Efficient parameter estimation for a highly chaotic system

J.D. Annan[1] and J.C. Hargreaves

Frontier Research System for Global Change

3173-25 Showa-machi Kanazawa-ku

Yokohama, Kanagawa 236-0001, Japan

and

Proudman Oceanographic Laboratory

Joseph Proudman Building

6 Brownlow Street, Liverpool, L3 5DA, UK

5 April 2004

[1]Corresponding author: `jdannan@jamstec.go.jp`

Abstract

We present a practical, efficient and powerful solution to the problem of parameter estimation in highly nonlinear models. The method is based on the ensemble Kalman filter, and has previously been successfully applied to a simple climate model with steady state dynamics. We demonstrate, via application to the well-known Lorenz (1963) model, that the method can successfully perform multivariate parameter estimation even in the presence of chaotic dynamics. Traditional variational methods using an adjoint model have limited applicability to problems of this nature, and the alternative of a brute force (or randomised) search in parameter space is prohibitively expensive for high dimensional applications. The cost of our method is comparable to that of integrating an ensemble to statistical convergence, and therefore this technique appears to be ideally suited for probabilistic climate prediction.

1 Introduction

Parameter estimation is increasingly recognised as having a critical rôle to play in climate prediction (Andronova & Schlesinger, 2001; Knutti *et al.*, 2002; Gregory *et al.*, 2002). Although short-term numerical weather prediction is essentially a problem of initial state estimation, the behaviour of a coupled ocean-atmosphere model on climatological time scales is determined much more by the details of its parameterisations than by the initial conditions of the model state variables. Furthermore, as well as determining the forecast mean, it is also the uncertainty in parameterisations that generates the range of uncertainty in climate forecasts (at least for a given scenario of external forcing, the determination of which is primarily a socio-economic rather than geophysical question).

Various methods have been previously used for parameter estimation in this field, but all of them have substantial drawbacks. In particular, direct perturbation methods based on an exhaustive or random exploration of parameter space (Dickinson & Gelinas, 1976; Dilks *et al.*, 1992) are hugely expensive in computational terms (the cost increases exponentially with the number of free parameters), and the direct application of variational methods using an adjoint model cannot cope with the hypersensitivity to infinitesimal perturbations that is characteristic of chaotic models under long integrations (Lea *et al.*, 2000, 2002). Recently, Annan *et al.* (2004) introduced an efficient method based on the ensemble Kalman filter, or EnKF (Evensen, 1994b), and applied it to the simultaneous estimation of 12 parameters in a low-resolution coupled atmosphere-ocean model. The method worked successfully both in the original identical twin testing, and also in a subsequent application using real data. Therefore, application to more detailed and realistic models seems justified. However, the previous testing was somewhat limited in its scope. One particular limitation which may be of some significance for application to more realistic models, is that the climate model used in those initial tests has a very low spatial resolution and highly diffusive atmosphere, so much so that (under the steady annual mean solar forcing which was used) the model converges to a true steady state with virtually no temporal variability let alone any internal chaotic dynamics. Under such conditions, an adjoint model could also be expected to work well, but the chaotic dynamics of a higher resolution ocean model is known to pose a much more difficult problem (Lea *et al.*, 2002).

In this paper, we investigate the applicability of the EnKF method to parameter estimation in the highly chaotic model of Lorenz (1963). As well as being widely used for assimilation and forecasting research, this model is particularly interesting as a test case for parameter estimation because it was the subject of the study by Lea *et al.* (2000), and therefore the development of a successful parameter estimation method would represent a significant advance on previous techniques. In Section 2, we outline the parameter estimation problem addressed in this paper, and present our method. Section 3 contains the results of the numerical experiments and some discussion, and we conclude the paper in Section 4.

2 Model and Methods

2.1 Parameter estimation in the Lorenz model

The Lorenz model (Lorenz, 1963) has been widely used for studies involving prediction and data assimilation in chaotic systems (eg Miller *et al.*, 1994; Evensen, 1994a). The model consists of 3 variables x, y and z, which evolve according to the equations

$$x' = \sigma(y - x) \tag{1}$$

$$y' = rx - y - xz \tag{2}$$

$$z' = xy - bz \tag{3}$$

where σ, b and r are three constant parameters that are generally given the values 10, 8/3 and 28 respectively. For these values, the model variables follow a highly chaotic orbit. All calculations in this paper were performed using a 4th order Runge-Kutta method (Press *et al.*, 1994, Section 16.1) with a time step of 0.01. The outputs of two typical model runs are shown in Figure 1.

Parameter estimation in this model was studied by Lea *et al.* (2000). They addressed the simplest possible problem, of the sensitivity of the climatological mean of a single variable $\overline{z}$ to changes in a single parameter r. We note in passing that, as these authors mention, the task of a 'sensitivity analysis' is generally not undertaken as an end in itself, but is instead generally pursued only as a means to the end of parameter estimation and the related tasks of data assimilation and the analysis of uncertainty. We are focussing directly on the parameter estimation problem itself.

The upper plot in Figure 2 shows the variation of $\overline{z}$ with r (which we may write as $\overline{z}(r)$ to emphasise the functional relationship), as calculated by a direct integration over a range of values of r for an interval of 10,000 time units, with initial conditions for x, y and z fixed. The approximately linear relationship is clear, although there is also what appears to be substantial numerical noise. In fact this variability is entirely deterministic, and is a feature of the sensitive dependence of model behaviour on initial conditions (including parameter values) which is characteristic of chaotic models. The lower plot shows the cost function given by $C(r) = \frac{1}{2}\left(\frac{\overline{z}(r)-25.55}{0.1}\right)^2$, which is typical of what we might seek to minimise by a parameter estimation scheme (assuming in this case a single observation defined by $\overline{z}_o = 25.55 \pm 0.1$). It is immediately clear why the adjoint model does not help in this task, since the local gradient of the cost function (which is what an adjoint model calculates) gives no useful indication of the underlying near-quadratic curve. If the integration period is lengthened, then the amplitude of the deterministic 'noise' decreases gradually (with the square root of the run length, as if it really was generated by a stochastic process), but the density of peaks and troughs increases exponentially, so the typical gradient gets ever steeper. If, on the other hand, one only performs a short integration, then the adjoint indicates the transient sensitivity rather than the desired climatological response. Using some somewhat ad-hoc methods based on ensemble averaging and carefully selecting the duration of the integration interval, Lea *et al.* (2000) were able to generate some helpful results for this problem, but concluded that method would probably be difficult to apply to more complex situations. These conclusions were confirmed by later work using an eddy-permitting ocean model (Lea *et al.*, 2002). Recently, Moolenaar & Stelten (2004) have extended these ideas further, finding some limited relationships between the short and long term behaviour of a forced version of the Lorenz model. However, the applicability of their ideas to realistic climate models with a range of chaotic time scales is not yet established.

The variables x and y oscillate around zero, and this is their climatological mean value which does not depend on any model parameters. However, the magnitude of their oscillations (as measured by the root mean square), and also the magnitude of the oscillations of z about its mean, do vary somewhat as the model parameters vary. In fact within reasonable limits, the macroscopic relationships are again quite close to linear (we exclude *a priori* extremely large parameter changes which qualitatively change the nature of the model), however the climatological diagnostics also display the chaotic noise on microscopic scales similar to that in Figure 2. Rather than restricting our attention to the simple univariate problem considered by Lea *et al.* (2000), we will attempt to estimate all three model parameters σ, b and r simultaneously from estimates of the climatological values of $\overline{z}$, $||x||$, $||y||$ and $||z-\overline{z}||$ where the norm used is the root mean square.

Figure 1 provides an example of how these climatological diagnostics of model output varies when parameters are changed. The two model runs (left and right) were performed using different sets of parameters. The lowest horizonal line on each side of the plot indicates $\overline{z}$, and the remaining horizonal lines indicate the amplitude of the oscillations of each variable.

2.2 The Ensemble Kalman Filter and parameter estimation

The ensemble Kalman filter (Evensen, 1994b) is an efficient Monte Carlo approximation to the optimal Kalman filter equations (Kalman, 1960). It has been widely used in near-operational forecasting, especially for short-term numerical weather and ocean prediction. A clear exposition of the theory and basic methodology is provided in Evensen (2003), along with a survey of recent applications, and need not be repeated here. Although the EnKF has generally been used for initial state estimation, parameter estimation can readily be included in the same framework, by the means of state space augmentation (Derber, 1989; Anderson, 2001). The idea here is that the parameters can be considered to be part of the model state alongside the conventional variables. In the simplest case of temporally constant parameters (as in our problem), the model equations (1)–(3) can be formally considered to be augmented by the trivial equations

$$\sigma' = 0 \tag{4}$$

$$b' = 0 \tag{5}$$

$$r' = 0 \tag{6}$$

(which of course need not be coded explicitly) and now the state of the model is defined by the sextuple (x, y, z, σ, b, r). Although this method has been known for some time, it does not appear to have been widely exploited, perhaps due to the added difficulties of combining parameter estimation with the initial value estimation problem of short-term forecasting, and the transient effects which may dominate the forecasts. For climatological estimation, no such problems have arisen in our work.

The state augmentation method also simplifies the assimilation of asynoptic and/or nonlinear observations, which is particularly helpful in the case of observations of climatological means and more complex diagnostics such as the root mean square amplitudes of oscillations that we are using in our application. As in the case of parameters, the model state is augmented by its prediction of the measurement. For an adjoint model, using such observations would require the linearisation of (and creation of the adjoint for) the nonlinear observation operator.

Our previous application of the EnKF for parameter estimation is described in Annan *et al.* (2004). In that application, which used a low resolution fully coupled global 2D atmosphere - 3D ocean model, a parallel supercomputer was used to integrate the ensemble members simultaneously, and a domain decomposition was implemented for the analysis step following the method of Keppenne (2000). In our application here, there is no geographical representation and the entire 100-member ensemble integration and analysis is performed on a single processor. However, the iterative process to converge to the (climatological) steady state solution is implemented in the same way. At each iteration step, the ensemble is inflated by a fixed factor (ie the distance of each member from the ensemble mean is increased by a fixed proportion), and then the observations are assimilated with an adjusted estimate of observational uncertainty. The ensemble is integrated again in order to recalculate the correct covariances (and rebalance the model state), and the cycle is repeated to convergence. With the appropriate adjustment to the observational uncertainty (which depends on the inflation factor), the system converges to the same solution as would be obtained in the linear case by a single standard analysis, but this iterative method appears to be more robust in nonlinear and high dimensional applications where the prior ensemble is very sparse and may with high probability contain no samples which are close to the posterior distribution. It is therefore especially helpful in situations where one wishes to use a highly ignorant prior, in order to ascertain what can be objectively determined from the data alone without the risk of overconfidence through the use of prior expert opinion which already takes observational results into account (Allen *et al.*, 2002). In fact, it is straightforward to apply this method without any prior constraints whatsoever, although in principle the resulting problem may not be well posed (in which case the ensemble will diverge).

2.3 Experimental details

In the experiments presented here, we used an ensemble of 100 members, and integrated for 1,000 time units in order to calculate the climatological statistics at each iteration. A set of synthetic observations were generated by a preliminary run of the model with parameters set to $\sigma = 11.5$, $b = 2.87$ and $r = 32.0$, and the 4 statistics so generated were given by $\overline{z} = 27.46$, $||x|| = 8.87$, $||y|| = 10.03$ and $||z - \overline{z}|| = 9.47$. Observational errors of 0.1 were assumed for all of these values (which exceeds the sampling error which arises from the variability due to different initial conditions). Prior distributions for the three parameters were given by $\sigma = 10 \pm 2$, $b = 8/3 \pm 0.5$ and $r = 28 \pm 4$, that is they were centred on the standard parameter choices for the Lorenz model, but only weakly constrained (and not inconsistent with the actual values used in the 'truth run'). Expansion factors of 2% or 10% were used, and the initial guesses for the parameter values were drawn from a range of distributions (not necessarily the prior distribution) in order to check that the convergence occurred irrespective of these factors, albeit at different rates. The initial values for the model variables were also chosen at random throughout the ensemble. Given the length of the model run, the transient effect as they converged to the attractor for their specific parameter set was overwhelmed by the length of the subsequent integration. A total of 300 iterations were performed for the experiments, but convergence was generally much more rapid than this.

3 Results and Discussion

Figure 3 shows the evolution of the parameter values during three separate experiments. In each plot, the thick lines indicate the ensemble means for the three experiments, and the thin lines indicate the ensemble width (one standard deviation). The different colours refer to different numerical experiments. For the black lines, an expansion factor of 10% was used. For the red and blue lines, an expansion factor of 2% was used. In the blue experiment, the initial guesses for the parameters in the ensemble were taken from a rather narrow distribution, and so it takes many iterations for the ensemble to expand and converge to the final solution. In the red experiment, the initial guess is much broader, and convergence is faster, comparable to when the larger expansion factor was used. The clear result is that each experiment has converged to the same solution, in terms of the both the ensemble's mean and also its width. Since the ensemble is finite and its members are by construction random samples from the posterior distribution, a modest amount of numerical noise remains in the converged results due to sampling error. An additional source of sampling error arises from the use of finite length model integrations to estimate the model's climate. The ensemble's mean parameter values are fully consistent with the known solution to the problem, being very close to the parameter values which were used to generate the synthetic observations, but overall they are biased slightly towards the means of the weak prior beliefs as linear estimation theory would predict. The ensemble's climate statistics (not shown) are also very close to the observational estimates, both in mean value and in variation about this mean, exceeding the observational error estimates slightly due to the additional chaotic noise of their sample statistics. Therefore, we can conclude that the method has worked correctly in these applications.

There is a significant sample correlation in the ensemble between σ and b, but r is essentially independent of the other two parameters, and as Figure 3 shows, it is constrained in a very narrow range close to 32 in the converged ensembles. Therefore, we examine in more detail a slice through the cost function, holding $r = 32$ fixed and allowing σ and b to vary. The cost function (transformed logarithmically for numerical convenience) in this plane has been plotted in Figure 4, and the relationship between σ and b can be seen by the orientation of the optimal (low cost) region. The diamonds indicate the projection of one of the sets of converged ensemble members onto this plane, and the dotted ellipse marks the one standard deviation contour of the probability distribution that they define (determined via a principal component analysis). The ellipse is very accurately aligned with the region of low cost, confirming the success of the experiment.

Several variations on this experiment were also performed, in order to check that our results were not fortuitous or unrepresentative in some way. A range of observational errors and priors (including no prior) were tested along with different expansion factors and initial states, and different integration intervals (down to as little as 10 time units) were also tested. The results were degraded somewhat for the smallest integration intervals where the samples of the climatology are very noisy, but other than that, the results were generally of similar quality to those presented here. For longer integration intervals, the statistics of the ensemble matched the observations even more precisely in both mean and range.

There is one situation which occasionally arises, in which one member (or a small subset) of the ensemble becomes widely separated from the bulk of the distribution. As a result, the sample covariances are rather distorted and the ensemble may not converge correctly, and the isolated member(s) may even wander off into a region of qualitatively inappropriate model behaviour. This problem seems to be well known among users of the EnKF (K. Jacobs, personal communication, and see also `http://www.nersc.no/~geir/EnKF/documentation.html`) although it has not to our knowledge been mentioned in the peer-reviewed literature. It has a very simple solution, which is to eliminate the offending isolated members and resample them from the distribution defined by the remaining ensemble. This problem has only occurred near the start of our experiments when the ensemble is initialised from a very wide distribution (such as a relatively ignorant prior) with the aim of contracting towards the posterior distribution, rather than starting from a narrow initial distribution and expanding to the final answer. In this situation, the Kalman gain is initially very large and ensemble members undergo substantial adjustments with a significant random component due to the perturbed observations (Burgers *et al.*, 1998). Once the ensemble is narrow, it does not seem to spread excessively. This may be another (admittedly minor) point in favour of our iterative solution method since the ensemble can be initialised from an arbitrary distribution much narrower than the assumed prior, thus greatly reducing the likelihood of these isolated ensemble members and nonphysical solutions *a priori*.

The success of this application contrasts strongly with the failure of the adjoint method in Lea *et al.* (2000). Gradient descent, and the calculation of the gradient itself, is very vulnerable to the vast number of local minima that are present in the cost function. However, the EnKF does not use the local gradient, instead the ensemble members sample the parameter space at a macroscopic scale, and the macroscopic features of the cost function dominate the small-scale chaotic variability. When using this method, the local chaotic variation really can be considered to be nothing more pernicious than truly random noise, and since its size shrinks with integration length, it can if desired be effectively banished from the system. However, in practice a substantial amount of chaotic noise can be tolerated without badly degrading the results. Therefore, the integration interval can be shortened to affordable levels. For example, we have successfully used an integration interval of 10 time units in this application, and even after repetition in the iterative process, the total computational demand of this ensemble integration is comparable to the cost of running an ensemble of the same size to statistical convergence. We therefore conclude that this method is an eminently practical and affordable solution to the problem of parameter estimation in computationally demanding numerical models even when chaotic dynamics are present, including (but not limited to) problems of climate prediction.

4 Conclusions

The iterative ensemble Kalmn filter for parameter estimation has been successfully applied to the Lorenz model. The method has proved capable of simultaneously estimating all three model parameters, under a wide range of conditions. The method exhibits no theoretical or practical difficulty in using nonlinear, asynoptic observations of the model state, which is of particular value for tuning the model's climate where means and other nontrivial diagnostics may be of interest. The computational cost is comparable to (ie a modest multiple of) the cost of merely integrating an ensemble to climatological convergence, which represents a substantial advance in efficiency when compared to the brute force and random search methods that have been frequently applied. This method appears to be ideally suited to the problem of probabilistic climate prediction.

References

Allen, M., J. Kettleborough & D. Stainforth, 2002: Model error in weather and climate forecasting, in *Proceeduings of the 2002 ECMWF predictability seminar*, pages 275–294, ECMWF, Reading, UK.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation, *Monthly Weather Review*, **129**, 2884–2902.

Andronova, N. G. & M. E. Schlesinger, 2001: Objective estimation of the probability density function for climate sensitivity, *Journal of Geophysical Research*, **108**(D8), 22,605–22,611.

Annan, J. D., J. C. Hargreaves, N. R. Edwards & R. Marsh, 2004: Parameter estimation in an intermediate complexity Earth System Model using an ensemble Kalman filter, *Ocean Modelling*, in press.

Burgers, G., P. J. van Leeuwen & G. Evensen, 1998: On the analysis scheme in the ensemble Kalman filter, *Monthly Weather Review*, **126**, 1719–1724.

Derber, J., 1989: A variational continuous assimilation scheme, *Monthly Weather Review*, **117**, 2437–2446.

Dickinson, R. P. & R. J. Gelinas, 1976: Sensitivity analysis of ordinary differential equation systems — a direct method, *Journal of Computational Physics*, **21**, 123–143.

Dilks, D. W., R. P. Canale & P. G. Meier, 1992: Development of Bayesian Monte Carlo techniques for water quality model uncertainty, *Ecological Modelling*, **62**, 149–162.

Evensen, G., 1994a: Inverse methods and data assimilation in nonlinear ocean models, *Physica D*, **77**, 108–129.

Evensen, G., 1994b: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, **99**(C5), 10143–10162.

Evensen, G., 2003: The ensemble kalman filter: theoretical formulation and practical implementation, *Ocean Dynamics*, **53**, 343–367.

Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott & N. A. Rayner, 2002: An observationally based estimate of the climate sensitivity, *Journal of Climate*, **15**(22), 3117–3121.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems, *J. Basic Engineering*, **82D**, 33–45.

Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter, *Monthly Weather Review*, pages 1971–1981.

Knutti, R., T. F. Stocker, F. Joos & G.-K. Plattner, 2002: Constraints on radiative forcing and future climate change from observations and climate model ensembles, *Nature*, **416**, 719–723.

Lea, D. J., M. R. Allen & T. W. N. Haine, 2000: Sensitivity analysis of the climate of a chaotic system, *Tellus*, **52A**, 523–532.

Lea, D. J., T. W. N. Haine, M. R. Allen & J. A. Hansen, 2002: Sensitivity analysis of the climate of a chaotic ocean circulation model, *Q. J. R. Meteorological Soc.*, **128**(586), 2587–2605.

Lorenz, E. N., 1963: Deterministic non-periodic flow, *Journal of the Atmospheric Sciences*, **20**, 130–141.

Miller, R. N., M. Ghil & F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems, *Journal of the Atmospheric Sciences*, **51**(8), 1037–1056.

Moolenaar, H. E. & F. M. Stelten, 2004: Finding the effective parameter perturbations in atmospheric models: the Lorenz63 model as a case study, *Tellus*, **56A**, 47–55.

Press, W. H., S. A. Teukolsky, W. T. Vetterling & B. P. Flannery, 1994: *Numerical recipes in Fortran: the art of scientific computing*, Cambridge University Press.

Figure captions

Figure 1 Two example model runs with different parameter values. The horizontal lines indicate the climatological statistics (mean of z, and amplitude of the oscillations of all threee model variables) of the model runs.

Figure 2 The 'climatological' sensitivity of $\overline{z}$ on r, and cost function given by $C(r) = \frac{1}{2}\left(\frac{\overline{z}(r)-25.55}{0.1}\right)^2$

Figure 3 Convergence of parameter distributions in three numerical experiments.

Figure 4 Cost function and ensemble results in the plane $r = 32$. The contours indicate the logarithm of the cost function, and the diamonds show the projection of the results of an EnKF experiment projected onto this plane. The ellipse indicates the one standard deviation width of the ensemble.

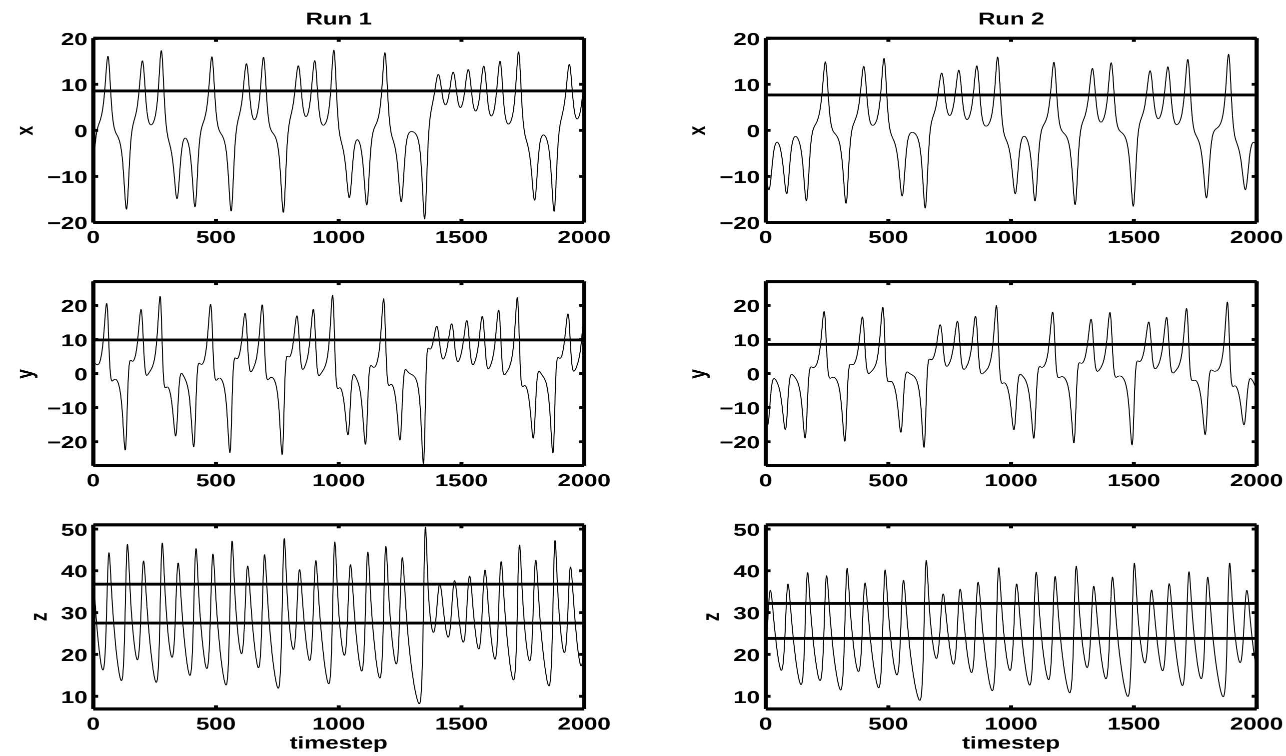

Figure 1: Two example model runs with different parameter values. The horizontal lines indicate the climatological statistics (mean of z, and amplitude of the oscillations of all threee model variables) of the model runs.

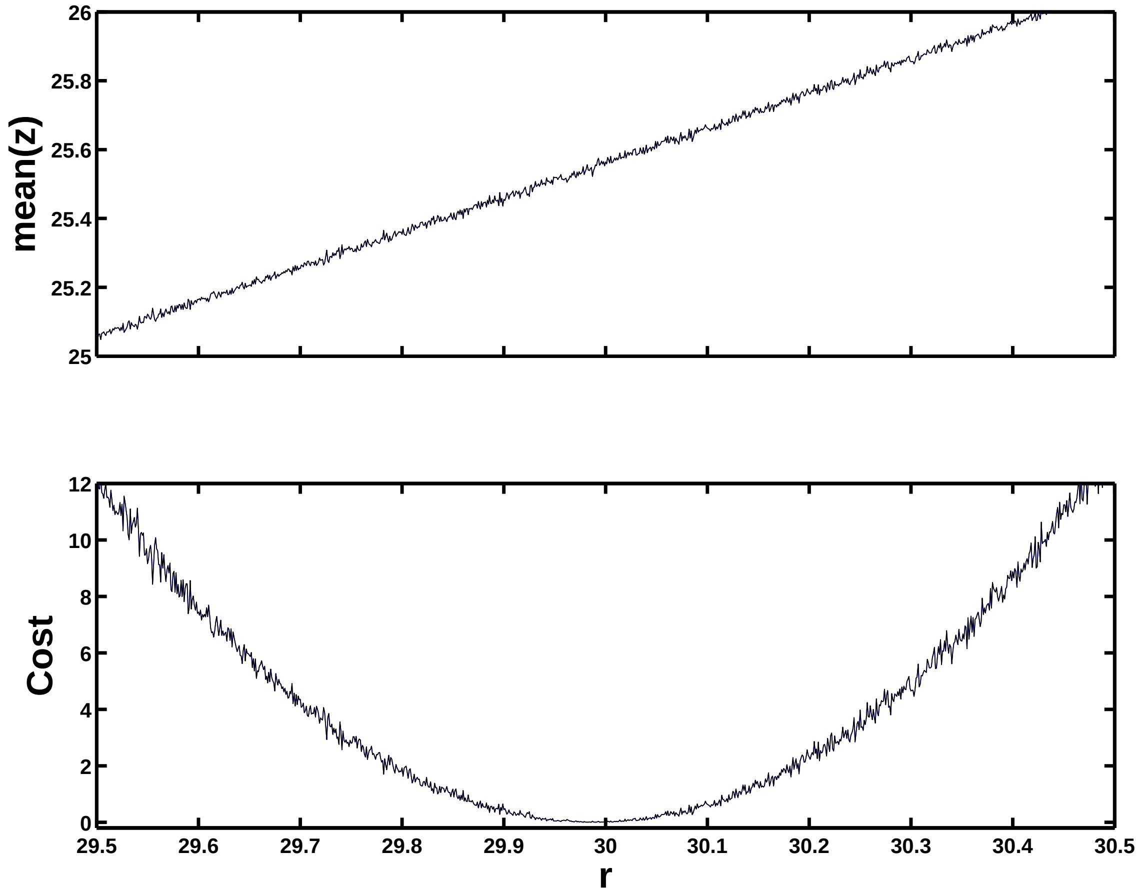

Figure 2: The 'climatological' sensitivity of $\overline{z}$ on r, and cost function given by $C(r) = \frac{1}{2}\left(\frac{\overline{z}(r)-25.55}{0.1}\right)^2$

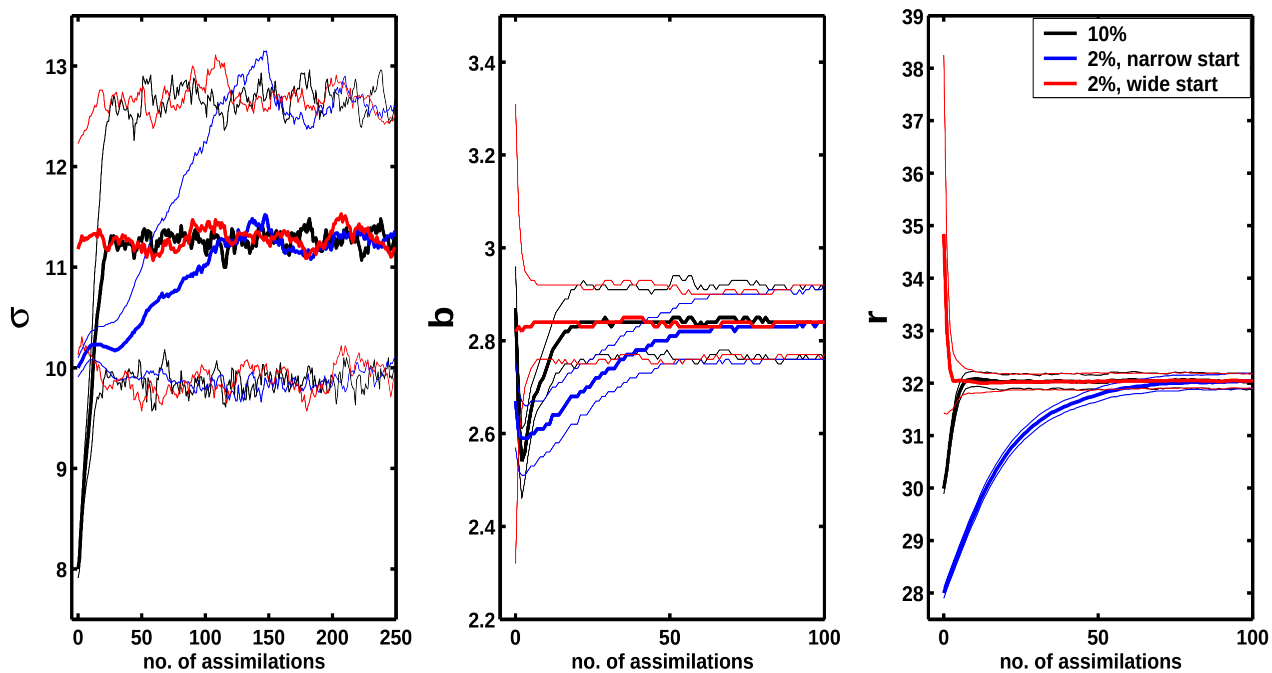

Figure 3: Convergence of parameter distributions in three numerical experiments.

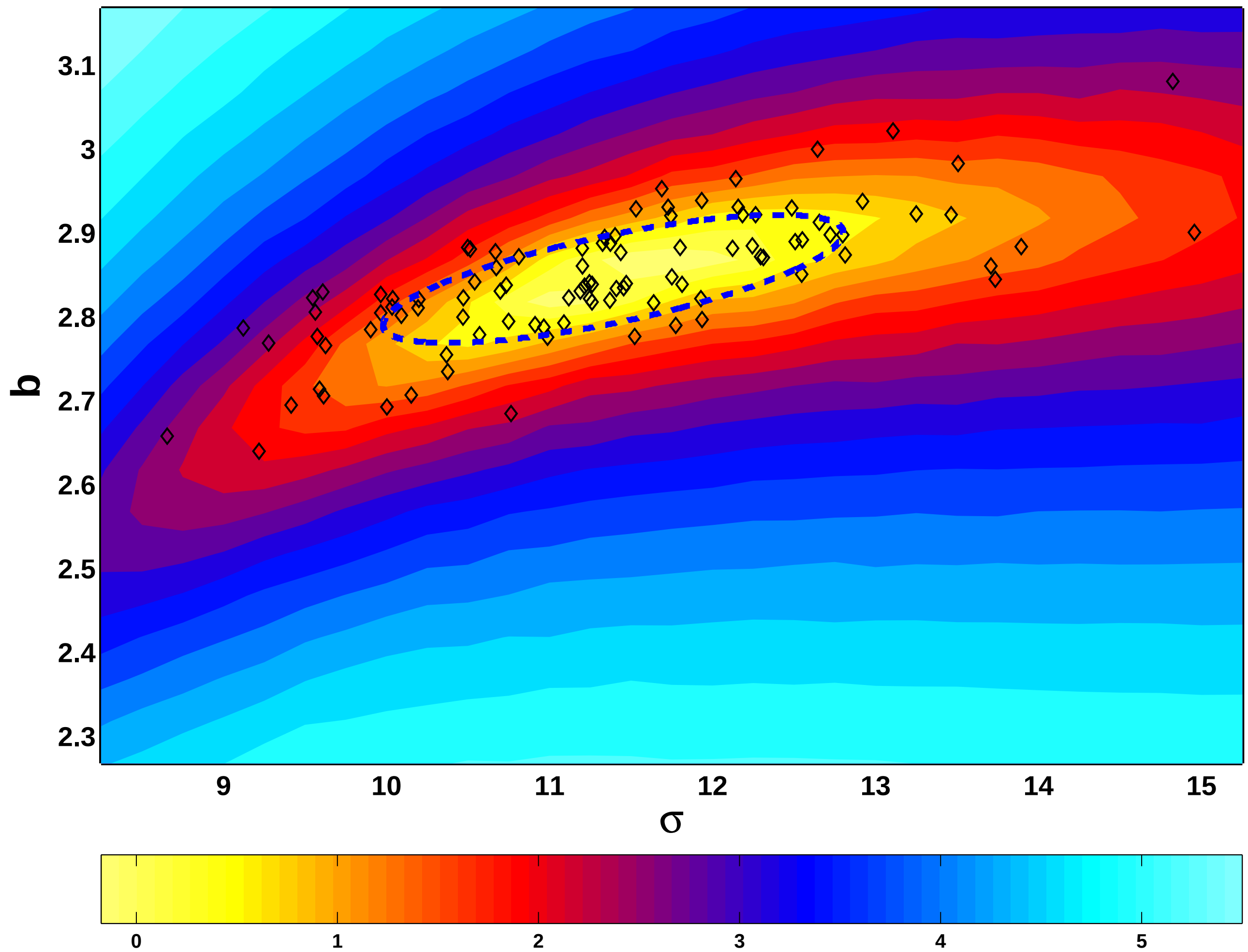

Figure 4: Cost function and ensemble results in the plane $r = 32$. The contours indicate the logarithm of the cost function, and the diamonds show the projection of the results of an EnKF experiment projected onto this plane. The ellipse indicates the one standard deviation width of the ensemble.