UCL Université catholique de Louvain

École polytechnique de Louvain (EPL)

Particle filters using sparse observations: applications in palaeoclimatology

Dissertation presented by
François Delcourt

for obtaining the master's degree in
Mathematical Engineering

Supervisors
Pierre-Antoine Absil, Hugues Goosse

Readerss
Pierre-Antoine Absil, Michel Crucifix, Hugues Goosse,

Academic year 2016-2017

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgium **www.uclouvain.be/epl**



Acknowledgements

I wish to express my sincere gratitude to my supervisors Pierre-Antoine Absil and Hugues Goosse for their continued support and the opportunity to work on this interesting subject.
I sincerely thank François Klein for his help with the off-line assimilation method and for letting me use some of his scripts.
I also thank Pierre-Yves Barriat for all the time and efforts he has put in helping me, especially during the implementation of the online experiment.

François Delcourt

Contents

Acknowledgement . . . 1
Introduction . . . 5

1 Filters for sequential data assimilation **7**
1.1 The Kalman filter (KF) . . . 8
1.2 The ensemble Kalman filter (EnKF) . . . 8
1.3 The sequential importance resampling particles filter (SIR) . . . 11
1.4 The non-linear ensemble transform filter (NETF) . . . 13
1.5 The equivalent weight particles filter (EWPF) . . . 14

2 Test on Lorenz 63 **17**
2.1 The Lorenz 63 system . . . 17
2.2 Criteria . . . 18
2.2.1 Root Mean Square Error (RMSE) . . . 18
2.2.2 Spread . . . 18
2.2.3 Correlation coefficient . . . 18
2.3 Experimental set up . . . 19
2.4 Results . . . 19
2.4.1 Ensemble evolution: EnKF . . . 23
2.4.2 Ensemble evolution: SIR . . . 23
2.4.3 Ensemble evolution: NETF . . . 23
2.4.4 Ensemble evolution: EW . . . 23
2.5 Impact of the ensemble size . . . 24
2.6 Impact of the interval between observations . . . 24
2.7 Selecting the method for larger scale applications . . . 25
2.7.1 Impact of the inflation parameter . . . 25

3 Offline assimilation **33**
3.1 The LOVECLIM model . . . 33
3.2 Method . . . 34
3.3 Results . . . 35

4 Online assimilation **40**
4.1 Experimental set up . . . 40
4.2 Results with 10 particles . . . 41
4.3 Results with 100 particles . . . 41

5 Conclusion **50**

A Filter's impact on prior pdf **52**

B Poor nudging scheme **57**

Introduction

Climatology is an active field of Earth science devoted to the understanding of physical mechanisms driving the climate. Confronted to risks arising from climate changes, a good understanding of the climate system is more crucial than ever before. As real scale experiments are impossible to conduct, climate's history constitutes a major source of informations on the impact of climate on life on Earth and on the system dynamics. Palaeoclimatology, the study of past climates, try to produce accurate reconstructions and improve our knowledge of the climate system. Although obtaining good climate reconstructions is in itself interesting this also allows future improvements in climate modelization. Historical records of direct measurement only cover a small portion of possible climate states, therefore limiting the validity of model based on those data alone. Reconstructions of different periods should provide new informations to improve models and assess their validities under different conditions. Moreover those reconstruction allows to inspect models abilities to capture phenomenon with diverse frequencies.

When modelling a system two main approaches are possible. The descriptive approach relies directly on empirical observations through regression and/or interpolation techniques in order to represent the system behaviour. In this framework, the system is treated as an entity only described by the set of quantities of interest. This method uses available informations to provide useful results without providing any insight on internal mechanics at play in the system. In the explanatory approach, the system has an in depth internal representation so that leading phenomena can be described through a priori knowledge on relevant physical laws. This method doesn't explicitly use informations from observations but provides a better insight on internal mechanisms. As palaeoclimatology deals with past events, the two approaches are available. Climatic models have already been produced for other purposes and a variety of proxies preserve information on past climate. Proxies can take different forms, for instance the ratio of different oxygen isotopes captured in ice layers can be used to reconstruct past temperatures [16]. Many other proxies are available and bring more informations on different aspects of past climates. Unfortunately those observations tend to be sparse, as only a fraction of all variables are observed. Moreover, proxies may not register the instantaneous state of a variable but instead an average of the variable over some period of time.

The data assimilation problem consists in finding a way to use observations within a model to improve the quality of its output. Variational methods try to use observations by minimizing a cost function designed to penalize a distance between observations and predicted states as well as enforcing physical constraints. Although those methods may be useful, they require to construct and minimize a cost function based on a complex model which can constitute non-trivial tasks. However, stochastic methods for sequential assimilation can directly use an existing model with few modifications. The model of your choice can be applied to some initial conditions and stopped when the simulation reach a time when an observation is available. At this point the filter can correct the simulation output using the observation and the simulation may resume.

In this work we concentrate our efforts on methods derived from particle filters. Particle filters, are a family of sequential data assimilation methods that use multiple independent simulations in order to build Monte-Carlo approximation of probability distributions of the system state. The update step of these methods is designed to use available observation and enforce Bayes'theorem. Particle filters, in their simplest forms are easily implemented and can theoretically be used in palaeoclimatology despite the potentially non-linear and non-Gaussian

model. Unfortunately, the robustness and quality of data assimilation based on particle filters depend on the number of independent simulations used by the filter. As the model complexity increases, the affordable number of simulations decreases and particle filters become less and less robust. The following question naturally arises, how can we modify a particle filter to improve its robustness when the number of particles is low? The goal of this work is to implement a method to assimilate sparse observations into an existing climatic model such as LOVECLIM [14] and compare our result with an existing particle filter implementation. First we will provide theoretical description of some sequential data assimilation methods. This will require the introduction of different notions such as ensemble, particles, prior and posterior distributions. Then we will test some of these filters trough twin experiments on the Lorenz 63 model. Using results from those experiments, a particular method will be selected. This filter will then be tested off-line using an available simulation of the southern hemisphere climate in pre-industrial conditions made with the model LOVECLIM. A comparison with an existing off-line assimilation method will validate the implementation. At last, the selected method will be adapted for online assimilation with the same model and a short twin experiment will be performed. This should validate the online implementation and allow quantitative assessment of the method in a real scale application.

Chapter 1

Filters for sequential data assimilation

In this chapter we will present a selection of sequential data assimilation methods from a theoretical perspective. Although this is not an exhaustive list we hope that this chapter will at least be an educational introduction to state of the art methods. We will start from the well known Kalman's filter, designed for linear-Gaussian model, then the ensemble Kalman filter will generalize the method to non-linear models. This will require the introduction of the concept of ensemble. Then, starting from Bayes'theorem, we will describe the sequential importance resampling filter, a method adapted to generic data assimilation problems. This filter is one of the simplest particle filter available and a good starting point to develop more advance techniques. Using elements from the ensemble Kalman filter and the sequential importance filter we will explain how to construct a non-linear ensemble transform filter that share some similarities with both of them. At last a variation of the sequential importance resampling filter will be presented in the form of the equivalent weight particle filter.

First lets describe some useful notations. We are interested in a model with state space $\mathcal{X} \subseteq \mathbb{R}^m$. The model for our application works with discrete time so the model dynamic can be describe as $f : \mathcal{X} \to \mathcal{X} : x(t) \to x(t+1)$, where $x(t)$ represent the model state at time t. The relationship between an observation $y(t)$ taken on the true system and the system state at the corresponding time step $z(t)$ can be express as $h : \mathcal{X} \to \mathcal{Y} : z(t) \to y(t)$. Where $\mathcal{Y} \subseteq \mathbb{R}^d$ is the observations space. Unfortunately, observations are often noisy, for clarity we consider that the function h is deterministic and we add a random noise ϵ, a random vector of appropriate dimensions, $y(t) = h(z(t)) + \epsilon$. This expression will become useful to construct twin experiments for testing the different methods while taking observation noise into account.

Starting from an initial condition $x(0)$, we can apply the model several times in order to obtain a first estimation for the system state at time t. This estimation $x(t)$ is the *a priori estimate* or the *forecast*. Assuming that on observation $y(t)$ is available at time t, the goal of a data assimilation method is to produce a better estimation $x(t)|y(t)$ called *a posteriori estimate* For clarity we will note x^f for a forecast and x^a for an a posteriori estimate.

Unfortunately, uncertainties on initial conditions and noises on observations add some difficulties to the situation. The simple fact that initial conditions may not be clearly defined put us in a probabilistic framework. Instead of estimating the system state $x(t)$ we will try to estimate its probability distribution. Once again we will have to make the distinction between the a priori and a posteriori probability distribution using the same notation as before.

1.1 The Kalman filter (KF)

The Kalman filter is a well known method that provide optimal result in a closed form as long as a few hypotheses hold. Optimality of the Kalman filter resides in the minimization of the squared error between the value of the system state and the value of the forecast.

- the model is linear Gaussian
- measurements come from a linear operator with Gaussian noise
- distribution of the initial condition is Gaussian

We will present the KF in a similar way as in [23]. Using the linear model A and the linear observation operator H we can note,

$$x(t) = Ax(t-1) + \gamma \tag{1.1}$$
$$y(t) = Hx(t) + \epsilon \tag{1.2}$$
$$\text{with: } \gamma \sim N(0, U) \tag{1.3}$$
$$\epsilon \sim N(0, R) \tag{1.4}$$

In this formulation, U and R are covariance matrices of appropriate dimensions. ϵ and γ are random vectors drawn from the corresponding Gaussian distribution. Starting from a Gaussian state distribution at time $t-1$, $x \sim N(\bar{x}(t-1), P(t-1))$ applying the model lead to the following prediction step:

$$\bar{x}^f(t) = Ax(t-1) \tag{1.5}$$
$$P^f(t) = AP(t-1)A^T + U \tag{1.6}$$
$$x^f(t) \sim N(\bar{x}^f(t), P^f(t)) \tag{1.7}$$

In this framework the a posteriori distribution will also be Gaussian. Therefore the update step only has to modify the mean and covariance. This update step can be written concisely by omitting the time notation as everything happen in the same time t.

$$K = P^f H^T (HP^f H^T + R)^{-1} \tag{1.8}$$
$$\bar{x}^a = \bar{x}^f + K(y - H\bar{x}^f) \tag{1.9}$$
$$P^a = P^f - KHP^f \tag{1.10}$$
$$x^a \sim N(\bar{x}^a, P^a) \tag{1.11}$$

If the model appears to be linear and deterministic the Kalman filter is still valid as the covariance matrix U can be set to zero in ours equations.

Although this filter proposes a closed form for the data assimilation problem, the underlying hypothesis are far too restrictive for application in palaeoclimatology. Indeed, in this field models are rarely linear and state distributions are not known to be Gaussian.

1.2 The ensemble Kalman filter (EnKF)

In order to deal with non linearity present in model we can use the Ensemble Kalman Filter as describe in [9], [10]. Instead of applying the model on the mean of the initial distribution, we

can use multiple simulations starting from initial conditions drawn from the best state distribution known at that time. Because computational resources are limited we use a finite number n of simulations. The set of simulated states, x_i $i \in \{1, 2, ..., n\}$, is the ensemble and n is the ensemble size. The idea behind the EnKF is to use this ensemble to evaluate quantities needed in order to compute the Kalman's update.

Because the model is no longer linear the forecast step in the KF (equation 1.5) doesn't stand. Applying a linear transformation to a Gaussian distribution result in a Gaussian distribution with modified moments, this is not true with a non linear transformation. However, mean and covariance can be approximated by the empirical ensemble mean and covariance. Therefore, starting from an ensemble representing the distribution of initial conditions we can apply the model to propagate the ensemble forward in time. Unlike the Kalman filter this step requires to apply the model on several initial conditions, fortunately this can be done in parallel as each ensemble member is independent. As soon as all ensemble members have reached a time when an observation is available, the empirical mean and covariance of the ensemble can be computed as:

$$\bar{x}^f = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.12}$$

$$P^f = E((x - \bar{x})(x - \bar{x})^T) \tag{1.13}$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \tag{1.14}$$

Then, we can compute the Kalman's gain as in the original KF (equation 1.5). Keeping the assumption of Gaussian distribution for the prior and for the posterior distribution, the mean and covariance of the posterior can be approximated using the KF update.

$$K = P^f H^T (H P^f H^T + R)^{-1} \tag{1.15}$$

$$\bar{x}^a = \bar{x}^f + K(y - H\bar{x}^f) \tag{1.16}$$

$$P^a = P^f - KHP^f \tag{1.17}$$

Those equations assume that the underlying distribution is Gaussian and update the ensemble mean and covariance, therefore this is enough to describe the a posteriori distribution. However, in order to use the EnKF in the following step we need to generate a new ensemble. This can be done by applying the Kalman's update to each ensemble member. However, this formulation requires the use of perturbed observations so that the covariance of perturbations match the prescribed covariance R. This can be done by adding a Gaussian noise with mean equal to zero and covariance R [9]. For infinitely large ensemble the two covariances would be equals. In [5], [10], [26], discussion on variation of the update equation can be found.

The filter can be summarized using matricial notations. This kind of notations will be useful later on. Each ensemble member can be stored in a column of a matrix. For instance, x_i is stored in the ith column of the matrix X. The distinction between prior and posterior ensemble is made with the subscript f or a. The covariance matrix can be express using the ensemble perturbation matrix X'. The ith column of the perturbation matrix is $x_i - \bar{x}$. Therefore the covariance matrix will be $P = \frac{1}{n-1}X'X'^T$. In the same way, the set of perturbed observations can be stored in the column of the matrix Y. As usual, we apply the model on each ensemble member independently: $X^f(t) = [f(x_1^f(t-1)), f(x_2^f(t-1)), ..., f(x_n^f(t-1))]$ until we reach an observation time. The analysis step is written in matricial notation without the time indications

as everything append at an observation time.

$$X'^f = X^f - \bar{X}^f \tag{1.18}$$

$$P^f = \frac{1}{n-1} X'^f X^{fT} \tag{1.19}$$

$$K = P^f H^T (H P^f H^T + R)^{-1} \tag{1.20}$$

$$X^a = X^f + K(Y - H X^f) \tag{1.21}$$

Although this is enough to understand and apply the EnkF, another formulation will become helpful in section 1.4. The idea behind this reformulation is to express the Ensemble Kalman update in term of the perturbation matrix and not in term of the covariance matrix. This will require the computation of a matrix square root, origin of the appellation of ensemble square root filter [9], [10]. We known that covariance matrices can be written as:

$$P^a = \frac{1}{n-1} X'^a X'^{aT} \tag{1.22}$$

$$P^f = \frac{1}{n-1} X'^f X'^{fT} \tag{1.23}$$

Using those expression in the update equation of a Kalman filter we have:

$$P^a = P^f - P^f H^T (H P^f H^T + R)^{-1} H P^f \tag{1.24}$$

$$P^a = \frac{1}{n-1} X'^f X'^{fT} - \frac{1}{n-1} X'^f X'^{fT} H^T [H \frac{1}{n-1} X'^f X'^{fT} H^T + R]^{-1} H \frac{1}{n-1} X'^f X'^{fT} \tag{1.25}$$

$$\tag{1.26}$$

At this point we can factorise in term of prior and perturbation both left and right

$$P^a = X'^f V X'^{fT} \tag{1.27}$$

$$V = \left[(n-1)I + (H X'^f)^T R^{-1} (H X'^f)\right]^{-1} \tag{1.28}$$

Then using the symmetric matrix square root of V in order to preserve the ensemble mean we get:

$$V = T T^T \tag{1.29}$$

$$X'^a = \sqrt{n-1} X'^f T \tag{1.30}$$

This last equation can be used to update the ensemble perturbations however the ensemble means still have to be updated. This update can be done by applying the Kalman update to the the ensemble mean. The a posteriori ensemble is generated by centring the ensemble perturbations around this a posteriori ensemble mean.

Note that the transformation is defined up to a mean preserving rotation Λ. This rotation matrix must satisfy a few properties:

- the matrix must be orthogonal $\Lambda\Lambda^T = I$
- the matrix must have 1 as an eigenvalue and e as the associated eigenvector. $\Lambda e = e$

The first property ensures that the covariance matrix obtain is the same with or without rotation is the same. $(T\Lambda)(T\Lambda)^T = TT^T = V$ The second property ensures that the ensemble mean is not affected by the rotation.

From an implementation point of view, the generation of the random rotation can be done as explain in appendix B of [26]. First, generate a random orthogonal matrix Ω of size $(n-1)\times(n-1)$. Generate an arbitrary $n\times n$ orthogonal base B with $\frac{1}{\sqrt{n}}$ as first column. In the base B the random rotation could be written as $\Lambda_B = \begin{pmatrix} 1 & 0 \\ 0 & \Omega \end{pmatrix}$. At last the final rotation can be obtain as $\Lambda = B\Lambda_B B^T$. Random orthogonal matrix can be obtain as a householder transformation based on a random vector. Orthogonalisation of a random base can be achieve trough QR factorisation. Note that this rotation take place in the ensemble space and not in the state space of the model. Therefore, the generation of the rotation is model independent and only require to know the ensemble size.

Note that the EnKF doesn't need a linear model but still assume Gaussian distributions of model states by construction. In general states distribution is non-Gaussian, however this can be seen as a tentative to describe the mean states and the associated uncertainty. This limitation can be relaxed in the following filter.

1.3 The sequential importance resampling particles filter (SIR)

This filter uses explicitly Monte-Carlo approximations of different distributions as well as observation likelihood in order to enforce Bayes'theorem through an appropriate weighting of the prior ensemble.
The Monte-Carlo approximation of a probability distribution function $p(x)$ is build from an ensemble of particles x_i $i = 1...n$ drawn independently from this pdf

$$p(x) \approx \frac{1}{n}\sum_{i=1}^{n} \phi_i(x_i) \tag{1.31}$$

Where $\phi_i(x_i)$ is a known pdf computed thanks to the ith particle. Classical choices are normal and delta distributions centred in x_i. Choosing normal distributions give a continuous pdf but the choice of the covariance can be tricky. Choosing the delta distribution get rid of this issue but the approximation become a discrete distribution. For simplicity let's use delta distributions to build the approximation of the prior distribution.

$$p(x) \approx \frac{1}{n}\sum_{i=1}^{n} \delta(x - x_i) \tag{1.32}$$

For an ensemble of infinite size the approximation is perfect, obviously such an ensemble is not realistic but the greater the ensemble the better the approximation. We want to produce a similar approximation of the posterior distribution form this ensemble such that Bayes theorem is respected.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{1.33}$$

The distribution $p(x)$ can be approximated as previously. Assuming that the distribution of noises on observations is known we have $p(y|x)$. The last term $p(y)$ is not known directly, however it can be said that $p(y) = \int p(y|x)p(x)dx$

At that point, we can use Monte-Carlo approximations of $p(x)$ and write:

$$p(x|y) \approx \frac{p(y|x)\frac{1}{n}\sum_{i=1}^{n}\delta(x-x_i)}{\int p(y|x)\frac{1}{n}\sum_{i=1}^{n}\delta(x-x_i)dx} \tag{1.34}$$

$$p(x|y) \approx \frac{p(y|x_i)}{\sum_{i=1}^{n}p(y|x_i)}\delta(x-x_i) \tag{1.35}$$

This is often written as $p(x|y) \approx \sum_{i=1}^{n} w_i\delta(x-x_i)$ where $w_i = \frac{p(y|x_i)}{\sum_{i=1}^{n}p(y|x_i)}$ are called weights.

In practice, the first step is to compute the observation likelihood with respect to each particles of the prior ensemble. Then weights are obtained after normalisation of the likelihood vector.

Although it can be tempting to use this approximation of the posterior distribution and assume independence of each particles in order to reapply this method at the next observation time with $w_i(t) = w_i(t-1)\frac{p(y|x_i)}{\sum_{i=1}^{n}p(y|x_i)}$ this will cause filter degeneracy. Filter degeneracy describe a phenomenon where all the weigh is concentrated in a few particles. The model must be applied to each particle and is often costly. Particles with negligible weight have little impact on the approximation therefore propagating those particles is a waste of resources. However simply dismissing them will cause the ensemble size to decrease, asymptotically, only one particle will have a non-zero weight and the posterior distribution will be a delta function. At this point filter divergence is unavoidable.

The idea in the Sequential Importance Resampling Particle Filter is to apply a resampling step in order to avoid filter degeneracy. After computing the weigh vector, a new ensemble is drawn from the posterior distribution by resampling with replacement amongst the current ensemble such that the probability of particle i to be chosen is equal to w_i. Then all weights are reset to $\frac{1}{n}$. At this point we have a new ensemble where each particle is equally weighted. This new ensemble and the the ensemble from which we resample are two Monte Carlo approximation of the same distribution. Filter degeneracy is avoided and the ensemble size is preserved, leading to a better usage of computational resources. However another problem may arise, particularly if the model do not have stochastic component. Because some particles of the new ensemble may be identical computational resources are wasted and the filter may diverge. An easy-fix solution is to add a small perturbation at each particles after resampling. It slightly perturbs the model dynamic but reintroduces diversity in the ensemble.

This filter doesn't require heavy hypothesis on the model or observations, it can be applied to a non linear model with any kind of prior distribution. Moreover, it can produce an estimation of nearly any distribution, even multi-modal as long as the ensemble size is large enough. However, in case of filter divergence there is no mechanism to recover. This fact encourage the use of ensemble as large as possible. Another drawback of this method is the stochastic noise introduce at the resampling step. A stochastic resampling can sometime misbehave, justifying deterministic resampling scheme.

1.4 The non-linear ensemble transform filter (NETF)

The filter described in [26], [27] can be seen as the equivalent of a square root formulation of the SIR. The formulation is similar to the square root formulation of the EnKF but it tries to reproduce the mean and ensemble covariance obtain with a SIR. Unfortunately it doesn't maintain higher order moments.
First let's describe the mean and covariance produce by a SIR. The mean can be computed as the weighted sum of each particles.

$$\bar{x}^a_{SIR} = \sum_{i=1}^{n} w_i x_i \tag{1.36}$$

$$\tag{1.37}$$

The covariance matrix can also be computed explicitly but it is slightly more complex. The first step is to start from the expression of the posterior mean in term of prior perturbation:

$$\bar{x}^a_{SIR} = X'^f w + \bar{x}^f w \tag{1.38}$$

$$P^a_{SIR} = \sum_{i=1}^{n} w_i (x^f_i - \bar{x}^a_{SIR})(x^f_i - \bar{x}^a_{SIR})^T \tag{1.39}$$

$$= \sum_{i=1}^{n} w_i [(x^f_i - \bar{x}^f) - X'w][(x^f_i - \bar{x}^f) - X'w]^T \tag{1.40}$$

$$= \sum_{i=1}^{n} w_i (x^f_i - \bar{x}^f)(x^f_i - \bar{x}^f)^T \tag{1.41}$$

$$- \sum_{i=1}^{n} w_i (x^f_i - \bar{x}^f) w^T X'^{fT} \tag{1.42}$$

$$- \sum_{i=1}^{n} w_i X'^f w (x^f_i - \bar{x}_f)^T \tag{1.43}$$

$$+ \sum_{i=1}^{n} w_i X'^f w w^T X'^{fT} \tag{1.44}$$

This expression can be written more clearly with matrices. The first term become $X'^f W X'^{fT}$ where W is a diagonal matrix where the element (i,i) is the weight of the ith particle, $W = diag(w)$. Because the sum of all weights must be equal to 1 we can write the last term as $X'^f w w^T X'^{fT}$. Second and third terms have similar notation with a minus sign. Therefore the covariance of the posterior is equal to:

$$P^a_{SIR} = \frac{n}{n-1} X'^f (W - w w^T) X'^{fT} \tag{1.45}$$

This expression allow to reformulate the particle filter as a square root filter where the posterior ensemble is generated through the application of a transformation on the perturbation matrix.

$$T = \sqrt{W - w w^T} \tag{1.46}$$

$$X'^a = \sqrt{n} X'^f T \tag{1.47}$$

In practice, after computing the weight vector w, we can compute the prior ensemble mean and the prior perturbation matrix. Then, using the weights we can compute the posterior mean and the posterior perturbation matrix. At last the two can be combined in order to obtain the

posterior ensemble.

Note that the matrix square root is not unique, therefore a random mean preserving rotation can be applied to generate a new transformation with the same properties. Such a rotation in the ensemble space is a $n \times n$ matrix Λ that verify the following properties:

$$\Lambda\Lambda^T = I = \Lambda^T\Lambda \tag{1.48}$$

$$\Lambda e = e \tag{1.49}$$

$$\tag{1.50}$$

Where e is a vector with all entries equal to one. This filter produce a posterior ensemble with the same mean and covariance as if a sequential importance resampling particles filter would have been apply. The new ensemble is equally weighted by construction. However, it doesn't guarantee anything for higher order moments. A method to construct the random rotation can be found in section 1.2.

In [26], the authors claim that this filter with the random rotation shows better stability than the SIR and contrarily to the EnKF no assumption on the prior distribution is made and the observation error doesn't have to be Gaussian as long as the likelihood can be computed. This square root formulation also help to apply ensemble inflation scheme design to artificially increase the ensemble spread to reduce the risk of filter degeneracy. This is done by multiplying the perturbation matrix by a scalar $\gamma > 1$ [26].

1.5 The equivalent weight particles filter (EWPF)

Another approach based on the notion of proposal density try to reduce both the risk of filter divergence and filter degeneracy encounter in SIR by perturbing the model dynamic while respecting Bayes'theorem. The Equivalent weight particles filter proposed by P.J. van Leeuwen and describe in [28], [2], [3] is a good example of such a method and deserve to be summarize here.

The idea is to propagate each particles independently with a modified model that will pull the particle toward the next observation. This is done by adding to the original model a nudging term, function of the observation and the particle position in the observation space. The nudging term can take different forms and is an important tuning element of this method, a classical choice is to add to the model a term inspired from the Kalman update as illustrated later in this section. In order to respect Bayes theorem, weights are updated to take into account the model relaxation. The modified model can be used to propagate the ensemble until the time step just before an available observation. The last step is design to decrease the final weight variance and therefore improve the behaviour of the resampling step.

First let's describe the model relaxation and the associated proposal density. The initial model can be represented as a conditional probability distribution $p(x(t)|x(t-1))$ either because the model has stochastic components or to represent modelisation errors. A modification of the model would therefore be translated as another pdf based on the next observation y, $q(x(t)|x(t-1)y)$. The weight modification require by Bayes' theorem can be express as in [28]

$$p(x(t)|y) = \frac{p(y|x(t))p(x(t)|x(t-1))}{p(y)} p(x(t-1)) \tag{1.51}$$

Integrating the past out of this formulation and multiplying the expression by $\frac{q(x(t)|yx(t-1))}{q(x(t)|yx(t-1))}$ give

$$p(x|y) = \frac{p(y|x(t))}{p(y)} \int p(x(t)|x(t-1)p(x(t-1)) \frac{q(x(t)|yx(t-1))}{q(x(t)|yx(t-1))} dx(t-1) \tag{1.52}$$

This operation requires the support of q to include the support of p, and not only to be greater as it has been unfortunately express in [28]. Using the particle representation of this expression we obtain

$$p(x|t) = \sum_{i=1}^{n} \frac{1}{n} \frac{p(y|x(t))}{p(y)} \frac{p(x(t)|x(t-1))}{q(x(t)|x(t-1)y)} q(x(t)|x(t-1)y) \tag{1.53}$$

Therefore using the relaxed model lead to another weighting method to describe the pdf obtain by the modified model with the particle ensemble.

$$p(x(t)|y) = \sum_{i=1}^{n} \frac{1}{n} \frac{p(y|x(t))}{p(y)} \frac{p(x(t)|x(t-1))}{q(x(t)|x(t-1)y)} \delta(x_i - x) \tag{1.54}$$

As previously explained, the unknown $p(y)$ result is a normalisation of the likelihood across the ensemble.
Although in this particular formulation we only consider one time step this can easily be extended to several successive time steps and the final weight expression would become

$$w_i = p(y|x(t)) \prod_{j=1}^{n} \frac{p(x(j)|x(j-1))}{q(x(j)|x(j-1)y)} \tag{1.55}$$

We now have to design the different proposal densities.
To pull particles toward the observation we need to construct a general formulation for the proposal densities used to reach the time just before observation. The nudging scheme can be related to the EnKF if the Gaussian distribution generated by a local EnKF is used as proposal. The resulting model apply to each particle take the following form.

$$x(t) = f(x(t-1)) + K(y - H * x(t-1)) \tag{1.56}$$

However, any kind of relaxation can be applied as long as the condition on the support of the proposal density is respected. Another kind of relaxation encounter in [29] is

$$x(t) = f(x(t-1-)) + \tau Q H^T R^{-1}(y - Hx(t-1)) \tag{1.57}$$

Where τ depends on the time between the current step and the next observation time. This formulation doesn't require a local computation of the EnKF and allow a control of the relaxation strength. The idea being to increase the relaxation strength as we get closer to the observation time using $\tau = \frac{t_{obs} - t}{\Delta t_{obs}}$

Other nudging schemes use conditions on the model natural dynamic ability to moves particles toward observations in order to better control the relaxation strength [1]. Although those variations seem intuitively interesting the increasing complexity of those schemes leave too much room to parameter tuning to constitute general method. Moreover this could affect the dynamic in an unforeseen way.

At last, the proposal density for the last time step is designed such that a fixed percentage of particle obtain a similar weight, thus decreasing weight variance and improving the resampling

step. Assuming that particle i has a accumulated a weight w_i^{rest} we want to choose a specific state $x_i(t)$ for this particle such that $w_i^{rest} p(y|x_i(t)p(x_i(t)|x_i(t-1))$ is set to a specific value. If the observation noise is Gaussian and we settle for a last step based on a Kalman filter-like expression for the last proposal $x(t) = f(x(t-1)) + \alpha_i K(y - Hf(x_i(t-1))$ we have to find the appropriate value of α_i. As shown in [1] we can use

$$\alpha_i = 1 + \sqrt{1 - b_i/a_i} \tag{1.58}$$

$$a_i = 0.5 d_i^T R^{-1} H K d_i \tag{1.59}$$

$$b_i = 0.5 d_i^T R^{-1} d_i - C - log(w_i^{rest}) \tag{1.60}$$

$$K = Q H^T (HQH^T + R)^{-1} \tag{1.61}$$

$$d_i = y - Hf(x_i(t-1)) \tag{1.62}$$

The value of C must be chosen so that a fixed percentage of particle may reach the corresponding weight. Therefore, for each particle compute the maximal possible weight at the last step

$$w^{target} = -log(w_i) = -log(w_i^{rest}) + \frac{1}{2} d_i^T (HQH^T + R)^{-1} d_i \tag{1.63}$$

Setting C to a percentile of this set of value enforce that the corresponding proportion of particles may reach the target weight. If a particle can not reach this value simply apply the initial model. At last a resampling step close the algorithm. As most particles have the same weight their are few risks of filter degeneracy.

Chapter 2

Test on Lorenz 63

In order to test different filter as honestly as possible we want to use a twin experiment set-up in which we assimilate pseudo-observations with different methods. In order to construct a twin experiment, we first have to choose a model. Starting from arbitrary initial conditions we can integrate the model on a given time-interval, this gave us a simulation, often called the truth, from which we can generate pseudo-observations as we wish. Starting from initial conditions similar to those used to generate the truth we can apply a data assimilation method using the same model and the set of observations previously generated. In a twin experiment we have access to the system trajectory from which assimilated data come from and the model dynamic is exactly the same as the system dynamic. This allows us to evaluate the assimilation methods performances and compare them with each others. At the end of this chapter we want to select an assimilation method for large scale application.

2.1 The Lorenz 63 system

A classical model for twin experiments on data assimilation techniques is the Lorenz 63 system. This model, proposed by Edward Lorenz in 1963 describe an atmospheric convection phenomenon:

$$\begin{cases} \dfrac{dx}{dt} = \sigma(y - x) \\ \dfrac{dy}{dt} = \rho x - y - xz \\ \dfrac{dz}{dt} = xy - \beta z \end{cases} \tag{2.1}$$

The physical interpretation of this system is not really relevant for our experimental set up. However the fact that this model present a chaotic behaviour for the following choice of parameters, $(\sigma = 10, \rho = 28\beta = 8/3)$, is interesting. Indeed, using a trivial model such as $\dfrac{dx}{dt} = -x$ would not be useful as any error would vanish.
Several parameters can be modified in order better understand filters behaviour. Because we concentrate on particle filters and ensemble filter the ensemble size is obviously an interesting parameter. We want to identify filter that behave properly even with a small number of particles. What constitute a "small number" depend on the dimension of the model. With only three state variables the Lorenz 63 model should require less particles than a model with thousands of state variables, all other things being equals. Another parameter that can be interesting is the time interval between observations as it could lead to filter divergence as the amount of available information decrease.

2.2 Criteria

We need objective criteria to compare different filters. Because we are working with Monte-Carlo approximations of probability distribution we need criteria that can capture the filter performance without being impacted by the nature of the approximation.

2.2.1 Root Mean Square Error (RMSE)

The root means square error between a true trajectory $f(t)$ and an estimation $g(t)$ over a time interval $[0, t]$ can be define as:

$$RMSE_{[0,T]} = \sqrt{\frac{1}{T}\int_0^T (f(t) - g(t))^2 dt} \tag{2.2}$$

Although this could be computed for each particle and averaged over the ensemble, this criteria would become sensible to outliers. Therefore we will evaluate the RMSE between the true trajectory and the average trajectory of the ensemble. Multidimensional extension is natural using euclidean distance.

$$RMSE_{[0,T]} = \sqrt{\frac{1}{T}\sum_t ||f(t) - \bar{g}(t)||^2} \tag{2.3}$$

This criterion capture the quality of the ensemble mean.

2.2.2 Spread

In order to measure the uncertainty in the ensemble, we compute the spread from the empirical ensemble covariance. With P^a the covariance matrix after assimilation and d the dimension of the state space.

$$SPREAD = \sqrt{\text{trace}(P^a)/d)} \tag{2.4}$$

Although this constitute an objective measurement of the ensemble spread it is not easy to capture the trade-off between under and over-confidence. An ensemble with a large spread is less susceptible to filter degeneracy but bring us fewer informations. In the other hand an ensemble that collapse into a too small region bring more informations but is more at risk to diverge.

2.2.3 Correlation coefficient

Given the truth and the reconstruction time series computing the correlation coefficient give us a simple measurement of the ability to reconstruct the truth run. For two time series X,Y the correlation matrix is define as

$$\rho = \frac{COV(X, Y)}{\sqrt{\sigma(X)^2\sigma(Y)^2}} \tag{2.5}$$

Element outside the main diagonal are the correlation coefficients between the corresponding data.

2.3 Experimental set up

We use the Lorenz 63 model with standard parameters. We integrate the model using Euler explicit with a time step of $dt = 0.01$. Simulation durations are set to $T = 5$ An observation is taken every 10 time steps. The observation is build as the first variable with a Gaussian noise $\epsilon \sim N(0, 2)$. Initial conditions for the "truth" generation are $(2, 8, -2)$ and the initial ensemble is drawn from a Gaussian distribution around the same conditions. The size of the ensemble is set to 15. See table 2.1. Data assimilation methods to be tested in our twins experiment are the ensemble Kalman filter (EnKF) the sequential importance resampling filter (SIR), the non-linear ensemble transform filter (NETF) and the equivalent weight particle filter (EW) moreover we also simulate ensemble without any assimilation methods (FREE) in order to have a baseline for comparison.

T	dt	dt_{obs}	H	R	n
5	0.01	$10dt$	$(1, 0, 0)$	2	15

Table 2.1: Set up for a standard twin experiment

2.4 Results

Twin experiments are stochastic process as we generate random noises for each pseudo-observation, for this reason, multiple twin experiments can give different results especially for chaotic system such as Lorenz 63. Therefore, we will do multiple twin experiments using same parameters in order to compute an average value for ours criteria. However, we want a reliable assimilation method especially for large scale applications. Therefore we also have an interest in the variability of results for the same experiment.

After 50 twin experiments using the parameters of table 2.1, we can compute the average of our criteria, these results can be found in table 2.2. One way to understand the variability of

	EnKF	SIR	NETF	EW	Free
RMSE	4.07	10.8	3.47	6.32	14.5
SPREAD	0.95	0.96	1.23	0.98	7.72
corr(x)	0.96	0.67	0.94	0.95	0.27
corr(y)	0.93	0.66	0.94	0.91	0.36
corr(z)	0.94	0.74	0.97	0.88	0.47

Table 2.2: Average results obtained for the standard twins experiment.

those results is to generate boxplot of the RMSE for each methods. In figure 2.1 we can see that without assimilation method the RMSE consistently take a high value while with assimilation the rmse may be lower. The large variability of the SIR can be explain by the filter inability to recover from filter divergence. In Best case scenario the SIR perform reasonably well, however the method lake of robustness. The Equivalent Weight particle filter is a clear improvement on the SIR, both in term of robustness and in average, even when the parametrization is sub-optimal. The NETF and the ENKF outperform all other methods. Another representation of those result can be found in figure 2.2 where correlation according to the variable x is confronted to the RMSE. With all assimilation methods, there exists a subset of experiments in which observations are successfully assimilated, this correspond to dots in the upper left corner of the figure 2.2. In those experiments the correlation is high and the rmse is small

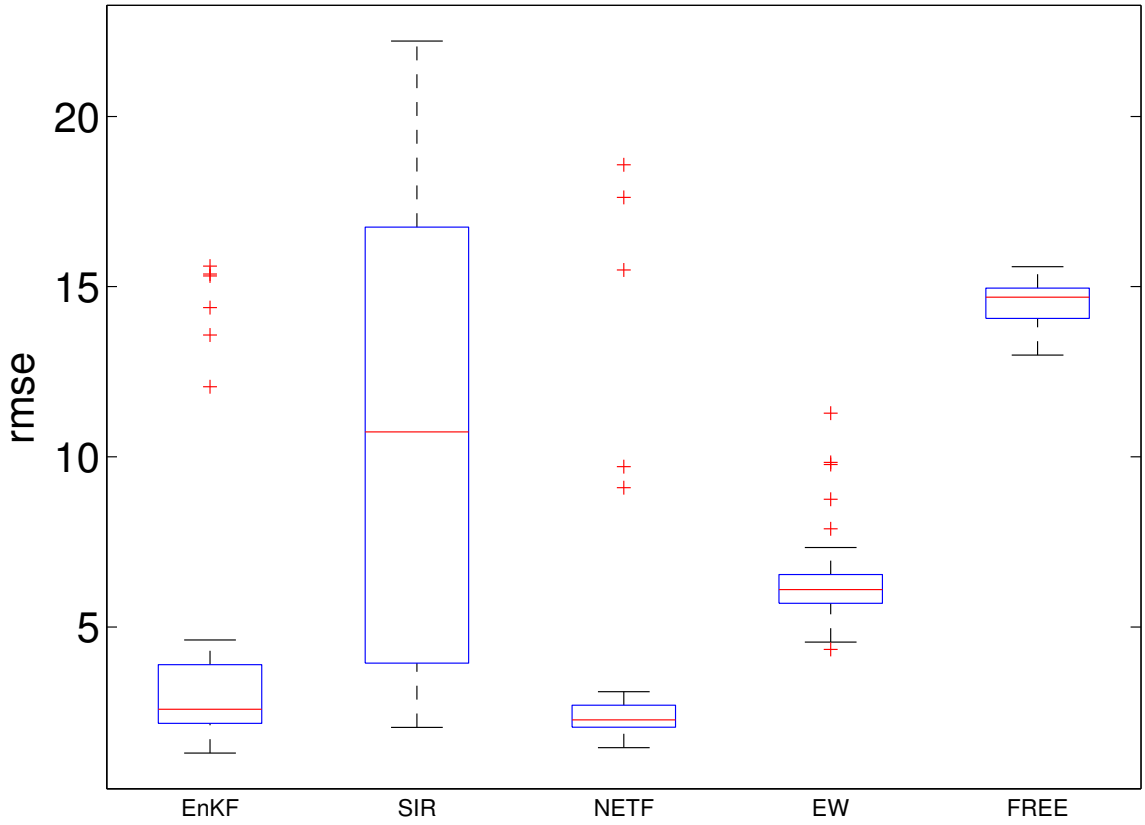


Figure 2.1: Boxplot of the RMSE obtain after 50 twin experiments with the standard parameters. From left to right: EnKF, SIR, NETF, EW, FREE

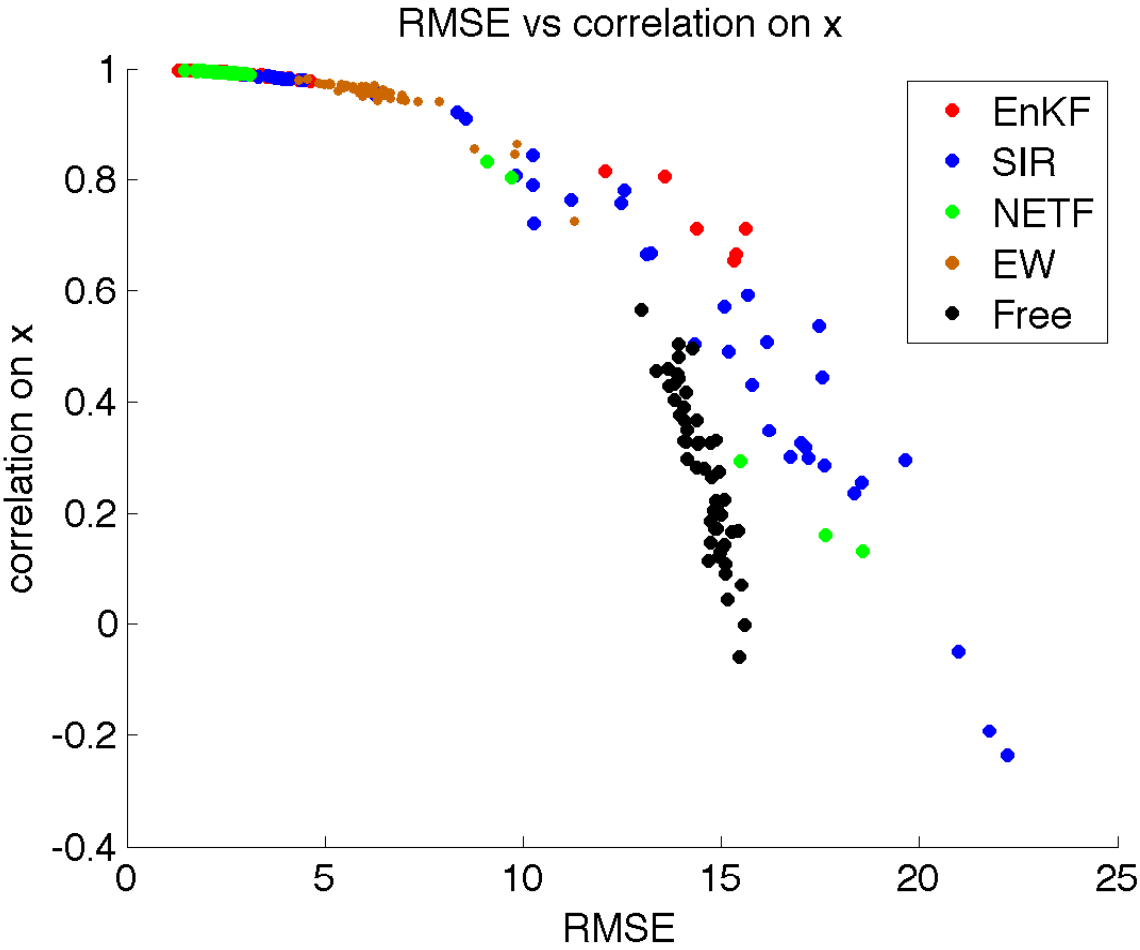


Figure 2.2: Correlation on x in function of RMSE for 50 twin experiments with standard parameters and different assimilation methods

in comparison with what we obtain without assimilation (black dots). However,in some case assimilations fails and results can even be worst than without assimilation. This is especially true in the case of the SIR.

The ensemble spread can be interpreted as a measure of uncertainty as a large ensemble spread correspond to a wider pdf. However,the spread must be large enough to protect the method from filter degeneracy therefore a large spread may increase the filter robustness. In figure 2.3 we compare, for our set of experiments, the relation between the RMSE and the spread. For all method the ensemble spread clearly smaller than the spread for free ensemble. In term of RMSE we observe the same things as in figure 2.1. We can observe that the ensemble spread for the NETF is larger than for the other methods without negatively impacting the RMSE. In experiment where the SIR obtain a good RMSE its ensemble spread is comparable with the spread obtain with the NETF and is slightly larger than with the EnKF. In term of spread there is few variability within a method.

We can now look at ensembles evolution during a twin experiment. In practice we will look at each filter separately on a particular twin experiment showing representative results. Evaluation of those criteria for this twin experiment can be found in table 2.3. Values in this table are similar to those of table 2.2, however they correspond to a particular twin experiment

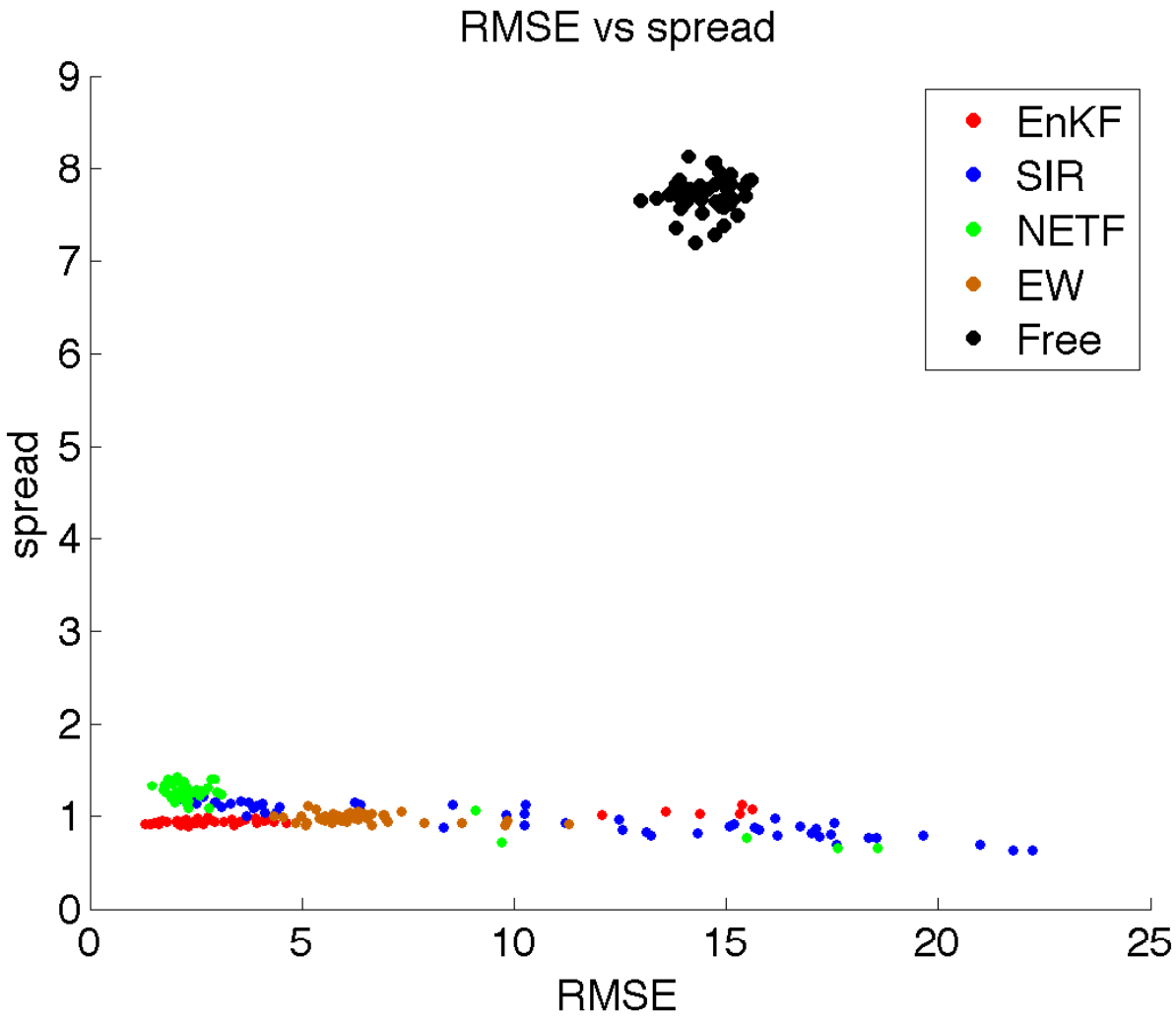


Figure 2.3: Ensemble spread in function of the RMSE for 50 twin experiments with standard parameters and different assimilation methods.

in which the NETF performed remarkably well while the SIR didn't.

	EnKF	SIR	NETF	EW	Free
RMSE	3.89	10.06	2.14	7.40	14.86
SPREAD	1.05	1.04	1.44	0.96	7.37
corr(x)	0.98	0.86	0.99	0.93	0.37
corr(y)	0.96	0.81	0.99	0.90	0.47
corr(z)	0.97	0.81	0.99	0.90	0.64

Table 2.3: Results obtained for one occurrence of the standard twin experiment.

2.4.1 Ensemble evolution: EnKF

The evolution of the ensemble during a twin experiment using the EnKF can be seen in figure 2.4. Particles closely follows the truth during most of the simulation. We can see the impact of the EnKF at different locations. For instance, Just before time=1 the ensemble starts to spread out, in the Lorenz 63 model it corresponds to a situation where particles tends to pass from one wing to another. The filter manage to compress the ensemble toward the good wing and avoid divergence. Another occurrence of this phenomenon is visible slightly before time=4, there the ensemble is on the wrong wing with respect to the observation but the filter repetitively move the ensemble toward the observation, eventually the ensemble move on the appropriate wing. At different point we can see that the ensemble describe a non Gaussian distribution. This is in conflict with the assumption on which the Kalman update is constructed. Therefore we can said that the EnKF is not theoretically appropriate for this data assimilation problem however in practice it produces good results.

2.4.2 Ensemble evolution: SIR

The evolution of the ensemble during a twin experiment using the SIR can be seen in figure 2.5. In this experiment the filter diverge,slightly before time=1. Observation noises trick the filter into resampling particle on the wrong wing of the butterfly. The filter has no means to detect this issue and produce an ensemble that doesn't correspond to the truth and is over confident. At one point the filter recover miraculously. This issue is not systematic as we can see around time=3.5, there the filter successfully drive the ensemble on the appropriate wing of the butterfly. Although theoretically appropriate for this kind of problem, the SIR is not able to recover and can frequently diverge.

2.4.3 Ensemble evolution: NETF

The evolution of the ensemble during a twin experiment using the NETF can be seen in figure 2.6. In this example the ensemble follow closeley the truth especially in term of ensemble mean. The greater ensemble spread typically encounter with the NETF is noticeable. From a theoretical perspective this filter is appropriate for this problem and can be seen as a relaxation of the SIR. The NETF recreate the same ensemble mean and covariance but its ability to create new particle preserve us from filter divergence. Moreover, in this experiment we used an inflation coefficient of 1.05, this explain the larger spread and prevent filter degeneracy.

2.4.4 Ensemble evolution: EW

The evolution of the ensemble during a twin experiment using the EW can be seen in figure 2.7. In this experiment the parametrization of the filter may very well be suboptimal. We

can see how this filter perturb the model dynamic, before the first observation all ensemble behave in the same way expect with this filter. The nudging term constantly pull the ensemble toward the next observation. One side effect of this nudging term is that the dynamic is altered, here the ensemble would have naturally move toward the first observation without any help but the nudging term accelerate this movement and the ensemble is slightly in advance of the truth. Near time=0.5 we are close to filter divergence, however the last steps of the filter manage to avoid this issue. However those same steps may sometimes produce strange behaviours, as we can see between time 3 and 4. Although the EW is an improvement on the SIR the parametrization is not obvious and may be counter-productive. This filter should not be recommended for a generic data-assimilation problem as different parameters may be application dependent.

2.5 Impact of the ensemble size

This parameter control the number of simulations needed for the data assimilation process and therefore impacts the required computational resources. As the ensemble size increases, the quality of Monte-Carlo approximation of the different probability distribution also increase as long as enough particles are not exact copies of each other. As long as we are working with a small system, such as Lorenz 63, we could afford large ensemble size without difficulties and the SIR would provide almost exact results. However, model used in climatology tends to be much larger and computational resources become a major bottleneck. In order to understand the impact of the ensemble size we run multiples set of 50 experiment with different ensemble sizes. Figures 2.8 represent the evolution of the RMSE and the correlation resulting of this procedure. It was generated by doing 50 twin experiment with different assimilation methods with the standard parameter but with different ensemble size and computing the mean of appropriate criteria. As expected the larger the ensemble the better the results, in general the RMSE decrease and the correlation increase. For very low ensemble size (less than 10 particles) the EnKF is the best filter. As the ensemble size increases, performances of the SIR and the NETF improve drastically. Around 20 particle, the NETF have the same performances as the EnKF and even start to outperform it. It is also useful to observe that the NETF systematically outperforms the SIR and quickly beats the EW. We can also observe that for large ensemble, the SIR achieves performances similar to the EnKF and the NETF.

2.6 Impact of the interval between observations

Previously, the time between two successive observations was set to 0.1. In order to investigate filter behaviour as observation become more and more rare we will run multiple set of 50 twin experiments with different observation periods. In all experiments we use 25 particle and others parameters are kept unchanged. The selected ensemble size is arbitrary, it correspond to a value for which the NETF performs relatively well. Intuitively, the more frequent observations are the better the results will be as we have access to more informations. This parameter controls the observation sparsity, however it is difficult to compare this sparsity between different applications. In figure 2.9 we represent the impact of the time interval between observations on different criteria in term of average over 50 independent twin experiments with same parameters. As expected, increasing the interval between observations decrease performances, the RMSE increases and correlations decrease. We can see that the EW is more affected by observations sparsity than other filters. One possible explanation could be that the nudging term pulls particles toward the observation too soon and becomes counter-productive. It is possible to take the time into account in the nudging term computation in order to avoid this

issue but this kind of modification is system dependent and complexifies the filter even more. For frequent observations, observation period inferior to 0.1, we see that the EW is still an improvement on the SIR and that the NETF and EnKF achieve similar performances. As the period between observations increases, the NETF start to outperforms the EnKF. As the observation period reaches 0.3, the SIR manage to beat the EnKF for a while. For sparser observations, the NEFT remains the best method.

2.7 Selecting the method for larger scale applications

At this point we have to choose a method for larger scale applications. We want to select a method that can works with a small ensemble and sparse observations and that is as robust as possible. The robustness criterion rules out the SIR as we have seen in figure 2.1. The EW is an improvement on the SIR but it require a lot of effort in term of parametrization, is greatly affected by the observations period and perturbs the model dynamics. Moreover the EnKF and the NETF tend to outperform the EW. At this point we have to choose between the EnKF and the NETF. Although the EnKF is better for very small ensemble it relies on unverified hypotheses on the prior distribution and is beaten by the NETF when the ensemble size is slightly greater. Moreover the NETF is less negatively affected by the period between observations. In following chapter we will therefore concentrate ours efforts on the NETF. However a last step is required before larger scale application, we have to understand the impact of ensemble inflation on the NETF.

2.7.1 Impact of the inflation parameter

In the NETF, ensemble inflation can be easily implemented and is known to improve the algorithm. It only require to multiply the ensemble perturbations by a scalar but this scalar is arbitrary. In order to understand what could constitute an acceptable value we can perform multiple sets of 100 twin experiments with different value for this parameter γ. In figure 2.10 we illustrate the relationship between ensemble spread and RMSE for different values of this parameter. We can see that results arranged themselves in two main groups, one with a large RMSE ans another with low RMSE. The first groups correspond to experiments where filter divergence occurred and is mostly made of experiment with low inflation parameter. This shows that a slight inflation protect the filter from divergence. In the groups of experiments with low RMSE another tendency appears, as the inflation parameter increase the ensemble spread also increase. This observation is quite natural and constitutes a first disadvantage of high γ value. Moreover, This cluster of experiments also shows that as the ensemble spread increase the RMSE also increase, adding another reasons to limit the inflation. In practice we recommend to set the inflation parameter somewhere between 1 and 1.05 in order to improve the filter robustness while limiting the ensemble spread.

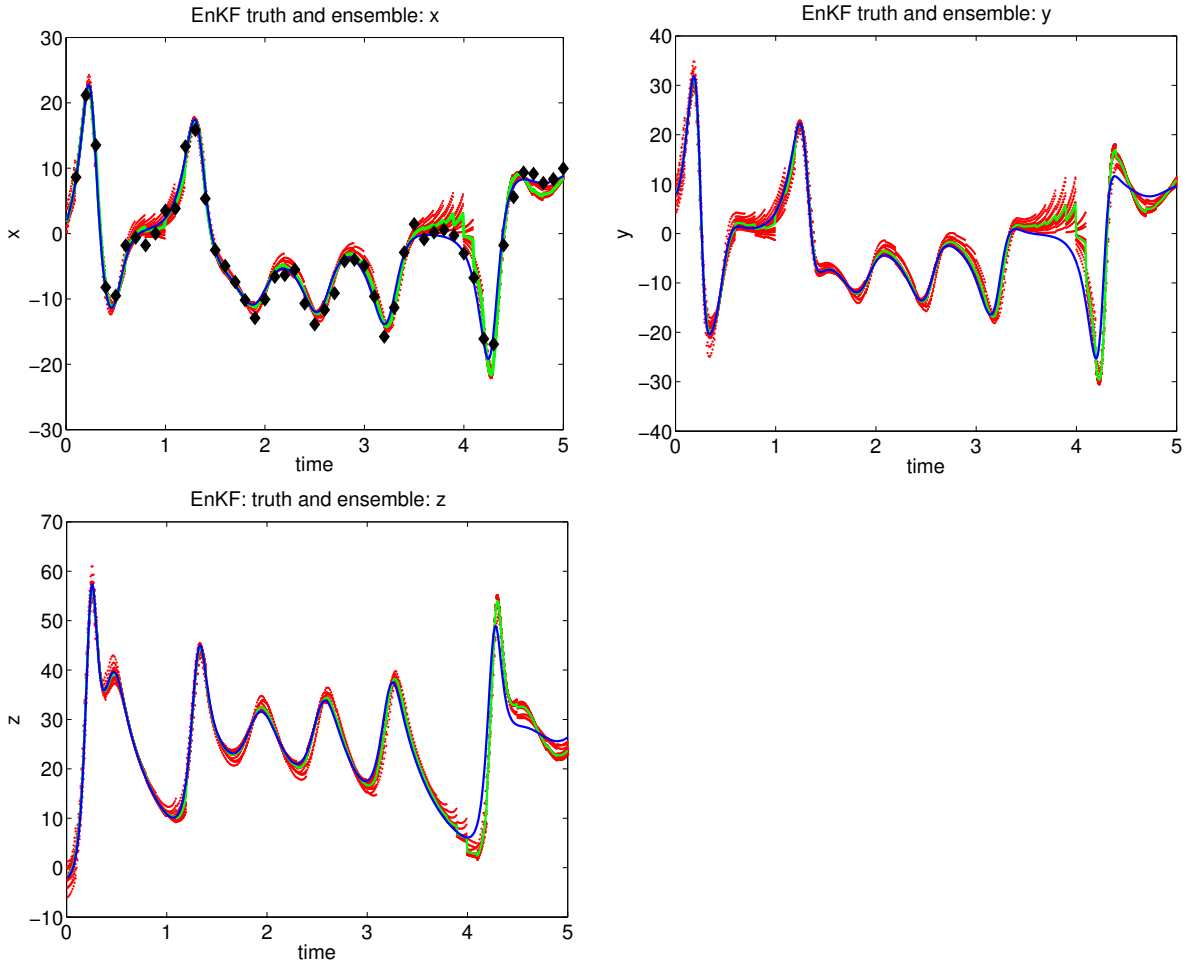


Figure 2.4: Evolution of the ensemble in a twin experiment with standard parameter using the EnKF. From left to right, evolution for variable x,y and z. In blue: the truth, in red: the ensemble, in green: the ensemble mean, in black: observations.

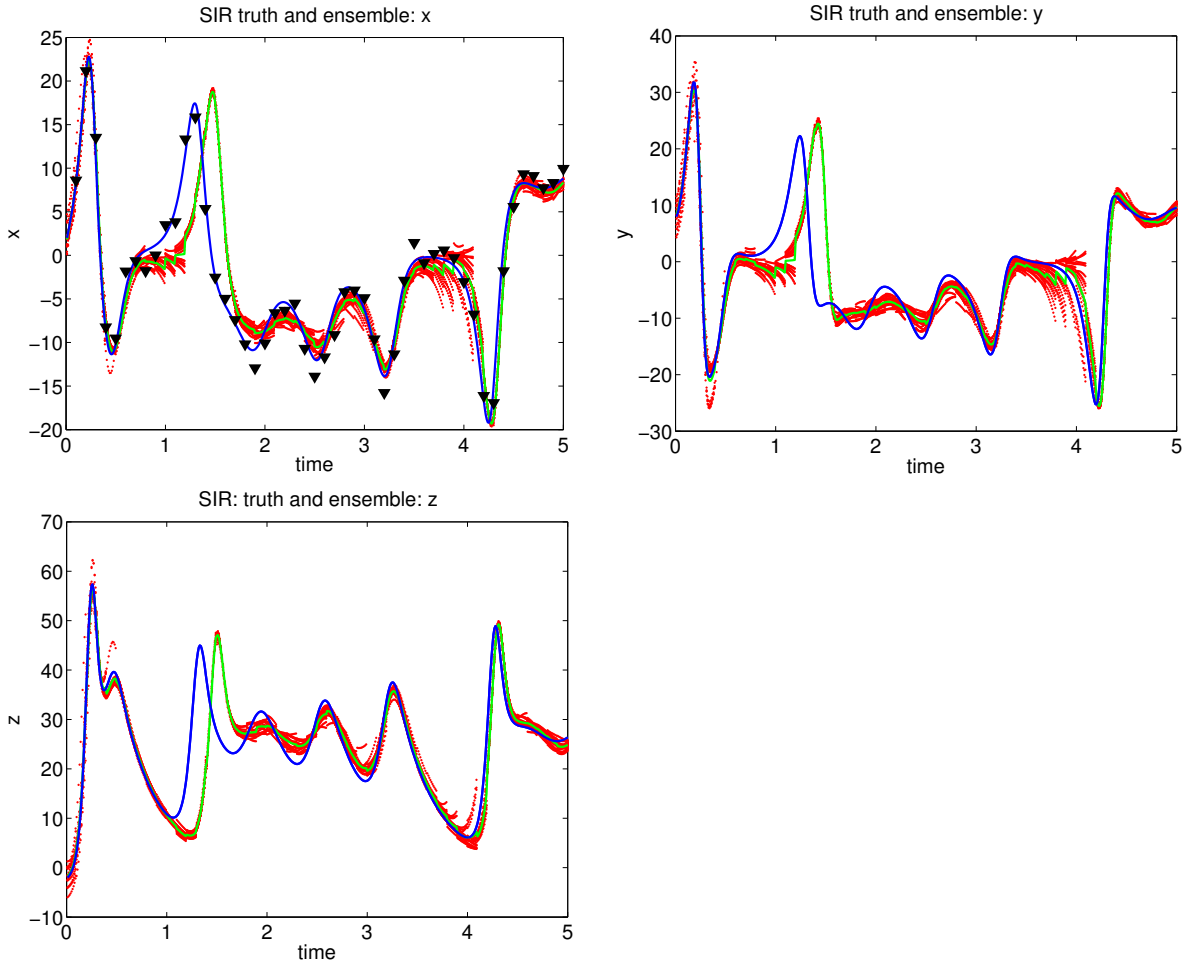


Figure 2.5: Evolution of the ensemble in a twin experiment with standard parameter using the SIR. From left to right, evolution for variable x,y and z. In blue: the truth, in red: the ensemble, in green: the ensemble mean, in black: observations.

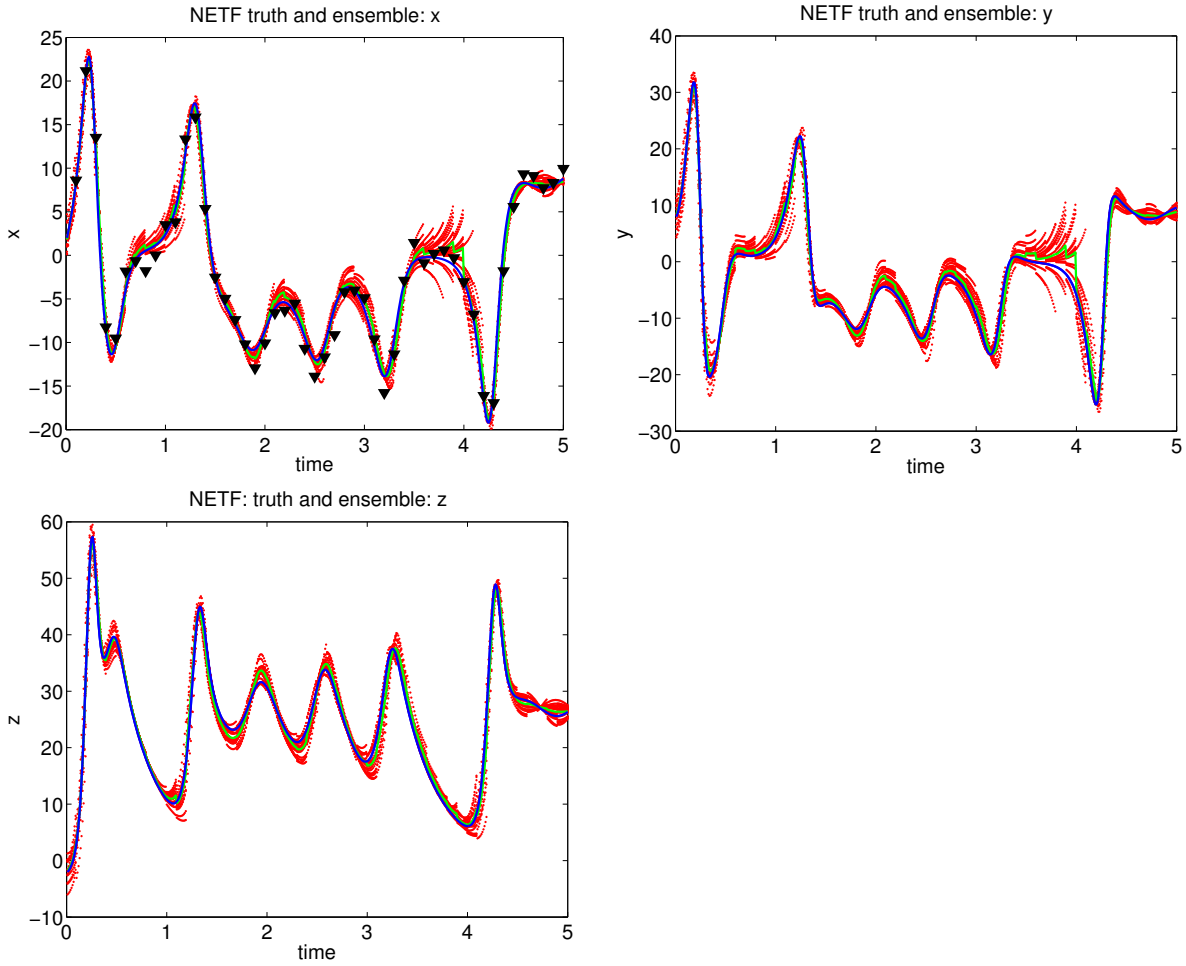


Figure 2.6: Evolution of the ensemble in a twin experiment with standard parameter using the NETF. From left to right, evolution for variable x,y and z. In blue: the truth, in red: the ensemble, in green: the ensemble mean, in black: observations.

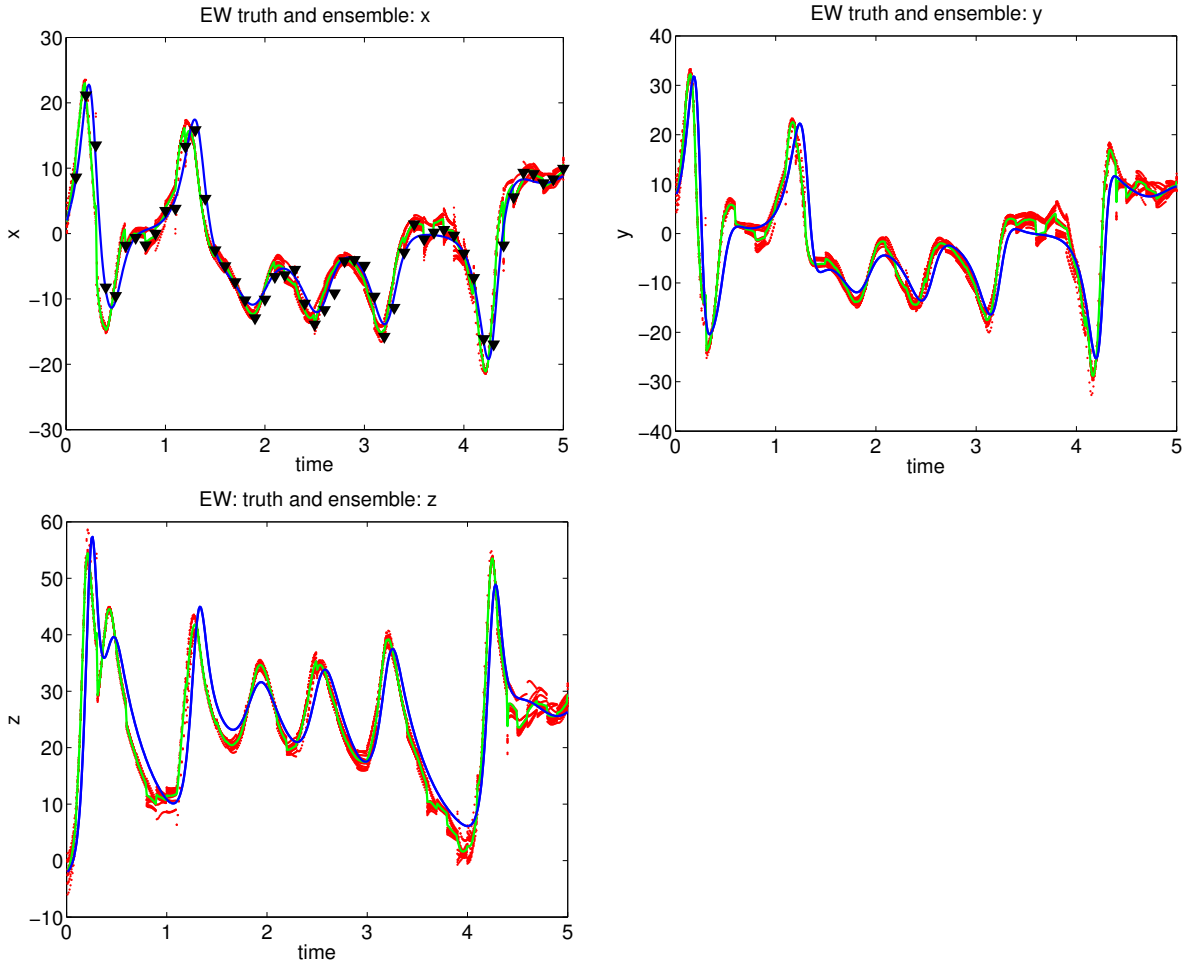


Figure 2.7: Evolution of the ensemble in a twin experiment with standard parameter using the EW. From left to right, evolution for variable x,y and z. In blue: the truth, in red: the ensemble, in green: the ensemble mean, in black: observations.

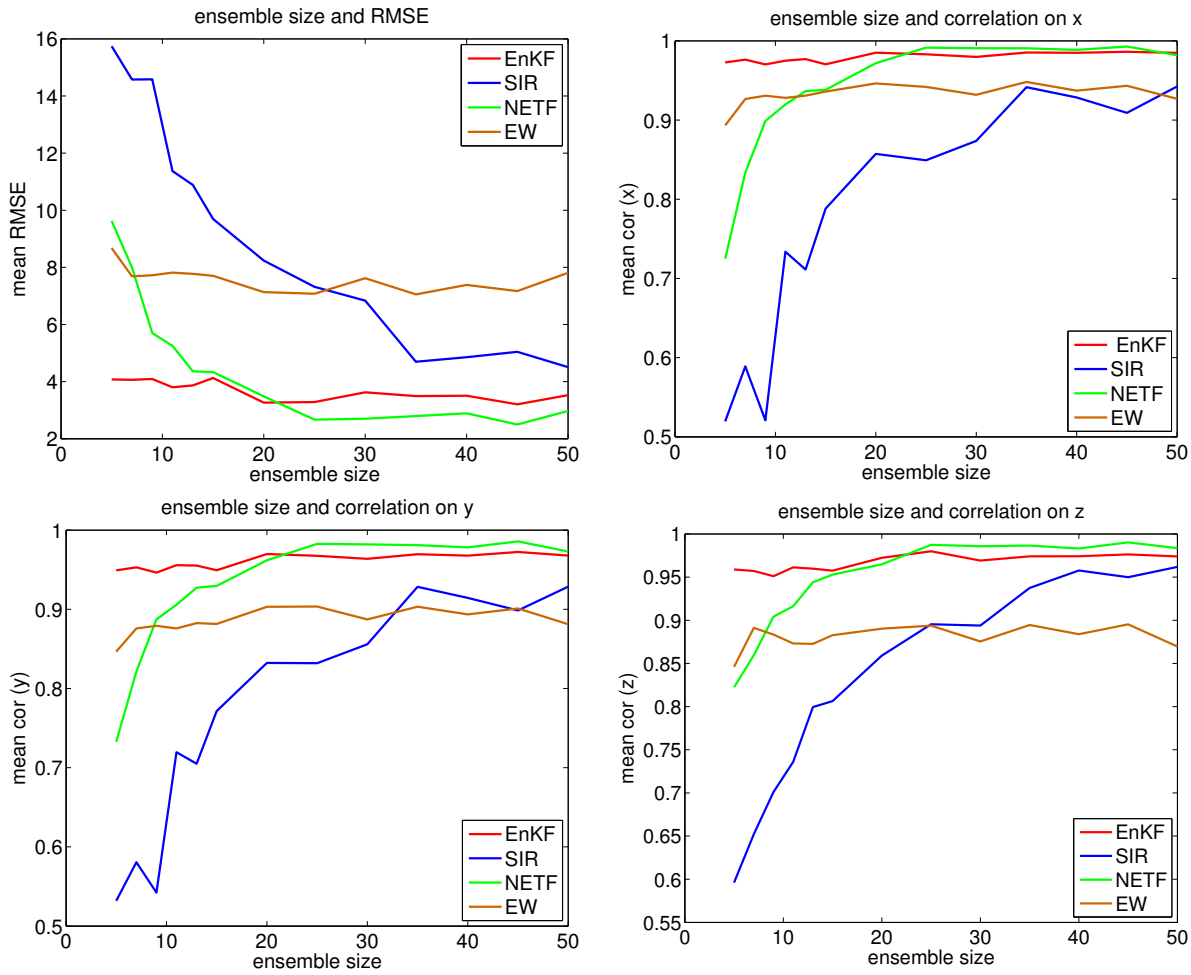


Figure 2.8: Impact of the ensemble size on different criteria for the EnKF, SIR, NETF, and EW. From left to right, impact on the RMSE, impact on the correlation on variable x, impact on the correlation on variable y and impact on the correlation on variable z

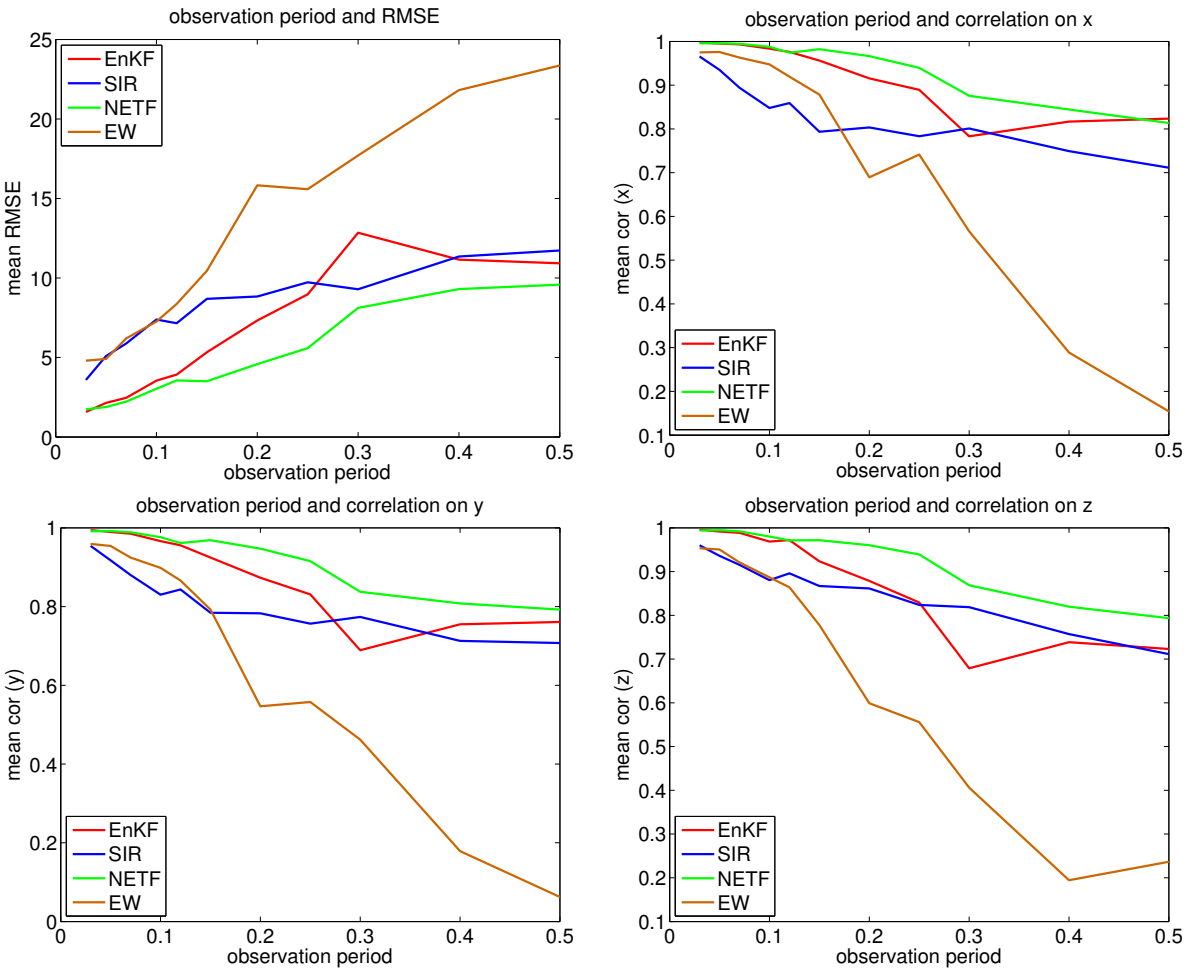


Figure 2.9: Impact of the observation period on different criteria for the EnKF, SIR, NETF, and EW. From left to right, impact on the RMSE, impact on the correlation on variable x, impact on the correlation on variable y and impact on the correlation on variable z

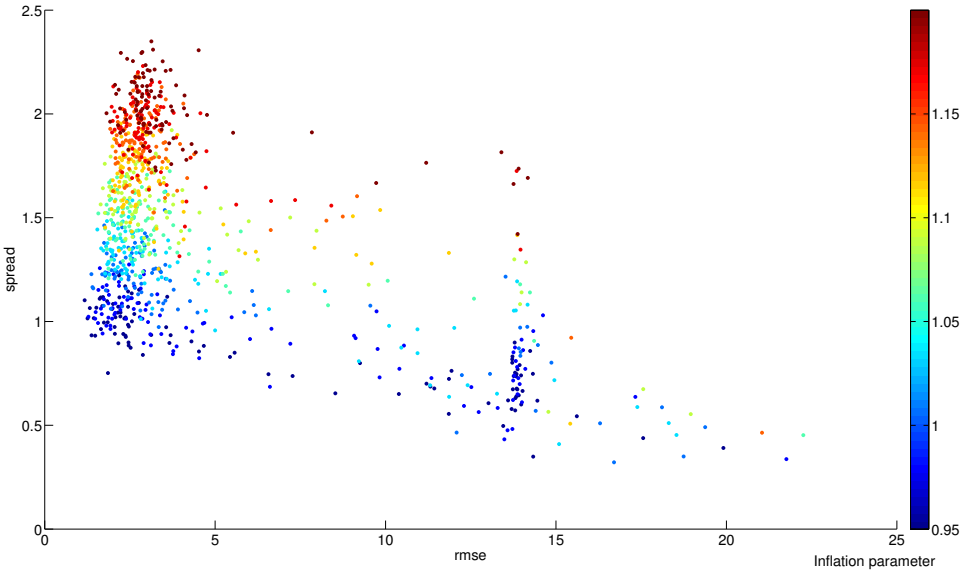


Figure 2.10: Impact of the inflation parameter on performances of the NETF. Ensemble spread and RMSE for twin experiments with different values of the inflation parameter.

Chapter 3

Offline assimilation

In this chapter, we will test a SIR and a NETF on a climatic model of intermediate complexity through off-line assimilation. The purpose of off-line assimilation is to save on computational resources by exploiting previous simulations while correctly assessing the impact of the assimilation method on a specific model. Although the assimilation step is in itself correctly represented we have to keep in mind that this method doesn't allow to capture the future behaviour of the posterior ensemble. Nonetheless it constitutes a good way to test the filter implementation and to assess some basic properties. In this work we will use the LOVECLIM [14] model to create twin experiments as LOVECLIM constitute a good example of climatic model. We will start this chapter by presenting the LOVECLIM model, then the method use for our offline experiments and at last we will present our results.

3.1 The LOVECLIM model

LOVECLIM is a model representing the Earth system obtained by coupling different models in charge of different aspects of the Earth system, in this section we will quickly present LOVECLIM based on an article in which the version 1.2 has been introduced [14].

- **Atmosphere:** LOVECLIM uses the atmospheric component of the model ECBilt. This model represent the atmosphere with, amongst other variables, temperatures, stream functions, vorticity and water content at different pressure horizons.
- **sea-ice and ocean component:** Using the model CLIO the ocean and the sea-ice component are represented using, amongst other variables, barotropic velocity, surface elevation, velocity, salinity, temperatures, turbulent kinetic energy, vertical diffusion and viscosity. Those variables are defined on a three dimensional grid.
- **continental biosphere component:** Using the model VECODE, the vegetation is represented via fraction of tree, grass and total vegetation. Using this representation of the biosphere an important part of the carbon cycle can be simulated.
- **the ocean carbon cycle component** Using the model LOCH, other parts of the carbon cycle can be simulated. Its mains variables contain dissolved carbon, total alkalinity, dissolved and particulate organic matter and oxygen.
- **the polar ice sheet component** The model AGISM is in charge of simulated ices sheets situated in Antarctica and Greenland. This require three main components, the ice flow itself, the interaction with the solid Earth and masses exchanges with the atmosphere and oceans.

The result is a three-dimensional model of intermediate complexity in the sense that its spatial resolution is not as high as in state of the art general circulation model and that some simplifications were made, especially in the atmosphere component [14]. However this model have the advantage to be sufficiently light in term of computational requirements so that we can afford to run several simulations in parallel. This particularity allows the use of assimilation method based on particle filters with reasonable ensemble sizes. It is clear that the dimension of the state-space of such a model is in no way comparable to the small Lorenz 63 model. For instance the surface temperature on the southern hemisphere is describe on a 32 by 64 grid. The salinity is represented on a 122 by 65 by 20 grid. For those two variables alone we have respectively 2048 and 158 600 dimensions needed in the state space. Therefore we can expect poorer result on the LOVECLIM model than on Lorenz 63 for the same ensemble size. Although the number of variable is large, the system dynamics as well as the spatial nature of the system may reduce the number of useful dimension, making any comparison difficult.

In mathematics, problems resulting form the high dimension of a state space are often refereed as the curse of dimensionality. One way to understand the curse of dimensionality from a sampling perspective is to consider the number of grid points required to sample an unit-hypercube, (an hyper-cube in which every edge has a length equal to one), with a given resolution. In the three-dimensional state space of the Lorenz 63 model, in order to fill an unit-cube with a resolution of 0.1, corresponding to 10 points on an edge, we need 10^3 grid point. For the surface temperature in LOVECLIM the state space has 2048 dimensions and sampling an unit-hypercube in this state-space with the same resolution would require 10^{2048} grid points. It is clearly impossible to have an ensemble large enough to sample uniformly the state space of LOVECLIM, at best we can only hope for a few hundred particles. Therefore Monte-Carlo representations of probability distributions on the LOVECLIM state space are rather crude. However, the system doesn't evolve freely in the state space and the effective dimension for a model is often lower than the dimension of the state space itself and depend of the dynamic of the system. For instance the Hausdorff dimension of a random walk on a three dimensional state space is almost surely 3 while the correlation dimension of the Lorenz 63 model near its attractor is slightly greater than 2 [22]. Although, the state space of the Lorenz 63 model is three dimensional, a trajectory generated by this model fill the space nearly in the same way as a two-dimensional surface. This reduce the ensemble size required for data assimilation as the system dynamic constrains the trajectory. Using a time series of surface temperatures produced by LOVECLIM over a 12000 months period we estimated the correlation dimension to be around 15. This first estimation doesn't take all variables into account and therefore doesn't represent the model in all its complexity. The computation of the correlation dimension is obviously affected by the limited data set. Nonetheless, we can compare this result with the 2048 dimensions of the state space for surface temperature to show that the dynamic affect the effective dimension of the system.

3.2 Method

We start from a simulation of the Earth system under pre-industrial conditions made with the model LOVECLIM over a period of 1000 years. Because there is no major modification in forcing throughout the simulations, each years can be seen as another realisation of the same simulated year. Therefore we can sample a set of years in order to create an ensemble representing the system state and evolution at a given time. To set up a twin experiment we can use a set of consecutive years to represent the truth and build pseudo observation based on an appropriate mask. Then by sampling a set of years amongst the simulation we can build an ensemble of initial conditions. Instead of applying a costly model to a potentially vast

ensemble one can just read results from the initial simulation, starting form the beginning of sampled years. This formulation of the off-line method keep a lot of similarities with an online assimilation method. However, the method can be understood differently. The long simulation can be divided in a set of simulated years corresponding to a Monte-Carlo approximation of the probability distribution of the system trajectory over a year. Sampling within the set of available realisations produce a prior ensemble that can be used by our filter. In our case data assimilation take place each years using annual average surface temperature. In order to reflect data sparsity, surface temperature measurements are not taken everywhere. In particular the grid used is 64 by 32 and measurement are taken in only 91 grid points concentrated in the Antarctica region.

An implementation of off-line assimilation with a SIR was already available and produce a weighted mean of the ensemble as a reconstruction of the system trajectory. Based on the same program, the NETF transformation have been computed. However, in off-line assimilation it is not possible to use newly generated particles as they are not part of the initial simulation therefore they can not be propagated trough time. For this reason we have to resample within the simulation in order to generate new ensemble and make sure that the process carry on. In order to produce a reconstruction of the system trajectory, we can use the NETF transformation and compute the ensemble mean. By construction, the NETF and the SIR should produce similar results in this set-up. Our main criterion in this experiment is the correlation coefficient between the reconstruction and the time series from which the pseudo observations were produced.

3.3 Results

Two different ensemble sizes were tested, 10 and 100 particles, as well as two filter, a SIR and a NETF. In order to assess each filter performance we concentrate on correlation of surface temperatures, first locally then averaged over Antarctica and over the southern hemisphere. The computation of this correlation is based on a single reconstruction, the ensemble mean after data assimilation. Because the two filters have theoretically the same ensemble mean we should not observe major differences.

The table 3.1 shows the covariance averaged over Antarctica in different cases.The table 3.2 shows the covariance averaged over the southern hemisphere in different cases. As expected, for the same ensemble size there are very little differences between methods. However we can clearly see that using 100 particles provides better results. This is true for the two regions but the effect is more visible when looking at the correlation over the southern hemisphere. We can compute a relative increase of 63% for the southern hemisphere and 28% for the Antarctica.

ensemble size	correlation coefficient with SIR	correlation coefficient with NETF
10	0.26	0.25
100	0.32	0.32

Table 3.1: Correlation coefficient averaged over Antarctica, based on the ensemble mean reconstruction with a SIR and a NETF for different ensemble sizes.

Considering the surface temperature alone on the 64 by 32 and the fact that we only have 91 observations we only have informations on 4.4% of the system state. Therefore we can ask ourself if the quality of the reconstruction is the same on every grid points. In order to

ensemble size	correlation coefficient with SIR	correlation coefficient with NETF
10	0.11	0.10
100	0.18	0.18

Table 3.2: Correlation coefficient averaged over the southern hemisphere, based on the ensemble mean reconstruction with a SIR and a NETF for different ensemble sizes.

better understand how the reconstruction behave far from observed point we can display the correlation coefficients on maps.

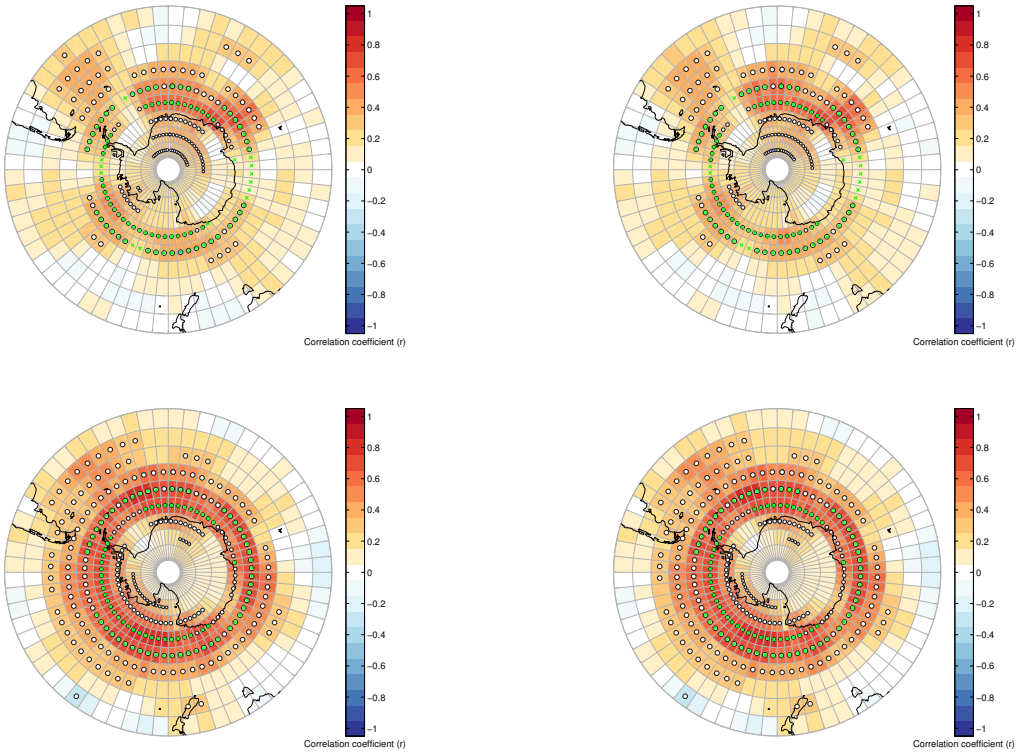


Figure 3.1: Maps of correlations coefficients, origins of assimilated points are in green, locations with significant correlation (0.05) are designated with a white circle. From left to right: SIR with 10 particles, NETF with 10 particles, SIR with 100 particles, NETF with 100 particles.

Figure 3.1 represent the correlation between the truth and the reconstruction on the southern hemisphere. The blue-red colouration represent the correlation coefficient, green cross are displayed where observation are available and white circle denote locations where the correlation is significant at a 5% level. Once again there is no major differences between methods for the same ensemble size. We can observe that observed points have a good correlation, at a significant level for most of them. In the other hand the correlation is lower for points far from observed locations. We can also notice the impact of the ensemble size, results with 100 particles are better even far from observed locations. The data assimilation process have to provide information on non-observed locations to be useful. In figure 3.1 we can clearly see that correlation coefficients become significant on some non-observed locations. For large ensemble size this phenomenon is more visible.

Furthermore we can look at the distribution of the surface temperature as represented by different ensembles. For clarity we only look at one location at a time and limit ourself to ensemble of size 10, similar results can be obtains with larger ensembles. In figure 3.2 blue bars represent prior particles weighted according to Bayesian weights, red bars represent posterior particles generated by the SIR, green bars correspond to the NETF without rotation and magenta bars correspond to the NETF with rotation. By representing prior particle like we did we can visualize both the prior ensemble and the information carried by the observation. Moreover it simplifies the comparison of the different methods. The set of bars with a particular colour represent therefore a probability distribution of the surface temperature. Note that we didn't add any perturbation to the ensemble generated by the SIR but taking this modification into account is trivial.

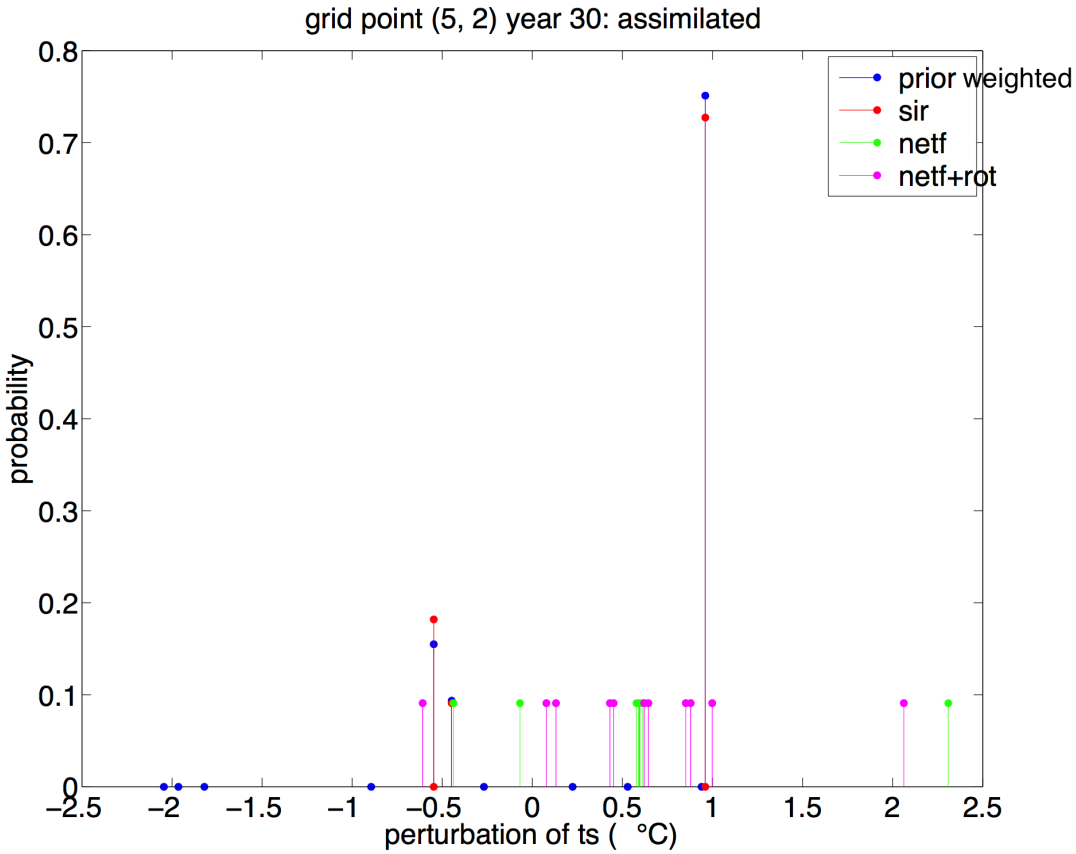


Figure 3.2: Probability distribution produce by different filter for the annual average surface temperature over the grid point (5,30) on year 30 of the simulation. In blue prior particle with Bayesian weight, in red particle after resampling, in green particles after NETF transformation and in magenta particles after NETF transformation and rotation.

Interpretation of such a graph is not easy as multiple observations are assimilated. A direct consequence is the complex distribution of the Bayesian weigh amongst the ensemble. Only three prior particles have a non-negligible weights, therefore the SIR generate an ensemble with multiple copies of those particles. The offline assimilation method doesn't allow us to judge if filter degeneracy may occurs in the following step however this illustrate the need for extra noise on particles. Particle after assimilation have the same weight but as some of the prior particles are sampled several time this is equivalent to an ensemble with few particles and non

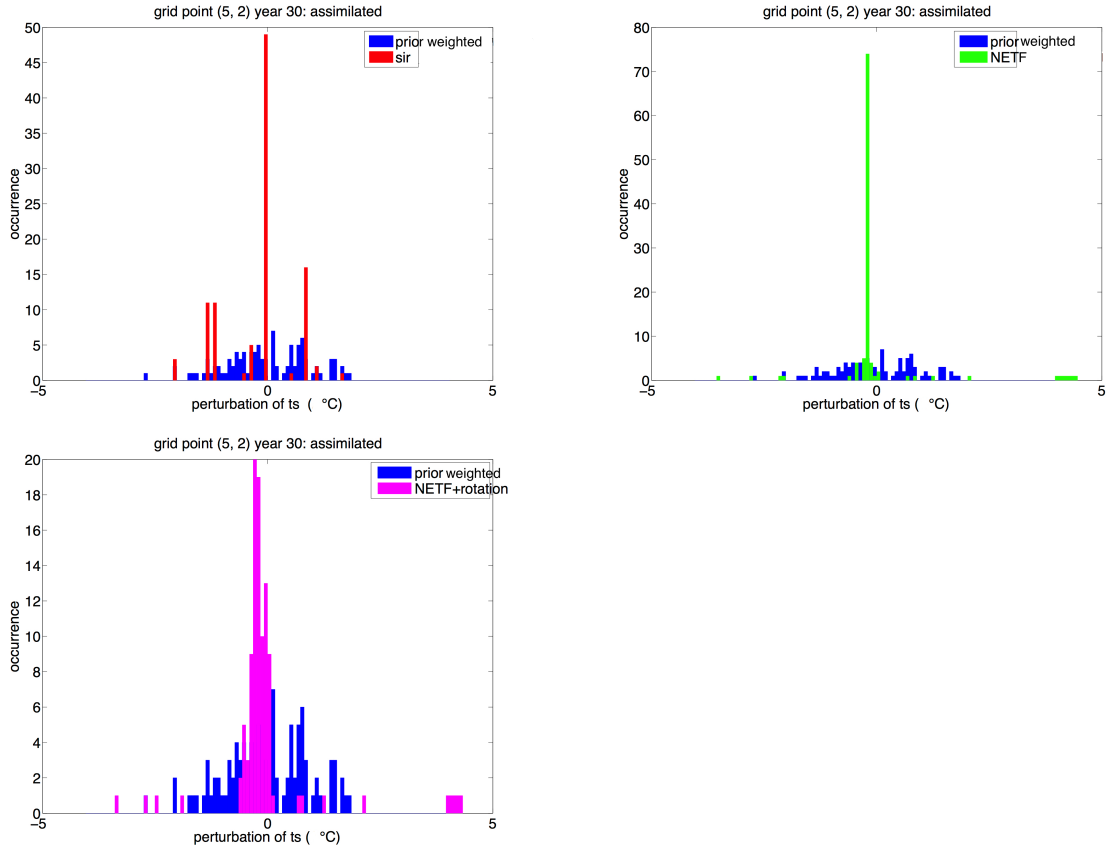


Figure 3.3: Histogram representation of ensemble for surface temperature at grid point (5,2) with different filters using 100 particles. In blue: prior ensemble, in red the SIR posterior ensemble, in green the NETF posterior ensemble and in magenta the NETF with rotation posterior ensemble.

uniform weights. The NETF produce an ensemble (in green) with uniform weights. However, posterior particles are not exact copies of prior particles. Therefore physical balances may not be respected. Unfortunately this filter produce a cluster of particles very close to each others, this may also be a waste of resources if the model doesn't scatter them enough in the future. Applying the mean and covariance preserving rotation generate the set of magenta particles. By definition the mean and covariance are the same as previously but the cluster of particles encounter with the NETF without rotation disappear and other region of the state space are now represented in the ensemble. This should protect the filter from filter degeneracy.

In the case of ensemble of 100 particles, similar observations can be done, however, for visibility we will use separated histogram, see figure 3.3.

This comparison help us to understand two issues with the SIR. First, it may produce an ensemble formed only by slightly perturbed copies of single particle therefore increasing the risk of filter degeneracy. Secondly it doesn't allow the posterior ensemble to incorporate state outside those represented in the prior ensemble. This last fact highlight the SIR inability to recover once filter divergence occurs. The NETF without rotation produce an ensemble with the same mean and covariance as the SIR but is not limited in the same way as the SIR. It can indeed create particles drastically different thus exploring a larger portion of the state-space. As previously explained, the NETF only guaranties that the mean and covariance obtained

with the NETF are the same as with the SIR. Thus the NETF with rotation carries the same informations, but it samples through the state space differently and is less likely to form inappropriate particle clusters.

Although the NETF give promising results, the off-lines assimilation doesn't allow us to assess the risk of filter degeneracy during an online assimilation. Therefore the next chapter will be devoted to online assimilation using the LOVECLIM model.

Chapter 4

Online assimilation

At this point we have selected a data assimilation method, the non-linear ensemble transform filter (NETF) and we have successfully implemented the method for offline assimilation on a climatic model. We have obtained results similar to those of a sequential importance resampling filter (SIR) in our offline twin experiment. In this chapter we will apply the NETF in an twin experiment using the model LOVECLIM directly. The set-up of the twin experiment is the same as in the previous chapter but this time the model is directly applied to the ensemble. Therefore we can use the NETF explicitly and compare it with the SIR. In particular, we will use pseudo data from a particular run of the model LOVECLIM. In order to reflect the sparsity of real observations we will only assimilate the surface temperatures measured on a fraction of the globe. Moreover, because some proxies only reflect time averaged quantities we will assimilate surface temperatures averaged over a year. For convenience we will do an assimilation step at the end of every simulated year. We will compare the results obtained with the NETF with those of the SIR for two ensemble sizes, 10 and 100 particles. In this twin experiment we have access to the simulation from which the assimilated data come from. Therefore, we can compute correlation coefficients between our reconstruction and the simulation of reference. In particular we will concentrate our efforts on the surface temperature for simplicity.

4.1 Experimental set up

A simulation of reference was produced using the model LOVECLIM. This simulation cover a 100 years period under pre-industrial conditions in the southern hemisphere. This simulations was used to produce sparse pseudo-observations. In particular, annual average surface temperature of 91 grid-points around the Antarctica constitute our pseudo-observations. Starting from similar initial conditions, we run an ensemble of simulations for one year, then an assimilation method is applied using the corresponding set of pseudo-observations. We repeat the process until all of the pseudo-observations have been assimilated. From the simulations set we compute a reconstruction of the surface temperature using a weighted mean of all particles. Then, we compute at each grid point the correlation coefficient for the surface temperature between the simulation of reference and the reconstruction. As we have seen in previous chapters, the non-linear ensemble transform filter (NETF) appears to be an interesting assimilation method for high dimensions and non-linear data assimilation problem. In a first time we will perform a twin experiment using the NETF without inflation and another with a slight inflation (inflation parameter set to 1.05). These two experiments will be conducted with an ensemble of 10 particles. We will compare the results in term of correlation with a reconstruction done using a sequential importance resampling filter with the same number of particles provided by François Klein. Based on those result we will select one of the two versions of the NETF and conduct another twin experiment using 100 particles and compare the correlations values with

Correlation coefficients	NETF no inflation	NETF inflation 1.05	SIR
Averaged over grid cells assimilated:	0.28	0.28	0.19
Averaged over Antarctica:	0.17	0.15	0.21
Averaged over South Hemisphere:	0.09	0.07	0.08

Table 4.1: Averaged correlation coefficients for the surface temperature between the simulation of reference and reconstructions obtained with different data assimilation methods using 10 particles.

those obtained with a SIR using the same number of particles.
Theoretically, the transformation of the NETF can be applied to every variables in the model, however, for simplicity of implementation we will only apply the transformation to a subset of oceanic variables. We will update the following variables, barotropic velocity, velocity, salinity and temperatures. The reasoning for a partial update is the following. Some variables have a shorter memory than other and all variables don't propagate informations in the same way in the system. For instance, atmospheric variables tends to have a faster dynamic than oceanic variables. Therefore discontinuities at the atmosphere-oceans interface induced by updating only oceanic variables should not be and issue as atmospheric variables can quickly adjust. Hopefully, the information coming from the observation will be explicitly assimilated by oceanic variables during our update. Then the system dynamics would transfer relevant informations to others variables without inducing strange behaviours.

4.2 Results with 10 particles

After assimilation of pseudo-data, we obtain a reconstruction of the surface temperature and we compute the correlation coefficient between the reconstruction and the simulation of reference at each grid points. This can be done from results of twin experiments using a NETF without inflation, with inflation and with a SIR. In figure 4.1 we can compare the resulting correlation coefficient obtained with each method.

Another way to compare the correlation coefficient is to take the average over specific regions, such as the set of locations where pseudo-data come from, over the Antarctica (all point at the south of the -60th parallel) and over all of the southern hemisphere, see table 4.1

Comparing the SIR with the NETF without inflation we can see that the NETF achieves better correlation on assimilated grid cell but not over the Antarctica, where in average correlation are comparable. We can observe that the inflation doesn't improve the performances of the NETF in this case, although it is not possible to judge if it is always the case with only two experiments we arbitrary decide to use the NETF without inflation in the next experiment. One raison for the lake of improvement may be that LOVECLIM is able to spread out an ensemble initially constituted of very similar particles. This idea will later be reinforced by the figure 4.2 where we see the evolution of the ensemble and simulation of reference in term of monthly surface temperature averaged over all assimilated location. Although the ensemble shows initially a small range (less than one degrees) after a few month the ensemble spread out and its range is closer to five degrees.

4.3 Results with 100 particles

Because we didn't notice any positive impact from the inflation we can save ourself some computation and only perform an experiment with the NETF without inflation. Using the same methodology we produce new maps of correlation coefficients, see figure 4.3

Correlation coefficients	NETF no inflation	SIR
Averaged over grid cells assimilated:	0.49	0.49
Averaged over Antarctica:	0.33	0.41
Averaged over South Hemisphere:	0.13	0.18

Table 4.2: Averaged correlation coefficients for the surface temperature between the simulation of reference and reconstructions obtained with different data assimilation methods using 100 particles.

Increasing the ensemble size improve the results as we can clearly see, the correlation becomes significant in more places and the average correlation coefficient over assimilated location increases.

In table 4.2 we can see that on assimilated grid cells the NETF perform as well as the SIR. However, in average the correlation coefficient is lower for the NETF (0.13) than for the SIR (0.18). The average over the Antarctica shows the same trend.

In order to visualize the system evolution we can use a principal components analysis as a dimensionality reduction tool. This method allows for a qualitative assessment of the system behaviour despite its dimensions by providing graphical representation of the both the ensemble and the simulation of reference. Starting from the set of simulated state of the system we can construct a correlation matrix, then we can compute the eigenvectors associated with the two largest eigenvalues. These vector constitutes a base of a two-dimensional subspace on which we can project all our simulated state. We can perform this task using monthly surface temperatures obtained with our ensemble and use different colouration schemes in order to understand the system. In the figure 4.4 red triangles represent states from the simulations of reference, blue-to green dots represent simulated states of the system where the colour gradient indicates the time since the beginning of the simulation. No particular colour pattern appears in this first analysis, indicating that there is no strong pattern of variation between years. For instance, if we had observed a colouration gradient going from the top to the bottom of the figure, it would have showed that there were a long term drift in the time series. The 6 patches observed in figure 4.4 correspond in fact to 12 partially overlapping patches, one per month. In the figure 4.5 we use a different colour for each month and twelve visible patches of colour appears, indicating that the system states vary more between months than between years and that no clear long term patterns can be seen with this small number of dimensions. Although all these observations seem to indicate that the filter behave properly, in the sense that the model dynamic is respected and that the ensemble and the simulation of reference behave similarly. Unfortunately there is no strict guarantee as the dimensionality reduction method doesn't show all possible features present in the data set. One way to measure the lost of information due to the dimensionality reduction is to inspect the eigenvalues of the covariance matrix. The ratio between the sum of the two largest eigenvalues and the sum of all eigenvalue was around 0.3 while 130 eigenvalues were necessary to reach 95 percent of the total variance on our time series. Using the same method of dimensionality reduction with annual surface temperature and allocating the same colour for each ensemble member of the same year we obtain the figure 4.6. No clear pattern emerges in this figure, which confirms the absence of long term drift in the simulation of reference and in the ensemble itself.

Another way to illustrate this method is to display the spatial average of each ensemble members and compare with the simulation of reference. In figure 4.2 we can compare the evolution of the ensemble and the reference based on the monthly value of the average surface temperature taken over assimilated locations. We can clearly identify the different month and observe that there were no clear need for ensemble inflation.

A single experiment may not be enough to conclude, therefore following remarks must be taken with caution. For small ensemble sizes the NETF outperformed the SIR and assimilated locations and manage similar performances in average. For larger ensemble, the NETF performed as well as the SIR on assimilated locations but the SIR is better on average. From this, the rule of thumbs would be to use the NETF in application where the ensemble size must be kept very small and where a lot of locations/variables are observed. In contrast the SIR should be preferred on application where the ensemble size is less limited and when observation are sparser. Although, we insist, this conclusion may very well be premature.

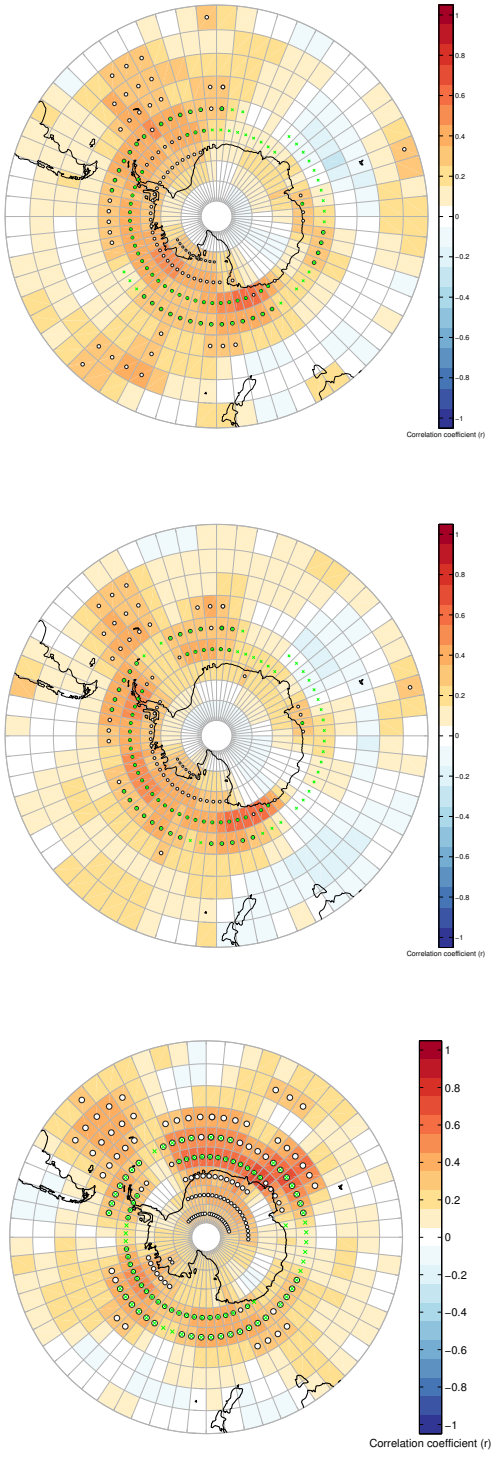


Figure 4.1: Maps of correlation coefficients, origins of assimilated data are in green, locations with significant correlation (0.05) are designated with a white circle. From top to bottom: NETF without inflation, NETF with inflation (1.05), SIR. Results obtained with 10 particles.

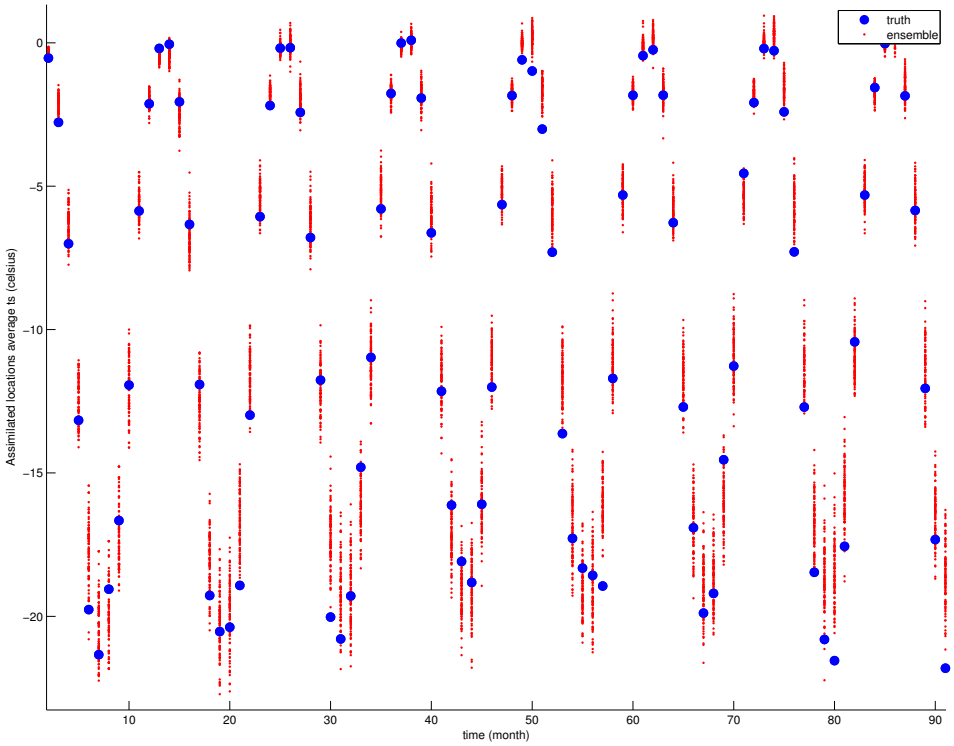
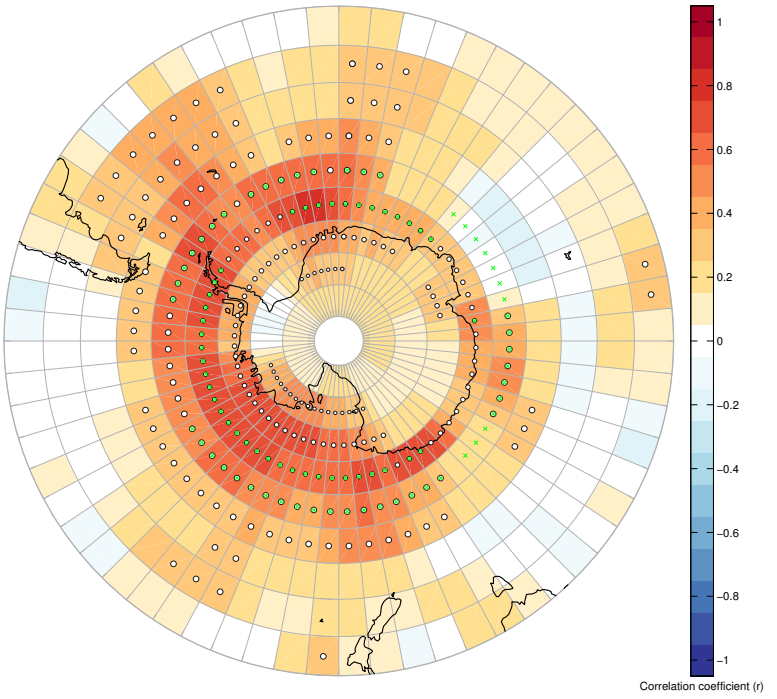


Figure 4.2: Evolution of the ensemble and simulation reference in term of monthly surface temperature averaged over all assimilated grid cells at the beginning of the twin experiment. In red, ensemble members, in blue the simulation of reference, (truth). Experiment done using 100 particles with the NETF without inflation.



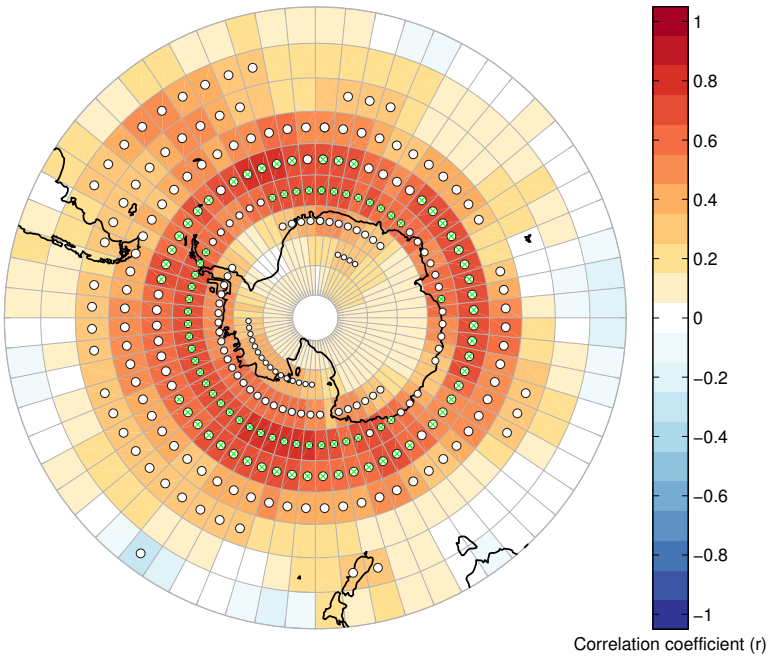


Figure 4.3: Maps of correlations coefficients, origins of assimilated data are in green, locations with significant correlation (0.05) are designated with a white circle. From top to bottom: NETF without inflation, SIR. Results obtained with 100 particles.

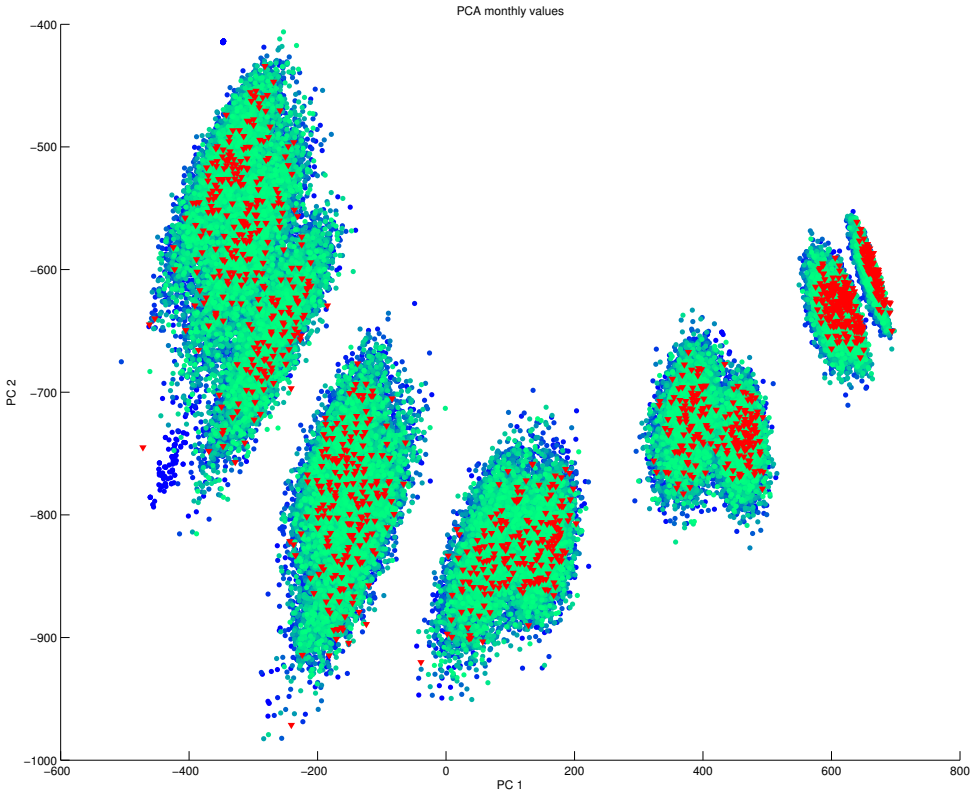


Figure 4.4: Principal components representation of the experiment with 1000 particles using the NETF. the analysis is based on annual surface temperature of each ensemble member. Red triangles represent states from the simulation of reference, dots represent ensemble members and the blue to green colouration correspond to the time in the simulation.

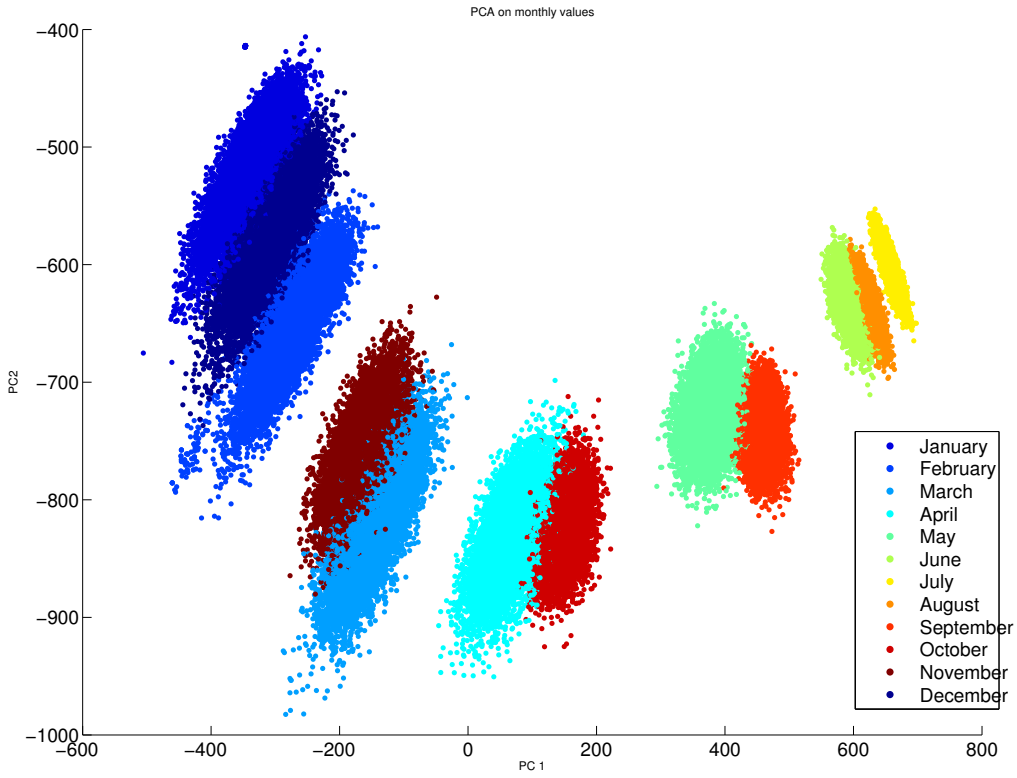


Figure 4.5: Principal components representation of the experiment with 1000 particles using the NETF. The analysis is based on monthly surface temperature of each ensemble member. Red triangles represent state from the simulation of reference, dots represent ensemble members. Each colours correspond to a particular month.

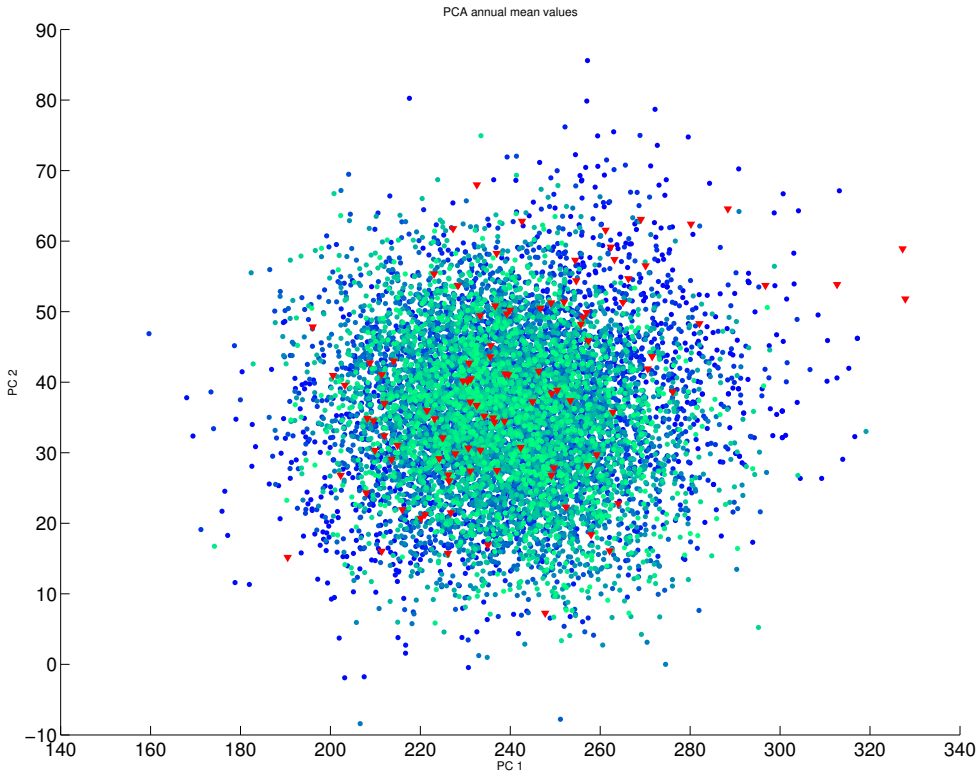


Figure 4.6: Principal components representation of the experiment with 1000 particles using the NETF. The analysis is based on annual surface temperature of each ensemble member. Red triangles represent state from the simulation of reference, dots represent ensemble members and the blue to green colouration correspond to the time in the simulation.

Chapter 5

Conclusion

In this work we have started by describing a selection of data assimilations methods from a theoretical point of view. In particular we concentrated our efforts on sequential data assimilation methods making use of ensemble. Those methods have the advantage to be able to use a model directly. Moreover, particles are considered independent between assimilation phases therefore allowing their computation in parallel.

Using twin experiments, we tested a selection of filters on the Lorenz 63 system in order to assess their performances. Then, we studied the impact of different parameters, such as the ensemble size and the period between successive observations. Using results from these experiments, we find that the Non linear ensemble transform filter, recently presented in [26] and [27], performed well even with relatively small ensemble and with sparse observations. The NETF outperform the sequential importance resampling filter both in terms of average results an robustness. Although this filter can be formulated as a square root filter, like the EnKF, it is completely compatible with non linear model and non Gaussian observation noises, unlike the EnKF. This filter is, by construction, a relaxation of the classical sequential importance resampling filter in the sense that it enforces the same ensemble mean and covariance but relaxes the Bayes theorem for higher order moments. For the nonlinear ensemble transform filter we also inspected the impact of the ensemble inflation and highlighted the trade off between uncertainty and robustness.

We then implemented the NETF on the LOVECLIM model using an offline assimilation algorithm. Using an already implemented particle filter as a baseline we validated our implementation. Moreover we saw that the two filters managed similar performances in this kind of experimental set up.

Finally, we adapted our implementation of the NETF for online assimilation with the model LOVECLIM. We ran twin experiments using 10 particles with the NETF with and without inflation and compared results with an already available particle filter. This step showed that our implementation is correct. It also showed that our filter can be useful for application in palaeoclimatology by producing reconstruction with superior correlation on assimilated variables than what was obtained with the existing particle filter for a small ensemble, (see table 4.1). A twin experiments using 100 particles and without inflation was then performed. In this last experiment, our filter obtained correlation values slightly inferior to those of the older particle filter except for assimilated variables (see table 4.2).

We have shown that the non-linear ensemble transform filter is better than the classical sequential importance resampling particle filter on a low dimensional systems, both in term of

average performances and in term of robustness. We have successfully implemented this filter for its utilization with the LOVECLIM model. Although we were not able to outperformed the classical particle filter, the NETF appears to be an acceptable data assimilation method for this kind of applications. A plausible justification for these results is the the fact that we only ran a single twin experiment. On Lorenz 63 our filter was more robust than the classical particle filter but we cannot generalize to experiment with LOVECLIM with only one experiment. Nonetheless, we suspect that the NETF with random rotation and eventually a small inflation is more robust and could be useful for real scale applications in palaeoclimatology.

Despite our efforts, some points still need to be inspected. In order to compare the robustness of the NETF and classical particle filter when using the LOVECLIM model, one could run multiple twin experiments using different but comparable set-ups and inspect performances variability. This would require numerous twin experiments and would be computationally intensive. Another method would be to compare the methods on a smaller set of pathological experiments where the classical particle filter is known to diverge. This would require less experiments to reach a conclusion but designing pathological experiments could be a non trivial task and a good understanding of underlying physical phenomena could be useful. The robustness of the NETF is drastically dependent on the presence of a mean and covariance preserving rotation in the ensemble space. For the moment, the construction of the rotation is based on randomly generated matrices. The corresponding degrees of freedom could potentially be exploited to improve the filter. Another interesting subject, from a physical perspective, would be to inspect how the NETF affects physical balances.

Appendix A

Filter's impact on prior pdf

The idea is to visualize the impact of different filters on prior distributions. We will work with two dimensional state prior distributions represented with particle ensembles and apply different method and qualitatively compare the resulting posterior distributions.
In figure A.1 the prior ensemble is presented as black dots, we assimilate a data $x = 0$ assuming Gaussian error with variance of $R = 2$ with different methods. We chose to try the EnKF, the SIR and NETF with and without rotation. As the prior distribution is multi-modal with $x = 0$ and $y = 0$ as axis of symmetry one may hope to obtain multi-modal posterior distributions.
On the first panel we can observe that the EnKF contracts the ensemble toward the subspace corresponding to the observation. Although the prior distribution is clearly non Gaussian the EnKF works under the assumption that the prior pdf correspond to a Gaussian distribution with the same mean and covariance, therefore the resulting posterior ensemble doesn't carry meaningful information other than the ensemble mean and covariance. In the second panel an importance resampling scheme is use to generate the posterior ensemble (blue dots). We can clearly see the impact of random perturbations added after resampling to preserve the ensemble diversity. The nature of the prior distribution is preserved but the filter cannot move the ensemble in a region of the state space where no particle was present previously.
In the third panel, the ensemble in green is generated with a NETF without rotation. The posterior ensemble mean and covariance correspond to those obtained with the SIR. however, particles are not just perturbed copies of prior particles. Although some kind of symmetry is preserved the NETF provide an ensemble that doesn't look like the SIR ensemble as only the first and second order moments are equivalent by construction. Therefore applying the random rotation, (that preserves both the mean and ensemble covariance), can be applied without regret. We obtain the ensemble represented in the last panel. The randomness present in the rotation result in a seemingly random ensemble, but per construction first and second order moment are equivalent to those obtained with the the SIR.

In the previous example, the observation correspond to a symmetry axis of the prior ensemble, and the observation error distribution was also symmetric therefore result were symmetric. Now let's move the observation to the left and generate the figure A.2. Note that in this case the EnKF procudes a bimodal distribution that seem to correspond to the prior mods but this is only a coincidence in this example. The SIR select copies of the most likely particles and the resulting ensemble preserve higher order features of the prior distribution. The NETF without rotation somewhat preserve theses features but the rotation completely erases higher order moments. However, due to the absence of guarantee on any other things than the mean and covariance, higher order features are equally unreliable in the two variants of the NETF. Therefore, as the random rotation is known to increase filter stability this variant should be preferred [26].

At last we can go back to the first example but use a different distribution for observation noises. If this distribution is skewed, the likelihood of each mode can be affected even if the observation doesn't directly favours one mode over the others. in our case we used a probability distribution favouring positive noises. Despite the observation being the vertical axis of symmetry of the ensemble, the most likely particles are on the left side of the observation. We already know that the EnKF doesn't provide a way to deal explicitly with this issue. In figure A.3 we can observe the impact of this modification. Because the EnKF doesn't take the modification into account the posterior is the same as in the first example. However, the SIR is heavily impacted and present a distribution akin to the posterior of the second example. This example highlights the necessity to take into account the non Gaussian characteristic present in the observation error distribution.

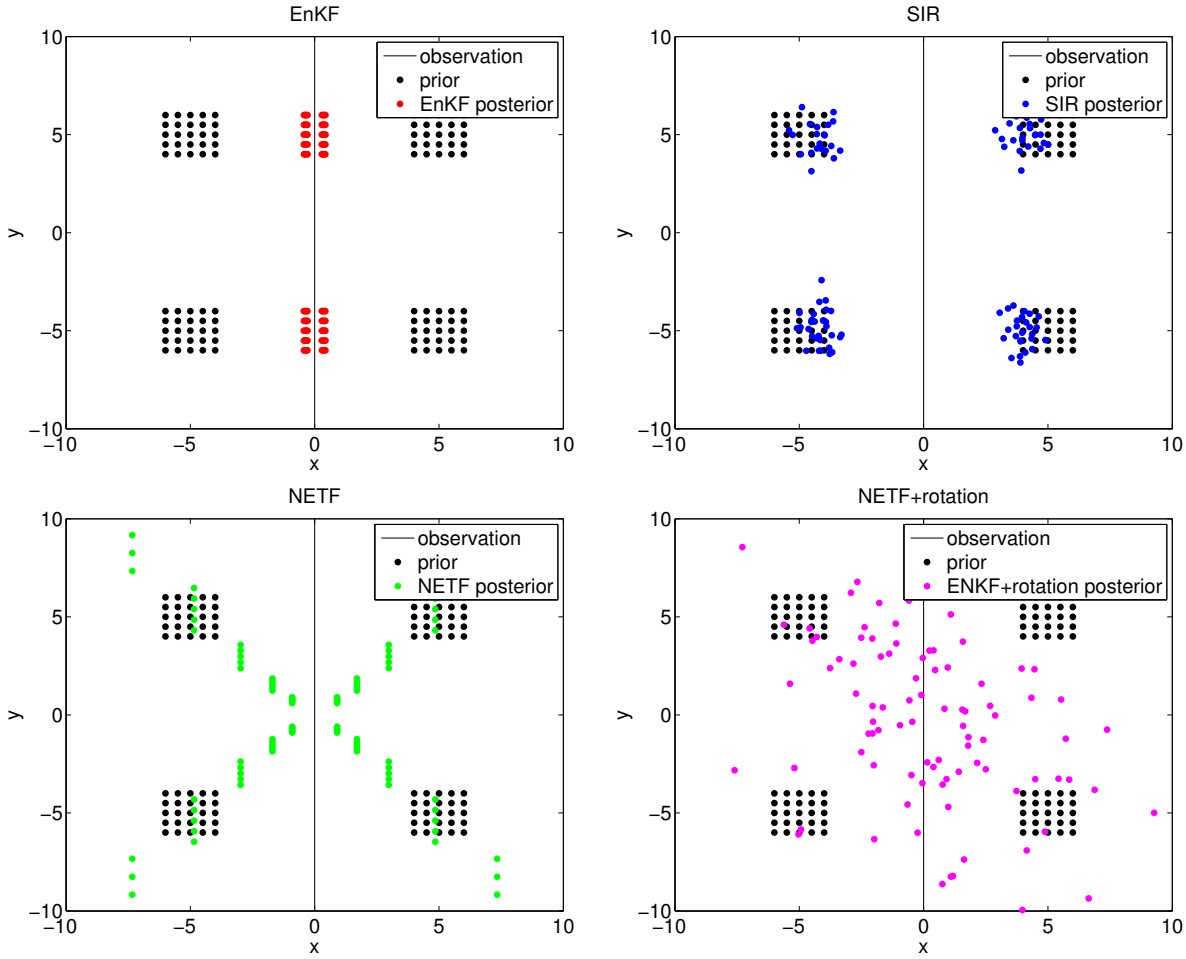


Figure A.1: Impact of different filters on a non trivial prior ensemble. Black dots: prior ensemble, Black line: observation assimilated, color dots: posterior ensemble with from left to right: EnKF, SIR, NETF without rotation and NETF with rotation.

In our example, the SIR produce a bimodal distribution, a property lost by the NETF, especially when the rotation is apply. The rotation take place in the ensemble space and replace each particle by an appropriate linear combination of particles. Although the NETF without rotation may sometimes seems to somewhat preserve other features than the mean and covari-

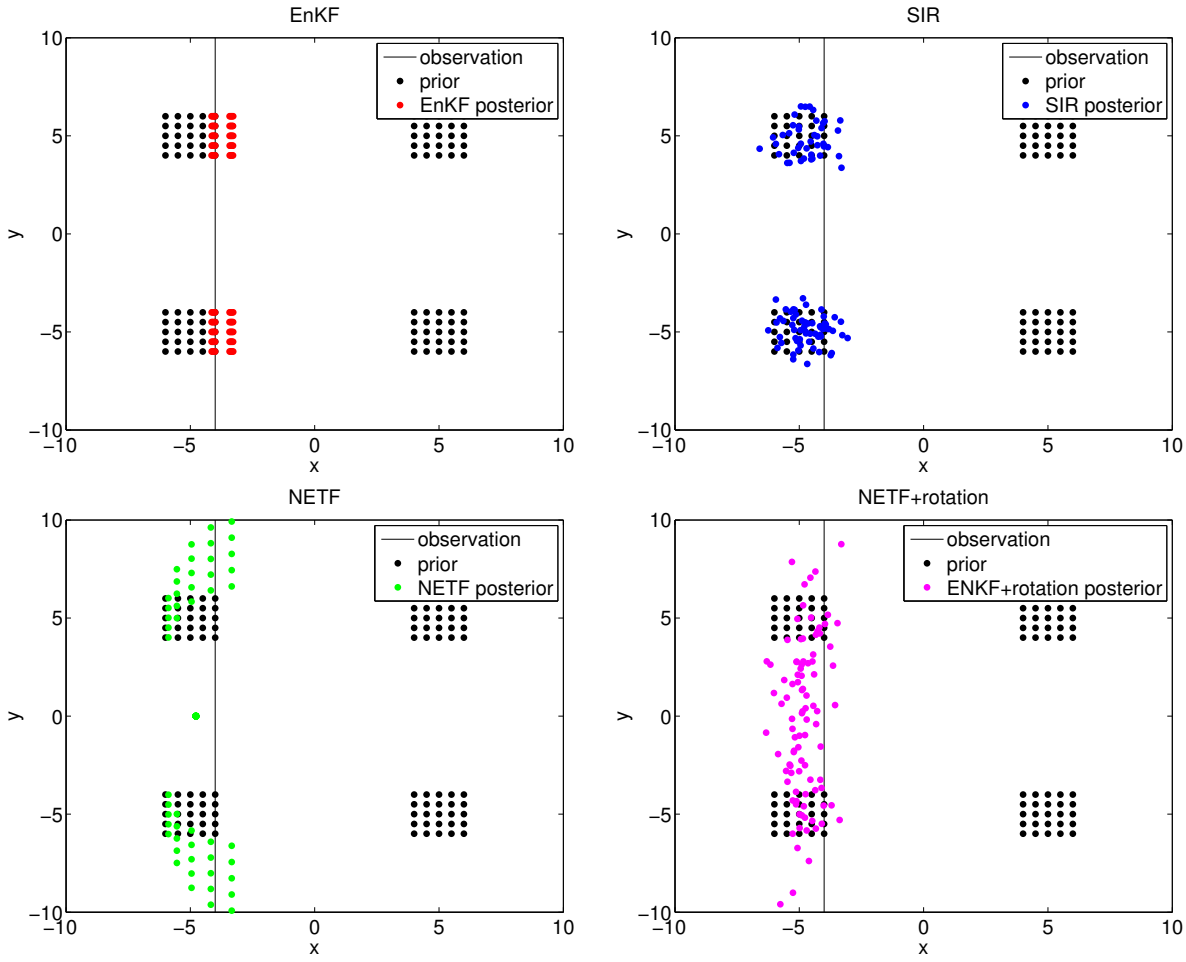


Figure A.2: Impact of different filters on a non trivial prior ensemble. Black dots: prior ensemble, Black line: observation assimilated, color dots: posterior ensemble with from left to right: EnKF, SIR, NETF without rotation and NETF with rotation.

ance, they are difficult to exploit and should not be seen as meaningful.

In figure A.4 particles of the prior ensemble are linked by dotted lines to their image after transformation by a NETF without rotation. We consider different prior distribution and observations. It appears that some particles are strongly modified and moved toward the subspace define by the observation. this tends to create set of particle in location where no prior particle were, thus creating modes unrelated with the prior distribution. In order to maintain appropriate mean and covariance, some particle are moved in strange directions. This last feature tends to get rid off higher order moments from the prior distribution. We can conclude this section by asserting that the NETF doesn't provide useful information on higher order moments and creates modes in the distribution. Therefore applying the rotation allow the filter to break those modes and sample trough the state space in a more homogeneous way. The SIR is able to preserve higher order moments, the NETF doesn't have this ability any more but allow some particles to move toward the observation while preserving the correct mean and covariance. Therefore using the ensemble mean as a final reconstruction would produce equivalent results. The contrary to the SIR the NETF is able to recover from filter divergence. The rotation decrease the risk of filter divergence by re-combining posterior particles in order

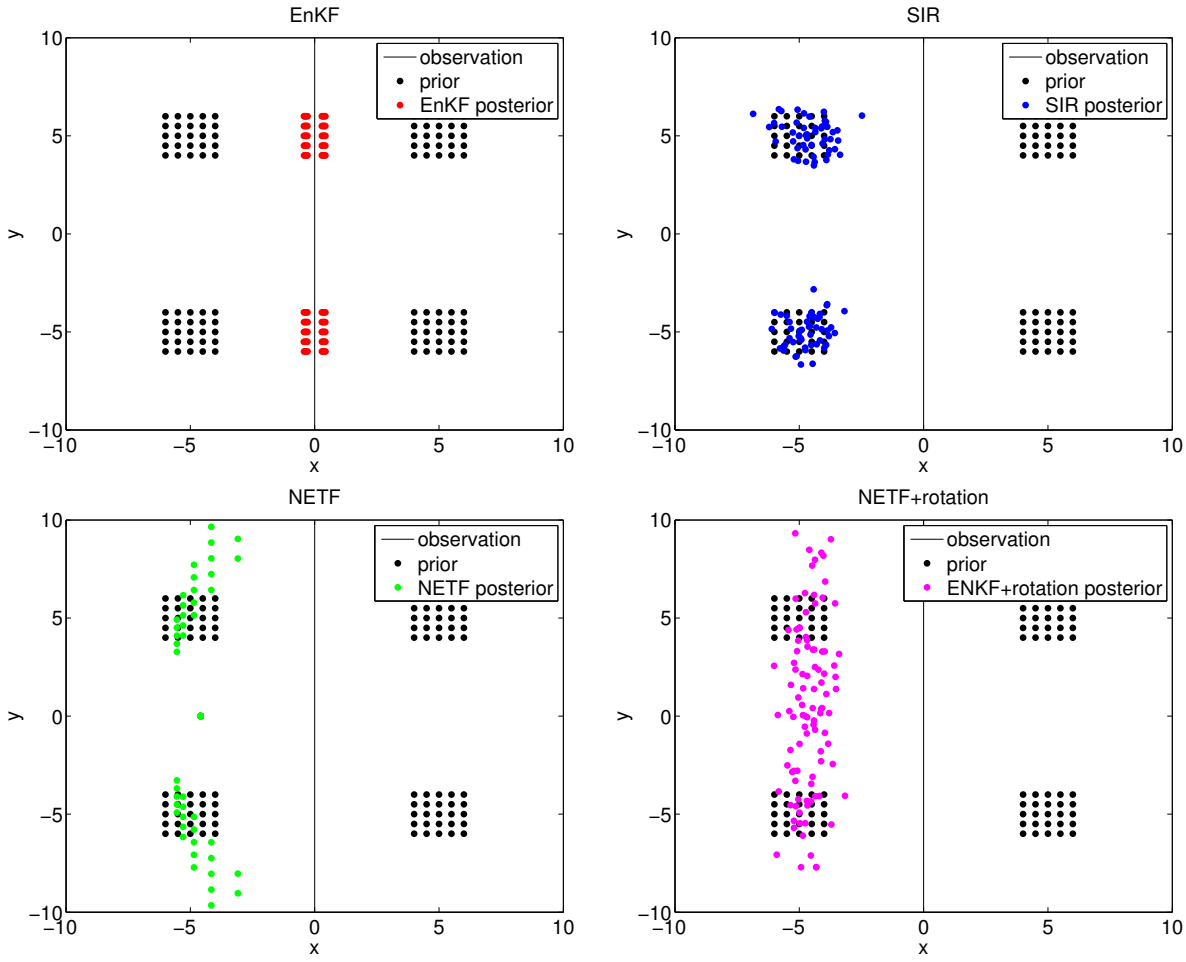


Figure A.3: Impact of different filters on a non trivial prior ensemble. Black dots: prior ensemble, Black line: observation assimilated, color dots: posterior ensemble with from left to right: EnKF, SIR, NETF without rotation and NETF with rotation.

to produce a new distribution with the correct moments. Moreover the rotation break away and scatter particle clusters created by the NETF transformation.

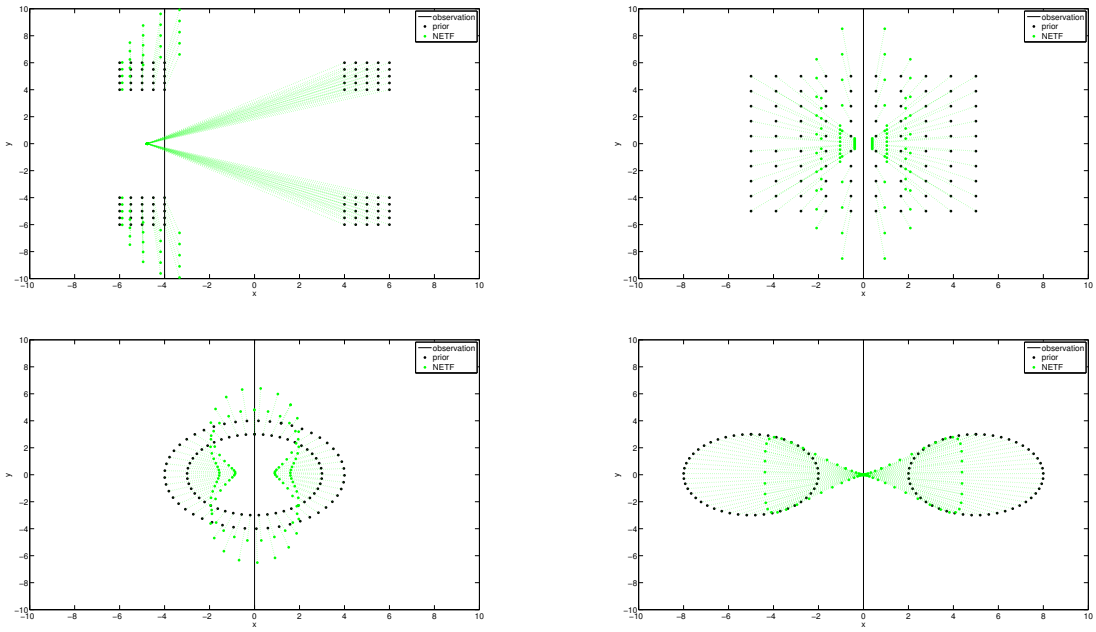


Figure A.4: Impact of the NETF transformation for different prior ensembles. Black line: the observation, Black dots: prior ensemble, Green dots: posterior ensemble generated with a NETF without rotation. Dotted green lines link prior particles to their images after transformation

Appendix B

Poor nudging scheme

One of the main issue with the EW particle filter or associated method is the definition of the nudging term. Although the idea of a model relaxation can be attractive it may be counter productive even in simple cases. For instance, consider an harmonic oscillator formed by a mass m and a spring with stiffness k. The corresponding differential equation can be written as

$$\frac{dv}{dt} = -(k/m)x \tag{B.1}$$

$$\frac{dx}{dt} = v \tag{B.2}$$

$$\tag{B.3}$$

where x is the mass position relative to the equilibrium point and v the speed of the mobile mass. We try to pull the position toward an observation y with a simple scheme $F\tau(t)(y-x)$ where $\tau = t/t_{obs}$. Numerical results can be seen in figure B.1. Observe how the simulation without nudging term performs better than other simulations. The new dynamics can easily be written as

$$\frac{dv}{dt} = -k/mx \tag{B.4}$$

$$\frac{dx}{dt} = v + F\tau(t)(y-x) \tag{B.5}$$

In this simple case the nudging term perturb a dynamics that would leads to good results if unperturbed. In general, such behaviour is hard to detect beforehand and nudging scheme must be carefully designed. Moreover, there is no guarantee that nudging method maintain system invariants and a poorly construct nudging scheme could move a particle outside the proper basin of attraction thus impacting eventual stability properties.
This simple example highlight the issue with nudging terms present in the EW particle filter, perturbing the model dynamics may have unfortunate repercussions. In [8], a nudging scheme is applied to LOVECLIM, a model of intermediate complexity, in a twin experiment. Pseudo-observations of surface temperature are used to constraint the model. A filter with a nudging proposal scheme showed a slight improvement on a SIR for observed variables and a clear improvement on the sea-ice concentration in term of correlation. The nudging on sea surface temperature took the following form

$$x(t) = f(x(t-1))/\alpha H^T(y - H(x(t-1))$$

Which correspond to a modification of the heat flux between the ocean and the atmosphere.
However, nudging scheme perturbs the system dynamics, in this case wrong patterns of sea surface salinity were observed as vertical ocean temperature profile and vertical salinity profile

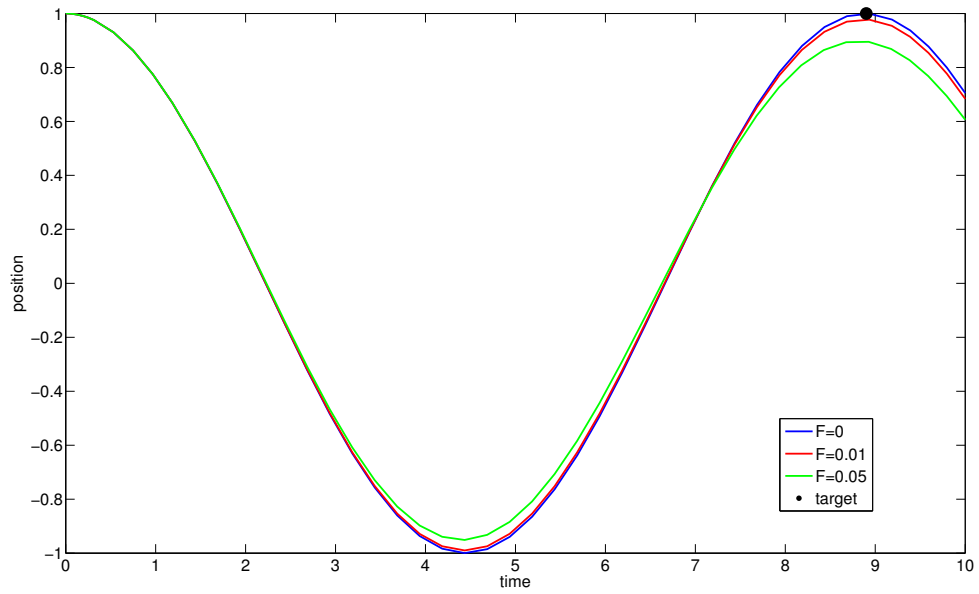


Figure B.1: Evoultion of an harmonic oscillator with different nudging strength.

were not consistent, leading to a bad vertical stratification.

However, in [3] an exploration of the impact of the equivalent weight particle filter have been realised on a single-layer ocean model and one of the conclusion is that keeping the nudging term smaller (in term of norm or energetic perturbation) than the perturbation coming from the model uncertainties. In these circumstances, dynamics is mainly impacted at the deterministic step just before observation time. Moreover, by construction the corresponding perturbation is comparable to the perturbation encounter in EnKF analysis steps.
To conclude, nudging scheme must be designed carefully and are application dependent. This is one of our justification for discarding the EW in our large scale application.

Bibliography

[1] M. Ades and P. J. Van Leeuwen. An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 139(672):820–840, 2013.

[2] M. Ades and P. J. van Leeuwen. The equivalent-weights particle filter in a high-dimensional system. *Quarterly Journal of the Royal Meteorological Society*, 141(687):484–503, January 2015.

[3] Melanie Ades and Peter Jan van Leeuwen. The Effect of the Equivalent-Weights Particle Filter on Dynamical Balance in a Primitive Equation Model. *Monthly Weather Review*, 143(2):581–596, February 2015.

[4] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188, 2002.

[5] Marc Bocquet, Carlos A. Pires, and Lin Wu. Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation. *Monthly Weather Review*, 138(8):2997–3023, August 2010.

[6] Nawinda Chustagulprom, Sebastian Reich, and Maria Reinhardt. A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):592–608, 2016.

[7] Sébastien Dirren and Gregory J. Hakim. Toward the assimilation of time-averaged observations: assimilation of time-averaged observation. *Geophysical Research Letters*, 32(4):n/a–n/a, February 2005.

[8] S. Dubinkina and H. Goosse. An assessment of particle filtering methods and nudging for climate state recostructions. *Clim. past*, 9:1141–1152, 2013.

[9] Geir Evensen. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, November 2003.

[10] Geir Evensen. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54(6):539–560, December 2004.

[11] Geir Evensen. Mean preserving rotations in SQRT schemes. 2007.

[12] Geir Evensen, Joakim Hove, Hilde Meisingset, Edel Reiso, Knut Sponheim Seim, Øystein Espelid, and others. Using the EnKF for assisted history matching of a North Sea reservoir model. In *SPE reservoir simulation symposium*. Society of Petroleum Engineers, 2007.

[13] Mohinder S. Grewal and James Kain. Kalman filter implementation with improved numerical properties. *IEEE transactions on automatic control*, 55(9):2058–2068, 2010.

[14] Goosse H., Brovkin V., Fichefet T., and et al. Description of the earth system model of intermediate complexity loveclim version 1.2. *Geoscientific Model Development*, 2010.

[15] Helga S. Huntley and Gregory J. Hakim. Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Climate Dynamics*, 35(6):995–1009, November 2010.

[16] Jouzel J., R.D. Koster, R.J. Suozzo, and G.L. Russell. Stable water isotope behavior during the last glacial maximum: A general circulation model analysis. *Journal of geophysical research*, 1994.

[17] Wolfgang Kurtz, Guowei He, Stefan J. Kollet, Reed M. Maxwell, Harry Vereecken, and Harrie-Jan Hendricks Franssen. TerrSysMP–PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface–subsurface model. *Geoscientific Model Development*, 9(4):1341–1360, April 2016.

[18] Tiancheng Li, Shudong Sun, Tariq Pervez Sattar, and Juan Manuel Corchado. Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches. *Expert Systems with applications*, 41(8):3944–3954, 2014.

[19] Pavel Sakov, Geir Evensen, and Laurent Bertino. Asynchronous data assimilation with the EnKF: asynchronous data assimilation with the ENKF. *Tellus A*, 62(1):24–29, January 2010.

[20] Pavel Sakov, Dean S. Oliver, and Laurent Bertino. An Iterative EnKF for Strongly Nonlinear Systems. *Monthly Weather Review*, 140(6):1988–2004, June 2012.

[21] Laura Slivinski and Chris Snyder. Exploring practical estimates of the ensemble size necessary for particle filters. *Monthly Weather Review*, 144(3):861–875, 2016.

[22] Steven H. Strogatz. *NonLinear Dynamics and chaos with applications to physics, biology, chemistry and engineering. second edition.* Westview press, 2015.

[23] Simo Särkkä. *Bayesian filtering and smoothing*, volume 3. Cambridge University Press, 2013.

[24] Michael K. Tippett, Jeffrey L. Anderson, Craig H. Bishop, Thomas M. Hamill, and Jeffrey S. Whitaker. Ensemble square root filters. *Monthly Weather Review*, 131(7):1485–1490, 2003.

[25] Julian Tödter and Bodo Ahren. Generalization of the ignorance score: Continuous ranked version and its decomposition. *Monthly Weather Review*, 140:2005–2017, June 2012.

[26] Julian Tödter and Bodo Ahrens. A Second-Order Exact Ensemble Square Root Filter for Nonlinear Data Assimilation. *Monthly Weather Review*, 143(4):1347–1367, April 2015.

[27] Julian Tödter, Paul Kirchgessner, Lars Nerger, and Bodo Ahrens. Assessment of a Nonlinear Ensemble Transform Filter for High-Dimensional Data Assimilation. *Monthly Weather Review*, 144(1):409–427, January 2016.

[28] P. J. Van Leeuwen. Particle filters for the geosciences. *Advanced Data Assimilation for Geosciences: Lecture Notes of the Les Houches School of Physics: Special Issue*, page 291, 2012.

[29] Mengbin Zhu, Peter Jan van Leeuwen, and Javier Amezcua. Implicit equal-weights particle filter: implicit equal-weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 142(698):1904–1919, July 2016.