**Earth System
Dynamics**

# A multi-model ensemble method that combines imperfect models through learning

**L. A. van den Berge**[1], **F. M. Selten**[1], **W. Wiegerinck**[2], **and G. S. Duane**[3]

[1]Royal Netherlands Meteorological Institute, Wilhelminalaan 10, 3732 GK De Bilt, The Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Geert Grooteplein 21,
6525 EZ Nijmegen, The Netherlands
[3]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO 80309, USA

**Abstract.** In the current multi-model ensemble approach climate model simulations are combined a posteriori. In the method of this study the models in the ensemble exchange information during simulations and learn from historical observations to combine their strengths into a best representation of the observed climate. The method is developed and tested in the context of small chaotic dynamical systems, like the Lorenz 63 system. Imperfect models are created by perturbing the standard parameter values. Three imperfect models are combined into one super-model, through the introduction of connections between the model equations. The connection coefficients are learned from data from the unperturbed model, that is regarded as the truth.

The main result of this study is that after learning the super-model is a very good approximation to the truth, much better than each imperfect model separately. These illustrative examples suggest that the super-modeling approach is a promising strategy to improve weather and climate simulations.

## 1 Introduction

There is a broad scientific consensus that our climate is warming due to anthropogenic emissions of greenhouse gasses (IPCC, 2007). Due to the large impacts of climate change on society there is a growing need to widely sample and assess the possible climate change related to the plausible scenarios for future emissions. At about a dozen climate

institutes around the world complex climate models have been developed over the past decades. Despite the improvements in the quality of the model simulations, the models are still far from perfect. For instance a temperature bias of several degrees in annual mean temperatures in large regions of the globe is not uncommon in the simulations of the present climate (IPCC, 2007).

Nevertheless these models are used to simulate the response of the climate system to future emission scenarios of greenhouse gasses. It turns out that the models differ substantially in their simulation of the response: the global mean temperature rise varies by as much as a factor of 2 and on regional scales the response can be reversed, e.g. decreased precipitation instead of an increase. It is not clear how to combine these outcomes to obtain the most realistic response. The standard approach is to take some form of a weighted average of the individual outcomes (Tebaldi and Knutti, 2007), but is this the best strategy?

We think we can do better by letting the models exchange information during the simulation instead of combining the results of the individual models afterwards. We propose to combine the individual models into one super-model by prescribing connections between the model equations. The connection coefficients are learned from historical observations. This way the super-model learns to combine the strengths of the individual models in order to optimally reproduce the historical climate. Is this approach feasible?

An example of combining models successfully is found in the study by Kirtman et al. (2003) in which they coupled two different atmospheric models to one ocean model. It turned out that the most realistic simulation in terms of the annual mean, annual cycle and interannual variability of sea surface temperatures over the tropical pacific was obtained by

coupling the momentum fluxes from one model and the heat and fresh water fluxes from the other to the ocean model.

Another indication that this approach might be feasible is found in the practice of data assimilation (Compo et al., 2006). It turns out that with a limited amount of information, the complete state of the atmosphere can be recovered. Synchronization in chaotic systems provides an explanation why this is at all possible, since linking chaotic systems with a signal from one system to the other is known to lead to synchronization of their states (Pecora and Carroll, 1990; Duane et al., 2006). Therefore we expect that in the super-modeling approach only limited information needs to be exchanged to effectively influence the combined behaviour of the imperfect models.

In this paper we use simple chaotic systems to develop and demonstrate the super-modeling approach. We regard the model with standard parameter values as ground truth and create imperfect models by perturbing the parameter values. Three imperfect models are connected and combined into a super-model. The strength of the connections are determined from data from the ground truth through a learning process. The learning process takes the form of the minimisation of a cost function that measures the difference between the truth and the super-model during short integrations.

In Sect. 2 the form of the connections is introduced, followed by the introduction of the cost function and the minimisation method. The super-modeling approach is applied to the Lorenz 63, Rössler and Lorenz 84 systems in Sects. 3 and 4. Discussion and conclusion of the method and ideas for improvement can be found in Sect. 5.

## 2 The super-modeling approach

To introduce the super-modeling approach we use the Lorenz 63 system (Lorenz, 1963). The Lorenz 63 system is often used as a metaphor for the atmosphere, because of its abrupt regime changes and unstable nature. The equations for the Lorenz 63 system are

$$
\begin{aligned}
\dot{x} &= \sigma \, (y - x) \\
\dot{y} &= x \, (\rho - z) - y \\
\dot{z} &= xy - \beta z.
\end{aligned}
\tag{1}
$$

The standard parameter values are $\sigma = 10$, $\beta = \frac{8}{3}$ and $\rho = 28$. Numerical solutions are obtained by a fourth order Runge-Kutta time stepping scheme, with a time step of 0.01.

### 2.1 Connecting imperfect models

Imperfect models are created by taking three copies of the Lorenz 63 system with perturbed parameter values. A super-model is created by the introduction of linear connection terms

$$
\begin{aligned}
\dot{x}_k &= \sigma_k \, (y_k - x_k) + \sum_{j \neq k} C_{kj}^x \left( x_j - x_k \right) \\
\dot{y}_k &= x_k \, (\rho_k - z_k) - y_k + \sum_{j \neq k} C_{kj}^y \left( y_j - y_k \right) \\
\dot{z}_k &= x_k \, y_k - \beta_k \, z_k + \sum_{j \neq k} C_{kj}^z \left( z_j - z_k \right), \quad k = 1, 2, 3,
\end{aligned}
\tag{2}
$$

where $k$ indexes the three imperfect models with perturbed parameter values $\sigma_k$, $\beta_k$ and $\rho_k$ and $C_{kj}^x$, $C_{kj}^y$ and $C_{kj}^z$ are referred to as connection coefficients.

Each variable of each model is connected to the other two models. This gives two connection coefficients for each of the variables and a total number of $2 \times 3 \times 3 = 18$ connection coefficients. These 18 coefficients will be learned from data that sample the truth. The solution of the super-model, denoted by subscript "s", is taken to be the average of the imperfect models

$$
\begin{aligned}
x_{\mathrm{s}} &= \frac{1}{3} \, (x_1 + x_2 + x_3) \\
y_{\mathrm{s}} &= \frac{1}{3} \, (y_1 + y_2 + y_3) \\
z_{\mathrm{s}} &= \frac{1}{3} \, (z_1 + z_2 + z_3).
\end{aligned}
\tag{3}
$$

Note that Eqs. (2) define a new dynamical system with three times the number of degrees of freedom. The super-model is not merely a sort of average system. Depending on the connections, it can have a very different dynamics. The super-model has the potential to outperform the ensemble averaged simulations of the individual models because it can display richer dynamical behavior. The learning must ensure that the behavior after learning is more realistic.

### 2.2 Cost function

We assume that we have a long time series of observations of the truth $\mathbf{x}_{\mathrm{o}}$. We pick initial conditions $\mathbf{x}_{\mathrm{o}}(t_i)$ from this long time series at $K$ times $t_i$, $i = 1, ..., K$, separated by fixed distances $d$. Short integrations of length $\Delta$ are performed with the super-model starting from these $K$ initializations (see Fig. 1). To measure the ability of the super-model to follow the truth we introduce the following cost function $F$, that depends on the vector of connection coefficients $\mathbf{C}$.

$$
F(\mathbf{C}) = \frac{1}{K\Delta} \sum_{i=1}^{K} \int_{t_i}^{t_i + \Delta} |\mathbf{x}_{\mathrm{s}} \, (\mathbf{C}, t) - \mathbf{x}_{\mathrm{o}}(t)|^2 \, \gamma^t \, \mathrm{d}t
\tag{4}
$$

The cost function is normalized by $\frac{1}{K\Delta}$, so that it represents the time averaged mean squared error. The factor $\gamma^t$ with $0 < \gamma \leq 1$ is introduced to give stronger weight to the errors close to the initial conditions. The idea behind this is that the Lorenz 63 system displays sensitive dependence on initial conditions. Trajectories diverge not only due to model imperfections, but also due to internal error growth: even a perfect model deviates from the truth if started from slightly
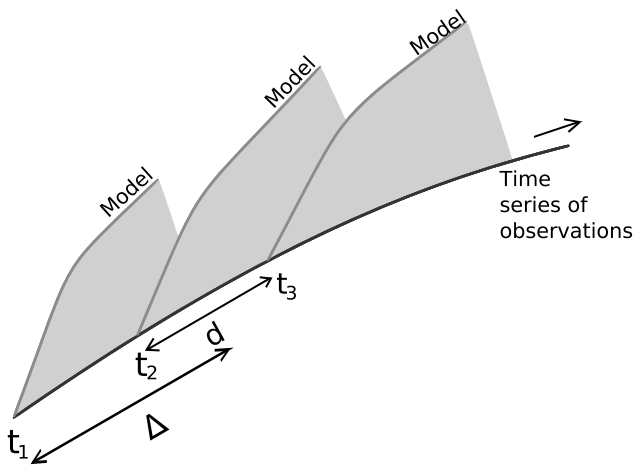
**Fig. 1.** The cost function is based on short integrations of the super-model starting from observed initial conditions of the truth at times $t_i$ and measures the mean-squared difference between the short evolutions of the super-model and the truth as indicated by the shaded areas. The short integrations span a time interval $\Delta$ and $d$ denotes the fixed time interval between the initial conditions $t_i$.

**Table 1.** Standard and perturbed parameters for the Lorenz 63 system.

|         | $\sigma$     | $\rho$     | $\beta$      |
|---------|--------------|------------|--------------|
| Truth   | 10           | 28         | $\frac{8}{3}$ |
| Model 1 | 13.25 (32 %) | 19 (32 %)  | 3.5 (31 %)   |
| Model 2 | 7 (30 %)     | 18 (36 %)  | 3.7 (39 %)   |
| Model 3 | 6.5 (35 %)   | 38 (36 %)  | 1.7 (36 %)   |

first for a small number of initializations. Subsequently we take this solution as the initial guess of the minimum for an increased number of initializations to find the minimum for this set. This process is repeated until we found that the minimum did not change any longer by increasing the number of initializations. This issue is discussed further in Sect. 3.

To avoid negative solutions for the connection coefficients we added an extra term in the cost function in case one of the coefficients becomes negative. This term is just the absolute value of the negative connection coefficient, which easily dominates the value of the cost function.

different initial conditions and leads to a non-zero cost function due to chaos. This implies that the cost function measures a mixture of model errors and internal error growth. Model errors dominate the inital divergence between model and truth, but at later times in the short term integrations internal error growth dominates. Since we wish to measure the model errors, the factor $\gamma^t$ discounts the errors at later times to decrease the contribution of internal error growth.

We base the choice of $\gamma$ on the doubling time of errors. From a large number of runs ($10^7$) from randomly perturbed initial conditions we estimate the average doubling time $\tau$ of the initial error. We choose $\gamma$ such that $\gamma^\tau = \frac{1}{2}$, so at time $\tau$ the weight is reduced to $\frac{1}{2}$. For the Lorenz 63 system $\tau = 0.75$, which gives $\gamma = 0.4$. The length of the short integrations is taken to be $\Delta = 1$, which is a little bit longer than the doubling time. By comparison the average time for one rotation in the Lorenz 63 system is 0.8. The separation $d$ between the initializations is 0.2 time units.

## 2.3 Minimisation

For a fixed choice of the number of initializations $K$ the cost function solely depends on the connection coefficients $C$ in Eq. (4). The super-model can be determined by finding a minimum in the cost function in the 18 dimensional space of $C$. For this we employ the Fletcher-Reeves-Polak-Ribiere Conjugate Gradient method (Fletcher and Reeves, 1963). It uses the gradient of the cost function to approach minima and stops when the gradient is (close to) zero.

We found it advantageous to make use of the dependence of the cost function on the number of initializations $K$ to avoid shallow local minima. We minimize the cost function

## 3 Results Lorenz 63

Three imperfect models are created by perturbing the standard parameter values as displayed in Table 1. The perturbed values differ from the standard values by 30 % to 40 % and in each imperfect model parameter values have been increased as well as decreased. With these perturbations the imperfect models behave quite differently from the truth as can be seen in Fig. 2. Both model 1 and 2 are attracted to a point, whereas model 3 has a chaotic solution that resembles the truth, but the attractor is displaced and larger. All models were initiated from the same state and the transient evolution towards the attractor is plotted as well.

By using bifurcation methods, it can be analytically shown that there exists a Hopf bifurcation for the Lorenz 63 system at $\rho_H = \frac{\sigma(3+\sigma+\beta)}{\sigma-1-\beta}$. This bifurcation marks different kinds of dynamical behaviour. Both model 1 and 2 have values for $\rho$ below the Hopf bifurcation, whereas model 3 has a value for $\rho$ that lies far above the Hopf bifurcation. For the truth the value of $\rho$ lies above the Hopf bifurcation as well, which is why model 3 and the truth have similar behaviour.

The minimisation procedure outlined above successfully identified a minimum of the cost function with a value of 0.02. By comparison the value of the cost function for an arbitrary choice of all connection coefficients equal to unity is 0.5. With the connection coefficients of this minimum we performed a long integration with the super-model and plotted the trajectory in Fig. 3. The attractor of the super-model is very close to the true attractor. While integrating the super-model, the imperfect models fall into an approximate synchronous behaviour due to the connections: the
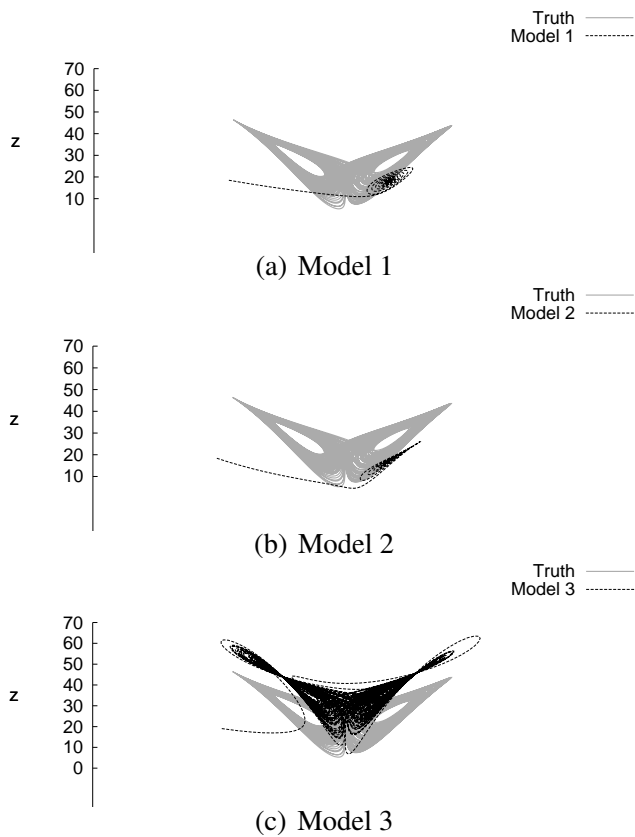
(a) Model 1

(b) Model 2

(c) Model 3

**Fig. 2.** Trajectories for the three unconnected imperfect models (black) and the standard Lorenz 63 system (grey). The trajectory for the imperfect models includes the transient evolution from the initial condition towards the attractor.



(a) Point of view 1

(b) Point of view 2

**Fig. 3.** Trajectories for the super-model (black) and the standard Lorenz 63 system (grey) from two different points of view.

temporal correlations between the $x$, $y$, and $z$ variables of the three models are in excess of 0.95 (not shown) and the sum of the time-mean distances between the three model states normalized by the sum of the standard deviations of $x_s$, $y_s$ and $z_s$ is 0.34. In particular the $z$-values of the third model are systematically larger than those of the other two models (see Fig. 4). The improvement in the behaviour of the connected imperfect model solutions as depicted in Fig. 4 (to be compared with Fig. 2) is a clear indication of the feasibility of super-modeling in the context of this low-dimensional chaotic system.

In addition to this minimum, we found that by choosing different initial values for the connection coefficients in the minimisation procedure different local minima in the cost function are obtained with values of the cost function that are of comparable magnitude. In the remainder of this section we will test the robustness of the learning process, research the issue of local minima and develop measures to determine the quality of the different super-model solutions.
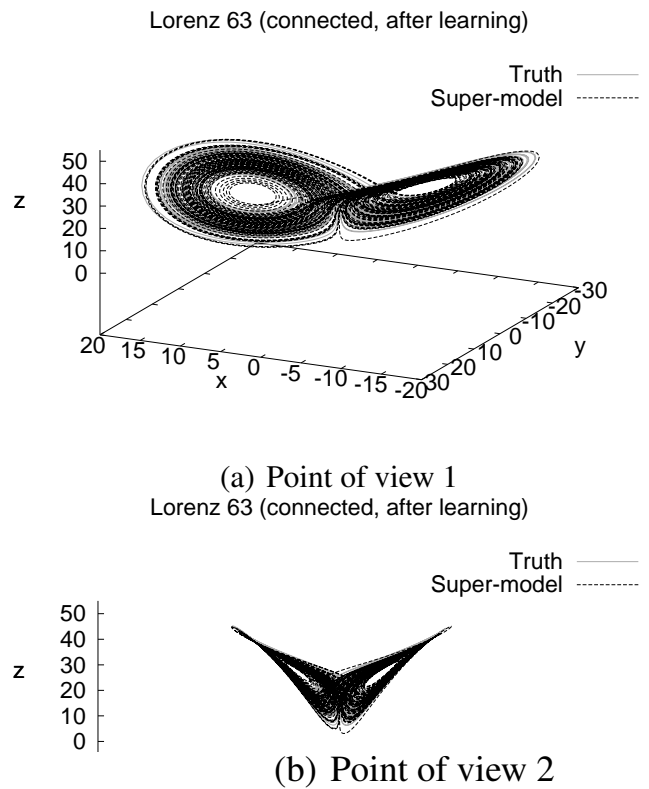
## 3.1 Robustness

The minimum of the cost function is determined on a limited period of observations of length $(K-1)\cdot d + \Delta$ that we refer to as the training set. We have chosen $K = 200$ to determine the minimum and subsequently evaluate the cost function using the $C$ values of this minimum for subsets of the training set of length corresponding to $K = 20$, 50, 100, 150. Cross sections of the cost function around the minimum can be created by changing one connection coefficient and keeping the others fixed at the values of the minimum. The cross sections for the different subsets are plotted in Fig. 5 for connection coefficients $C_{23}^y$ and $C_{21}^z$, since these are typical examples.

In Fig. 5a the cost function for $K = 200$ displays one well defined minimum $C_{23}^y = 10.1$. The position of the minimum does not change much when the cost function is calculated using the different subsets. The minimum becomes more pronounced as the training set is enlarged. The values of the cost function monotonically converge and $K = 200$ seems a reasonable size of the training set. Figure 5b does not show a well defined minimum for any $K$. Note that the scale is smaller than in Fig. 5a. The values of the cost function do not change much in the different subsets and the slopes are very flat. Changing connection coefficient $C_{21}^z$ apparently does not change the quality of the solutions much. A family

Lorenz 63 (model 1, connected, after learning)



(a) Model 1

Lorenz 63 (model 2, connected, after learning)



(b) Model 2

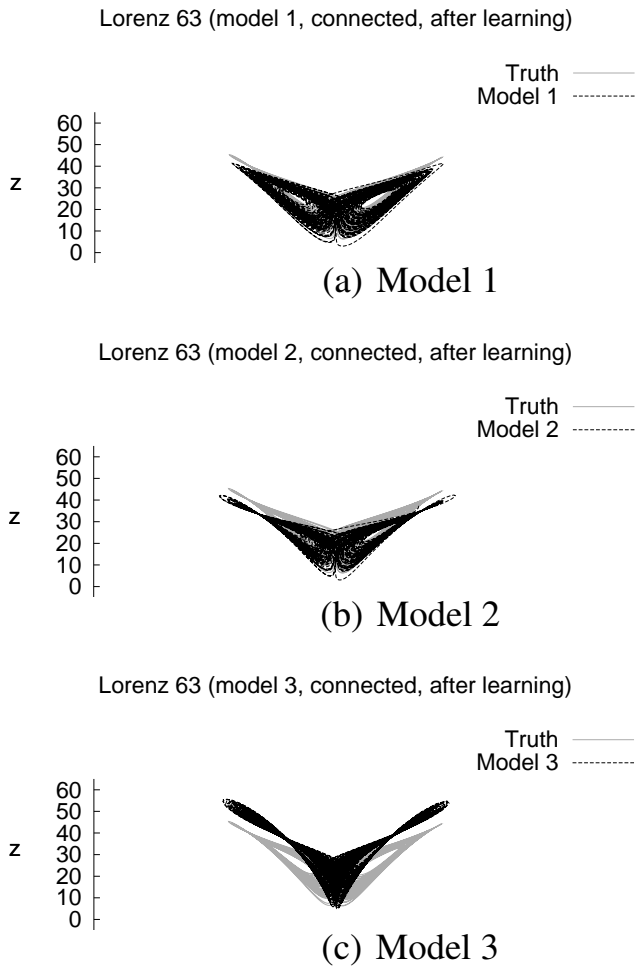Lorenz 63 (model 3, connected, after learning)



(c) Model 3

**Fig. 4.** Trajectories for the three connected imperfect models with connections determined by the learning process (black) and the standard Lorenz 63 system (grey).

**Table 2.** The connection coefficients of two super-model solutions of the Lorenz 63 system and their differences.

|  | Super-model 1 | Super-model 2 | Difference |
|---|---|---|---|
| $C_{12}^x$ | −0.01 | 1.52 | 1.53 |
| $C_{13}^x$ | 4.81 | 0.03 | −4.78 |
| $C_{21}^x$ | 5.69 | 13.28 | 7.59 |
| $C_{23}^x$ | 13.75 | 14.99 | 1.24 |
| $C_{31}^x$ | 17.64 | 21.51 | 3.87 |
| $C_{32}^x$ | −0.01 | 1.09 | 1.10 |
| $C_{12}^y$ | 7.67 | 3.53 | −4.14 |
| $C_{13}^y$ | 18.14 | 27.36 | 9.22 |
| $C_{21}^y$ | 3.64 | 0.00 | −3.64 |
| $C_{23}^y$ | 10.06 | 6.50 | −3.56 |
| $C_{31}^y$ | 2.71 | 3.89 | 1.18 |
| $C_{32}^y$ | 9.79 | 6.93 | −2.86 |
| $C_{12}^z$ | 5.47 | 3.95 | −1.52 |
| $C_{13}^z$ | 4.03 | 12.24 | 8.21 |
| $C_{21}^z$ | 10.72 | 3.50 | −7.22 |
| $C_{23}^z$ | 13.54 | 2.20 | −11.34 |
| $C_{31}^z$ | 8.70 | 2.89 | −5.81 |
| $C_{32}^z$ | 1.50 | 3.85 | 2.35 |

### 3.2 Local minima

In the previous section we noted that there is a large family of super-model solutions with similar values of the cost function connected to the minimum found by the minimisation. The minimisation was repeated starting from random values for the connection coefficients between [0, 10] that were drawn from a uniform probability distribution. In this way we found other minima that are distinct in many more connection coefficients. For one of these minima, the connection coefficients are shown in Table 2, together with the values for the first minimum. In the fourth column the difference between the connection coefficients of minima 1 and 2 indicates that the minima are clearly distinct and do not belong to the same family of solutions.

A plot of the attractor of the second super-model solution in its phase space (not shown) looks almost exactly the same as the plots of the first super-model solution in Figs. 3 and 4. The value of the cost function for the second solution is slightly lower (0.003 instead of 0.02) and is a first indication that the second solution might be better. In the next section we will use various measures to evaluate the quality of these two super-model solutions.

### 3.3 Quality measures

The cost function is a measure of the quality of the short term behaviour of the super-model in which the initial conditions play a role as is the case in weather predictions. To evaluate
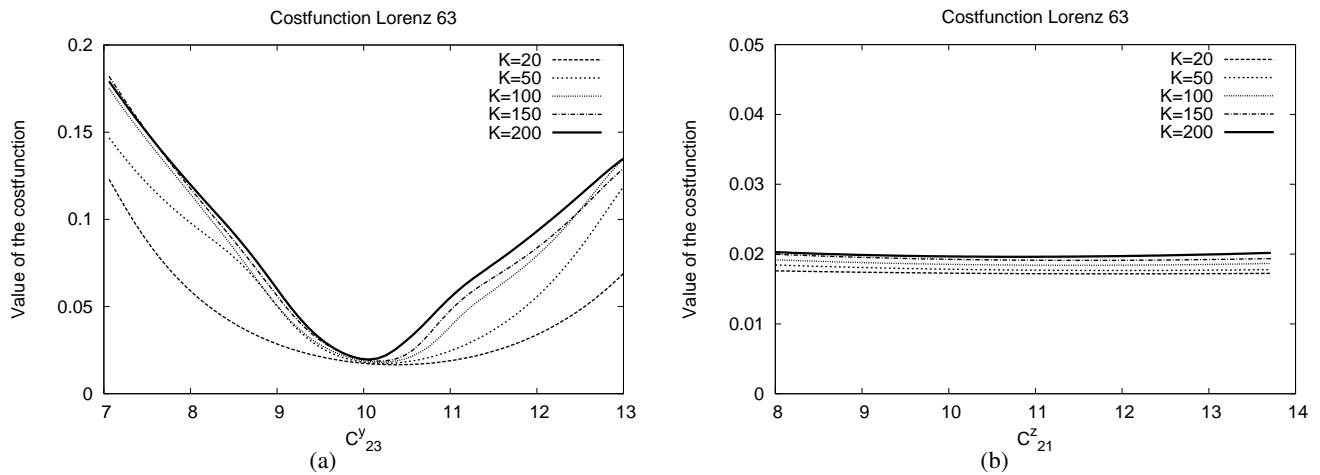
of super-models of similar quality can be found by changing connection coefficient $C_{21}^z$ between 8 and 14.

Ideally the super-model found by the learning process is not dependent on the training set. To test whether $K = 200$ is large enough for this to be true the cost function is plotted in Fig. 6 for different periods of observations: the training set and independent sets of the same size that were obtained from a longer consecutive integration of the truth. Again the cross sections for connection coefficients $C_{23}^y$ and $C_{21}^z$ are shown (Fig. 6). In Fig. 6a the position and value of the minimum remain close to that of the training set. In Fig. 6b the cost function is flat for all sets of observations. We conclude that with $K = 200$ the cost function verifies rather well on independent data, so $K = 200$ seems a reasonable size of the training set.

**Fig. 5.** Cross section of the cost function for the super-model of the Lorenz 63 system calculated for different subsets of the original training set that was based on $K = 200$ initializations. The subsets vary in the number of initializations, i.e. $K = 20, 50, 100, 150$. A cross sections is created by changing connection coefficients $C_{23}^{y}$ in (**a**) and $C_{21}^{z}$ in (**b**) and keeping the other coefficients fixed at the values of the minimum found by the learning process using the training set.

**Table 3.** Mean, standard deviation (SD) and covariance for the three unconnected imperfect models of the Lorenz 63 system. The values for the first two models are calculated analytically. Statistics for model 3 are based on 500 runs of 5000 time units. Between brackets the 95 % error estimation is given.

|  | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Mean $x$ | ±7.94 | ±7.93 | 0.003 (0.002) |
| Mean $y$ | ±7.94 | ±7.93 | 0.003 (0.010) |
| Mean $z$ | 18.00 | 17.00 | 34.23 (0.030) |
| SD $x$ | 0 | 0 | 7.628 (0.002) |
| SD $y$ | 0 | 0 | 9.416 (0.010) |
| SD $z$ | 0 | 0 | 8.765 (0.030) |
| Cov. $xy$ | 0 | 0 | 58.19 (0.036) |
| Cov. $xz$ | 0 | 0 | 0.007 (0.44) |
| Cov. $yz$ | 0 | 0 | 0.012 (0.68) |

**Table 4.** Mean, standard deviation (SD) and covariance for the truth and for the two super-models of the Lorenz 63 system. Statistics are based on 500 runs of 5000 time units. Between brackets the 95 % error estimation is given.

|  | Truth | Super-model 1 | Super-model 2 |
|---|---|---|---|
| Mean $x$ | −0.006 (0.22) | 0.007 (0.21) | −0.000 (0.25) |
| Mean $y$ | −0.006 (0.22) | 0.007 (0.21) | −0.000 (0.25) |
| Mean $z$ | 23.549 (0.02) | 22.93 (0.02) | 23.19 (0.03) |
| SD $x$ | 7.924 (0.005) | 7.717 (0.003) | 7.812 (0.005) |
| SD $y$ | 9.011 (0.008) | 8.791 (0.009) | 8.723 (0.009) |
| SD $z$ | 8.623 (0.025) | 8.596 (0.016) | 8.549 (0.032) |
| Cov. $xy$ | 62.786 (0.07) | 58.952 (0.05) | 60.6416 (0.08) |
| Cov. $xz$ | −0.020 (0.76) | 0.023 (0.74) | 0.000 (0.88) |
| Cov. $yz$ | −0.016 (0.61) | 0.021 (0.65) | −0.001 (0.69) |

the quality of the super-model beyond the range that is influenced by the initial conditions, different measures can be used as in climate simulations.

The most straightforward measures are the different moments of the probability density function of the states in phase space, such as the mean, variance and covariance of the state variables. Since these do not take into account the temporal evolution through phase space, we will also evaluate the ability of the super-model to reproduce the autocorrelation functions of the state variables. As a final measure we will check the ability of the super-model to synchronize with the truth at the end of this section.

**Mean, standard deviation and covariance**

The calculation of these statistics is based on 500 runs of 5000 time units of the truth, the imperfect models and both super-models. An error estimation of these quantities is based on the spread of the 500 estimates of each quantity. The results for the imperfect models are given in Table 3 and for the truth and both super-models in Table 4.

For the parameter values of model 1 and 2 the attractor reduces to two stable point attractors. The $x$, $y$ and $z$ values of these fixed points can be calculated analytically from Eqs. (1) by setting the time derivatives to zero. Since the system settles in one of these point attractors depending on the initial condition, the mean values are equal to these values.
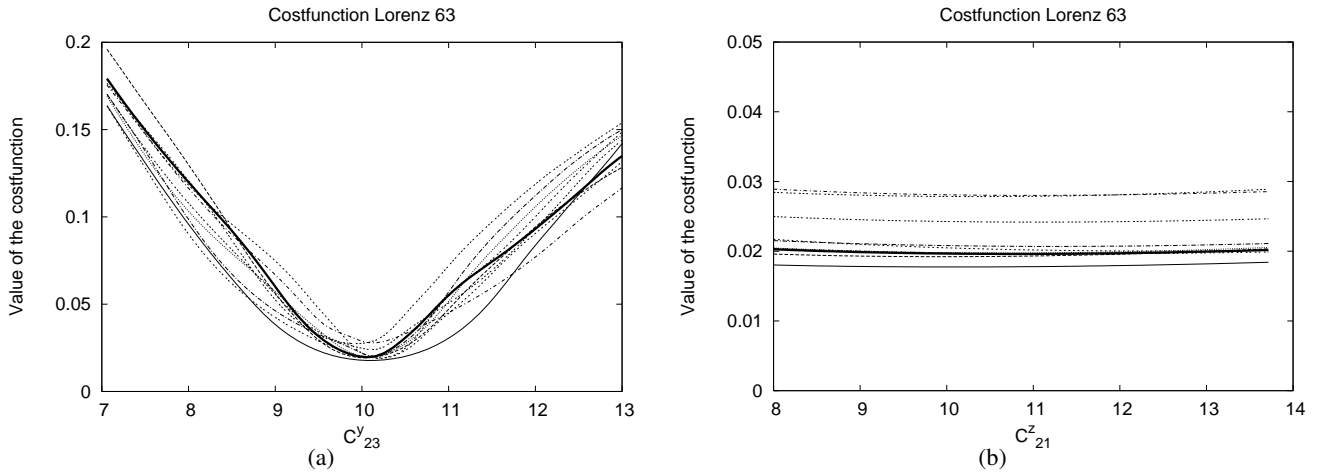
**Fig. 6.** As in Fig. 5, except that the cost function is calculated for the training set with $K = 200$ initializations (thick line) and 9 additional independent sets of observations of the same length (thin lines) that were taken from a consecutive longer integration of the truth.

The statistics of the chaotic solution of model 3 (see Table 3) differ substantially from the truth (see Table 4), especially the mean value of $z$ is much larger.

Both super-models (see Table 4) have statistics that are close to that of the truth with the largest differences of order 5 % in the covariance between $x$ and $y$. The second super-model is somewhat closer to the truth, especially in the co-variance of $x$ and $y$.

### 3.4 Autocorrelation

The autocorrelation is a statistical measure of the temporal evolution. It gives an indication of the memory and time scales present in a system. The plots in Fig. 7 are based on 100 runs of 3000 time units and the shading corresponds to the 95 % error range of the autocorrelation of the truth.

Both super-models capture the fast decorrelation of $x$ and $y$ and the slow decorrelation of $z$ well, but the second super-model is closer to the truth. It also better represents the dominant time scale which is most apparent in the autocorrelation of $z$. After 9 oscillations super-model 1 runs out of phase with the truth somewhat, whereas super-model 2 is still in phase.

### 3.5 Synchronization with the truth

Pecora and Carroll (1990) have shown that limited information exchange between two identical Lorenz systems can lead to synchronization of the model states even when the systems are initialized from very different initial conditions and differ slightly in parameter values. The ability to synchronize with the truth is another measure of the quality of a model. In this section we will compare how well the super-models compare to a perfect model in this respect.

Yang et al. (2006) extended the study of synchronized Lorenz systems, re-interpreted in the context of data

assimilation. Following Yang et al. (2006) we add a so-called simple nudging term to the equations of the $y$ variable for each of the three connected imperfect models as in Eqs. (5). This term "nudges" the actual values of $y_k$ to the observed value $y_o$ and the value of parameter $n$ determines the strength of the nudging.

$$\dot{x}_k = \sigma_k \left( y_k - x_k \right) + \sum_{j \neq k} C_{kj}^x \left( x_j - x_k \right)$$

$$\dot{y}_k = x_k(\rho_k - z_k) - y_k + \sum_{j \neq k} C_{kj}^y \left( y_j - y_k \right) + n \left( y_o - y_k \right) \quad (5)$$

$$\dot{z}_k = x_k \, y_k - \beta_k \, z_k + \sum_{j \neq k} C_{kj}^z \left( z_j - z_k \right) \quad k = 1, 2, 3$$

We take the following definition of synchronization:

**Definition 1** *A model is synchronized with the truth if the RMS difference between the model state and observed true state at $t = t_0$ is smaller than $\delta$ and remains smaller than $\epsilon$ for $t \to \infty$.*

$\epsilon$ is chosen larger than $\delta$, since synchronized systems often deviate somewhat during short extreme excursions of the trajectory, but remain synchronized. As a measure of synchronization we use the minimum strength of the nudging $n$ for which synchronization is accomplished independent of the initial condition, for integer $n$. For practical purposes we choose a time interval of $T = 1000$ time units during which the models must remain within $\epsilon$ distance of each other.

How quickly systems synchronize very much depends on the initial conditions (Yang et al., 2006), therefore we check synchronization for 100 restarts from different initializations. By trial and error we found that two identical Lorenz systems with standard parameter values (what we call the truth) synchronize using $n = 3$, $\delta = 2$ and $\epsilon = 4$ for all 100 initializations.

To compare the two super-model solutions the same set of 100 initializations are used. The first super-model needs
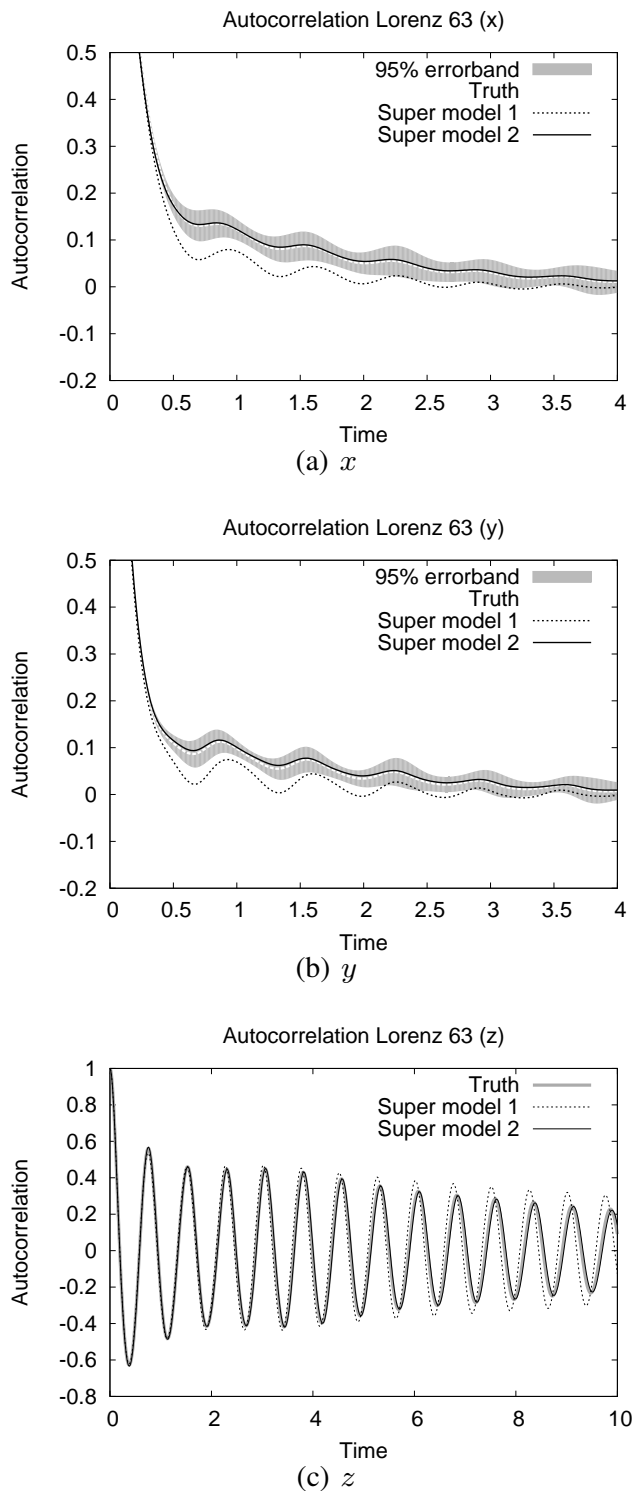
(a) $x$



(b) $y$



(c) $z$

**Fig. 7.** Autocorrelation as a function of delay time for $x$, $y$ and $z$ for the standard Lorenz 63 system and both super-models. The shaded area indicates the 95 % error band for the autocorrelation of the truth, based on 100 runs of 3000 time units.
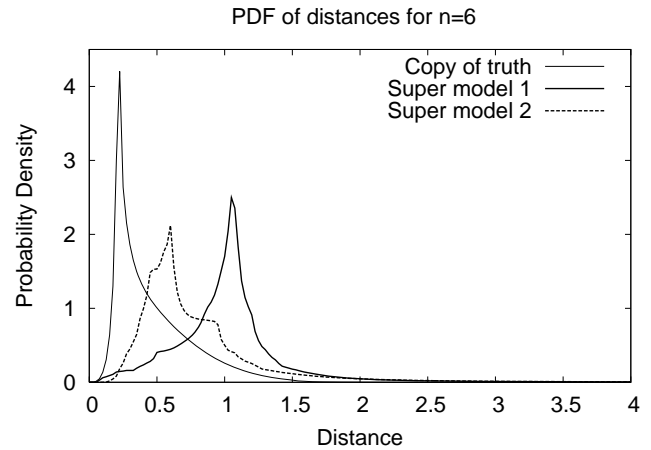


**Fig. 8.** Probability density function of the distance between the truth and the models while nudging the models to the truth in the $y$-variable with strength $n = 6$. Plots are for a copy of the truth (thin solid), the first (thick solid) and the second super-model (thick dashed). The PDF is calculated from a simulation of length $10^5$ time units and is normalized to one.

a nudging strength of $n = 11$ in order to synchronize with the truth. The second super-model needs a somewhat larger value $n = 13$. Using the same experimental setup, we found that the imperfect models individually are not able to synchronize with the truth at all. Both super-models need a stronger nudging than the perfect model. In this measure, the first super-model is closer to the truth, despite the fact that the mean temporal evolution, as measured by the autocorrelation, is more faithfully captured by the second super-model that also has a smaller cost function value. However, if we calculate the probability density function of the distance between the truth and the super-model from a $10^5$ time units long integration of the super-model nudged to the truth as in Eqs. (5) for $n = 6$, we find that more often the second super-model remains closer to the truth than the first super-model (see Fig. 8). Nevertheless, the second super-model needs a slightly larger nudging strength to synchronize with the truth than the first because for nudging values larger than $n = 5$, it has larger probability, albeit small, of exceeding the threshold $\epsilon = 4$. For $n = 6$ the distance between the second super-model and the truth is larger than 4 during 1.6 % of the time whereas it is 1.3 % for the first. For $n = 10$ it is 0.27 % for the second and 0.075 % for the first. This probability becomes small enough to meet the synchronization criterium of Definition 1 for $n = 13$ for the second super-model, whereas for the first this happens for $n = 11$. We conclude that the interpretation of the ability of a model to synchronize with the truth as a measure of the quality of a solution is not so straightforward. It serves more as a measure of the stability of the model.

All measures indicate that the second super-model is closer to the truth than the first. It turns out that the value

**Table 5.** Standard and perturbed parameters for the Rössler system.

|  | $a$ | $b$ | $c$ |
|---|---|---|---|
| Truth | 0.2 | 0.2 | 5.7 |
| Model 1 | 0.26 (30 %) | 0.14 (30 %) | 7.5 (32 %) |
| Model 2 | 0.12 (40 %) | 0.28 (40 %) | 7.4 (30 %) |
| Model 3 | 0.27 (35 %) | 0.12 (40 %) | 4. (30 %) |

of the cost function is indeed a good indication of the quality of the solution and that the approach of minimizing the cost function is a fruitful strategy.

## 3.6 Simulating climate change

In order to check whether the super-model is also able to simulate climate change, for instance the response of the truth to a parameter perturbation, we doubled the parameter $\rho$ in the true system and in the imperfect models in the super-model. The response of the attractor is an increase in size, the shape remains very similar (see Fig. 9). Although the connection coefficients are learned for $\rho = 28$, the super-model quite accurately simulates the attractor for $\rho = 56$. The mean $z$-value increases with a factor of 2.2 for both the truth as well as the super-model. The response is practically the same for both super-models.

## 4 Results Rössler and Lorenz 84

In this section the super-modeling approach is applied to the Rössler and the Lorenz 84 systems. Both display chaotic behaviour for standard parameter settings, but the attractors are quite different.

### 4.1 Rössler

The Lorenz 63 attractor is also called a *butterfly*, because of its shape. As a simplification of this example of chaos to one where the attractor only has one "wing", the Rössler equations were proposed (Rössler, 1976). The time evolution is less chaotic than in the Lorenz 63 system, since it lacks the irregular transitions between two unstable points. The equations are

$$\dot{x} = -(y + z)$$
$$\dot{y} = x + ay \qquad (6)$$
$$\dot{z} = b + z(x - c).$$

The parameter values for the truth are Rösslers values: $a = 0.2$, $b = 0.2$ and $c = 5.7$. The values for the parameters for the three imperfect models can be found in Table 5. The parameter perturbations applied are again of the order 30 % to 40 % and in each of the imperfect models parameters have been decreased as well as increased.
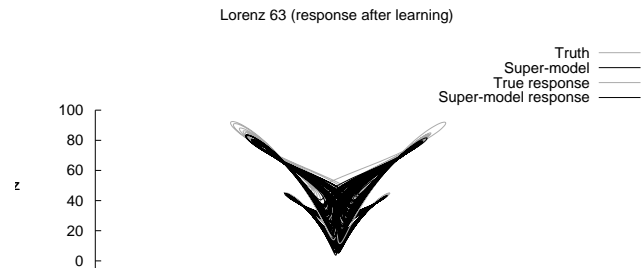


**Fig. 9.** Trajectories for the second super-model (black) and the standard Lorenz 63 system (grey). The larger attractor corresponds to $\rho = 56$, the smaller to the standard value $\rho = 28$. The super-model was trained on the standard parameter value.

With these parameter perturbations marked changes occur in the attractor as can be seen in Fig. 10. The attractor of imperfect model 1 is still chaotic and has a similar shape but the amplitude of the irregular oscillations is larger. Imperfect model 2 and 3 have a periodic attractor of different shapes.

To determine the super-model we first need to choose values for the different parameters in the cost function. For the Rössler system the time it takes for initial errors to double is on average 6.7. Following the same procedure as for the Lorenz 63 system we set $\gamma = 0.9$ and $\Delta = 12$ time units. The number of initializations in this case is $K = 300$.

We minimized the cost function by varying the connection coefficients of the super-model. This minimum is plotted in Fig. 11 in a cross section along $C_{23}^x$. The value at the minimum is approximately 0.0001, which is much lower than a typical value of the cost function (0.004 for all connection coefficients equal to 1). To check the robustness of this minimum with respect to the limited size of training set, we calculated the cost function for 9 additional sets of 300 initializations, that were taken from a longer simulation of the truth. The figure shows that 300 initializations are enough to reliably estimate the cost function. This minimum is not unique. By changing the initial values of the connection coefficients in the minimisation procedure, we found different minima with similar values of the cost function as was the case for the Lorenz 63 system. Here we evaluate the quality of this minimum only.

With the connection coefficients of this minimum, we integrated the super-model and plotted the trajectory of the three connected imperfect models in Fig. 12. The three models fall into an approximate synchronous behaviour, but especially the amplitudes of the excursion in the $z$ direction are different with model 3 making the largest excursions. The temporal correlations between the $x$, $y$, and $z$ variables of the three models are in excess of 0.99 (not shown) and the sum of the time-mean distances between the three model states normalized by the sum of the standard deviations of $x_s$, $y_s$ and $z_s$ is 0.18. The super-model solution, which is defined as the average of the three imperfect models, is plotted in Fig. 13 for two points of view. Visually the attractor of the super-model
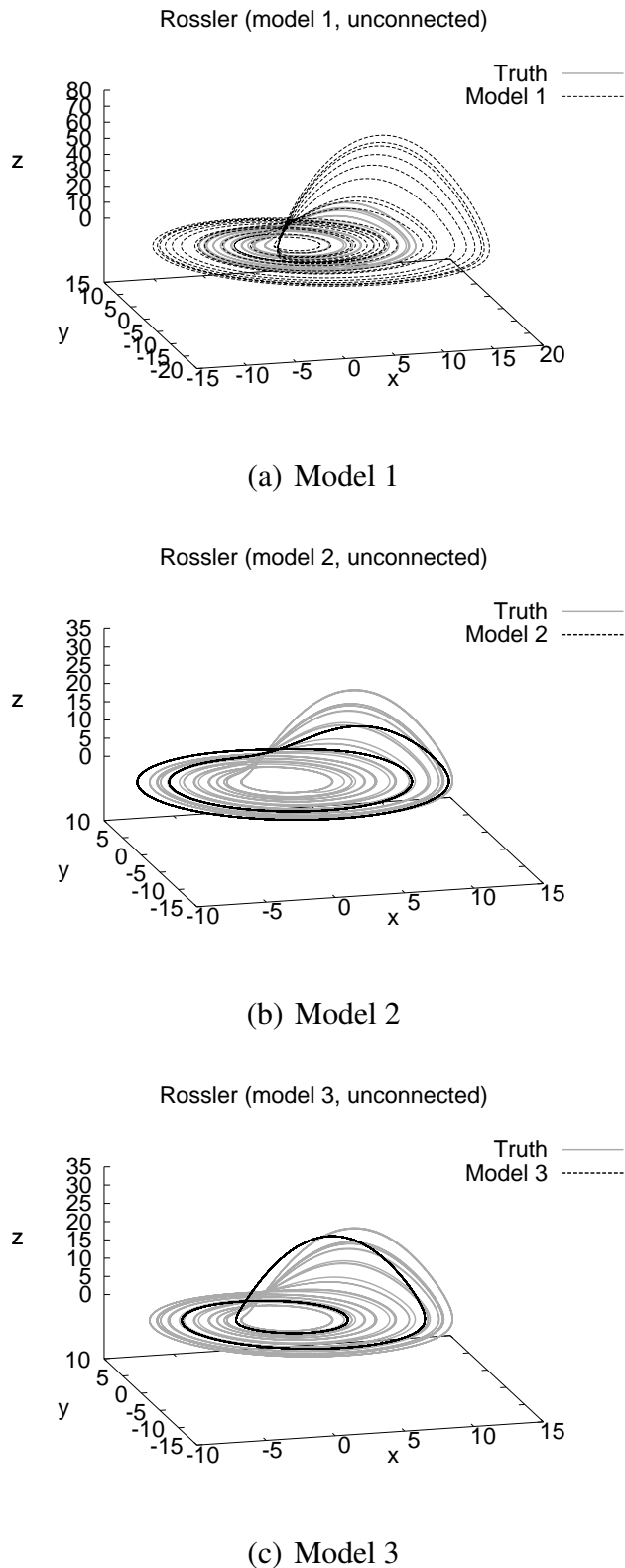
Rossler (model 1, unconnected)



(a) Model 1

Rossler (model 2, unconnected)



(b) Model 2

Rossler (model 3, unconnected)



(c) Model 3

**Fig. 10.** Trajectories for the three unconnected imperfect models (black) and the standard Rössler system (grey). Note the different scales on the axes. The truth is the same in all three plots.

**Table 6.** Mean, standard deviation (SD) and covariance for the three unconnected imperfect models of the Rössler system. The 95 % error estimation based on 500 runs of 5000 time units is given between brackets.

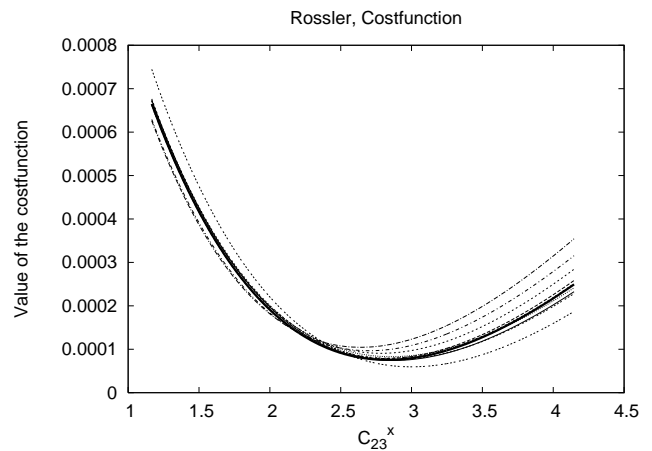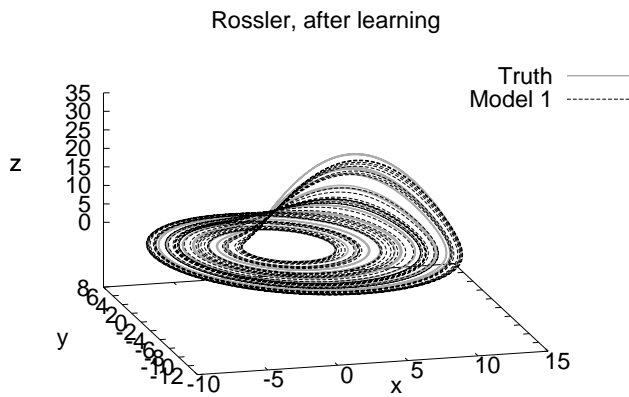|  | Model 1 | Model 2 | Model 3 |
| --- | --- | --- | --- |
| Mean $x$ | 0.417 (0.082) | 0.085 (0.0009) | 0.34 (0.0009) |
| Mean $y$ | $-1.603$ (0.099) | $-0.710$ (0.0009) | $-1.26$ (0.0009) |
| Mean $z$ | 1.603 (0.230) | 0.710 (0.0015) | 1.26 (0.0022) |
| SD $x$ | 6.759 (0.082) | 6.659 (0.0009) | 4.463 (0.0008) |
| SD $y$ | 6.567 (0.099) | 6.400 (0.0009) | 4.080 (0.0009) |
| SD $z$ | 6.853 (0.229) | 1.787 (0.0015) | 3.896 (0.0022) |
| Covariance $xy$ | -11.21 (0.33) | $-4.492$ (0.005) | $-4.49$ (0.004) |
| Covariance $xz$ | 11.21 (0.33) | 4.916 (0.006) | 4.49 (0.004) |
| Covariance $yz$ | -0.35 (0.39) | 2.784 (0.004) | 2.06 (0.003) |



**Fig. 11.** Cross sections of the cost function for the super-model of the Rössler system for the training set (thick line) with length corresponding to $K = 300$ and 9 additional independent sets of the same length (thin lines). Cross sections are created by changing the connection coefficient $C_{23}^x$ and keeping the others fixed.

is very similar to the true attractor. We will apply the same measures as for the Lorenz 63 system to check the quality of the super-model.

First we compare the means, standard deviations and covariances for the unconnected imperfect models in Table 6 and for the super-model and the truth in Table 7. The super-model turns out to be closer to the truth than the best imperfect model (model 3). Its statistics almost fall within the 95 % error bounds of the true values.

To compare the temporal behaviour we calculated the autocorrelation functions as plotted in Fig. 14 for the truth and the super-model. They indicate a strongly periodic behaviour with a long decorrelation time scale. For all three variables the autocorrelation function is close to and sometimes within the 95 % error band, again indicating that the super-model is a very good approximation of the truth.
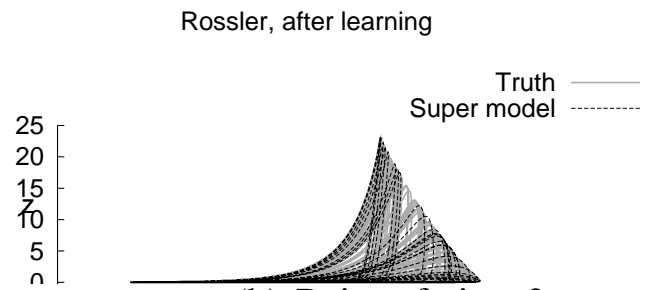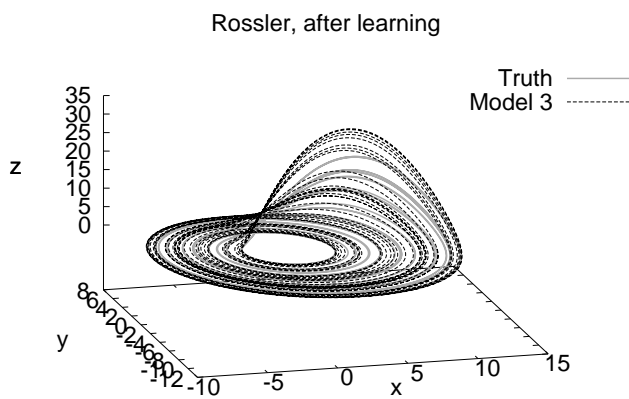
(a) Model 1
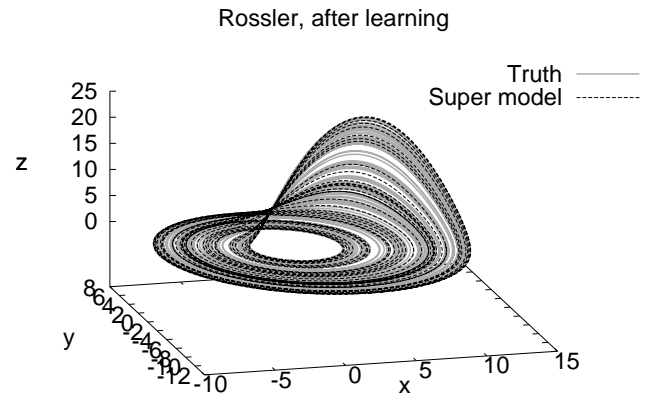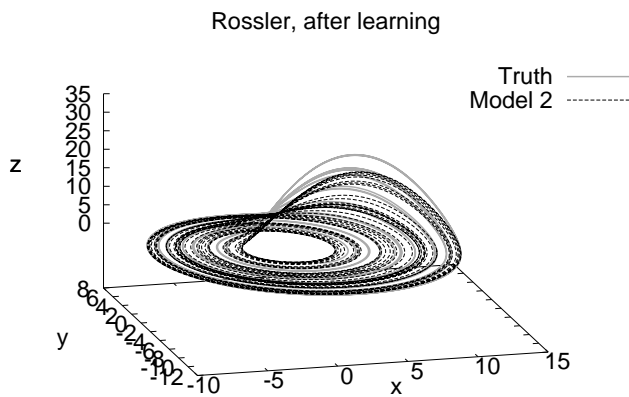


(b) Model 2



(c) Model 3

**Fig. 12.** Trajectories for the three connected imperfect models with connections determined by the learning process (black) and the standard Rössler system (grey).



(a) Point of view 1



(b) Point of view 2

**Fig. 13.** Trajectories for the super-model (black) and the standard Rössler system (grey) for two different points of view.

Finally we look at the minimum nudging strength needed to enable synchronization with the truth. We use the same definition of synchronization as for the Lorenz 63 model with the following values for the parameters $\delta = 0.05$, $\epsilon = 0.4$ and $T = 1000$ time units. When the nudging term is applied to the $y$ variable only, we find that the standard Rössler system synchronizes with a copy of itself for a nudging strength equal to $n = 1$. The super-model also synchronizes when nudging only the $y$ variable, but it needs a stronger nudging of $n = 2$. It outperforms model 3 in this measure; even by replacing the $y$ variable with the true value (which corresponds effectively to an infinitely large nudging strength), synchronization does not occur.

To conclude, also in the case of the Rössler system, super-model solutions can be found by combining imperfect models that give a very good approximation to the truth. This may not be surprising since the Rössler system is less chaotic than the Lorenz 63 system (note the long autocorrelation time-scale in Fig. 14) and more regular behaviour is presumeable easier to reproduce. On the other hand, a more chaotic system has richer dynamics (more time-scales, instabilities etc.) thus the connected models have more degrees of freedom to mimick the truth. Beforehand it is hard to predict whether
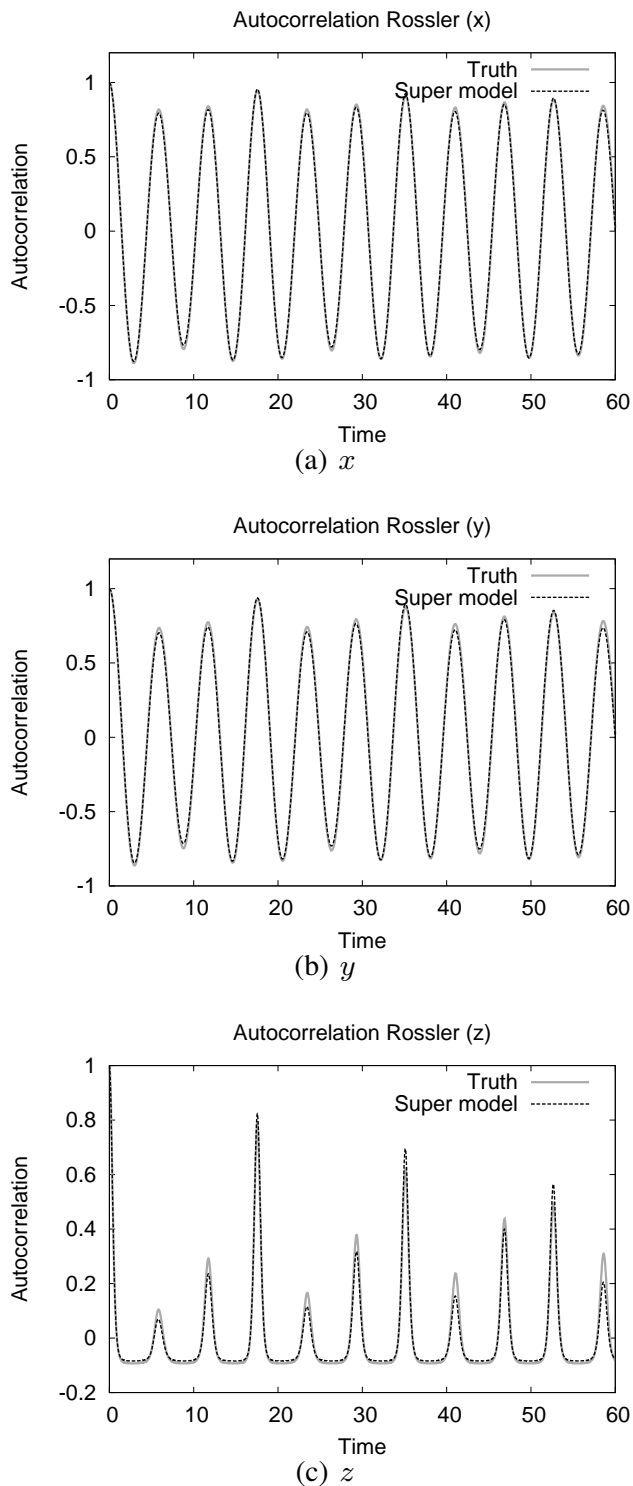
Autocorrelation Rossler (x)

(a) $x$



Autocorrelation Rossler (y)

(b) $y$



Autocorrelation Rossler (z)

(c) $z$

**Fig. 14.** Autocorrelation for the super-model (black) and the standard Rössler system (grey). The thickness of the grey line corresponds to the 95 % error band for the truth, based on 100 runs of 3000 time units.

**Table 7.** Mean, standard deviation (SD) and covariance for the truth and super-model of the Rössler system. The 95 % error estimation based on 500 runs of 5000 time units is given between brackets.

|                 | Truth           | Super-model     |
| --------------- | --------------- | --------------- |
| Mean $x$        | 0.177 (0.003)   | 0.175 (0.003)   |
| Mean $y$        | −0.886 (0.009)  | −0.878 (0.009)  |
| Mean $z$        | 0.886 (0.009)   | 0.874 (0.009)   |
| SD $x$          | 5.16 (0.04)     | 5.10 (0.03)     |
| SD $y$          | 4.84 (0.03)     | 4.82 (0.02)     |
| SD $z$          | 2.84 (0.04)     | 2.95 (0.03)     |
| Covariance $xy$ | −4.693 (0.05)   | −4.702 (0.04)   |
| Covariance $xz$ | 4.693 (0.05)    | 4.644 (0.04)    |
| Covariance $yz$ | 2.183 (0.12)    | 2.025 (0.19)    |

more chaos helps or hurts, so we test the super-modeling approach also on the more chaotic Lorenz 84 system.

### 4.2 Lorenz 84

The Lorenz 84 system was proposed by Lorenz as a toy model for the general atmospheric circulation at midlatitudes (Lorenz, 1984). The model equations are

$$\dot{x} = -y^2 - z^2 - ax + aF$$
$$\dot{y} = xy - bxz - y + G \tag{7}$$
$$\dot{z} = bxy + xz - z.$$

The $x$ variable represents the intensity of the globe-encircling westerly winds and $y$ and $z$ represent a travelling large-scale wave that interacts with the westerly wind. Parameters $F$ and $G$ are forcing terms representing the average north-south temperature contrast and the east-west asymmetries due to the land-sea distribution respectively.

The standard parameter values for the truth are $a = \frac{1}{4}$, $b = 4$, $F = 8$ and $G = 1$, for which the model displays chaotic behaviour (van Veen, 2001). In Table 8 the perturbed parameter values of the imperfect models are given. The perturbations are again about 30 % and in each imperfect model parameters have been decreased as well as increased.

With these parameter perturbations the attractor of the imperfect models differs substantially from the truth (see Fig. 15). Both model 1 and 3 have periodic attractors, whereas model 2 has a point attractor (the transient evolution towards the point attractor is shown for model 2). The periodic behaviour corresponds to the wave traveling periodically around the hemisphere.

Following the same procedure as before to find the parameters used in the cost function we found $\gamma = 0.5$ and $\Delta = 2.2$ time units, based on the average time it takes for initial errors to double (on average 1.1 time units). However with these values the minimisation algorithm did not produce a well defined minimum of the cost function. The high value of the
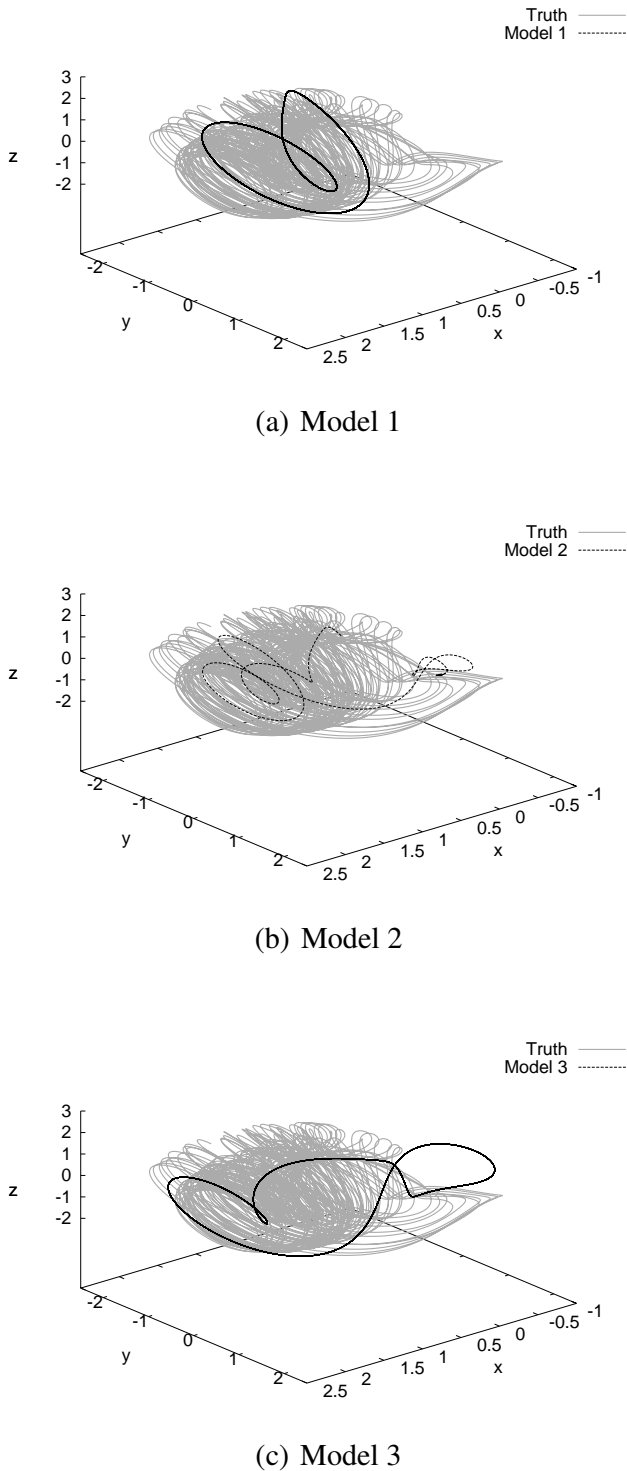
(a) Model 1



(b) Model 2



(c) Model 3

**Fig. 15.** Trajectories for the three unconnected imperfect models (black) and the standard Lorenz 84 system (grey).

**Table 8.** Standard and perturbed parameters for the Lorenz 84 system.

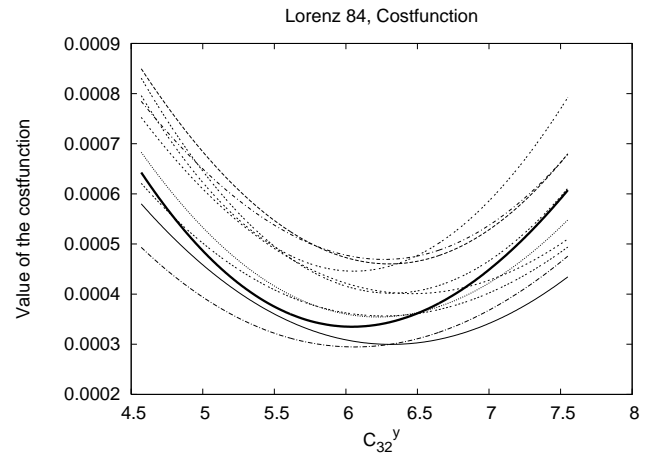|  | $a$ | $b$ | $F$ | $G$ |
|---|---|---|---|---|
| Truth | 0.25 | 4 | 8 | 1 |
| Model 1 | 0.33 (32 %) | 5.2 (30 %) | 10.4 (30 %) | 0.7 (30 %) |
| Model 2 | 0.18 (28 %) | 5.2 (30 %) | 5.6 (30 %) | 1.3 (30 %) |
| Model 3 | 0.18 (28 %) | 2.7 (33 %) | 10.4 (30 %) | 1.3 (30 %) |



**Fig. 16.** Cross section of the cost function for the super-model of the Lorenz 84 system for the training set (thick line) with length corresponding to $K = 200$ and 9 additional independent sets of the same length (thin lines). Cross sections are created by changing the connection coefficient $C_{32}^{y}$ and keeping the others fixed.

autocorrelation function of $x$ (0.6 at 8 time units, see Fig. 18) indicates that the initial conditions still have an impact on the evolution after 8 time units. Therefore we decided to increase $\Delta$ to 8 and $\gamma$ to 0.8. In addition it turned out that it was easier to find good minima using the amoeba minimisation algorithm (Nelder and Mead, 1965) instead of the conjugate gradients minimisation. The amoeba method does not need gradient information and is less susceptible to getting stuck in local minima. The training set is based on $K = 200$ initializations, each 0.2 time units apart selected from a long simulation of the truth.

Starting from different initial values for the connection coefficients we found different minima of the cost function. A cross section through the best minimum that we found is shown in Fig. 16. The value at this minimum is approximately 0.0003, which is again much lower than the value of the costfunction for all connection coefficients equal to 1 (0.08). To check the robustness the cost function is evaluated on 9 additional independent sets of 200 initializations. In all 9 sets the minimum is reproduced around the same value of the connection coefficient. The same is true for cross sections of the other connection coefficients (not shown).
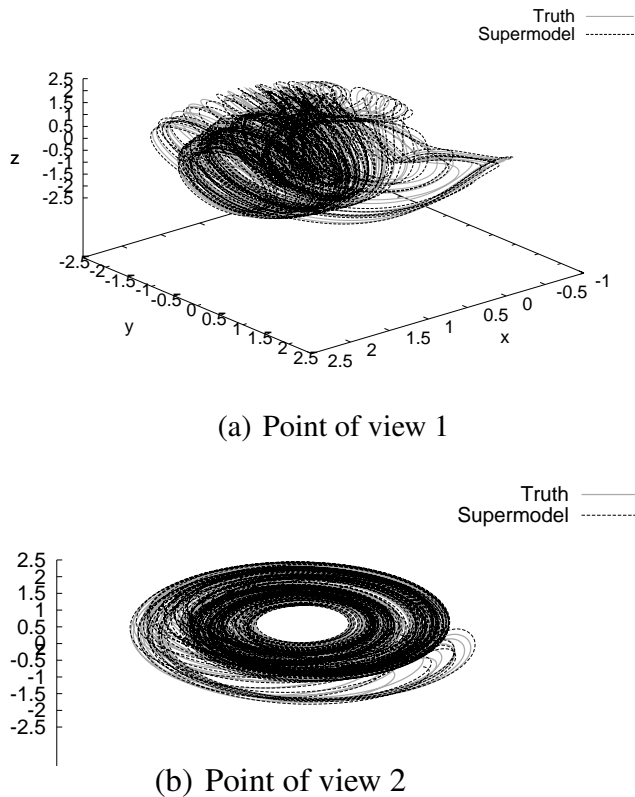
(a) Point of view 1



(b) Point of view 2

**Fig. 17.** Trajectories for the super-model (black) and the standard Lorenz 84 system (grey) for two points of view.

With the connection coefficients for this minimum, we integrated the super-model and plotted the trajectory in Fig. 17. A visual comparison with the truth indicates a very good agreement. In this case the three imperfect models are almost perfectly synchronized (not shown). The synchronization is stronger in this case as compared to the other two systems. The temporal correlations between the $x$, $y$ and $z$ variables of the three imperfect models in the super-model are in excess of 0.99 and the sum of the time-mean distances between the three model states normalized by the sum of the standard deviations of $x_s$, $y_s$ and $z_s$ is only 0.03. The model trajectories stay very close together on average. The reason for this might be found in the high value of several connection coefficients (for instance $C_{32}^x = 115$, $C_{23}^y = 147$ and $C_{31}^z = 169$). Such high values make synchronization easier since these connection terms in the equations bring the model solutions closer together. Maximum values of the connection coefficients in the other two systems are a factor of 10 smaller.

Again we use the same measures to evaluate the quality of the super-model solution. The mean, standard deviation and covariance for the truth and the super-model are presented in Table 9. These statistics are in excellent agreement.

In order to evaluate the temporal behaviour we compare the autocorrelation functions in Fig. 18. Up to a delay time of 14 time units the autocorrelation functions of the truth are

**Table 9.** Mean, standard deviation (SD) and covariance for the super-model and the standard Lorenz 84 system. The 95 % error estimation based on 500 runs of 5000 time units is given between brackets.

|  | Truth | Super-model |
|---|---|---|
| Mean $x$ | 1.015 (0.008) | 1.013 (0.008) |
| Mean $y$ | 0.060 (0.018) | 0.058 (0.017) |
| Mean $z$ | 0.271 (0.005) | 0.272 (0.004) |
| SD $x$ | 0.589 (0.014) | 0.596 (0.014) |
| SD $y$ | 0.919 (0.002) | 0.920 (0.002) |
| SD $z$ | 0.908 (0.002) | 0.906 (0.002) |
| Covariance $xy$ | −0.053 (0.018) | −0.050 (0.022) |
| Covariance $xz$ | −0.038 (0.004) | −0.039 (0.003) |
| Covariance $yz$ | −0.075 (0.006) | −0.063 (0.005) |

well reproduced by the super-model both in shape as well as in periodicity.

The Lorenz 84 system with standard parameter values synchronizes with the truth for a strength of the nudging term $n = 1$ in the $y$ variable only, using $\delta = 0.1$, $\epsilon = 0.5$ and $T = 1000$ in Definition 1. The super-model also synchronizes with the truth, but it needs a larger nudging of $n = 4$. None of the imperfect models is able to synchronize with the truth, when the nudging is in the $y$ variable only.

Concluding this section, super-model solutions can be found that reproduce the true system very well and outperform the individual imperfect models for the Lorenz 84 system as well. For this system, the minimisation process was found to be more sensitive to the length of the short integrations $\Delta$ and the discount parameter $\gamma$, requiring the use of the more robust amoeba minimisation procedure.

## 5 Conclusion and discussion

In this study we developed and tested a novel multi-model ensemble approach in which imperfect models of an observable system are combined into a single super-model by letting the models exchange information during the simulation. The information exchange takes the form of linear connections with weights that are learned from historical data such that the super-model minimizes the mean squared errors in short simulations initialized from past observed states. The main result of this study is that it is possible to construct super-models in the context of simple low-dimensional chaotic systems that outperform the constituent imperfect models.

This result motivates an alternative strategy to the weather and climate prediction problem. Current practice is that existing imperfect models of the climate system are integrated independently of one another, starting from observed initial conditions to provide forecasts for the future. To gauge
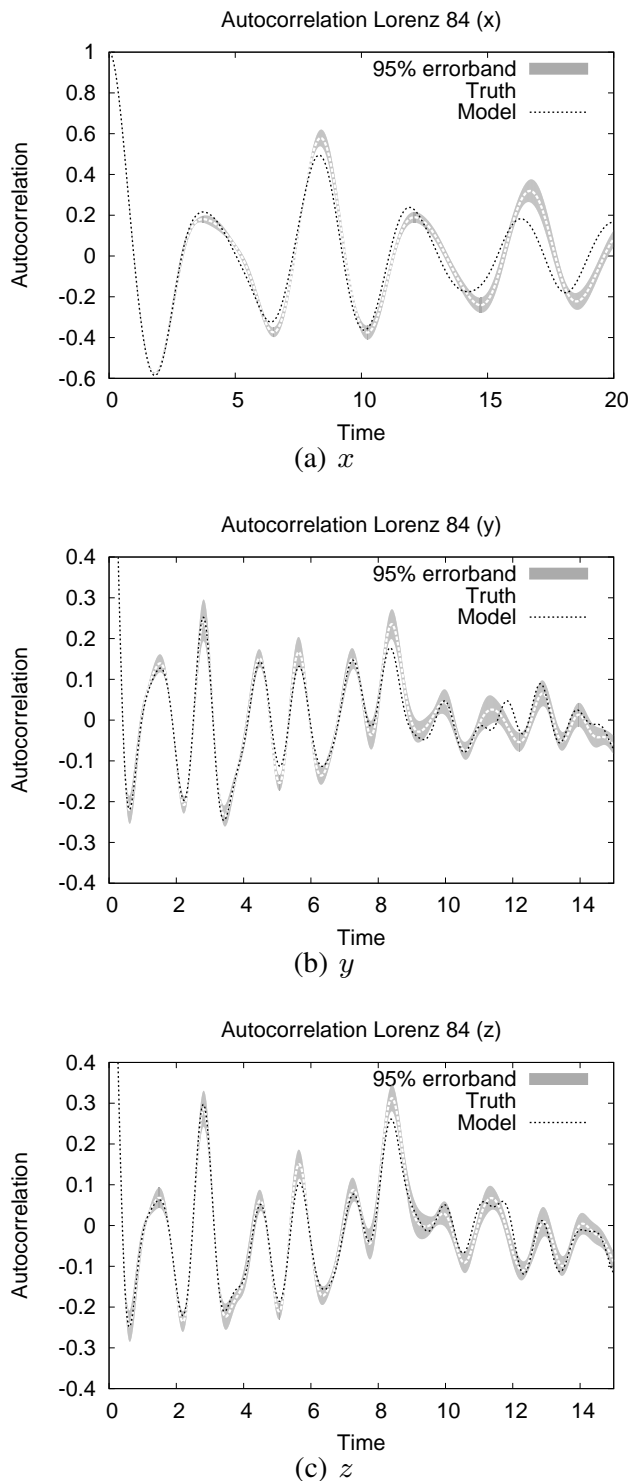
(a) $x$



(b) $y$



(c) $z$

**Fig. 18.** Autocorrelation for the super-model (black) and the standard Lorenz 84 system (white) The shaded area corresponds to the 95 % error band for the truth based on 100 runs of 3000 time units.

model uncertainty, the outcomes of the different models are combined into a single estimate of the probability density function of climate variables. This study indicates that better estimates of the true probability function can be obtained if the models are taught, using past observations, to combine the strengths of each into a single forecast of the probability density function.

In case the models synchronize perfectly, all models follow the same trajectory and no information is lost by ensemble averaging. A sudden loss of synchronization and increase of spread between the individual model solutions in the super-model might indicate a loss of predictability during that period. Also, the distance from the ensemble mean might be an indication of model error, the larger the distance, the larger the model error. We note that, since the models are coupled (rather than independent), the super-model is not an ensemble in the classical sense of probability theory. But in addition to evaluating the ensemble mean, valuable information might be obtained from the spread between the individual model realizations.

A large gap exists between the simple, chaotic systems of this study and the complex, state-of-the-art climate models. Many questions need to be addressed in order to apply the same approach to these models. There is the practical limitation of computer capacity to enable the parallel execution of an ensemble of state-of-the-art models that need to exchange information at every time step. In the study of Kirtman et al. (2003) two atmospheric models were coupled to one ocean model so in principle it should be feasible to couple several atmospheric models to several ocean models. Computational grids in the various climate models differ, so regridding should be part of the information exchange. Regridding is standard practice in the information exchange between the atmosphere and ocean components of climate models.

An important issue is the choice of state variables for the connections and the number of connections. In this study all state variables were connected and had similar dynamics. In the climate models the different state variables are driven by different physical processes and display distinct dynamical behaviour at various time scales. It is not clear a priori which state variables should be connected. In addition the number of connections that can be learned on the basis of historical data is limited and therefore careful choices for the connections need to be made. Optimizing many parameters leads to the risk of overfitting and failure to simulate the behavior outside the training set. In the implementation of the super-modeling approach with complex weather or climate models one must strive to keep the number of connections as few as possible in order to limit the number of parameters that need to be optimized. One possible approach would be to not connect the state variables, but the various parameterized physical processes that contribute to the tendency of the state variables. Most of the model uncertainty resides in these parameterized processes, so it makes sense to direct the learning to these processes.

The introduction of connections between models introduces non-physical terms in the equations and might disrupt physical balances present in the original equations. On the true climate attractor, the physical balances are naturally obeyed. In case the models synchronize with the observations and each other, the physical balances are obeyed also, and the precise implementation of the connections does not matter. With regard to the approximate balances that are observed in the atmosphere, like the hydrostatic or the geostrophic balance, we note that if the connections between the state variables remain sufficiently weak, these balances will be restored continuously by dynamical adjustments within each individual model.

An additional complicating factor for the learning phase is the difference in time-scale between atmosphere and ocean. Adjustments in the atmosphere have a short time-scale, but the ocean adapts to these changes on a much longer time-scale. Through its influence on the atmosphere, the ocean introduces longer time-scales in the atmosphere as well. So short integrations during the learning phase do not probe these effects. This might hamper the learning. On the other hand, there are well documented examples that certain model errors develop very quickly. Rodwell and Jung (2008) is an example of how a particular model error, namely the amount of aerosols over the Sahara, leads to remote biases in winds and precipitation through a sequence of fast atmospheric processes. It is a nice example of how difficult it can be to diagnose the origin of a particular model bias. But at the same time it illustrates that model biases develop quickly and it implies that improving climate models on the short time-scales (as in the super-model approach) could be a fruitful approach to reduce model biases. Jung (2005) shows that most of the climatogical errors in the circulation are already developed after ten days of integration when starting from observed states. In addition in Jung et al. (2009) it is reported that a change in the deep convection scheme, which describes fast turbulent processes on the time-scales of hours, led to a marked reduction of the climatological model bias. See further Palmer and Weisheimer (2010) for a more theoretical view on the origin of systematic model errors.

A main result of this study is that super-model solutions are not unique. However, the different super-models have similar quality and therefore this does not pose a severe problem and makes finding a suitable super-model solution easier. The existence of quite distinct super-model solutions of good quality remains a bit of a mystery.

A hierarchy of earth system models of intermediate complexity (EMIC's) could be used to address these various issues. The EMIC's resemble the state-of-the-art climate models in their structure, but differ in that the parameterization schemes for the physical processes are much less elaborate, fewer processes are explicitly modeled and the spatial resolution is much coarser. It has already been demonstrated that two different quasi-geostrophic channel models will synchronize with only limited connections (Duane and Tribbia,

2001, 2004). A fruitful strategy might be to start from a relatively simple climate model and add to the complexity in small steps and address a specific issue at each step. In a similar fashion as in this study a ground truth model is assumed at each step and an ensemble of imperfect models is created by perturbing parameters and/or using different formulations for unresolved processes.

An alternative learning strategy that is explicitly based on synchronization is outlined in the study by Duane et al. (2007). In this strategy the super-model equations contain nudging terms to the truth as in our Eqs. (5) and additional evolution equations are formulated for the parameters. The moment the super-model synchronizes with the truth the parameters stop updating. This alternative learning strategy leads to a particularly simple learning law that could be useful in the implementation of the super-model approach using more complex models. The strategy has been demonstrated with Lorenz systems (Duane et al., 2009). Quinn et al. (2009) present a learning strategy based on the minimisation of a similar cost function by varying parameters and the initial state of a single model simulation with an additional nudging term to the truth that allows the model to synchronize with the truth during the learning. In this sense it resembles the approach by Duane et al. (2007). Our approach is different in that we consider an ensemble of short model simulations from different initial states, do not nudge to the truth and only vary the parameters and not the initial state to minimize the cost function.

The main caveat is that the super-model is trained on historical data and in a climate prediction problem is subsequently applied to simulate the response of the system to an external forcing. It is not guaranteed that the super-model will also simulate this response more realistically, since the response was not part of the training. Therefore the super-model approach is more likely to be successful in weather- and seasonal predictions since the cases to be predicted remain closer to the cases present in the training set.

The problem is not peculiar to the super-modeling approach, but arises with climate models generally. Climate models contain numerous empirical parameters with values that are optimized to reproduce observed historical behavior as close as possible. Parameter values are adjusted such as to reduce the errors in the climatological mean fields using historical observations (Severijns and Hazeleger, 2005; Medvigy et al., 2010). It is not guaranteed that this automatically implies that the response to a future greenhouse gas forcing will be simulated more realistically. Other processes might play a role there (Klocke et al., 2011).

On the other hand, there are also more optimistic arguments indicating that super-modeling might work as a strategy for improving climate change predictions. In this study we obtained the encouraging result that for the Lorenz 1963 system the super-model was able to accurately predict the change of the attractor after doubling parameter $\rho$. This result gives hope that the super-modeling strategy applied to

climate models will help in improving predictions of the climate response to future greenhouse gas forcing.

To conclude, as argued above, the super-modeling approach might work or might fail in the context of complex climate models. This remains to be seen. We propose to further study the merits of the super-modeling approach using a similar hypothetical model setting as in this paper with a hypothetical perfect model simulating the truth and a set of imperfect models simulating state-of-the-art approximations of this hypothetical truth. For this super-modeling research, model classes of increased complexity – both for the hypothetical truth as for the imperfect model approximations – are then to be explored.

Finally, we wish to stress that we believe that long-term improvements in climate predictions must come from improving the description of the physical processes on the basis of dedicated process studies and observational databases. This is a slow, but necessary process. In the meanwhile, given a set of imperfect models, we could try to improve predictions by combining the strengths of the individual models either through multi-model or super-model approaches.

Edited by: A. Alessandri

# References

Compo, G., Whitaker, J., and Sardeshmukh, P.: Feasibility of a 100-Year Reanalysis Using Only Surface Pressure Data, B. Am. Meteorol. Soc., 87, 175–189, 2006.

Duane, G. and Tribbia, J.: Synchronized chaos in geophysical fluid dynamics, Phys. Rev. Lett., 86, 4298–4301, 2001.

Duane, G. and Tribbia, J.: Weak Atlantic-Pacific teleconnections as synchronized chaos, J. Atmos. Sci., 61, 2149–2168, 2004.

Duane, G. S., Tribbia, J. J., and Weiss, J. B.: Synchronicity in predictive modelling: a new view of data assimilation, Nonlin. Processes Geophys., 13, 601–612, doi:10.5194/npg-13-601-2006, 2006.

Duane, G., Yu, D., and Kocarev, L.: Identical synchronization, with translation invariance, implies parameter estimation, Phys. Lett. A, 371, 416–420, 2007.

Duane, G., Tribbia, J., and Kirtman, B.: Consensus on Long-Range Prediction by Adaptive Synchronization of Models, EGU General Assembly, Vienna, Austria, 2009.

Fletcher, R. and Reeves, C.: Function minimization by conjugate gradients, Comput. J., 6, 149–154, 1963.

IPCC: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, UK and New York, USA, 2007.

Jung, T.: Systematic errors of the atmospheric circulation in the ECMWF forecasting system, Q. J. Roy. Meteorol. Soc., 131, 1045–1073, 2005.

Jung, T., Balsamo, G., Bechtold, P., Beljaars, A., Köhler, M., Miller, M., Morcrette, J.-J., Orr, A., Rodwell, M., and Tompkins, A. M.: The ECMWF model climate: Recent progress through improved physical parametrizations, available from the ECMWF website: http://ecmwf.intECMWF, last access: 29 June 2011, Seminar Proceedings on Parameterization of Subgrid-scale Processes, 233–249, 2009.

Kirtman, B., Min, D., Schopf, P., and Schneider, E.: A new approach for CGCM sensitivity studies, COLA Technical Report, 154, 50 pp., 2003.

Klocke, D., Pincus, R., and Quaas, J.: On constraining estimates of climate sensitivity with present-day observations through model weighting, J. Climate, in press, doi:10.1175/2011JCLI4193.1, 2011.

Lorenz, E.: Deterministic Nonperiodic Flow, J. Atmos. Sci., 20, 130–140, 1963.

Lorenz, E.: Irregularity, a fundamental property of the atmosphere, Tellus A, 36, 98–110, 1984.

Medvigy, D., Walko, R. L., Otte, M. J., and Avissar, R.: The Ocean-Land-Atmosphere Model: Optimization and evaluation of simulated radiative fluxes and precipitation, Mon. Weather Rev., 138, 1923–1939, 2010.

Nelder, J. and Mead, R.: A simplex method for function minimization, Comput. J., 7, 308–313, 1965.

Palmer, T. N. and Weisheimer, A.: Diagnosing the Causes of Bias in Climate Models – Why it is so Hard?, available from the ECMWF website: http://ecmwf.int, last access: 29 June 2011, ECMWF Seminar Proceedings on Diagnosis of Forecasting and Data Assimilation Systems, 1–13, 2010.

Pecora, L. and Carroll, T.: Synchronization in Chaotic Systems, Phys. Rev. Lett., 64, 824–824, 1990.

Quinn, J. C., Bryant, P. H., Creveling, D. R., Klein, S. R., and Abarbanel, H. D. I.: Parameter and state estimation of experimental chaotic systems using synchronization, Phys. Rev. E, 80, 016201, doi:10.1103/PhysRevE.80.016201, 2009.

Rodwell, M. and Jung, T.: Understanding the local and global impacts of model physics changes: an aerosol example, Q. J. Roy. Meteorol. Soc., 134, 1479–1497, 2008.

Rössler, O.: An equation for continuous chaos, Phys. Lett. A, 75, 397–398, 1976.

Severijns, C. A. and Hazeleger, W.: Optimizing parameters in an atmospheric general circulation model, J. Climate, 18, 3527–3535, 2005.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philos. T. Roy. Soc. A, 365, 2053–2075, 2007.

van Veen, L.: Active and passive ocean regimes in a low-order climate model, Tellus A, 53, 616–628, 2001.

Yang, S., Baker, D., Li, H., Cordes, K., Huff, M., Nagpal, G., Okereke, E., Villafañe, J., Kalnay, E., and Duane, G. S.: Data Assimilation as Synchronization of Truth and Model: Experiments with the Three-Variable Lorenz System, J. Atmos. Sci., 63, 2340–2354, 2006.