lambda) = \vec{0} \\ &\vec{\nabla}_\lambda L\,(\bar{F}^o_{ij}, \bar{F}^t_{ij}, \lambda) = \vec{0}. \end{aligned} \tag{47}$$

for every λ.

Hence, λ can be adjusted in such a way that also

$$(\vec{\nabla}_{F^t_{ij}} L)_{nor.}\,(\bar{F}^o_{ij}, \bar{F}^t_{ij}, \lambda) = \vec{0}. \tag{48}$$

since:

$$(\vec{\nabla}_{F^t_{ij}} L)_{nor.}(\bar{F}^o_{ij}, \bar{F}^t_{ij}, \lambda) = (\vec{\nabla}_{F^t_{ij}} J)_{nor.}(\bar{F}^o_{ij}, \bar{F}^t_{ij}) + \lambda\, \vec{\nabla}_{F^t_{ij}} E(\bar{F}^o_{ij}, \bar{F}^t_{ij}) \tag{49}$$

and since both λ and $(\vec{\nabla}_{F^t_{ij}} J)_{nor.}$ have the same dimension, and $\vec{\nabla}_{F^t_{ij}} E$ span the whole vector space of normal gradients. Then, from (48) and (49), one can find a $\bar{\lambda}$ which satifies:

$$(\vec{\nabla}_{F^t_{ij}})_{nor.}(\bar{F}^o_{ij},\bar{F}^t_{ij}) + \bar{\lambda}\,\vec{\nabla}_{F^t_{ij}} E\,(\bar{F}^o_{ij},\,\bar{F}^t_{ij}) = \vec{0}. \tag{50}$$

Thus we have proved that:

if there exist a $(\bar{F}^o_{ij},\,\bar{F}^t_{ij})$ with the following property:

$$J\,(\bar{F}^o_{ij},\,\bar{F}^t_{ij}) = \min\,[\,J(F^o_{ij},\,F^t_{ij}) \;\mid\; E(F^o_{ij},\,F^t_{ij}) = 0\,] \tag{51}$$

then there is a $\bar{\lambda}$ which satisfies:

$$dL\,(\bar{F}^o_{ij},\,\bar{F}^t_{ij},\,\bar{\lambda}) = 0.$$

Appendix 2.

The possibility of adding a third term to the cost function J, J_3, was also considered. The aim of this third part of J was to keep smoothness in the field of the wave variances at initial time t = 0. The expression for this J_3 is, then, as follows:

$$J_3 = \frac{\alpha}{2} \sum_{i=1}^{20} \sum_{j=1}^{20} [F^0_{i+1\ j} + F^0_{i-1\ j} + F^0_{i\ j+1} + F^0_{i\ j-1} - 4\ F^0_{ij}]^2 \qquad (52)$$

where α represents a measure for the smoothness of the initial field of wave variances.

In this way, the term which has to be added to the expression of the gradient of J is the following:

$$\frac{\partial J}{\partial F^0_{ij}}_{new} = \frac{\partial J}{\partial F^0_{ij}}_{old} + \alpha \left[(F^0_{i\ j+2} + F^0_{i\ j-2} + F^0_{i+2\ j} + F^0_{i-2\ j}) \right.$$

$$+ 2\ (F^0_{i+1\ j+1} + F^0_{i+1\ j-1} + F^0_{i-1\ j+1} + F^0_{i-1\ j-1}) \qquad (53)$$

$$\left. - 8\ (F^0_{i\ j+1} + F^0_{i\ j-1} + F^0_{i+1\ j} + F^0_{i-1\ j}) + 20\ F^0_{ij} \right]$$

The results with this new cost function

$$J = J_1 + J_2 + J_3 \qquad (54)$$

are similar to the ones we obtained with the former one (8) (see Table 2).

Looking at the Table 2, we realize that the effect of the assimilation is, in this case, more remarkable at later times than at the beginning, for verification points.
This implies that the impact of the assimilation remains longer, when this new cost function is considered.

TABLE 2.

time	rms.error before assimilation obs.-mod.	rms.error before assimilation ver.-mod.	rms.error after assimilation obs.-mod.	rms.error after assimilation ver.-mod.
0 h.	0.	0.	0.048	0.662
3 h.	0.878	0.383	0.103	0.395
6 h.	1.260	1.563	0.236	1.392
9 h.	1.866	0.995	0.460	1.857
12 h.	0.734	2.393	0.782	2.251
15 h.	2.085	0.954	2.000	0.763
18 h.	1.543	2.638	1.117	2.386