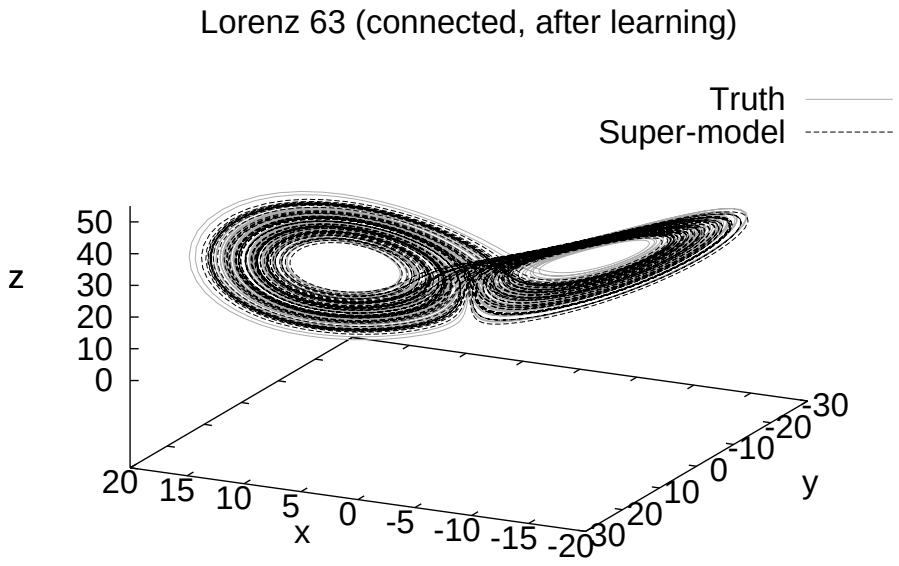
Universiteit Utrecht

Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Verkeer en Waterstaat

Improving simulations by combining imperfect models through learning



Author:
Leonie van den Berge
UU: 3020053

Supervisors:
Frank Selten (KNMI)
Heinz Hanßmann (UU)

August 30, 2010

Abstract

With the uprising climate problems like the warming due to greenhouse gasses, the need for future climate scenarios is growing fast. Many state of the art climate models exist, but they are imperfect and give a large variaty of future scenarios. In this thesis we try to find a way to improve the predictions of climate models by combining them. Multiple imperfect models will exchange information during the simulations and are combined into a super-model. The idea is not only that the imperfect models can use each others strengths, but also that synchronization between the models can occur, such that a mutual prediction results. The objective is to find a way to let models exchange information such that the resulting super-model gives a better prediction than any of the separate imperfect models. A learning process is developed to objectively determine the exchange of information between the models. The approach is tested on small chaotic dynamical systems that have similar properties as the atmosphere. The system with standard parameter values is regarded as truth and three imperfect models are created by perturbing these parameter values. By using the small chaotic dynamical systems we can use the information of the truth to see how well the approach works.

Contents

Chapter 1

Introduction

Our planet is warming due to the concentration of greenhouse gasses in the atmosphere. How much it will warm and the effect it will have on the total climate system can only be predicted by models. Although several good climate models have been developed, the models are not able to simulate the historical climate exactly. The quality of the different climate models can be assessed by comparing simulations over a historical period with the observations over that period. An example of historical values can be seen in figure 1.1, which shows the mean error in the annual mean air surface temperature between 1980 and 1999 [IPCC, 2007]. The figure is taken from the Fourth Assessment report of the Intergovernmental Panel on Climate Change, a panel consisting of scientists that periodically report on the scientific consensus regarding climate change. Error are typically 1 to $3°C$, but regionally can be as large as 4 to $5°C$.

Another problem is that the climate models do not agree on the respons of the climate system to the future scenarios of greenhouse gas emissions, which can be seen in figure 1.2, which was also taken from the IPCC Fourth Assessment report. Here the temperature and precipitation change are shown in respons to three different future scenarios: one with strong emissions of greenhouse gasses (upper panel), one with moderate emissions (middle panel) and one where emissions of greenhouse gasses are reduced immediately (lower panel) [Nakicenovic, 2000]. Man different models have calculated the respons to these scenarions and from figure 1.2 it can be seen that the spread is very large. This shows how difficult it is to make a prediction based on these models. As all models have some piece of information about the truth, but none of them captures the whole truth, it is hard to decide which model is closest to the truth. A solution to this problem is combining the different climate models to find one future prediction. At the moment the different outcomes of climate models are already combined by taking a weighted average [Tebaldi and Knutti, 2007], but in this way the models are combined afterwards and the models cannot benefit from the strengths of the other models. Moreover it is not clear how the weights should be determined.

We think that this can be improved by letting the models exchange information during the simulations and combining them into a super-model. In this way the models can use each others strengths. The objective is to find a combination of models for which the prediction of the truth outperforms the prediction made by any of the separate models. We want to find this combination by using a learning process that makes use of information of the truth to ensure an objective combination.

1.1 Supporting research

The basis of this approach is synchronization of chaotic systems. Chaotic systems can synchronize with each other when linked through one common signal [Pecora and Carroll, 1990]. If we let the models exchange information it is possible that they synchronize to each other. This would lead to a mutual

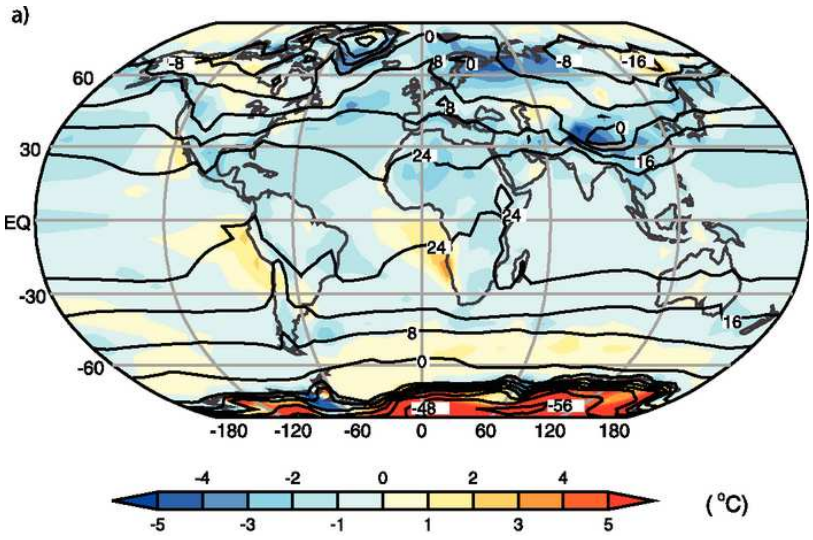


Figure 1.1: Observed climatological annual mean sea surface temperature (SST) and, over land, surface air temperature (labelled contours). The colours indicate the multi-model mean error in these temperatures, simulated minus observed. This is figure 8.2 (a) in the Fourth Assessment report of the IPCC [IPCC, 2007].

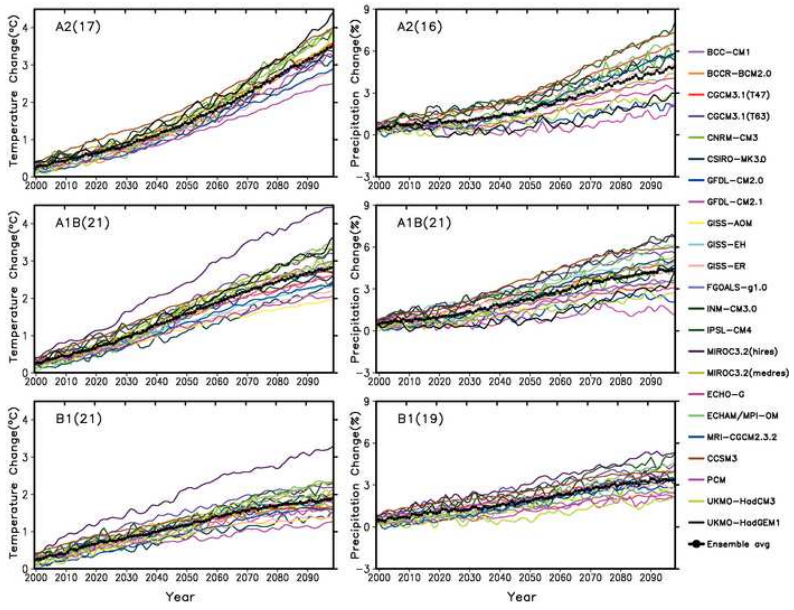


Figure 1.2: Time series of globally averaged surface warming (surface air temperature change, $°C$) (left) and precipitation change (%) (right) from the various global coupled models for the scenarios A2 (top), A1B (middle) and B1 (bottom) [Nakicenovic, 2000]. Numbers in parentheses following the scenario name represent the number of simulations shown. Values are annual means, relative to the 1980 to 1999 average from the corresponding 20^{th}-century simulations, with any linear trends in the corresponding control run simulations removed. Multi-model (ensemble) mean series are marked with black dots. This is figure 10.5 from the Fourth Assessment report of the IPCC [IPCC, 2007].

prediction. The hope is that the synchronized state is closer to the truth, which would improve the predictions. Synchronization has already been applied in data assimilation [Yang et al., 2006]. Throughout the world measurements are performed, but the measurements are not spread over the whole world and are irregular in time. Some areas lack measurements like the oceans. Data assimilation fills these gaps by combining measurements and models, such that the models synchronize with the measurements and a full estimate of the atmospheric state is obtained.

Exchanging information between multiple atmospheric models has been done before by for instance Kirtman and Shukla [Kirtman et al., 2003]. They coupled an ocean model to two atmospheric models. One of the models provided the heat flux and the other the momentum flux to the ocean model. It turned out that the predictions for a certain combination of these models were better for the coupled model than for each of the models separately. This supports the idea that by exchanging information during simulations can lead to an improved prediction.

1.2 Goal

The main goal of this thesis is to find a way of exchanging information between existing imperfect climate models and combining them into a super-model that outperforms the individual models. A learning proces will be developed and used to determine the exchange of information using observational data. The approach is tested on certain small chaotic dynamical system that have similar properties as the atmosphere. To test the quality of the resulting super-models measures will be introduced that can give an indication of how well the super-model approximates the truth for longer periods of time.

1.3 Outline

First we will introduce the different small chaotic dynamical systems that are used to test our approach in section 2. Then the method of letting models exchange information and the learning process are explained in detail in chapter 3. In chapter 4 we will explain how to choose certain parameters that are used in the learning process. To compare the quality of the super-model solutions some measures will be introduced in chapter 5. These first four chapters will give some background information on the article written for Earth System Dynamics [van den Berge et al., 2010] (chapter 6), that contains a much shorter description of the approach and the main results of this thesis. Variations of the approach are discussed in chapter 7 and some more results can be found in chapter 8. The discussions and conclusions can be found in chapter 9 and section 6.5 respectively.

Chapter 2

Dynamical systems

To illustrate and develop the super-modelling approach we make use of small chaotic dynamical systems. In this thesis a dynamical system is taken to be the solution of an initial value problem as in equation (2.1), where $x \in \mathbb{R}^3$ is dependent on time t and $f : \mathbb{R}^3 \times \mathbb{R}_+ \mapsto \mathbb{R}^3$ is a (non linear) function depending on parameters α. $x(t_0) = x_0$ gives the initial condition that determines the solution [Verhulst, 2000].

$$\dot{\vec{x}} = f_\alpha(\vec{x}, t),\ \vec{x}(t_0) = \vec{x}_0 \tag{2.1}$$

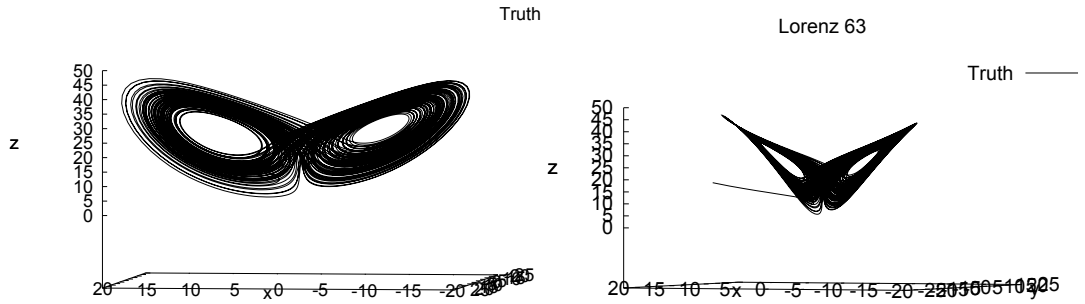
The dynamical systems that are used to test the approach are all chaotic. This means that the solution of the system is sensitively dependent on the initial conditions. Starting two solutions from two slightly different initial conditions will cause the solutions to deviate and at some time the solutions will be totally different. As climate models are also chaotic [Lorenz, 1984], this was an important property for our test systems. The approach was tested on three chaotic dynamical systems, namely the Lorenz 63, Rössler and Lorenz 84 systems that we will introduce in this section.

2.1 Lorenz 63

The Lorenz 63 system was proposed by Edward Lorenz in 1963 [Lorenz, 1963] and is given by equations (2.2). The equations were obtained from a fluid dynamics model of the atmosphere by a Galerkin approximation, which is a very crude simplification. The Lorenz 63 system has no physical meaning left, but is very well known because Lorenz analyzed this particular system in his article on deterministic chaos. It is used for the most extensive testing in this study. The standard parameter values of the Lorenz 63 system are $\sigma = 10$, $\rho = 28$ and $\beta = \frac{8}{3}$.

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z \end{aligned} \tag{2.2}$$

The Lorenz 63 system has been studied a lot and one of the advantages is that a bifurcation analysis can be done analytically [Doedel et al., 2006]. Bifurcation theory studies the behaviour of systems depending on values of the parameters. Changing the value of one of the parameters of the system results in different behaviour, for instance from a fixed point to a periodic orbit [Verhulst, 2000]. In our approach we will make use of the bifurcation theory that is known for the Lorenz 63 system. One of the features we will be using is a subcritical Hopf bifurcation at $\rho_H = \frac{\sigma(3+\sigma+\beta)}{\sigma-\beta-1}$ as a boundary between different sorts of behaviour. At a subcritical Hopf bifurcation two unstable periodic solution corresponding to the two stable fixed points exist for $\rho < \rho_H$, which vanish at $\rho = \rho_H$. For $\rho > \rho_H$ the two unstable fixed points



(a) Point of view 1 (b) Point of view 2

Figure 2.1: Two points of view of the Lorenz 63 attractor.

appear and a chaotic attractor arises, which is shown in figure 2.1 for the standard parameter values. An attractor is a bounded set that attracts all solutions within its domain of attraction. This means that any solution started from a random initial condition in its domain of attraction approaches this set arbitrarily close.

The Lorenz 63 attractor is also called *butterfly* because of its shape with two 'wings'. Each wing contains an unstable fixed point in the middle around which the solution circles. A solution can circle around one of the unstable fixed points for a long period before it makes a transition to the other wing. This aperiodic behaviour is one of the main reasons that the Lorenz 63 system is so chaotic and it decreases the predictability of the system.

2.2 Rössler

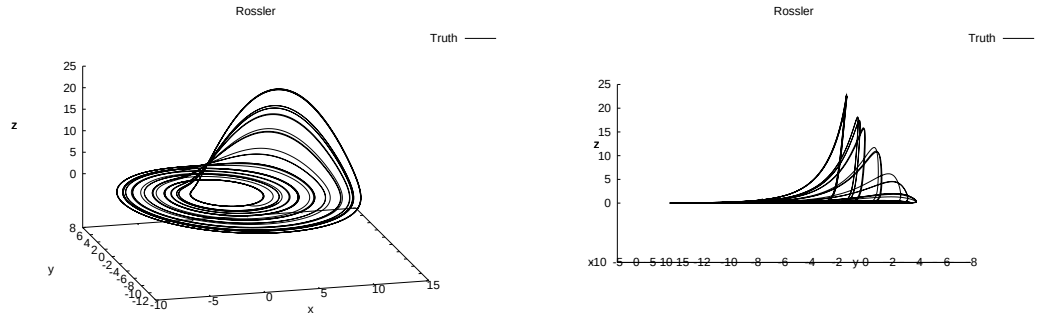
The Rössler equations were proposed by O.E. Rössler in [Rössler, 1976] to be a simplification of the Lorenz 63 system. The attractor of the Rössler system has only one wing instead of the two wings of the Lorenz 63 attractor. The equations are given in (2.4), where the standard parameter values are $a = 0.2$, $b = 0.2$ and $c = 5.7$.

$$\begin{aligned} \dot{x} &= -(y + z) \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c) \end{aligned} \tag{2.3}$$

In figure 2.2, the attractor of the Rössler system is shown with indeed only one circle instead of the two wings visible in the Lorenz 63 system. The lack of the second wing makes this system less chaotic as the transitions from one wing to another were one of the reasons for the very chaotic nature of the Lorenz 63 system. On the other hand the Rössler system has a very rich bifurcation diagram [Barrio et al., 2009] and is therefore still a good test case for the approach.

2.3 Lorenz 84

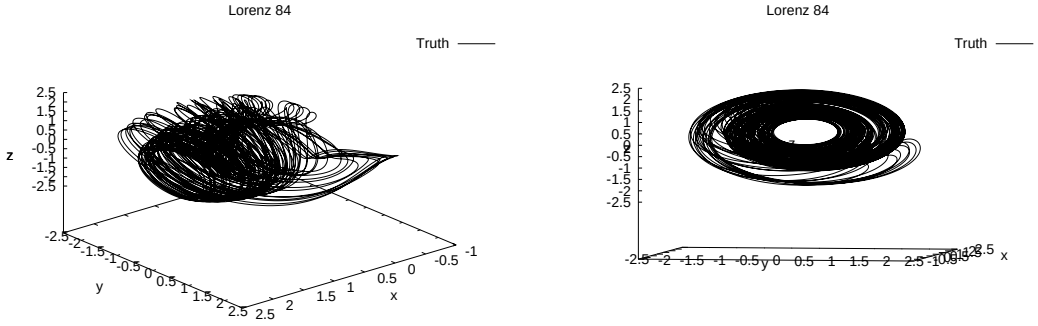
The Lorenz 84 model was also proposed by E.N. Lorenz in 1984 [Lorenz, 1984] and is given by the equations (2.4). This system has a physical meaning as x represents the intensity of the globe-encircling westerly wind current and y and z represent a travelling large-scale wave that interacts with the westerly wind. The standard parameter values are $a = \frac{1}{4}$, $b = 4$, $F = 8$ and $G = 1$, which gives the model unstable chaotic behaviour corresponding to a winter [van Veen, 2002].



(a) Point of view 1

(b) Point of view 2

Figure 2.2: Two points of view of the Rössler attractor.



(a) Point of view 1

(b) Point of view 2

Figure 2.3: Two points of view of the Lorenz 84 attractor.

$$\begin{aligned} \dot{x} &= -y^2 - z^2 - ax + aF \\ \dot{y} &= xy - bxz - y + G \\ \dot{z} &= bxy + xz - z \end{aligned} \tag{2.4}$$

The attractor of the Lorenz 84 system is shown in figure 2.3. The waves spiral in the y-z plane, but move parallel to the x axis from negative to positive values of x in the inner part of the attractor and back again along the outside. The attractor is more complicated than the Lorenz 63 and Rössler attractors and also has a chaotic nature, making it a good testcase for the approach too.

Chapter 3

Method

The objective of the approach is to find a way to let imperfect models exchange information and combine them into a super-model, such that the solution for the super model is a better approximation of the truth than each of the imperfect models on their own. To study whether it is possible to do this we test the approach on the small chaotic dynamical systems introduced in section 2, for which we create a truth and imperfect models that predict it. The advantage of this is that we can calculate all information of the truth. When applying this approach to larger climate models the information of the truth (reality) will exist of measurements and will therefore not be as complete. The measurements are not available everywhere and a good spread only exists for the last 50 years. Other advantages are that the attractor of the systems can be plotted and calculations take only a short time with respect to real climate models.

In this test we have all information about the truth, so we will use this to see how good the approach is. In the learning process we do need a small amount of information about the truth as there is no other way to train the super-model. More about the quality of the super-model solutions can be found in chapter 5, where the measures of quality of a resulting solution will be introduced.

In this chapter we will discuss how to let the imperfect models exchange information during simulations and we will address the learning process that is used to determine the information exchange.

3.1 Creating truth and imperfect models

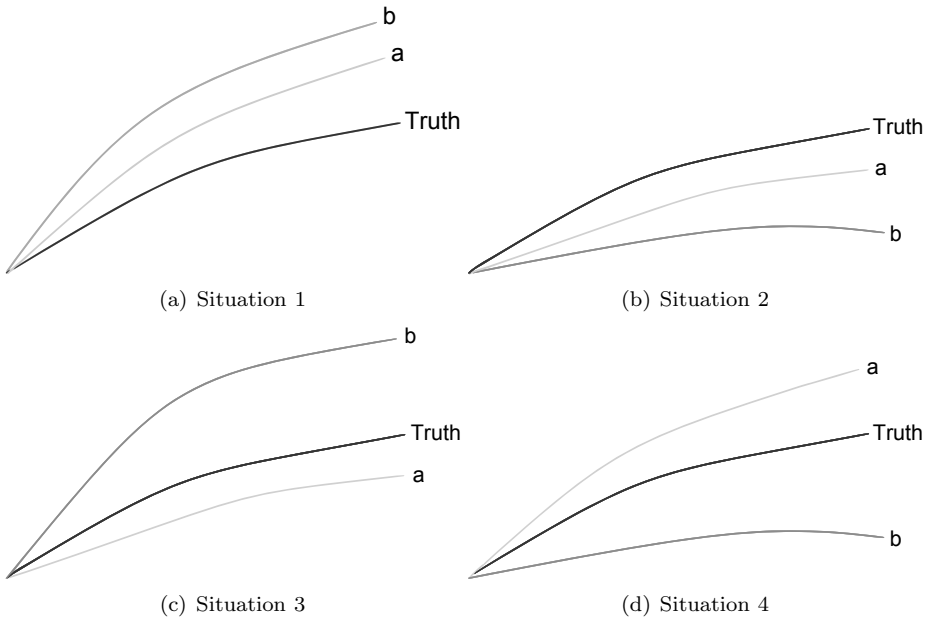
For the small chaotic dynamical systems we first need to create a truth and imperfect models that can predict it. The truth is taken to be the dynamical system with the standard parameter values α (equation (3.1)). The imperfect models, indexed by k, have the same equations as the truth and only differ in the parameter values α_k (equation (3.2)). These parameter values are perturbed to get different behaviour for the imperfect models, but still the imperfect models contain information about the truth. We have chosen to use three imperfect models throughout this thesis, but different numbers of imperfect models can be used as well.

$$\dot{\vec{x}} = f_\alpha(\vec{x}, t) \tag{3.1}$$

$$\dot{\vec{x}}_k = f_{\alpha_k}(\vec{x}_k, t) \; k = 1, 2, 3 \tag{3.2}$$

3.2 Connecting imperfect models

Next the imperfect models need to exchange information during the simulations. This is done by adding a *nudging term* to the equations of each of the models connecting it to the other two. For the Lorenz



(a) Situation 1 (b) Situation 2

(c) Situation 3 (d) Situation 4

Figure 3.1: The four situations that can occur with model a being closer to the truth than model b.

63 system the equations for model k are shown in equation (3.3). The nudging term is a linear term consisting of the difference between two models $(x_j - x_k)$ and a connection constant C_{kj} for each of the variables and is added for both other imperfect models such that each model is connected to the other two models. The larger the connection constant the more the nudging pulls model k to model j in a certain variable. The connection coefficients will be determined by the learning process to get an objective way of choosing them.

$$\begin{aligned} \dot{x}_k &= \sigma_k(y_k - x_k) + \sum_{j \neq k} C^x_{kj}(x_j - x_k) \\ \dot{y}_k &= x_k(\rho_k - z_k) - y_k + \sum_{j \neq k} C^y_{kj}(y_j - y_k) \\ \dot{z}_k &= x_k y_k - \beta_k z_k + \sum_{j \neq k} C^z_{kj}(z_j - z_k) \end{aligned} \tag{3.3}$$

The connection coefficients should be chosen such that models that are far away from the truth are pulled closer by the models that are already closer to the truth itselves. On the other hand the bad models should not attract the good models, so these connection coefficients should be small. The sign of the connection coefficients might be just as important. Taking a negative sign of a connection constant means that we repel the solution of a model from the other model instead of attracting it. For this reason we choose to keep the connection coefficients positive. The next illustration is used to motivate this choice.

Assume that we have two models, a and b that are connected to each oter and let's focus on the x variable only. For some point in time we have that $x_a = x + \epsilon_a$ and $x_b = x + \epsilon_b$, where x represents the truth and the ϵ_i can be both positive or negative. Now assume that $|\epsilon_b| > |\epsilon_a|$, implying that x_a is closer to the truth than x_b. There are four situations that agree with this, which can be found in figure 3.1. Now we will show what happens with positive and negative connection coefficients in all four situations. In all situations it is positive to attract model b to model a, as model a is closer to the truth. It is different for model a. In situations 1 and 2 model a should not be attracted to model b as this would

	Positive connection		Negative connection	
	Model a	Model b	Model a	Model b
1	$-$	$+$	$+$	$-$
2	$-$	$+$	$+$	$-$
3	$+$	$+$	$-$	$-$
4	$+$	$+$	$-$	$-$

Table 3.1: For each of the four situations shown in figure 3.1 it is shown what the effect is of a positive and a negative connection constant. A (possibly) positive effect is indicated with a + and a negative effect with a $-$.

repel the model from the truth for sure. Therefore repelling model a from model b might have a positive effect, though only if the repulsion is not too strong. For situations 3 and 4 attracting model a to model b can have a positive effect if the attraction is not too strong as it is posible that model a will come closer to the truth than before. Repelling model a from model b would cause model a to deviate more from the truth and would therefore not be desirable.

In table 3.1 the results are summarized, where a + indicates a (possibly) positive effect and $-$ a negative effect. Most plusses can be found for the positive connection and for model b, a lower quality model, negative connections would have a negative effect in all situations. This supports the choice for positive connection coefficients. We can argue that the learning process will choose the best strategy and will therefore keep the connection coefficients positive anyway. However the learning process sometimes led to negative connection coefficients. Since we do not want this to happen we will force the connection coefficients to be positive. In addition we prefer the connection coefficients to be positive, since we want the models to synchronize, which is less likely to happen if the models repel each other in one or several variables.

This whole reasoning holds for a certain moment in time, but how the connection coefficients influence the asymptotic behaviour of the connected super-model cannot be deduced on the basis of this reasoning. In effect connecting the models together creates a new dynamical system, that can have a very rich bifurcation structure depending on the connection coefficients.

Keeping the connection coefficients positive can be achieved in several ways. One of these ways is not to use the simple nudging terms, but to change equations (3.3) to equations (3.4), where the exponents of the connection coefficients are taken, such that the nudging terms will be positive at all times. When using this approach we found that the solution sometimes tended to infinity, causing the method to break down, which will be further discussed in section 7. The method that is currently used is to do penalize negative connection coefficients in the cost function, that is introduced in the learning process. More about this can be found in section 3.4.

$$\begin{aligned} \dot{x}_k &= \sigma_k(y_k - x_k) + \sum_{j \neq k} \exp(C^1_{kj})(x_j - x_k) \\ \dot{y}_k &= x_k(\rho_k - z_k) - y_k + \sum_{j \neq k} \exp(C^2_{kj})(y_j - y_k) \\ \dot{z}_k &= x_k y_k - \beta_k z_k + \sum_{j \neq k} \exp(C^3_{kj})(z_j - z_k) \end{aligned} \tag{3.4}$$

3.3 Super-model

At this point we still have three equations for each of the three imperfect models. The imperfect models will therefore be combined in one super-model by taking a simple mean for each of the variables, as shown in equation (3.5), where the subscript s indicates the super-model and 1, 2 and 3 the three imperfect models.

$$\begin{aligned} x_s &= \frac{1}{3}(x_1 + x_2 + x_3) \\ y_s &= \frac{1}{3}(y_1 + y_2 + y_3) \\ z_s &= \frac{1}{3}(z_1 + z_2 + z_3) \end{aligned} \tag{3.5}$$

The super-model could also be chosen to be determined by a weighted average as shown in equation (3.6). In that case the weights w_1, w_2 and w_3 could be determined by the learning process as well. In this study we take the weights equal to 1.

$$\begin{aligned} x_s &= \frac{1}{3}(w_1 x_1 + w_2 x_2 + w_3 x_3) \\ y_s &= \frac{1}{3}(w_1 y_1 + w_2 y_2 + w_3 y_3) \\ z_s &= \frac{1}{3}(w_1 z_1 + w_2 z_2 + w_3 z_3) \end{aligned} \tag{3.6}$$

3.4 Cost function

The objective of the learning process is to find the connection coefficients that make the super-model a good approximation of the truth on longer time scales, but as chaos plays such a large role in the small dynamical systems and in the climate models, it needs to be taken into account in our learning process. We have chosen to define a cost function that is a measure of the proximity of the super-model to the truth. For climate prediction it is important to have good predictions for longer time scales, but chaos ensures that the solutions will deviate more and more if we take longer periods of time. This means that the cost function would be dominated by chaos instead of by the model errors, which we want to measure. Therefore shorter trajectories of Δ time units are taken to compare truth and super-model.

The systems we use are strongly dependent on the initial conditions. Therefore it is important that the super-model is not trained on just one initial condition, but on multiple. Therefore the difference between super-model and truth is measured for multiple initial conditions t_i. The situation is depicted in figure 3.2, where the grey areas are the areas that we want to minimize using the cost function.

The cost function F can be found in equation (3.7), where $\vec{x}_o$ is (the observed state of) the truth and $\vec{x}_s$ refers to the super-models. For each of the K initial conditions t_i the difference is integrated for Δ time units. The term $\frac{1}{K\Delta}$ is a normalisation term.

$$F(\vec{C}) = \frac{1}{K\Delta} \sum_{i=1}^{K} \int_{t_i}^{t_i+\Delta} (\vec{x}_s(C,t) - \vec{x}(t))^2 \gamma^t dt \tag{3.7}$$

The only new term is γ that was also chosen to make sure that chaos does not dominate the cost function. γ is chosen between 0 and 1 such that the difference between truth and super-model at later time steps will be taken less into account than the difference in the beginning. How γ, Δ and K are chosen is

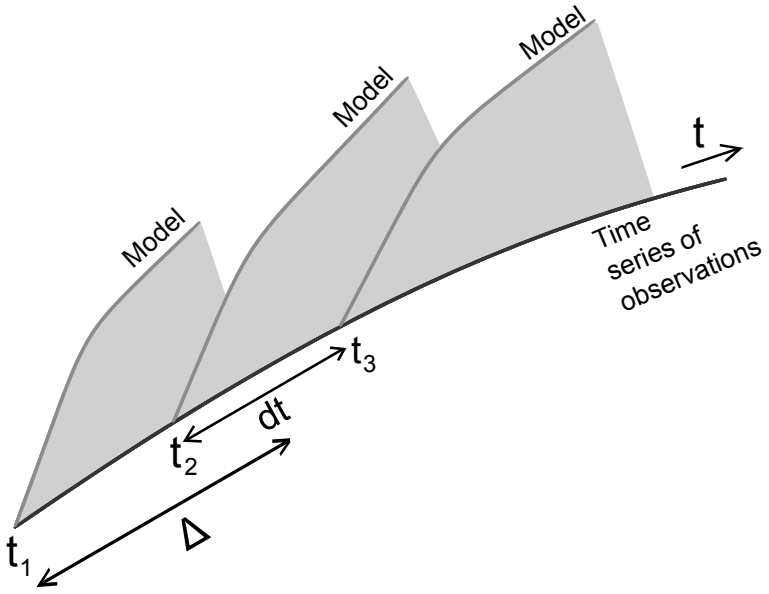


Figure 3.2: The situation for the learning process, where for short trajectories of Δ time units and for multiple initial conditions t_i the difference between truth and super-model is calculated.

discussed in chapter 4.

As mentioned before we keep the connection coefficients positive by manually making negative connection coefficients less appealing. This is done by adding an extra term to the cost function when (one of) the connection coefficients become negative. The extra term is the absolute value of the negative connection constant $|C^x_{kj}|$. This term is generally much larger than the values of the cost function, forcing the minimization method to use positive connection coefficients.

3.5 Minimization methods

By minimizing the cost function introduced in the previous section we determine all 18 connection coefficients (6 for each model). This learning process needs to be executed only once after which the connection coefficients will not change anymore. By taking the connection coefficients to be positive this space has been bounded from below, but is still very large. There are different minimization methods that can be used. Two of the minimization methods that were used in this approach are the Fletcher-Reeves conjugate gradient method and the amoeba or simplex method.

3.5.1 Conjugate gradients minimization

The Fletcher-Reeves conjugate gradient minimization starts from an initial guess and searches for a minimum in the direction of the gradient at the initial guess [Fletcher and Reeves, 1963]. The slope is followed downwards until an interval is found that contains a minimum. After that the exact minimum is calculated. This minimization technique searches for a minimum of the cost function F by searching for a new set of connection coefficients of the shape $\vec{C} + a\nabla F$, where a is a scalar. The disadvantage of searching only in the direction of steepest descent is that the method could fall into a local minimum preceding a more global minimum as in figure 3.3.

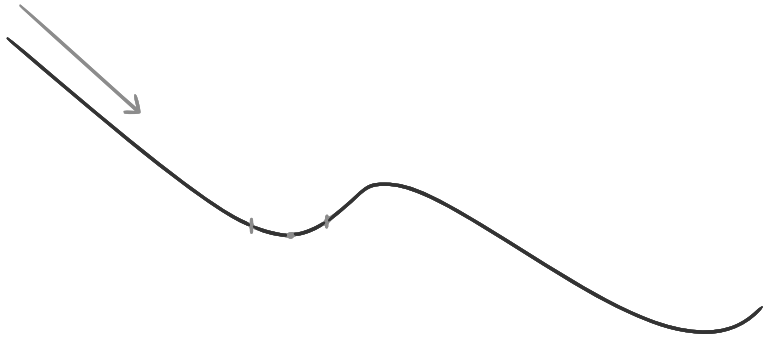


Figure 3.3: The conjugate gradients method may not succeed in finding a more global minimum of the cost function.

3.5.2 Adding initial conditions

To find a more global minimum we minimize the cost function for an increasing number of initial conditions. In this way the minimization is applied to multiple different initial conditions as well, but gradually. This method of adding multiple initial conditions can be compared to *simulated annealing* [Cerny, 1985], which is a flexible method for finding an approximate minimum. The idea behind simulated annealing is to use thermodynamics for the optimization. Cooling a system means that the system will try to find an equilibrium with lower energy. The idea is that cooling a system fast will give a larger chance of staying in a local minimum of energy. If the system is cooled slowly, also called annealing the system, the system will be able to find global minima.

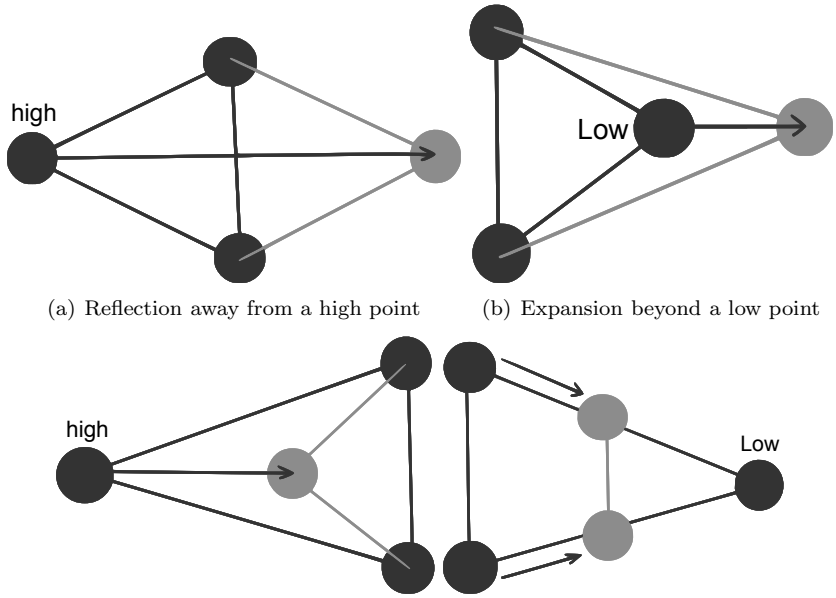
The approach of adding initial conditions is similar to simulated annealing in the sense that with fewer initial conditions the cost function contains fewer local minima, a situation that can be compared with the initial phase of the simulated annealing method, where the large perturbations effectively smooth the cost function. Increasing the number of initial conditions and starting the minimization from the value of the minimum in the cost function found for fewer initial conditions enhances the probability of finding a global minimum.

3.5.3 Amoeba minimization

In addition to using a method that only approaches minima from one side, we also used the amoeba or simplex method [Nelder and Mead, 1965] that explores the landscape of the cost function to find a global minimum. The amoeba is a little animal with $N+1$ feet that can feel the temperature (the value of the cost function) in an N dimensional space to find a cool spot (minimum).

The amoeba effectively uses only three of its legs: the one where the function has the highest value (P), the second highest value (Q) and the lowest value (R). Its basic moves are expansion, reflection and contraction as shown in figure 3.4 for $N = 2$, such that the amoeba has only three feet. The first step is to reflect away from P as shown in figure 3.4(a). The reflected foot will be denoted by P_{ref}. After that the moves are determined on the situation that arises. If P_{ref} has a smaller value than the coolest foot R, then expansion is used as shown in figure 3.4(b). If P_{ref} is still higher than Q, then we contract away from it as shown in figure 3.4(c). If this move decreases the value P_{ref}, then we determine which feet at this point are the highest, second highest or lowest and start with reflecting a foot again. If P_{ref} increases or doesn't change we first try the other contraction method as in figure 3.4(d), before determining P, Q and R again. An example can be found in figure 3.5.

The advantage of using this method is that we can use it to find a more global minimum. Still it is necessary to repeat the method a couple of times to let it converge and it seems that the minima are still



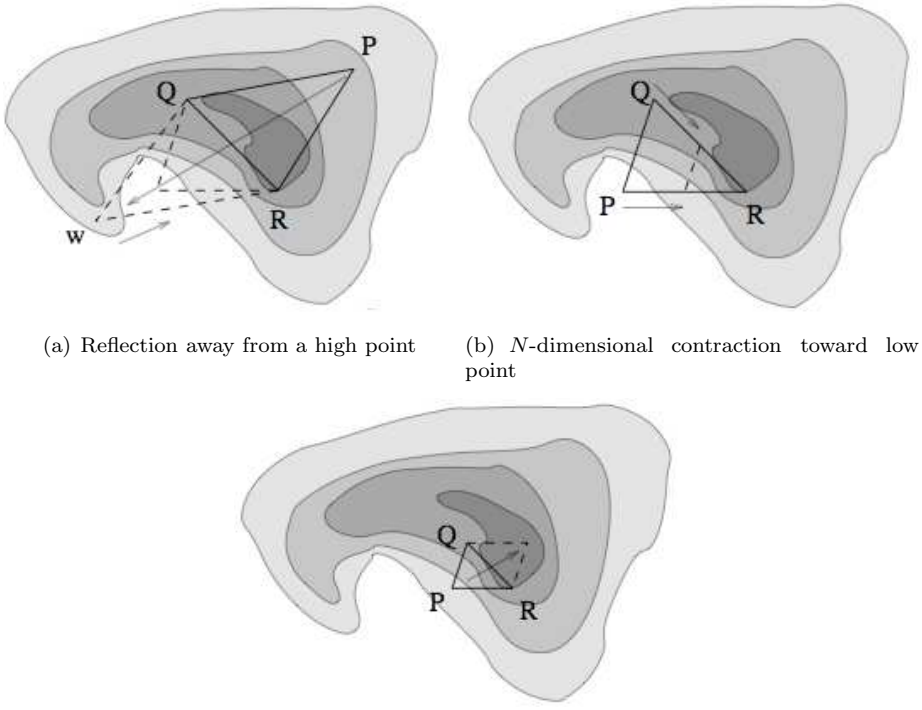
(a) Reflection away from a high point

(b) Expansion beyond a low point

(c) 1-dimensional contraction away from high point

(d) N-dimensional contraction toward low point

Figure 3.4: The basic moves of the amoeba minimization in $N = 2$ dimensions.



(a) Reflection away from a high point

(b) N-dimensional contraction toward low point

(c) Reflection away from a high point

Figure 3.5: An example of finding a minimum with the amoeba method for $N = 2$. Darker colors indicate a lower value of the function. P indicates the foot of the amoeba on the highest value, Q on the second highest and R the one on the lowest value.

dependent on the initial conditions. Therefore the minimization method can still be improved by using another minimization method.

Chapter 4

Determining parameters

The cost function (equation (3.7)) that is used to determine the connection coefficients is not only dependent on the connection coefficients. The cost function also depends on the choice of several parameters, such as γ, Δ and the set of initial conditions used. In this section we will motivate the choices we made for the parameters. In chapter 8 some attention will be paid to the influence of the parameters on the learning process.

4.1 Determining γ

As explained in section 3.4, γ is a constant between 0 and 1, such that in the cost function the difference between truth and model will be given less weight at later time steps. γ is chosen to prevent that chaos dominates the cost function, such that the value of the cost function is determined mainly by the quality of the super-model. The choice for γ should therefore depend on how chaotic the system is.

A measure of how chaotic a system is is the number of time steps in which errors double. This doubling time T can be calculated by starting a perfect model from a slightly perturbed initial condition and measuring at which time step this initial error is doubled. Repeating this for many different perturbed initial conditions gives an average number of time steps for doubling errors. This information was used in chaper 6 to find a value for γ by just prescribing a certain value ($\frac{1}{2}$) for γ^T, but the doubling time can also be used to determine the fraction of chaos present in different time steps. γ can then be chosen to be the fraction of the error that is not due to chaos.

To calculate the fraction of chaos present in the system, we will find an expression for the total error growth and the error growth due to chaos in every time step. We know that the error growth due to chaos is dependent on the previous errors. The doubling time T can be used to make up an expression for the error growth due to chaos. For a perfect model error growth is purely due to chaos and is given by equation (4.1), where f_i is the initial error. At each multiple of the doubling time T the error is doubled, resulting in the factor $2^{t/T}$, but as the initial error is not due to chaos we subtract it from the factor $2^{t/T}$.

$$f_c(t) \quad = \quad (2^{t/T} - 1) f_i \tag{4.1}$$

For an imperfect model there is also error growth due to flaws in the model. This means that the error at each time step consists of a modelling error and an error due to chaos. For simplicity we take an average modelling error f_M in each time step, which results in equation (4.2).

$$f_m(t) = t f_M \tag{4.2}$$

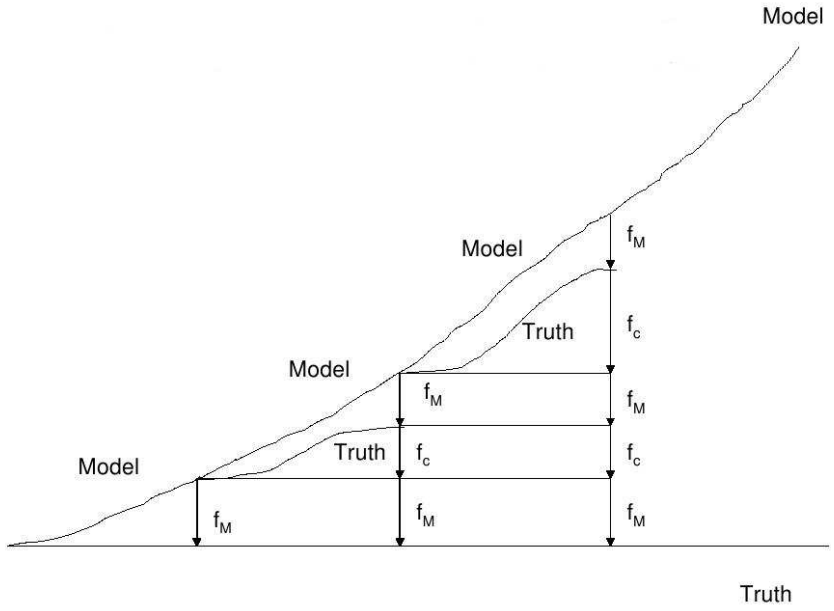
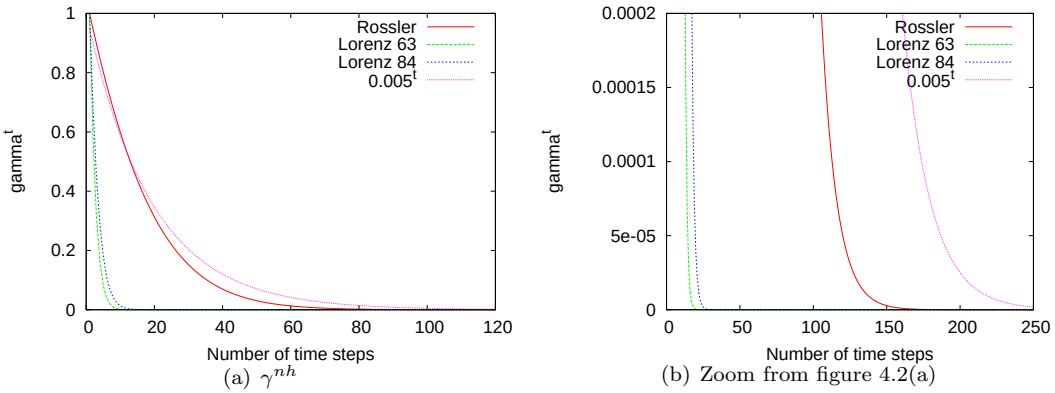


Figure 4.1: Growth of errors due to flaws in the model f_M and chaos f_c.

The appearance of the modelling errors makes the error due to chaos not only dependent on the initial error but also on the errors that arise during the other time steps. The situation can be found in figure 4.1, where the truth and the model are started from the same initial condition. At each time step a copy of the truth can be used to find the error due to chaos, but this is just an imaginary experiment as we can use equation (4.1) to find the size of the errors due to chaos.
We know from equation (4.1) that errors due to chaos grow with a factor of $2^{1/T}$ in each time step. So let's see what happens if we introduce a new error in each time step. Let E be the total error depending on the number of time steps. The first error is a modelling error in time step 1. Going to the next step we get an error due to chaos with initial error f_M, so that will be $(2^{1/T} - 1)f_M$. We also get a new modelling error.

The third step is where it gets more interesting. Here we get two new errors due to chaos. The first one is due to the first modelling error and the second due to the modelling error arising in time step 2. This second error will produce the same term as we had in the equation for $E(2)$, the other is just the next error as in equation (4.1). In each step an extra term is added, resulting in equation (4.3) for n time steps.

$$\begin{aligned} E(0) &= 0 \\ E(1) &= f_M \\ E(2) &= 2f_M + (2^{1/T} - 1)f_M \\ E(3) &= 3f_M + (2^{1/T} - 1)f_M + (2^{2/T} - 1)f_M \\ &\dots \\ E(n) &= nf_M + \sum_{i=1}^{n-1} (2^{i/T} - 1)f_M \end{aligned} \tag{4.3}$$



(a) γ^{nh}

(b) Zoom from figure 4.2(a)

Figure 4.2: γ^{nh} from equation (4.4) for each of the three small dynamical systems, compared to the value of 0.005^{nh}.

This gives us a formula for the total error and we also have the formula for the modelling error and the error due to chaos. This results in the following equation for γ^{nh}, which is just the fraction of the error that is due to modelling errors at time step n. h is the size of the time steps.

$$\gamma^{nh} = \frac{nf_M}{nf_M + \sum_{i=1}^{n-1}(2^{i/T} - 1)f_M} = \frac{n}{n + \sum_{i=1}^{n-1}(2^{i/T} - 1)} \tag{4.4}$$

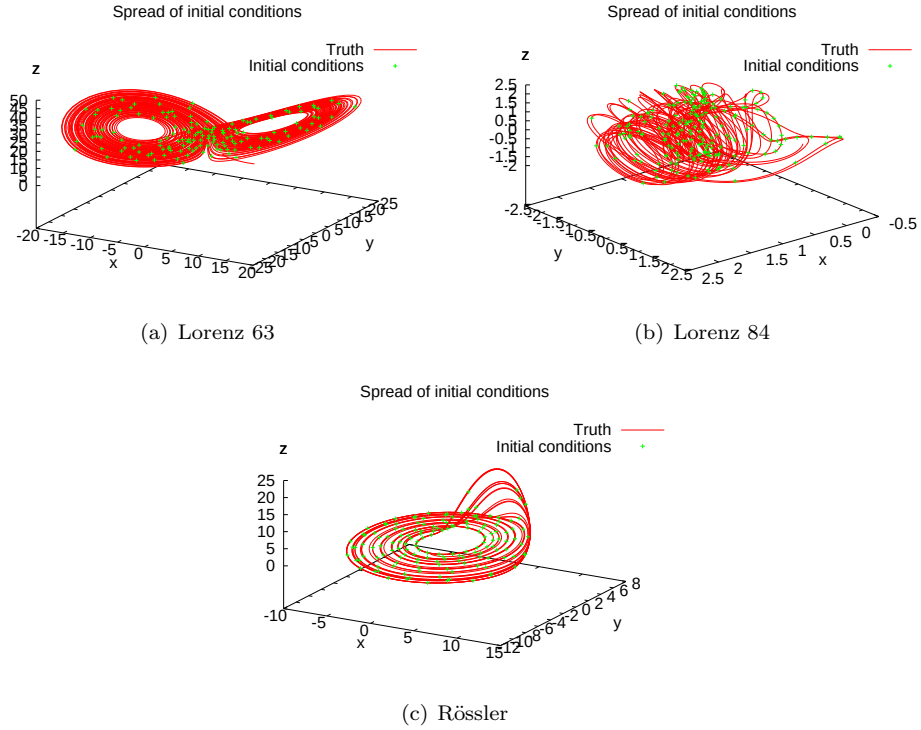
The problem of this expression is however that the chaos grows so fast that the resulting γ is so small, that learning process is limited to the first three time steps, which is not reasonable. This can be seen in figure 4.2, where the γ^{nh} is plotted for the three different systems, along with the line belonging to 0.005^{nh} (so $\gamma = 0.005$). The line for $\gamma = 0.005$ is close to the γ^{nh} for the Rössler system, but still not steep enough for both Lorenz systems. The zoom in figure 4.1 shows even more clearly that these calculations lead to values of γ that are too small, such that too little information is taken into account. This was the reason that this calculation was not used in finding a value for γ.

4.2 Determining Δ

As mentioned in section 3.4, Δ must be chosen in accordance with γ. It makes no sense to choose a small γ and a large Δ as γ^t will be (close to) zero long before the Δ time steps are reached. However, it can make a difference to choose Δ small if γ is large. In that case the difference between super-model and truth is only taken into account almost fully for a relatively short time. In this way Δ may have a lot more influence on the cost function than γ has. Looking again at figure 4.2 it may not be such a bad idea to choose Δ small, but γ relatively large, because of the large steepness in the beginning of the curves. In this approach γ is still taken rather large compared to figure 4.2 as it is a larger than the doubling time for each of the systems (table 4.1), which are the values based on chaos. For Lorenz 84 this turned out not to work as too little information was included in the cost function, such that the cost function was very flat and no minima could be found. Therefore we chose $\Delta = 800$, which is in agreement with the time scales present in the Lorenz 84 system (figure 8.7 shows that significant autocorrelation values are still present at $\Delta = 800$ time steps). For the larger Δ we also chose a larger γ of 0.8.

	Lorenz 63	Lorenz 84	Rössler
doubling time	0.75	1.08	6.67
Δ	1	2.2	12

Table 4.1: The number of time units in which an initial error doubles and the value of Δ for each of the three systems.



(a) Lorenz 63

(b) Lorenz 84

(c) Rössler

Figure 4.3: The spread of the initial conditions used in the learning process for each of the three systems.

4.3 Determining the initial condititons

Attention needs to be paid to determining the set of initial conditions used. To make sure that the dependence on these initial conditions is minimal the number and the distance between these initial conditions has to be chosen. Important is that the number of initial conditons is large enough to have a minimal dependence of the cost function on the initial conditions. The distance between initial conditions will have to be such that the attractor of the system is covered. For the climate system this is of course not known and all available observations will have to be used.

Plotting the cost function for different sets of initial conditions can indicate whether enough initial conditions are used. If enough initial conditions are used the dependence on the set of initial conditions should be small, implying that the cost function should look similar for a different set of initial conditions. This was used in chapter 6 to show that enough initial conditions were used. We could even argue whether too many initial conditions are used as the connection coefficients usually don't change anymore after we've minimized the cost function for about 50 to 100 initial conditions.

The spread of initial conditions can be seen in figure 4.3 for each of the three systems. The attractors are clearly covered by the initial conditions, which indicates that the initial conditions are chosen well.

Chapter 5

Measures

With the small dynamical systems that are used in this study, the solution of super-model and truth can be compared to check how realistic a certain solution is. This is however not possible for larger climate models for which this approach will be used eventually. Therefore some measures are introduced from which the quality of the solution can be detemined, that are also applicable to larger climate models.

There are two different sorts of measures in meteorology: measures for weather predictions and measures for climate predictions. For weather it is important that the prediction is close to the truth, but it can only be close for a short time as chaos will cause the model to deviate from the truth. Therefore weather measures determine the proximity of the model to the truth for shorter time scales. An example of these measures is the cost function. Climate predictions are made for much longer times. For these time scales it is impossible for the model to stay close to the truth due to error growth. Therefore climate measures determine how similar the statistical behaviour of the model is to the truth. An example of this is comparing the attractor for truth and model. As our approach will be used on climate models, the second sort of measures will be used to determine the quality of the super-model.

5.1 Attractor

One of the simplest measures used is visually comparing the attractors of super-model and truth. If the attractors (nearly) coincide this implies that the solutions for the long term are in the same regime. If an attractor is different the behaviour is less similar and therefore the super-model has a lower quality. For higher dimensional systems like climate models, a visual comparison of the attractor is impossible, so we need a computational measure that characterizes the attractor as introduced in the next section.

5.2 Mean, standard deviation and covariance

The mean, standard deviation and covariance are statistical measures, that can be used on systems of arbitrary size. They are therefore usable in larger systems too. The standard deviation σ (equation (5.2)) will give the spread of the solution with respect to the mean μ (equation (5.1)). The covariance (equation (5.3)) measures the linear dependency between the variables.

$$\mu(x) = \frac{1}{n}\sum_{i=1}^{n} x(i) \tag{5.1}$$

$$\sigma(x) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x(i)-\mu(x))^2} \tag{5.2}$$

$$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x(i)-\mu(x))(y(i)-\mu(y)) \tag{5.3}$$

As the mean, standard deviation and covariance are properties of an attractor, this again shows the behaviour on large time scales. Not only for this reason, but also to get a lower error it is important to calculate the values for a long period of time, 5.000 time units. An error indication is calculated by repeating the experiment a large number of times (500), giving an average value, and measuring the spread of with respect to this average by taking the standard deviation. The error estimation is equal to twice the standard deviation, which gives the range in which approximately 95% of all outcomes are situated if we assume a normal distribution. This error estimation can indicate that the errors are large, in which case the comparison of the values can only be done roughly, but if it is small the comparison can be more precise.

5.3 Autocorrelation

The autocorrelation is again a statistical measure, similar to the covariance. The autocorrelation is the correlation between a function and the function itself at a later point in time. Therefore the autocorrelation is a function of the difference between the two points in time, also called the delay time τ in equation (5.4). For a periodic function taking the delay time equal to a period gives a correlation of 1 as all points shifted a period in time give the same values. The autocorrelation will be lower for delay times that are not a multiple of a period. For our systems we expect that the autocorrelation function converges to 0 for larger delay times due to chaos.

$$ac(x,\tau) = \frac{1}{n\sigma(x)^2}\sum_{i=1}^{n}(x(i)-\mu(x))(x(i+\tau)-\mu(x)) \tag{5.4}$$

5.4 Synchronization

The third measure is the use of *synchronization*. Chaotic systems can synchronize if they are linked with a common signal [Pecora and Carroll, 1990], which is one of the reasons that we expect that the approach works (section 1.1). In figure 5.1 a super-model is shown that synchronizes with the truth. This means that the super-model is at all times close to the truth, but is not necessarily equal.

Synchronization can take place between chaotic systems by using simple nudging terms as in the connection of models in section 3.2, but this time not all variables will be connected. For the Lorenz 63 system for instance only the y variable is connected to the truth, for which the equations are shown in equation (5.5). This simple nudging term gives model k information about the truth, which is needed for synchronization. How easily a model can synchronize, for instance with a low value of the nudging strength n (section 5.4.1) or in a short time (section 5.4.2), indicates how much information the super-model itself contains about the truth. Therefore the super-model is a better approximation of the truth if it more easily synchronizes with the truth.

$$\begin{aligned}
\dot{x}_k &= \sigma_k(y_k - x_k) + \sum_{j\neq k} C^x_{kj}(x_j - x_k) \\
\dot{y}_k &= x_k(\rho_k - z_k) - y_k + \sum_{j\neq k} C^y_{kj}(y_j - y_k) + n(y_o - y_k) \\
\dot{z}_k &= x_k y_k - \beta_k z_k + \sum_{j\neq k} C^z_{kj}(z_j - z_k)
\end{aligned} \tag{5.5}$$

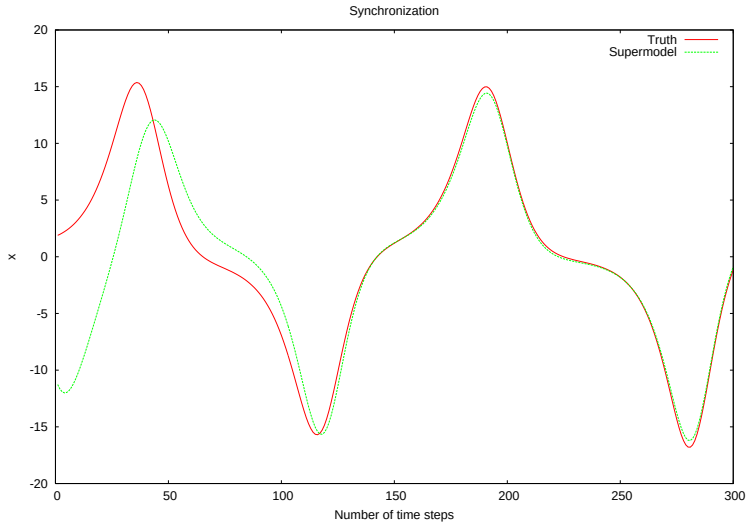


Figure 5.1: Synchronization of the super-model to the truth.

A definition of when a model is synchronized with the truth is definition 1. Synchronization depends on two tolerances δ and ϵ and the number of time units T that the super-model should be close to the truth. It means that the solutions of super-model and truth at some point should be closer together than δ and should stay closer than ϵ from that point onwards.

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

The two tolerances were chosen since the difference between super-model and truth can be a little larger for a short time at some points, but the solutions remain synchronized. This can be seen for example in figure 5.1 at $t = 280$, where the distance temporarily is a little larger. To take this into account the second tolerance ϵ, which is larger than δ, was introduced. This kind of definition for synchronization resembles, but is different from, the definition of *Lyapunov stability* (definition 2) [Arnold, 1989], which is a type of stability where a mapping stays close to a fixed point forever, but not necessarily as close as at the beginning. The important thing in Lyapunov stability is that the mapping stays in the proximity of the fixed point for ever, which is also the case for synchronization. The difference is that in the definition of synchronization there is a lower bound to the value of ϵ, for which the definition holds.

Definition 2 *A fixed point x_0 of the map A is Lyapunov stable if $\forall \epsilon > 0$, $\exists \delta > 0$, such that if $|x - x_0| < \delta$, then $|A^n x - A^n x_0| < \epsilon$ for all $0 < n < \infty$ [Arnold, 1989]*

Whether a model is synchronized or not with a certain nudging strength n depends on the two tolerances δ and ϵ and the time T that the model must stay synchronized. In all experiments T is chosen to be 1000 time units. The dependence on δ is not strong as well. It should not be chosen so small that the tolerance is never reached, but changing it otherwise does not have much effect on synchronization. Synchronization does depend strongly on ϵ. It turns out that there is a certain threshold for ϵ for which the system synchronizes more easily. If ϵ is below this threshold the model will need much higher nudging strength than when ϵ is above this threshold.

Choosing the parameters is not straight forward, but it can be done by using the information about the truth. In this case we know the truth very well, so we can do this, but it may be hard to do this for larger climate models. The parameters should at least be chosen such that a copy of the truth is able to synchronize with a low nudging strength. Still it may be necessary to increase the tolerances if the

models do not synchronize at all. As said there are two ways of using synchronization as a measure and we will address both in the next two sections.

5.4.1 Calculating nudging strength

The first way of measuring the quality of a super-model using synchronization is to find the lowest nudging strength that leads to synchronization with the truth with just one particular variable nudged for a particular choice of ϵ and δ. This nudging strength can be found by increasing the value of the nudging strength untill synchronization takes place. For each nudging strength 500.000 time steps are used to check synchronization. Usually a model that is able to synchronize, synchronizes much earlier implying that this is a good choice. We chose only integer nudging strengths are taken in this approach, but also decimal numbers can be used.

The sensitive dependence on initial conditions is decreased by repeating the experiment many times. We should also take into account which variable is nudged. For the Lorenz 63 model synchronization was found not to work for the z variable, but it did work well for both x and y [Yang et al., 2006]. This implies that the synchronization may also depend on which variable is nudged, so for comparison the variable that is nudged is taken to be the same.

The nudging strength that is needed can be compared with that of other models and the truth. The lower the nudging strength the better the model is. When a model has a nudging strength close to that of the truth, this indicates that the model is a good approximation. The lowest nudging strength needed can also be used for measuring of the average number of time steps that are needed for synchronization as described in the next section.

5.4.2 Calculating the average number of time steps

How fast a model synchronizes with the truth can be a measure of quality as well. As with measuring the lowest nudging strength needed for synchronization, the number of time steps is dependent on the variables that are nudged and even more so on the initial conditions. This leads to a large spread in the number of time steps needed for synchronization. There is also an additional dependence on the nudging strength, making this measure less precise than measuring the nudging strength. To use this measure, the nudging strength should at least be high enough for the model to reach synchronization.

5.5 cost function

In the introduction of this chapter we already mentioned that the cost function is a measure for weather predictions. As we are interested in climate predictions, using this measure in the learning process seems to be an odd choice. Especially since a good weather prediction is not necessarily a good climate prediction. So why do we use such a measure to try and predict the climate and wouldn't it be better to use a different measure in the learning process?

The reason that a measure for weather predictions is used instead of a measure for climate predictions has to do with chaos. As explained in section 3.4 the trajectories are chosen short to keep the influence of chaos on the cost function small. Another advantage is that only a small amount of measurements are needed. If we would use a cost function that would measure the quality for climate predictions the solutions should be compared for a larger period of time and for this many more measurements are needed. For the small chaotic dynamical systems this is no problem, but it is when applying the approach to larger climate models were measurements would be used. There are measurements available, but we have measurements with a good cover over the globe for only the last 50 years. This is enough to train for weather predictions, that range only over 10 days or so, but probably too short for climate predictions. By using data assimilation as described above, we can produce a more global coverage from data from

longer than 50 years ago. But the question remains if we have enough data to use in the learning process if we take a different cost function.

The results show that the cost function is indeed a good measure for climate predictions as well, as can be seen in chapter 6. It also seems that the other measures give better values for super-models that have lower values for the cost function than the super-models with higher values for the cost function, suggesting that the use of the cost function in the learning process enables to predict the long term statistics of the system.

Chapter 6

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

By 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]Dept. of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO 80309, US

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 climate simulations.

6.1 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]. They constructed what they called an interactive ensemble 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, Whitaker, and Sardeshmukh, 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 minimization of a cost function that measures the difference between the truth and the super-model during short integrations.

In section 6.2 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 section 6.3 and 6.4. Discussion and conclusion of the method and ideas for improvement can be found in section 6.5.

6.2 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 metaphore for the atmosphere, because of its abrupt regime changes and unstable nature. The equations for the Lorenz 63 system are

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z. \end{aligned} \tag{6.1}$$

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.

6.2.1 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

$$\begin{aligned} \dot{x}_k &= \sigma_k(y_k - x_k) + \sum_{j \neq k} C^x_{kj}(x_j - x_k) \\ \dot{y}_k &= x_k(\rho_k - z_k) - y_k + \sum_{j \neq k} C^y_{kj}(y_j - y_k) \\ \dot{z}_k &= x_k y_k - \beta_k z_k + \sum_{j \neq k} C^z_{kj}(z_j - z_k), \quad k = 1, 2, 3, \end{aligned} \tag{6.2}$$

where k indexes the three imperfect models with perturbed parameter values σ_k, β_k and ρ_k and C^x_{kj}, C^y_{kj} and C^z_{kj} 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

$$\begin{aligned} x_s &= \frac{1}{3}(x_1 + x_2 + x_3) \\ y_s &= \frac{1}{3}(y_1 + y_2 + y_3) \\ z_s &= \frac{1}{3}(z_1 + z_2 + z_3). \end{aligned} \tag{6.3}$$

6.2.2 Cost function

We assume that we have a long time series of observations of the truth $\vec{x}_o$. We pick initial conditions $\vec{x}_o(t_i)$ from this long time series at K times t_i, $i = 1, \cdots, K$, separated by fixed distances d. Short integrations of length Δ are performed with the super-model starting from these K initializations (see figure 6.1). 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 $\boldsymbol{C}$.

$$F(\boldsymbol{C}) = \frac{1}{K\Delta} \sum_{i=1}^{K} \int_{t_i}^{t_i + \Delta} |\vec{x}_s(\boldsymbol{C}, t) - \vec{x}_o(t)|^2 \gamma^t dt \tag{6.4}$$

The cost function is normalized by $\frac{1}{K\Delta}$, so that it represents the time averaged mean squared error. The factor γ^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 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 internal error growth dominates. Since we wish to measure the model errors, the factor γ^t discounts the errors at later times to decrease the contribution of internal error growth.

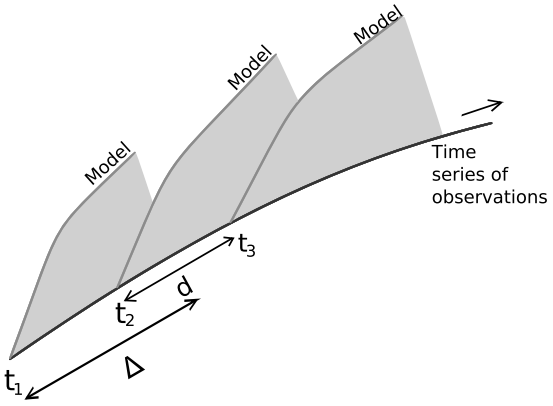


Figure 6.1: The cost function is based on short integrations of the super-model starting from observed initial conditions of the truth at times t_i and measures the mean-squared difference between the short evolutions of the super-model and the truth as indicated by the shaded areas. The short integrations span a time interval Δ and d denotes the fixed time interval between the initial conditions t_i.

We base the choice of γ 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 τ of the initial error. We choose γ such that $\gamma^\tau = \frac{1}{2}$, so at time τ 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.

6.2.3 Minimisation

For a fixed choice of the number of initializations K the cost function solely depends on the connection coefficients $\boldsymbol{C}$ in equation (6.4). The super-model can be determined by finding a minimum in the cost function in the 18 dimensional space of $\boldsymbol{C}$. 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 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 section 6.3.

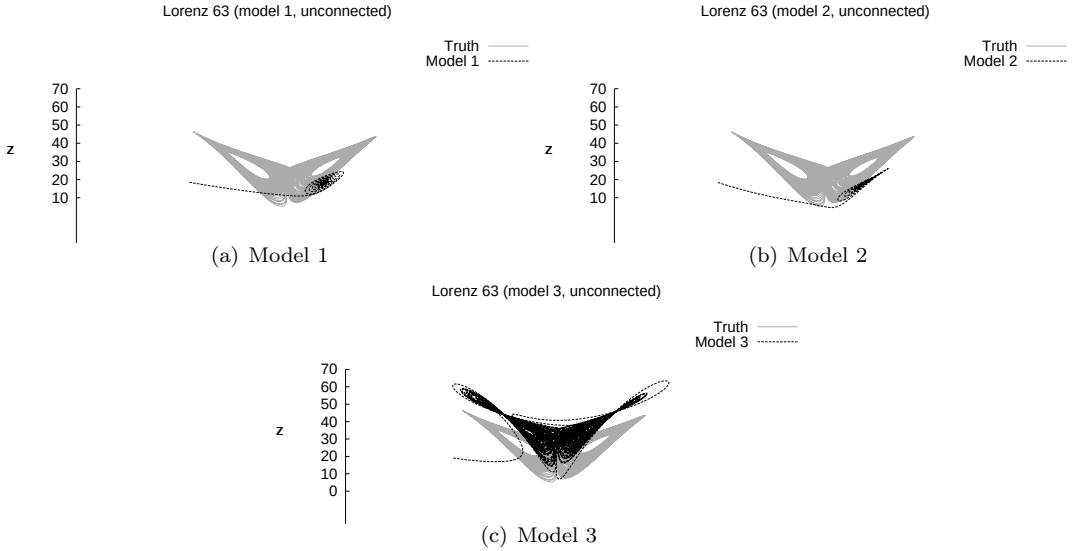
To avoid negative solutions for the connection coefficients we added an extra term in the cost function in case one of the coefficients becomes negative. This term is just the absolute value of the negative connection coefficient, which easily dominates the value of the cost function.

6.3 Results Lorenz 63

Three imperfect models are created by perturbing the standard parameter values as displayed in table 6.1. 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 figure 6.3. 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

	σ	ρ	β
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%)

Table 6.1: Standard and perturbed parameters for the Lorenz 63 system.



(a) Model 1

(b) Model 2

(c) Model 3

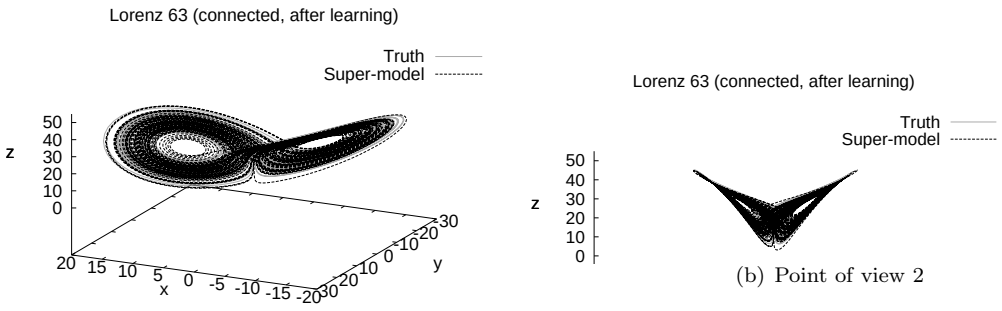
Figure 6.2: Trajectories for the three unconnected imperfect models (black) and the standard Lorenz 63 system (grey). The trajectory for the imperfect models includes the transient evolution from the initial condition towards the attractor.

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 ρ below the Hopf bifurcation, whereas model 3 has a value for ρ that lies far above the Hopf bifurcation. For the truth the value of ρ lies above the Hopf bifurcation as well, which is why model 3 and the truth have similar behaviour.

The minimization 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 figure 6.3. 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 improvement in the behaviour of the connected imperfect model solutions as depicted in figure 6.4 (to be compared with figure 6.3) is a clear indication of the feasibility of super-modeling in the context of this low-dimensional chaotic system.

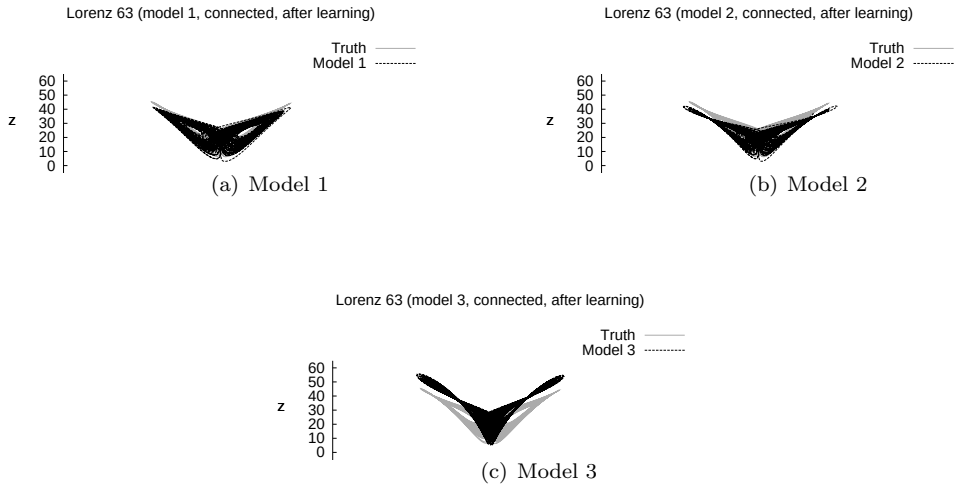
In addition to this minimum, we found that by choosing different initial values for the connection coefficients in the minimization procedure different local minima in the cost function are obtained with values of the cost function that are of comparable magnitude. In the remainder of this section we will test the



(a) Point of view 1

(b) Point of view 2

Figure 6.3: Trajectories for the super-model (black) and the standard Lorenz 63 system (grey) from two different points of view.



(a) Model 1

(b) Model 2

(c) Model 3

Figure 6.4: Trajectories for the three connected imperfect models with connections determined by the learning process (black) and the standard Lorenz 63 system (grey).

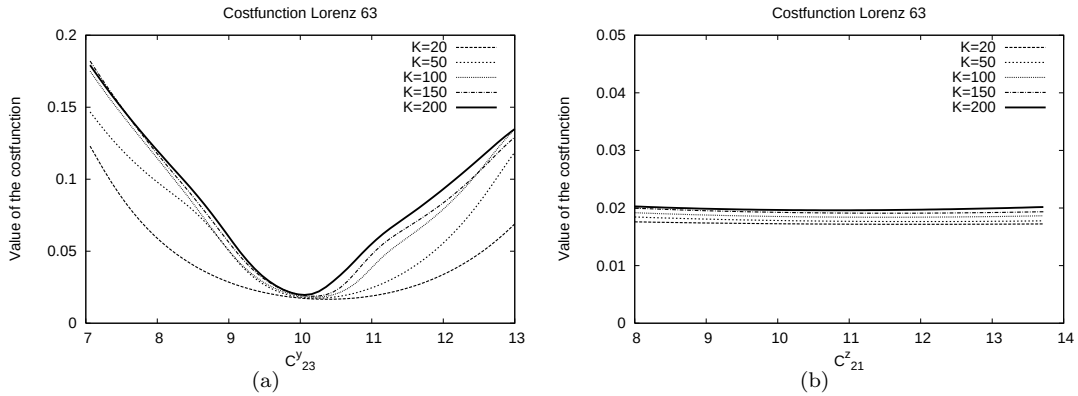


Figure 6.5: Cross section of the cost function for the super-model of the Lorenz 63 system calculated for different subsets of the original training set that was based on $K = 200$ initializations. The subsets vary in the number of initializations, i.e. $K = 20, 50, 100, 150$. A cross sections is created by changing connection coefficients C^y_{23} in (a) and C^z_{21} in (b) and keeping the other coefficients fixed at the values of the minimum found by the learning process using the training set.

robustness of the learning process, research the issue of local minima and develop measures to determine the quality of the different super-model solutions.

6.3.1 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 $\boldsymbol{C}$ 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 figure 6.3.1 for connection coefficients C^y_{23} and C^z_{21}, since these are typical examples.

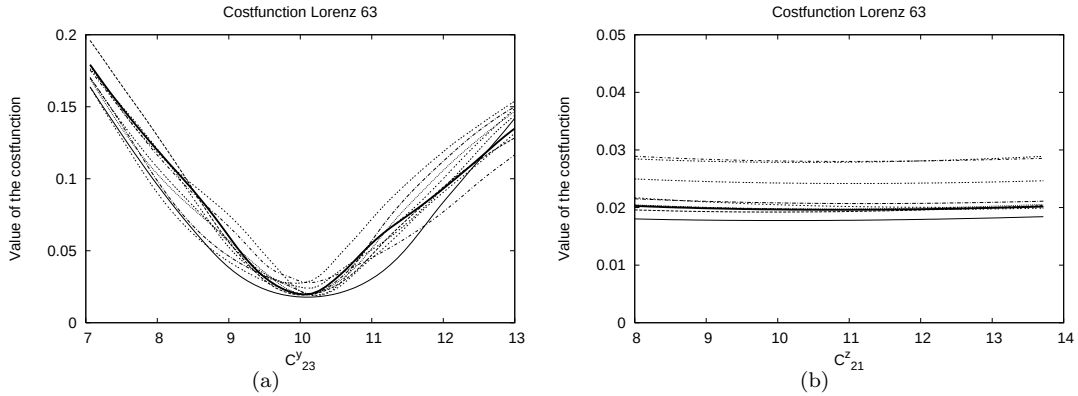


Figure 6.6: As in figure 6.3.1, except that the cost function is calculated for the training set with $K = 200$ initializations (thick line) and 9 additional independent sets of observations of the same length (thin lines) that were taken from a consecutive longer integration of the truth.

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

Table 6.2: The connection coefficients of two super-model solutions of the Lorenz 63 system and their differences.

In figure 6.5(a) the cost function for $K = 200$ displays one well defined minimum $C^y_{23} = 10.1$. 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 6.5(b) does not show a well defined minimum for any K. Note that the scale is smaller than in figure 6.5(a). The values of the cost function do not change much in the different subsets and the slopes are very flat. Changing connection coefficient C^z_{21} apparently does not change the quality of the solutions much. A family of super models of similar quality can be found by changing connection coefficient C^z_{21} 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 figure 6.6 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^y_{23} and C^z_{21} are shown (figure 6.6). In figure 6.6(a) the position and value of the minimum remain close to that of the training set. In figure 6.6(b) 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.

6.3.2 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 minimizationncd c. The minimization 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 6.2, 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.

	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 6.3: 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 5.000 time units. Between brackets the 95% error estimation is given.

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 figures 6.3 and 6.4. 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.

6.3.3 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 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 5.000 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 6.3 and for the truth and both super-models in table 6.4.

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 equation (6.1) 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. The statistics of the chaotic solution of model 3 (see table 6.3) differ substantially from the truth (see table 6.4), especially the mean value of z is much larger.

Both super-models (see table 6.4) 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 covariance of x and y.

	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)

Table 6.4: 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 5.000 time units. Between brackets the 95% error estimation is given.

6.3.4 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 figure 6.7 are based on 100 runs of 3.000 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 is lagging the truth somewhat, whereas super-model 2 is still in phase.

6.3.5 Synchronization with the truth

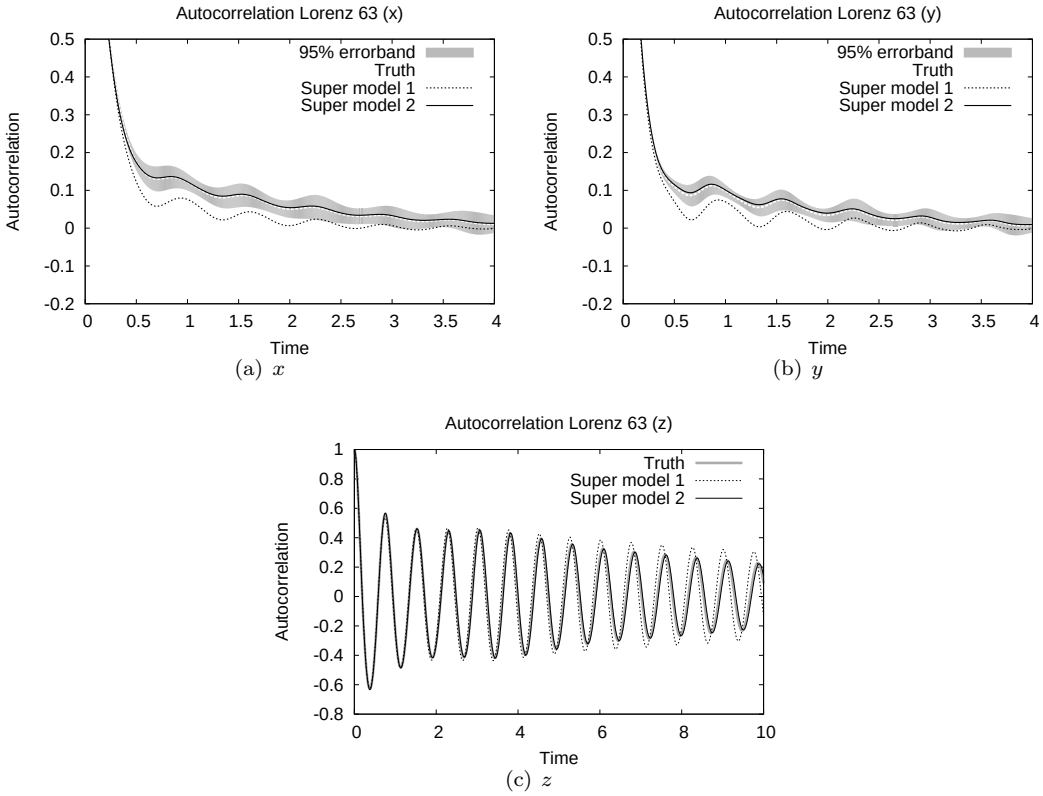
Pecora and Carroll [1990] have shown that limited information exchange between two identical Lorenz systems can lead to synchronization of the model states even when the systems are initialized from very different initial conditions and differ slightly in parameter values. The ability to synchronize 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 (equation (6.5). 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.

$$\begin{aligned} \dot{x}_k &= \sigma_k(y_k - x_k) + \sum_{j \neq k} C^x_{kj}(x_j - x_k) \\ \dot{y}_k &= x_k(\rho_k - z_k) - y_k + \sum_{j \neq k} C^y_{kj}(y_j - y_k) + n(y_o - y_k) \\ \dot{z}_k &= x_k y_k - \beta_k z_k + \sum_{j \neq k} C^z_{kj}(z_j - z_k) \; k = 1, 2, 3 \end{aligned} \tag{6.5}$$

We take the following definition of synchronization:

Definition 3 *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 δ and remains smaller than ϵ for $t \to \infty$.*



(a) x (b) y (c) z

Figure 6.7: 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 3.000 time units.

ϵ is chosen larger than δ, 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 ϵ distance of each other.

How quickly systems synchronize very much depends on the initial conditions [Yang, Baker, Li, Cordes, Huff, Nagpal, Okereke, Villafañe, Kalnay, and Duane, 2006], therefore we check synchronization for 100 restarts from different initial conditions. 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 $\epsilon = 4$.

To compare the two super-model solutions the same set of 100 initial conditions are used. The first super-model needs a nudging strength of $n = 11$ in order to synchronize with the truth. The second super-model needs a somewhat larger value $n = 13$. 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 temporal evolution, as measured by the autocorrelation, is more faithfully captured by the second super-model. However, if we reduce the time interval T during which the models must remain synchronized to 100 time units, we find that the second supermodel sometimes remains synchronized with a nudging strength of only $n = 4$, whereas the first super-model always needs at least $n = 5$ to remain synchronized during 100 time units. We conclude that the second super-model needs larger nudging to

	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%)

Table 6.5: Standard and perturbed parameters for the Rössler system.

remain synchronized over larger time periods due to the existence of particular trajectories of the truth for which the simulation errors of super-model 2 are larger than the simulation errors of super-model 1 and super-model 2 is needs larger nudging to remain synchronized during these trajectories than super-model 1. The probability of occurence of these trajectories in arbitrary time intervals of 100 time units is below 1, but it is 1 for 1000 time units.

Almost all measures indicate that the second super-model is closer to the truth than the first. It seems that the value of the cost function is indeed a good indication of quality of the solution and that the approach of minimizing the cost function is a fruitful strategy.

6.4 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.

6.4.1 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

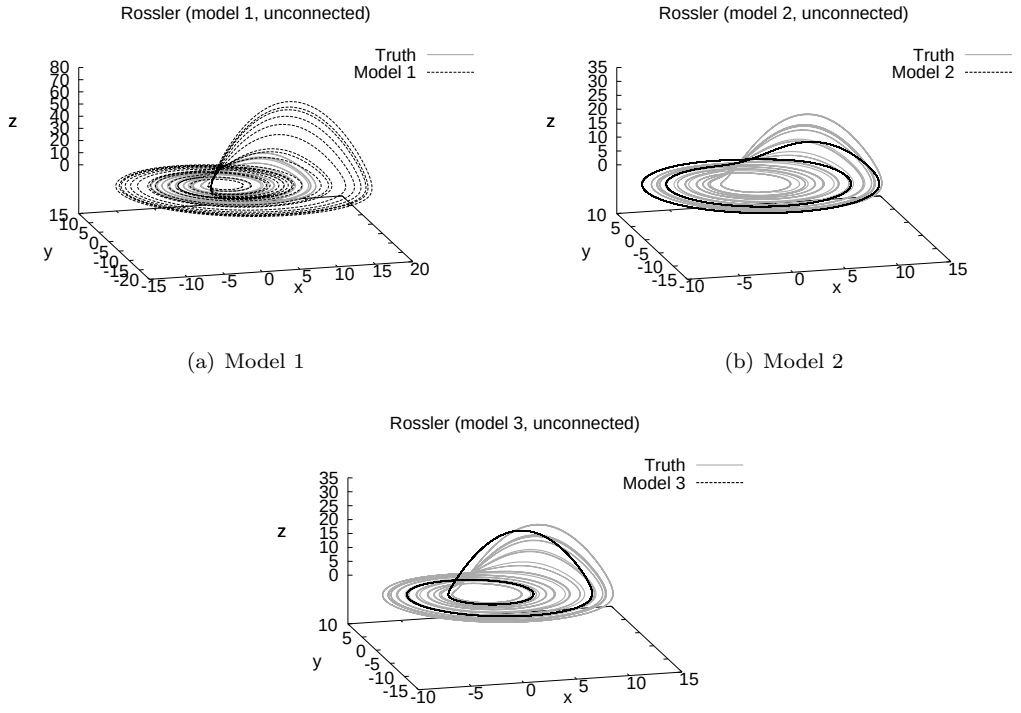
$$\begin{aligned} \dot{x} &= -(y+z) \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x-c). \end{aligned} \tag{6.6}$$

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 6.5. 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.

With these parameter perturbations marked changes occur in the attractor as can be seen in figure 6.8. 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 figure 6.9 in a cross section along C^x_{23}. The value at the minimum is approximately



(a) Model 1

(b) Model 2

(c) Model 3

Figure 6.8: 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.

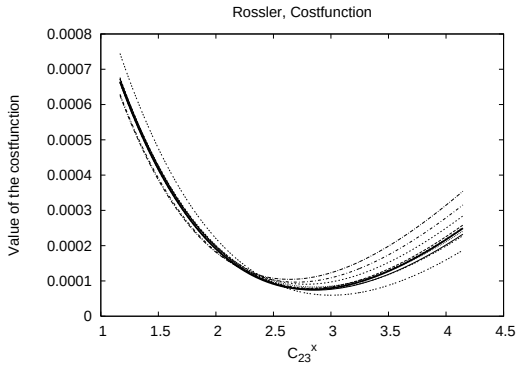
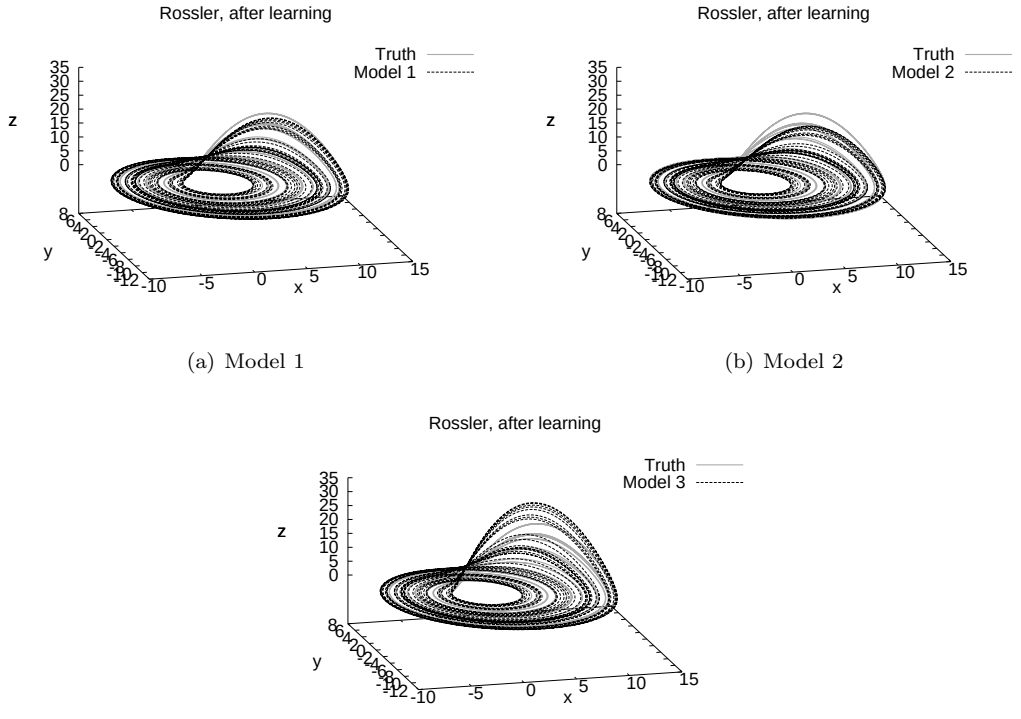


Figure 6.9: Cross sections of the cost function for the super-model of the Rössler system for the training set (thick line) with length corresponding to $K = 300$ and 9 additional independent sets of the same length (thin lines). Cross sections are created by changing the connection coefficient C^x_{23} and keeping the others fixed.



(a) Model 1

(b) Model 2

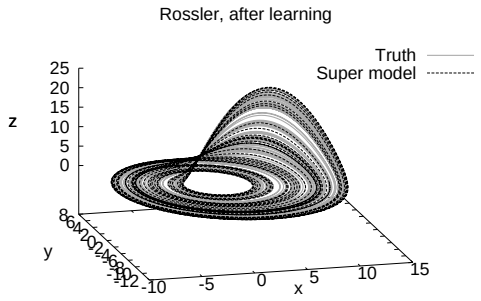
(c) Model 3

Figure 6.10: Trajectories for the three connected imperfect models with connections determined by the learning process (black) and the standard Rössler system (grey).

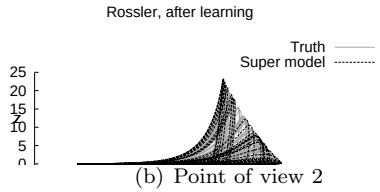
0.0001, which is much lower than a typical value of the cost function (0.004 for all connection coefficients equal to 1). To check the robustness of this minimum with respect to the limited size of training set, we calculated the cost function for 9 additional sets of 300 initializations, that were taken 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 minimization 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 figure 6.10. 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 super-model solution, which is defined as the average of the three imperfect models, is plotted in figure 6.11 for two points of view. Visually the attractor of the super model 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 6.6 and for the super-model and the truth in table 6.7. 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.



(a) Point of view 1



(b) Point of view 2

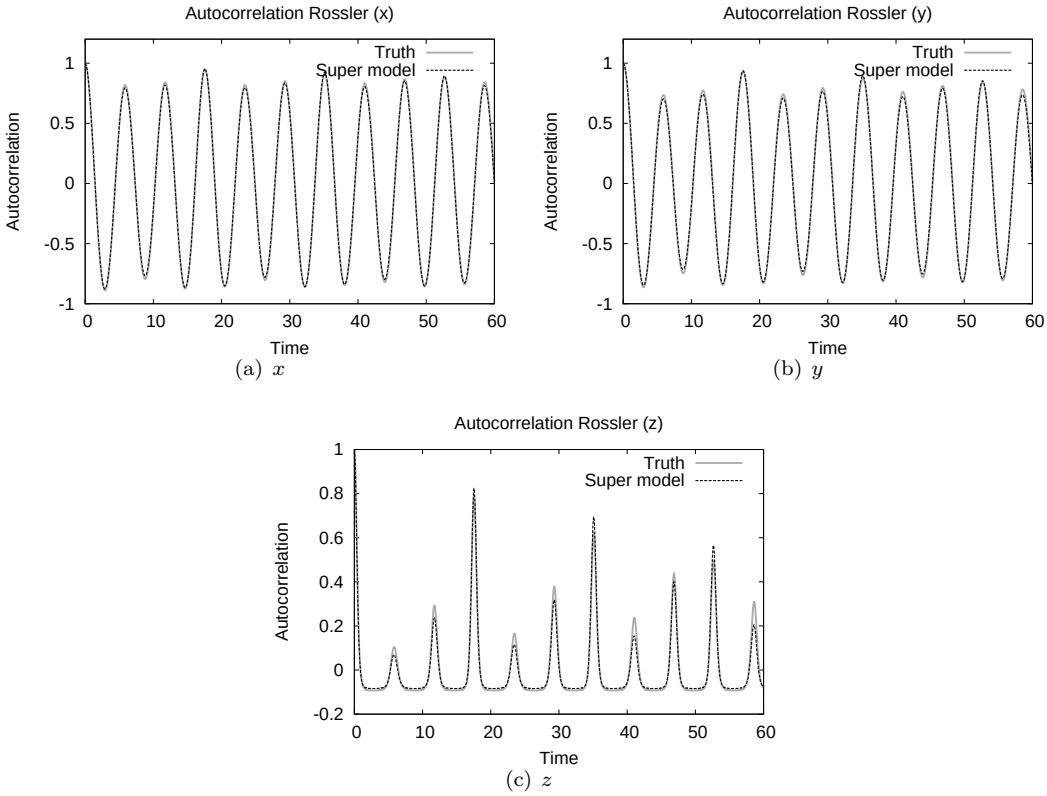
Figure 6.11: Trajectories for the super-model (black) and the standard Rössler system (grey) for two different points of view.

	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)

Table 6.6: 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 5.000 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)

Table 6.7: 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 5.000 time units is given between brackets.



(a) x (b) y (c) z

Figure 6.12: Autocorrelation for the super-model (black) and the standard Rössler system (white). The shaded area corresponds to the 95% error band for the truth, based on 100 runs of 3.000 time units.

To compare the temporal behaviour we calculated the autocorrelation functions as plotted in figure 6.12 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.

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 $T = 100$ 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 figure 6.12) 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 more chaos helps or hurts, so we test the super-modeling approach also on the more chaotic 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%)

Table 6.8: Standard and perturbed parameters for the Lorenz 84 system.

6.4.2 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

$$\begin{aligned} \dot{x} &= -y^2 - z^2 - ax + aF \\ \dot{y} &= xy - bxz - y + G \\ \dot{z} &= bxy + xz - z. \end{aligned} \tag{6.7}$$

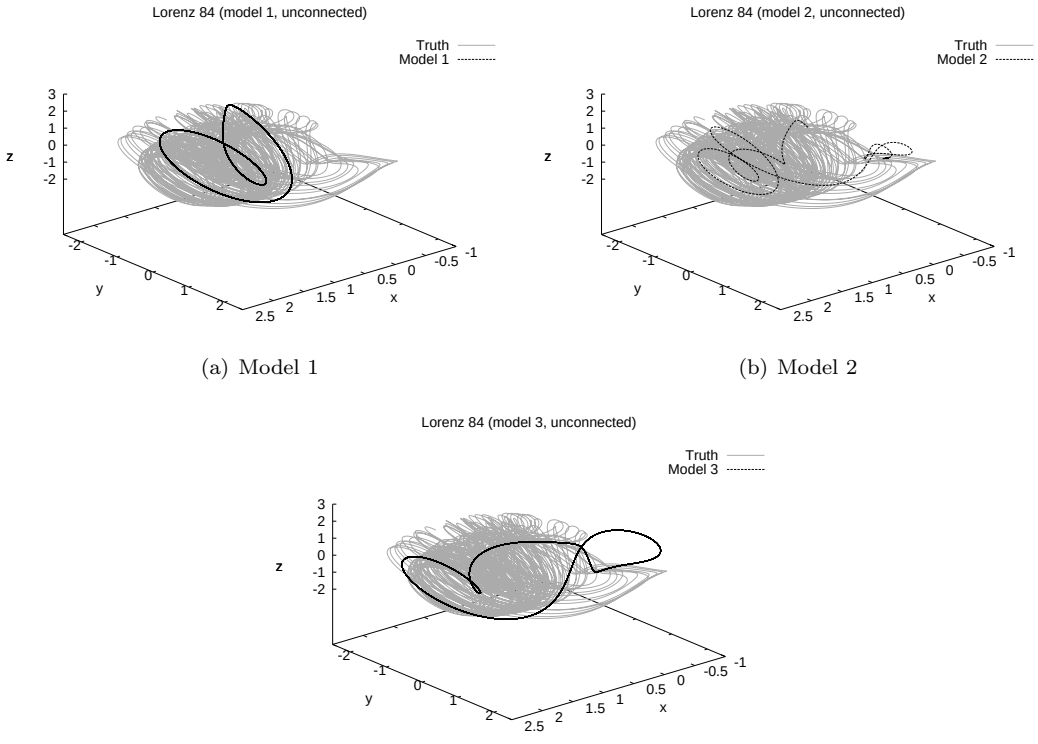
The x variable represents the intensity of the globe-encircling westerly winds and y and z represent a travelling large-scale wave that interacts with the westerly wind. Parameters F and G are forcing terms representing the average north-south temperature contrast and the east-west asymmetries due to 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, 2002]. In table 6.8 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 differ substantially from the truth (see figure 6.13). 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 minimization algorithm did not produce a well defined minimum of the cost function. The high value of the autocorrelation function of x (0.6 at 8 time units, see figure 6.16) indicates that the initial conditions still have an impact on the evolution after 8 time units. Therefore we decided to increase Δ to 8 and γ to 0.8. In addition it turned out that it was easier to find good minima using the amoeba minimization algorithm [Nelder and Mead, 1965] instead of the conjugate gradients minimization. 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 figure 6.14. 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) Model 1

(b) Model 2

(c) Model 3

Figure 6.13: Trajectories for the three unconnected imperfect models (black) and the standard Lorenz 84 system (grey).

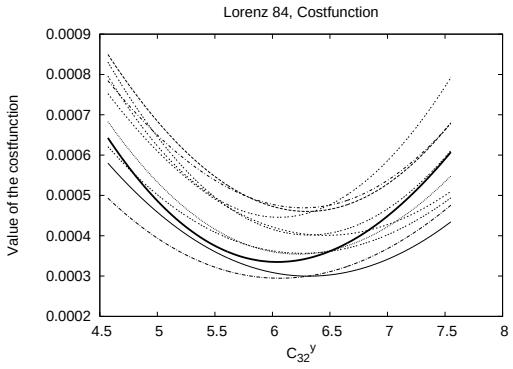
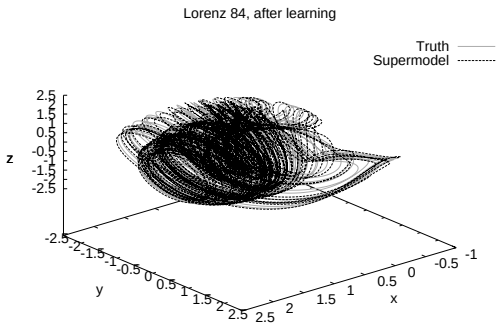
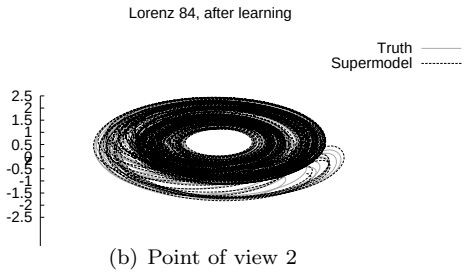


Figure 6.14: Cross section of the cost function for the super-model of the Lorenz 84 system for the training set (thick line) with length corresponding to $K = 200$ and 9 additional independent sets of the same length (thin lines). Cross sections are created by changing the connection coefficient C_{32}^{y} and keeping the others fixed.



(a) Point of view 1



(b) Point of view 2

Figure 6.15: Trajectories for the super-model (black) and the standard Lorenz 84 system(grey) for two points of view.

	Truth	Minimum
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)

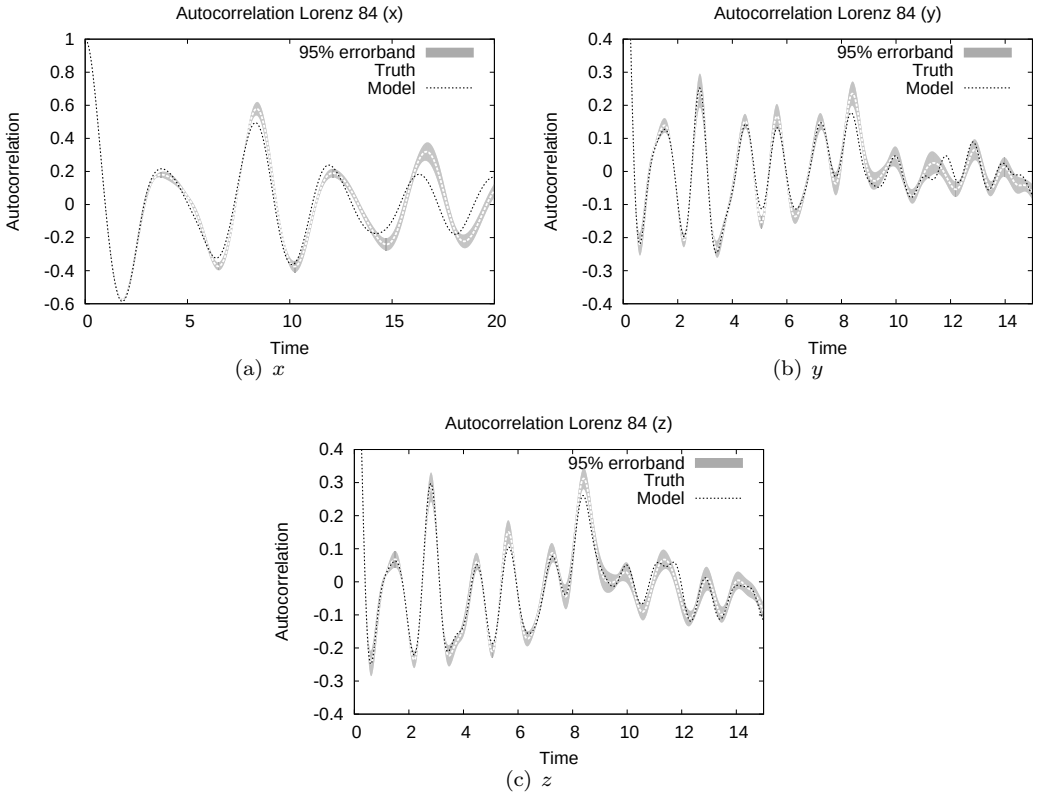
Table 6.9: Mean, standard deviation (SD) and covariance for the super-model and the standard Lorenz 84 system. The 95% error estimation based on 500 runs of 5.000 time units is given between brackets.

With the connection coefficients for this minimum, we integrated the super-model and plotted the trajectory in figure 6.15. 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 reason for this might be found in the high value of several connection coefficients (for instance $C^x_{32} = 115$, $C^y_{23} = 147$ and $C^z_{31} = 169$). 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 6.9. These statistics are in excellent agreement.

In order to evaluate the temporal behaviour we compare the autocorrelation functions in figure 6.16. Up to a delay time of 14 time units the autocorrelation functions of the truth are 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 $T = 100$ in definition 3. 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.



(a) x (b) y (c) z

Figure 6.16: 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 3.000 time units.

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 minimization process was found to be more sensitive to the length of the short integrations Δ and the discount parameter γ, requiring the use of the more robust amoeba minimization procedure.

6.5 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 past 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 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.

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 interactive ensemble 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. One possible approach would be to not connect the state variables, but the various parametrized physical processes that contribute to the tendency of the state variables. Most of the model uncertainty resides in these parametrized processes, so it makes sense to direct the learning to these processes.

A hierarchy of earth system models of intermediate complexity (EMICs) 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 time step and an ensemble of imperfect models is created by perturbing parameters and/or using different formulations for unresolved processes.

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 indications that fast atmospheric processes are the primary cause of model systematic errors [Rodwell and Jung, 2008].

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 equation (6.5) and additional evolution equations are formulated for the parameters. If the super-model synchronizes with the truth the parameters stop updating. This alternative learning strategy leads to a particularly simple learning law that would 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].

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.

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 therefore not guaranteed that the super-model will also simulate this response more realistically. The problem is not peculiar to the super-modeling approach, but arises with climate models generally, since they are ẗunedön historical data and are used to predict the climate response to a change in greenhouse gas concentrations. In the super-modeling approach the issue can be addressed in the perfect model setting using climate models of various complexity.

Chapter 7

Variations in the method

The final method described in the previous chapter turned out to work quite well. Earlier variations of the method that we will discuss in this chapter had some problems.

7.1 Procedure

Two different variations will be compared to the final approach. The first variation is changing the connection coefficients to the exponents of the connection coefficients as in section 3.2. As the connections will stay positive due to the exponents, the extra term in the cost function (section 3.4) is omitted. The second variation is different from the final approach in that during the minimization the cost function is not minimized for an increasing number of initial conditions, but for all initial conditions at once, which is basically the last step of the final approach.

To test the methods 100 minimizations of the cost function are initialized from random sets of initial connection coefficients in the range of 4 to 10. For a fair comparison the same set of initial connection coefficients is used for all experiments. To take into account the exponents in the first variation, the natural logaritm of the initial connection coefficients is taken as initial connection coefficients for this variation. For each of these initial conditions the program will try to find a super-model. If the solution gets too large (larger than 10^{25}) the program breaks the loop and counts a floating point exception, which is an error that arises when the numbers get too large (tend to infinity). The value 10^{25} is much larger than a solution for the Lorenz 63 system, as the x and y variables vary typically between -25 and 25 and the z variable between 0 and 45. This means that a floating point exception cannot be avoided if the solution gets this large.

Super-model solutions will be easier to find if model imperfections are small. Therefore we consider different magnitudes of the parameter perturbations in the imperfect models, as detailed in the next section.

7.2 Different perturbations

The three sets of perturbed parameters are given in table 7.1. The first set of parameters have perturbations between 1.5 and 2 %, the second set has perturbations of 9 to 10 % and the third set up to 20%. The individual imperfect models can be found in figure 7.1. The figure shows that for the small perturbations of up to 2%, the attractor is still very similar to that of the truth, but for perturbations of 10% the differences are clearer and for a 20% perturbation the solutions are even less comparable to that of the truth. This supports the idea that for larger perturbations of the parameter values, it will

	σ	ρ	β
Truth	10	28	$\frac{8}{3}$
Model 1	10.2 (2%)	27.44 (2%)	2.71 (1.5%)
Model 2	9.8 (2%)	28.42 (1.5%)	2.83 (2%)
Model 3	10.15 (1.5%)	28.56 (2%)	2.61 (2%)
Model 1	11 (10%)	25.2 (10%)	2.91 (9%)
Model 2	9 (10%)	30.52 (9%)	2.93 (10%)
Model 3	10.9 (9%)	30.8 (10%)	2.4 (10%)
Model 1	12 (20%)	22.4 (20%)	3.17 (19%)
Model 2	8 (20%)	33.32 (19%)	3.2 (20%)
Model 3	11.9 (19%)	33.6 (20%)	2.13 (20%)

Table 7.1: Three sets of perturbed parameters for the three imperfect models for the Lorenz 63 system. The sets of perturbations are about 2%, 10% and 20%.

	Final method	Variation 1	Variation 2
1 to 2%	100	89	99
9 to 10%	100	92	98
19 to 20%	100	55	85

Table 7.2: The number of runs that give a solution for the three different minimization methods for three different sets of imperfect models shown in table 7.1. A total number of 100 runs was calculated.

be harder to find a good set of connection coefficient.

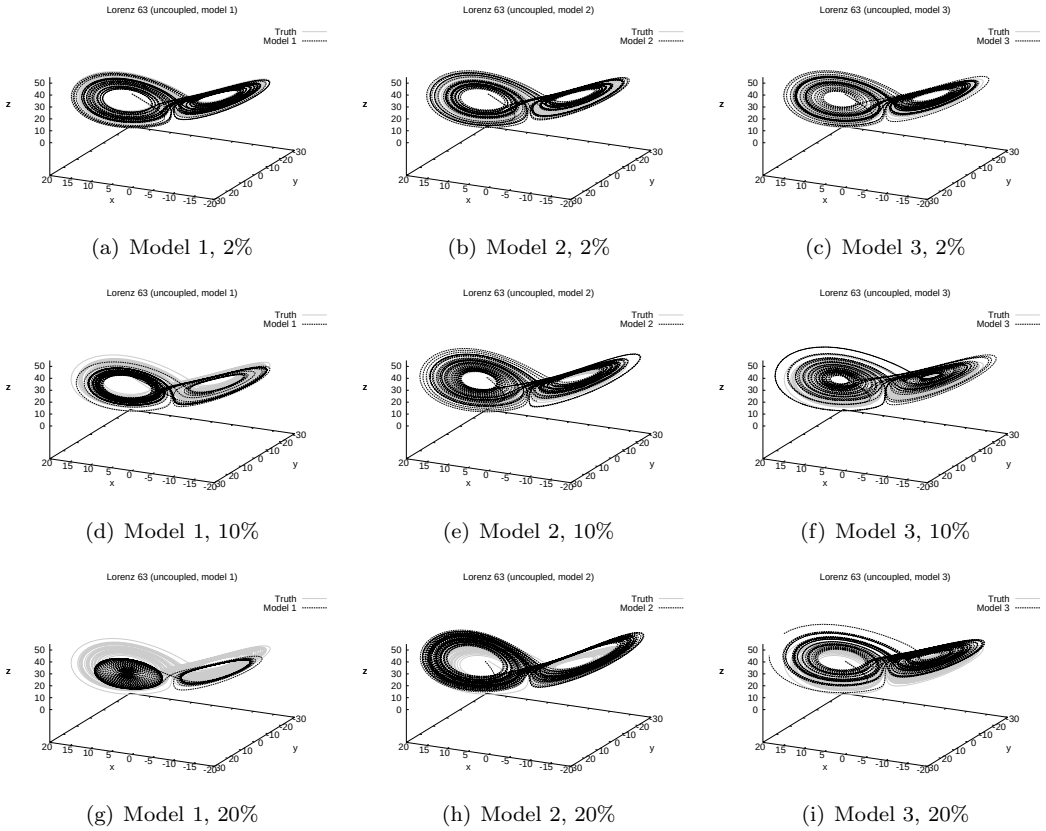
In table 7.2 the number of times that a minimum is found in the cost function is shown for all three minimization methods and for all three sets of perturbed parameter values. All solutions found for each of the three methods and each of the parameter sets were good solutions, judged on the basis of looking at the attractor. For both variations the number of found minima decreases with increasing size of perturbation of the parameter values, which is not the case for the final method. This suggests that the final method is the best method of the three. Variation 1 that makes use of the exponents, encounters most floating point exceptions of all methods, especially for the larger perturbations.

We also checked whether the minima that were found were unique. This was done by checking whether the difference between all connection coefficients was less than a certain tolerance. The results for different tolerances can be found in table 7.3, which contains the percentage of unique super-models.

The final method and variation 2 were able to find all different minima, whereas variation 1 found a smaller range of minima. This may indicate an advantage of using variation 1, since apparently the method finds the same minimum for different initial conditions. The chance that this is a more global minimum is larger than for the other two methods.

	Final method			Variation 1			Variation 2		
Tolerance	0.1	0.25	0.5	0.1	0.25	0.5	0.1	0.25	0.5
1 to 2%	100%	100%	100%	98%	46%	12%	100%	100%	84%
9 to 10%	100%	100%	100%	95%	55%	18%	100%	100%	100%
19 to 20%	100%	100%	100%	100%	76%	25%	100%	100%	100%

Table 7.3: The percentage of the runs that resulted in a unique super-model. Unique means in this context that at least one of the connection coefficients has a distance of more than the tolerance to any of the other minima. The results are shown for different tolerances, namely 0.1, 0.25 and 0.5.



(a) Model 1, 2% (b) Model 2, 2% (c) Model 3, 2%

(d) Model 1, 10% (e) Model 2, 10% (f) Model 3, 10%

(g) Model 1, 20% (h) Model 2, 20% (i) Model 3, 20%

Figure 7.1: Solution for the three perturbed unconnected models (black) and the truth (grey) for the Lorenz 63 system for different imperfect models from table 7.1.

Chapter 8

Results

Some results are already shown in chapter 6, but in this section we include some extra results. First we will study the influence of the choice of parameters on the learning process. The second topic deals with a model for which we expect certain values for the connection coefficients to work well. Do these values indeed work well and are they the values found by the learning process? In the last topic we will study the connection of three models that are similarly wrong. Can these models be connected such that the solution is still a good approximation of the truth?

8.1 The influence of parameters

Chapter 4 was devoted on how we should choose the parameters γ, Δ and the initial conditions. Do these parameters have as much influence on the learning process as we might think, or does the choice of parameters have little effect on the learning process? To answer these questions we varied the different parameters and tried to find an effect on the learning process (mainly on the cost function) and on the super-models found for each of the three small chaotic dynamical systems.

8.1.1 Varying γ and Δ

γ and Δ are used to limit the influence of chaos. In this section we will study the influence of γ and Δ on the learning process and thus on the found super-models. The values for γ and Δ are given in table 8.1 for the Lorenz 63 system. We roughly tested 5 situations to see what the effect on the learning process is. All these situations were tested for the same set of 10 different initial connection coefficients and in all situations 10 unique minima were found. In all these situations the cost function is evaluated on the same set of $K = 150$ initial conditions.

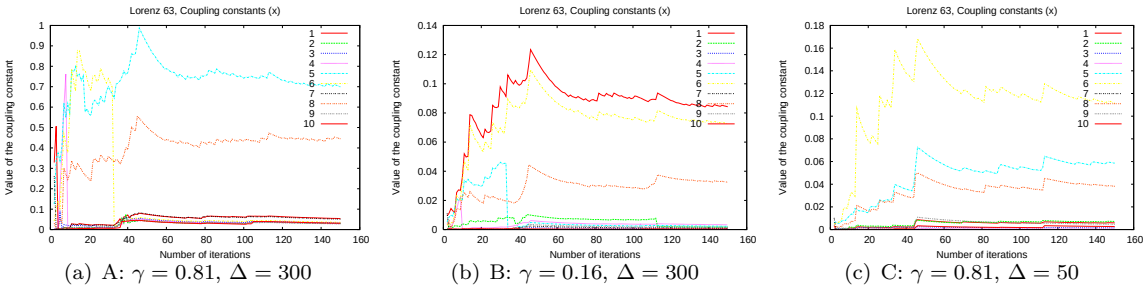
A γ and Δ large

B γ small and Δ large

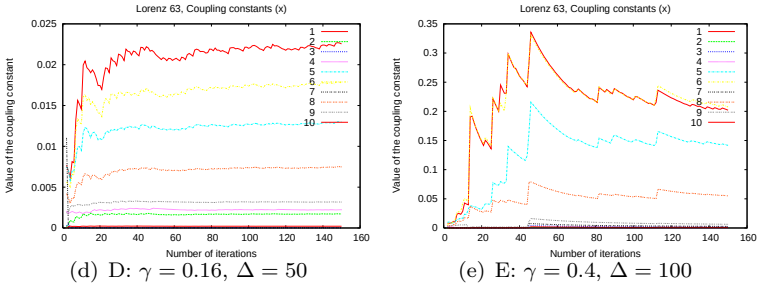
C γ large and Δ small

D γ and Δ small

E the standard γ and Δ



(a) A: $\gamma = 0.81$, $\Delta = 300$ (b) B: $\gamma = 0.16$, $\Delta = 300$ (c) C: $\gamma = 0.81$, $\Delta = 50$



(d) D: $\gamma = 0.16$, $\Delta = 50$ (e) E: $\gamma = 0.4$, $\Delta = 100$

Figure 8.1: The cost function for all 5 combinations of γ and Δ for 10 different sets of initial connection coefficients during the learning process.

Situation	γ	Δ
A	0.81	300
B	0.16	300
C	0.81	50
D	0.16	50
E	0.4	100

Table 8.1: 5 different combinations, A to E, of γ and Δ used in the Lorenz 63 system.

In figure 8.1 the value of the found minimum of the cost function is plotted during the learning process for all 5 situations. The figures for all situations do not differ much. The shapes are similar, since the same sets of initial connection coeficients are used for each of the situations. Most of the lines stay low, and seem to converge to a certain value almost immediately, while only 2 or 3 have a higher value and show more variations, though convergence is found in these cases as well. Notice that for situation D the values for the cost function are much lower than for the other minima. As we haven't placed a normalisation term for γ, this could cause these small values, but comparing to situation B, where the value of γ is just as small and Δ is even larger, the cost function is much smaller than the higher lines shown, indicating that these higher lines must represent local minima that may not be very good.

Some of the minima were found with different values for γ and Δ just by starting from the same initial connection coefficients. This only happened when the connection coefficients barely changed from the initial conditions. For initial connection coefficients that did lead to a large change in the learning process, the different values of γ and Δ gave different minima. This shows that using different parameters does have an effect on the learning process.

Looking at the solutions of the resulting super-models (not shown), we found that the differences are quite small, though in some cases the solutions are visually of less quality. For the smaller value of $\gamma = 0.16$ in combination with $\Delta = 50$ for instance, we found for example a solution with a point attractor (figure 8.2). Also one or two other solutions for $\gamma = 0.16$ had a lower quality, suggesting that taking

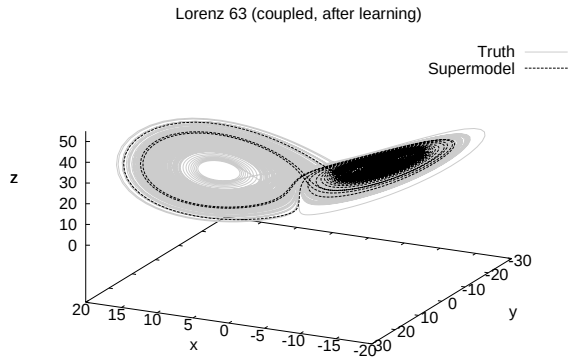


Figure 8.2: The solution for one of the super-models found with $\gamma = 0.16$ and $\Delta = 50$, which converges to a point.

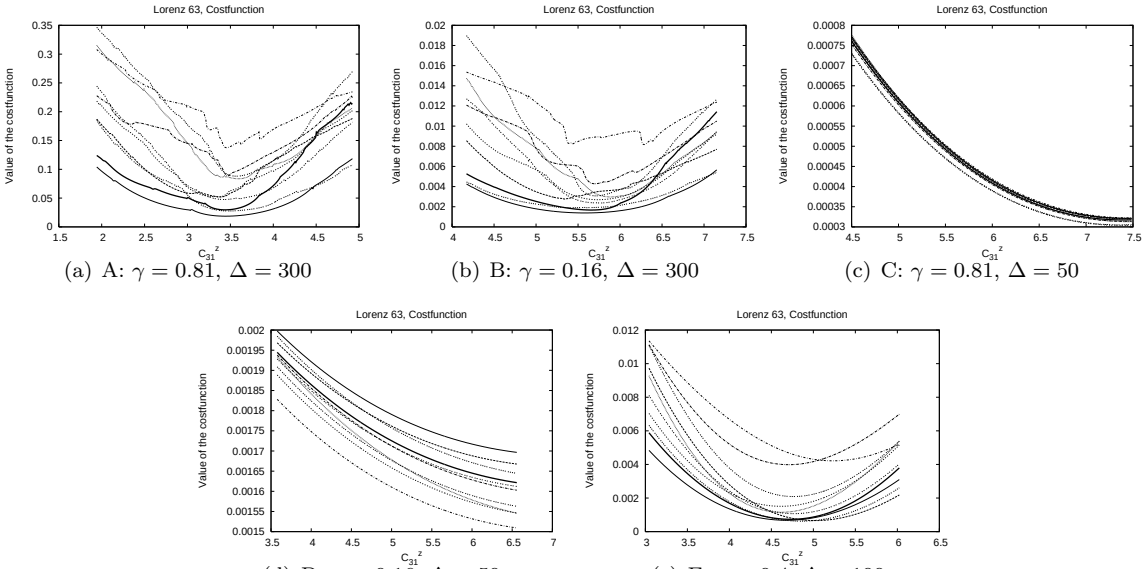
a small γ and a small Δ does not give enough information.

As we can find different super-models by just changing parameters, the cost function should look different for different combinations of γ and Δ. In figure 8.3 a cross section of the cost function is shown for one of the minima found from the same initial connection coefficients for each of the 5 situations. The figure is only for a cross section through C^z_{31}, but all other cross sections look similar. Notice that the thick black line representing the cost function for the training set is smooth in most situations. For $\Delta = 50$ and for the standard parameter values this smoothness can also be found in the cost function for different sets of initial conditions, but this is not the case for $\Delta = 300$, where there are many local minima and the cost function for different sets of initial conditions differ substantially. We expect that this is primarily caused by chaos being more dominant in the cost function. The figure for situation C and D also show that the learning process was not able to find a true minimum, as it stopped on the way to it. Note that for situation A, B and E the minimum is found. The reason that the minimum is not found in situation C and D may be because the slope of the cost function is so flat.

This experiment indicates that the choice of γ and Δ is not very important for the learning process for the Lorenz 63 system, though it seems that Δ should not be chosen too large, as this causes the number of local minima to grow and that the normalisation term should be altered in case of low values of Δ. For the Rössler system choosing γ and Δ is even less important than for the Lorenz 63 system. This can be expected as the attractor of the Rössler system is much simpler and less chaotic than that of the Lorenz 63 system. It turns out that for the Lorenz 84 system, the choice of γ and Δ is more important, which is probably due to the fact that the attractor has a more complicated shape.

The problem with the Lorenz 84 system is that the connection coefficients for many combinations of γ and Δ do not change during the learning process, when we take the conjugate gradient method described in section 3.5.1 and 3.5.2. Therefore the amoeba method was used (section 3.5.3). The cost function, which was very small, was not the cause of this, as it was in the case of Lorenz 63. It seems that for the Lorenz 84 system something else is going on. Still it is striking that solutions for some of the super-models for which no actual minimum is found behave so well, like the super-model solution shown in figure 8.4, for which also the statistical measures show good results.

What seems to happen in the Lorenz 84 system is that the learning process is not able to change the connection coefficients for some combinations of γ and Δ. For these values the learning process probably gets too little information, which can be explained by the more complicated shape of the attractor. Maybe the shape of this attractor should be taken into account when choosing γ and more importantly



(a) A: $\gamma = 0.81$, $\Delta = 300$ (b) B: $\gamma = 0.16$, $\Delta = 300$ (c) C: $\gamma = 0.81$, $\Delta = 50$

(d) D: $\gamma = 0.16$, $\Delta = 50$ (e) E: $\gamma = 0.4$, $\Delta = 100$

Figure 8.3: The cost function for connection constant C_{31}^z for all 5 combinations of γ and Δ for 10 different sets of initial conditions. The darkest line represents the cost function of the training set, the thinner lines represent different sets of initial conditions.

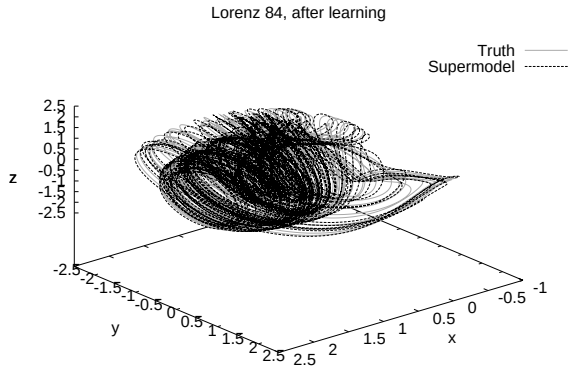
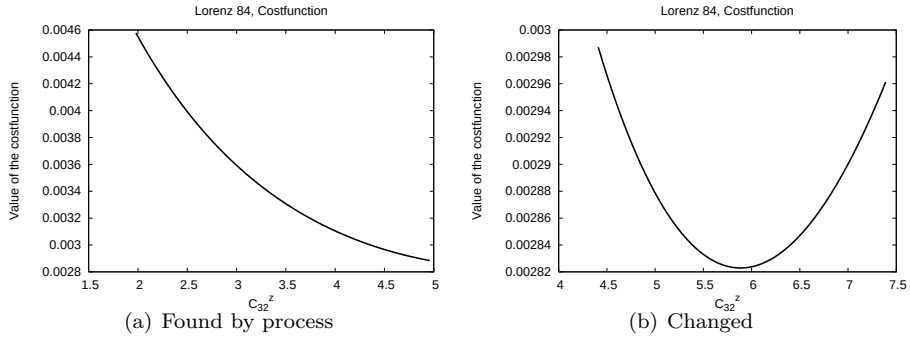


Figure 8.4: Solution for a super-model of the Lorenz 84 system, for which the cost function does not have a minimum.



(a) Found by process (b) Changed

Figure 8.5: The cost function for connection constant C_{32}^z. Left for the value found by the learning process and right the changed value, such that the cost function has an actual minimum for this connection constant.

Δ. The Lorenz 63 system needs approximately 0.7 to 0.8 time units to circle around one of the wings. For the Rössler system one circle is approximately 5 time units. These values are comparable to the values chosen for Δ and it might therefore help to choose Δ comparable in size to one circle on the Lorenz 84 system as well, which is about 5 time units. Therefore the value for Δ will be chosen to be 8 time units, instead of the 2 time units, which was apparently too small. Also γ is taken larger to get more information into the cost function ($\gamma = 0.8$). The autocorrelation function also indicates that there is still information of the initial condition left at $\tau = 8$ time units.

For these values the Lorenz 84 system was able to find a good minimum and therefore it seems that choosing γ and Δ just based on chaos might not be the best strategy in all situations. The properties of the attractor need to be taken into account too. Overall it seems that small values of the cost function make it more difficult to find minima for the conjugate gradients method, but in some cases this can be solved by changing the minimization term. In other cases the amoeba method might be a solution, which was able to find a minimum for the Lorenz 84 system.

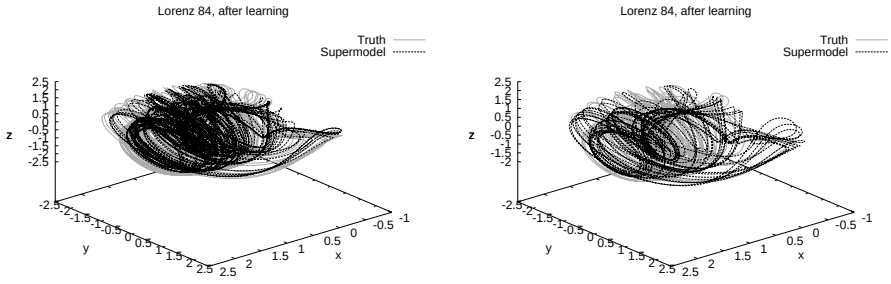
8.1.2 Manually changing connection coefficients

For the Lorenz 84 system it seeemed clear that in many cases no minimum of the cost function was found. Therefore we decided to check what happens to the solution if we manually change one of the connection coefficients towards a minimum. The results for this on the cost function are shown in figure 8.5, where the minimum is situated in the middle of the figure. On the left the situation is shown for the 'minimum' found by the process for $C_{32}^z = 3.47$. This sort of figures were found for most connection coefficients. On the right we have manually changed C_{32}^z to 5.9.

In figure 8.6 the solutions for the two super-models are shown. By eye it seems that the changed super-model is not as good as the original, but it still is a rather good approximation of the truth. By using the different measures we will find out whether the solutions are equally good or whether one of them is better.

Mean, standard deviation and covariance

The mean standard deviation and covariance can be found in table 8.1.2 for the originally found super-model and the changed super-model. The values for the original super-model are closer to the truth for all variables, which we might have expected from the shape of the attractor. Still this seems to be an odd result as the value of the cost function is lower for the changed super-model. Notice however that



(a) Found by process (b) Changed

Figure 8.6: The solutions for the super-model found by the learning process and the super-model that was created by changing the connection constant C^z_{32}, such that the cost function has a minimum for it as shown in figure 8.5.

	Truth	Found by process	Changed
Mean x	1.015 (0.008)	0.959 (0.012)	0.912 (0.015)
Mean y	0.063 (0.018)	0.058 (0.024)	0.153 (0.032)
Mean z	0.271 (0.005)	0.231 (0.007)	0.180 (0.014)
SD x	0.589 (0.014)	0.569 (0.021)	0.647 (0.022)
SD y	0.919 (0.002)	0.903 (0.002)	0.916 (0.004)
SD z	0.908 (0.002)	0.855 (0.002)	0.865 (0.004)
Cov xy	-0.053 (0.018)	-0.058 (0.022)	-0.142 (0.027)
Cov xz	-0.038 (0.003)	-0.009 (0.004)	0.028 (0.009)
Cov yz	-0.075 (0.006)	-0.049 (0.005)	-0.074 (0.007)

Table 8.2: Mean, standard deviation (SD) and covariance (Cov) for the originally found set of connection coefficients (Found by process) and the set with one of the connection coefficients changed (Changed) for the Lorenz 84 system. Between brackets the 95% error estimation is indicated.

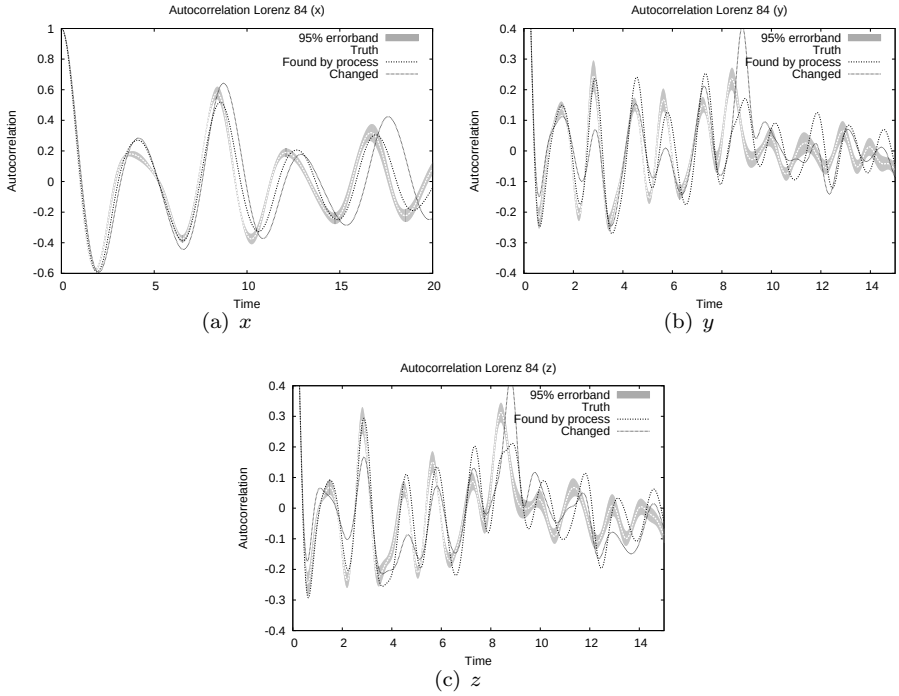
the values of the changed super-model are close to the truth as well, so it is still a good approximation, but it may not be better than the original super-model.

Autocorrelation function

The autocorrelation for both the changed and original super-models are plotted in figure 8.7. It seems that the autocorrelation is a good approximation for both sets as well, but again the original super-model is closer for all variables. The changed minimum is still quite good, but it is clearly not as good as the original.

Synchronization

For synchronization with the truth we found for the default values of the tolerances ($tol = 0.1$ and $tol_2 = 0.5$) that both super-models were not able to synchronize with the truth for nudgings under 30. Even increasing tol_2 up to 1 (which is already quite large), did not help the synchronization. Therefor this measure could not be used for comparison in this case and we can only use the other measures to compare the solutions. Still as both models were unable to synchronize the quality of the two super-models is similar in this measure as well.



(a) x (b) y

(c) z

Figure 8.7: The autocorrelation functions of the two super-models and the truth. The grey shaded area indicates the 95% error band.

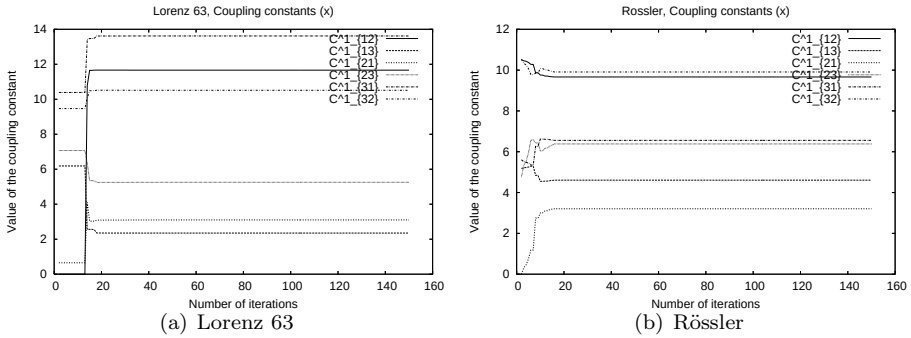
From the measures it seems that changing the connection constant to a minimum is not necessarily better, but the differences in quality are small. This suggests that changing the connection coefficients to a minimum will not very much improve the quality of the super-model as apparently all super-models in that range are reasonably good. This can also be explained by the flat slopes of the minima, suggesting that any value for the connection constant gives a low value for the cost function.

8.1.3 The set of initial conditions

Another important parameter is the training set. The number of initial conditions in the training set is large enough as it seems that adding more initial conditions for both Lorenz 63 and Rössler are not changing if we add more initial conditions (figure 8.8). This cannot be checked in this way for the Lorenz 84 system as it is not obtained by adding initial conditions, but plots of the cost function for different sets of initial conditions (figure 8.9) show that the dependence on the training set is small. This implies that also for the Lorenz 84 system enough initial conditions are used.

Changing the number of initial conditions will have little effect on the solutions, but can taking different sets of initial conditions change the learning process? The initial conditions are now nicely spread over the attractor (figure 4.3), but judging by the small amount of initial conditions needed to actually change the connection coefficients, the spread of the initial conditions may have little influence as well. To test whether this is indeed the case we took new sets of initial conditions for the Lorenz 63, Rössler and Lorenz 84 systems, such that the initial conditions are accumulated in a small part of the attractor.

The new spread is shown in figure 8.10. For the Lorenz 63 system all initial condtiions are positioned on the tip of the right wing. For the Rössler system the initial conditions are positioned on a line that does not cross the top of the attractor and for the Lorenz 84 attractor we have chosen all initial conditions in



(a) Lorenz 63 (b) Rössler

Figure 8.8: The values of the connection coefficients for the Lorenz 63 and Rössler systems for the standard values for γ, Δ and K during the learning process.

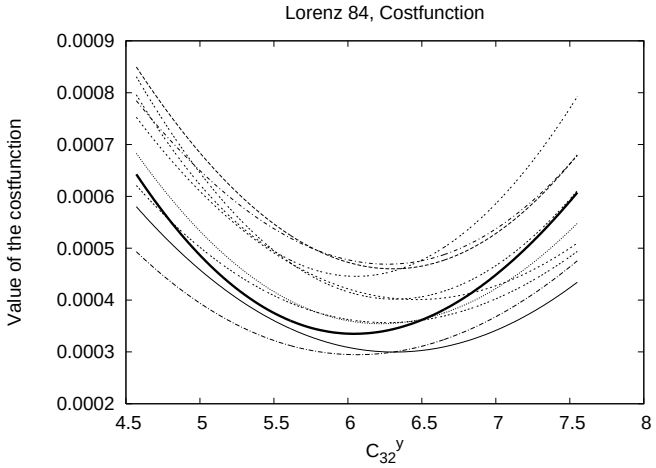
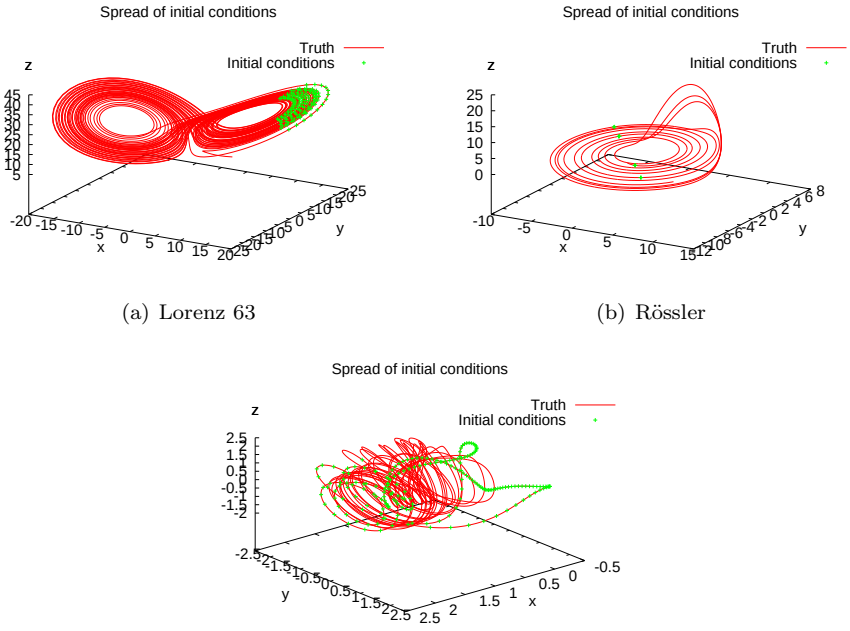


Figure 8.9: A cross section of the cost function for the Lorenz 84 system through connection constant C^x_{12} for the standard values for γ, Δ and K. The thick line is the training set and the thin lines represent different independent sets of initial conditions.



(a) Lorenz 63

(b) Rössler

(c) Lorenz 84

Figure 8.10: The new spread of the initial conditions for the Lorenz 63, Rössler and Lorenz 84 attractors.

the inner circles of the attractor.

It turns out that taking a different spread of initial conditions indeed has a low influence on the systems. Especially the Lorenz 63 and 84 systems have by visually inspecting the attractors equally good solutions. The learning process is also able to find these for the new spread. Also for the Rössler system it seems that the difference in the spread does not even help that much as for equal initial connection coefficients the minima are really close together. Apparently the dependence on the spread of initial conditions is not large for the small dynamical systems.

We did find that in some cases low values of the cost function were found for periodic orbits. These orbits run over the whole attractor and therefore give a low value of the cost function. This can happen even for an evenly spread set of initial conditions and is therefore hard to recognize. It can however be seen in the autocorrelation functions that result in a periodic function as well. It seems however difficult to use the spread to avoid these periodic orbits.

8.2 Expected outcomes

The current way of combining models is by just taking combinations of the outcomes of imperfect models, determined by how good the resulting solution is, but also by taking a higher weight for the models that we think work well. In this section we want to find out whether it is possible to do a similar thing with the connection coefficients. A combination of models is introduced that have one parameter in common with the truth. For this combination an expected value is introduced.

	σ	ρ	β
Truth	10	28	$\frac{8}{3}$
Model 1	10	25.2 (-10%)	2.93 (+10%)
Model 2	11 (+10%)	28	2.4 (-10%)
Model 3	9 (-10%)	30.8 (+10%)	$\frac{8}{3}$
Model 1	10	22.4 (-20%)	3.2 (+20%)
Model 2	12 (+20%)	28	2.13 (-20%)
Model 3	8 (-20%)	33.6 (+20%)	$\frac{8}{3}$
Model 1	10	19.6 (-30%)	3.47 (+30%)
Model 2	13 (+30%)	28	1.87 (-30%)
Model 3	7 (-30%)	36.4 (+30%)	$\frac{8}{3}$

Table 8.3: Three sets of perturbed parameters for the three imperfect models for the Lorenz 63 system. All imperfect models have exactly one parameter in common with the truth. The other parameters are perturbed for respectively 10, 20 and 30 %.

	x	y	z
C_{12}	0	large	?
C_{13}	0	?	large
C_{21}	large	0	?
C_{23}	?	0	large
C_{31}	large	?	0
C_{32}	?	large	0

Table 8.4: The expected size of the connection coefficients belonging to the combination of models in table 8.3.

8.2.1 Random initial conditions

We tested with the combination of models given in table 8.3 for the Lorenz 63 system. Here each of the models has exactly one parameter in common with the truth and the other two are perturbed with the same value, but with different signs. As each parameter appears for precisely one variable, we expect that the model might have more skills when it comes to predicting the course of this variable. This is a very crude statement as the variables are all dependent on each other, so the situation is probably more complicated. Therefore the expected values might not be good at all.

With the above reasoning we arrive at expected values shown in table 8.4. The information of the models that have the correct parameter for a certain variable should be valued high. Model 1 has for instance the correct value for σ and should therefore perform better on the x variable. Therefore the information from model 1 should be high valued in models 2 and 3, whereas the influence of model 2 and 3 on model 1 for the x variable should be close to zero. Nothing can be said however for the connection between models 1 and 2 as they have an identical perturbation.

We now choose a random set of connection coefficients in the range of 0 to 10, for which the resulting connection coefficients after minimization are shown in table 8.5. The sets of connection coefficients do have some similarities, caused by the fact that the perturbations in the sets of models only differ in size, but some large differences exist as well, indicating that the models do not behave similar due to larger perturbations in parameters.

A comparison with the expected values shows that the size of some of the connection coefficients is as expected, but most of the connection coefficients are not comparable with the expected values. In some cases the coefficients that we expected to be large and the ones that we expected to be close to zero are

	10%	20%	30%	Expected
C^1_{12}	8.7	8.1	4.9	0
C^1_{13}	5.8	6.5	4.2	0
C^1_{21}	5.8	3.8	1.5	large
C^1_{23}	8.9	8.2	7.1	?
C^1_{31}	4.1	4.4	4.5	large
C^1_{32}	12.7	12.1	8.8	?
C^2_{12}	7.4	4.3	9.7	large
C^2_{13}	0.9	6.0	9.3	?
C^2_{21}	4.7	5.1	8.1	0
C^2_{23}	9.6	4.2	8.0	0
C^2_{31}	6.2	4.7	7.4	?
C^2_{32}	5.5	5.5	10.9	large
C^3_{12}	3.7	1.9	1.4	?
C^3_{13}	0.8	3.2	1.2	large
C^3_{21}	16.7	18.6	12.5	?
C^3_{23}	9.8	9.3	7.9	large
C^3_{31}	0.0	4.9	0.0	0
C^3_{32}	3.3	6.8	3.8	0

Table 8.5: The value of the connection coefficients found by the learning process started from random initial connection coefficients and the expected size of the connection coefficients belonging to the combination of models in table 8.3.

even approximately of the same size. It can be that the learning process found a different minimum, than the minimum that we expected. The chance of this happening is rather large as many different minima exist. Therefore in the next section we will not take random initial connection coefficients, but coefficients that are already close to the values we would expect to give a good super-model.

8.2.2 An expected minimum as initial condition

Looking at the ranges of the found minimum in table 8.5 we chose the initial connection coefficients to be as in table 8.6. The expected values are just used as initial connection coefficients and not as a super-model as the problem is rather complicated and it is not exactly clear how large we should choose the connection coefficients. By starting with the expected values we can however be quite sure to find a minimum close to the expected values. The resulting connection coefficients are shown in table 8.6 as well. Indeed most of the connection coefficients are still close to the expected value, but there are some exceptions, for instance C^y_{23} and C^3_{31}. To see whether the expected values give an equally good or even better super-model, we will use the measures introduced in chapter 5 again.

8.2.3 Comparing super-models

In figure 8.11 the solutions of the resulting super-models are plotted. This figure shows that the attractors of the solutions for the super-models with the expected values are shifted to above and are therefore less close to the truth than the other solutions. This suggests that the solutions for expected values are not as good as those for the random values. The values for the cost function give a similar result as the values for the unexpected values are 0.0002, 0.001 and 0.017 for a perturbation of 10, 20 or 30% respectively, but for the expected values this was 0.03, 0.11 and 0.11 respectively.

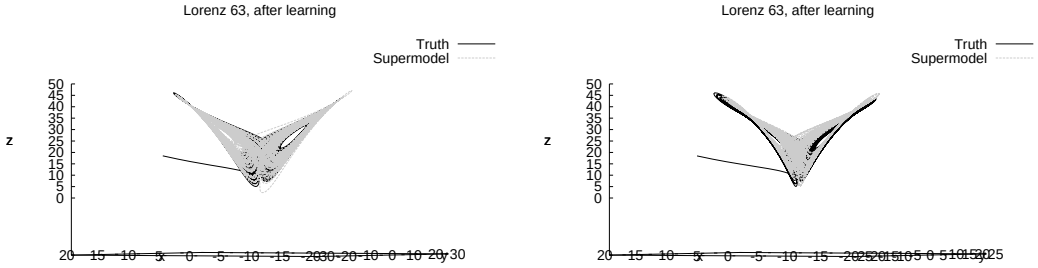
In table 8.7 the mean, standard deviation and covariance can be found. Due to the large errors it is not possible to use all results to a full extend, but for instance the standard deviations and the covariance for x and y have smaller errors and can therefore be used to compare the solutions. The values are all quite

	Initial	10%	20%	30%	Expected
C^1_{12}	0	0.3	0.4	1.0	0
C^1_{13}	0	0.2	0.3	1.4	0
C^1_{21}	10	9.8	9.6	9.3	large
C^1_{23}	2	2.0	2.0	2.1	?
C^1_{31}	10	9.9	9.7	9.2	large
C^1_{32}	2	2.0	1.8	2.0	?
C^2_{12}	10	4.3	7.2	5.7	large
C^2_{13}	2	6.0	9.8	8.2	?
C^2_{21}	0	1.6	1.1	1.5	0
C^2_{23}	0	9.9	10.0	9.6	0
C^2_{31}	2	2.9	2.5	3.5	?
C^2_{32}	10	10.6	10.5	11.2	large
C^3_{12}	2	0	0	0	?
C^3_{13}	10	8.1	8.2	8.7	large
C^3_{21}	2	2.1	2.5	2.7	?
C^3_{23}	10	10.0	10.1	10.5	large
C^3_{31}	0	2.4	3.2	3.4	0
C^3_{32}	0	0.2	0.8	0.2	0

Table 8.6: The value found of the connection coefficients found by the learning process started from the values in the column 'Initial' and the expected size of the connection coefficients belonging to the combination of models in table 8.3.

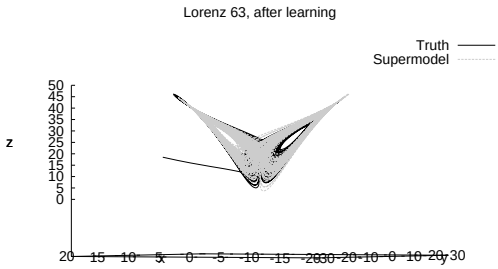
		10%		20%		30%	
	Truth	Random	Expected	Random	Expected	Random	Expected
Mean x	0	0.006 (0.22)	-0.003 (0.15)	-0.0003 (0.21)	0.003 (0.15)	0.0005 (0.19)	-0.003 (0.10)
Mean y	0	0.006 (0.22)	-0.003 (0.16)	-0.0003 (0.21)	0.003 (0.16)	0.0005 (0.19)	-0.003 (0.11)
Mean z	23.55	23.58 (0.02)	23.94 (0.008)	23.64 (0.02)	24.72 (0.04)	23.56 (0.001)	24.64 (0.007)
SD x	7.92	7.96 (0.004)	7.65 (0.001)	7.93 (0.003)	7.55 (0.002)	7.87 (0.002)	7.36 (0.003)
SD y	9.01	9.03 (0.008)	9.01 (0.007)	9.05 (0.008)	9.10 (0.007)	9.13 (0.007)	9.16 (0.006)
SD z	7.62	8.65 (0.022)	8.82 (0.009)	8.72 (0.021)	9.11 (0.005)	8.95 (0.014)	9.65 (0.005)
Cov xy	62.79	63.17 (0.07)	60.67 (0.02)	63.07 (0.06)	60.27 (0.02)	62.96 (0.04)	58.80 (0.04)
Cov xz	0	0.02 (0.75)	-0.008 (0.50)	-0.001 (0.72)	0.01 (0.49)	0.002 (0.68)	-0.01 (0.35)
Cov yz	0	0.02 (0.61)	-0.009 (0.52)	-0.002 (0.60)	0.01 (0.56)	0.002 (0.60)	-0.01 (0.46)

Table 8.7: The mean, standard deviation (SD) and covariance (Cov) for the two super-models for each of the perturbations. The values between brackets indicate the 95% error estimation. The first super-model was found from a random set of initial connections (Random), whereas the second was found by starting with an expected minimum of the cost function (Expected).

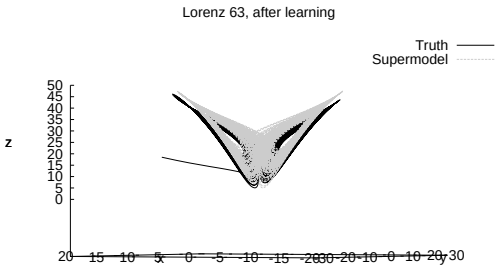


(a) Random (10%)

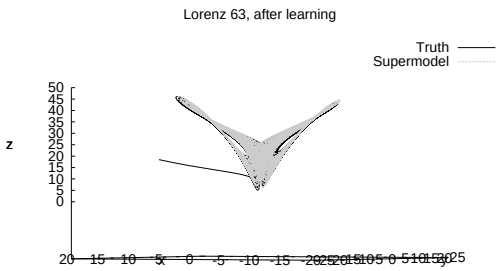
(b) Expected (10%)



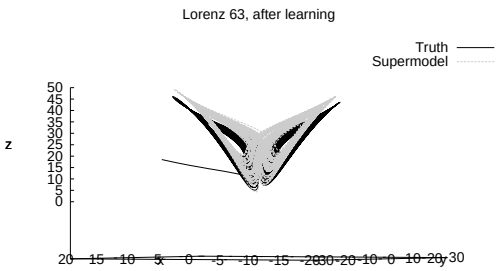
(c) Random (20%)



(d) Expected (20%)

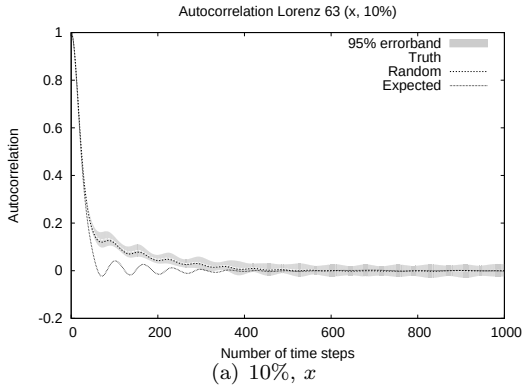


(e) Random (30%)

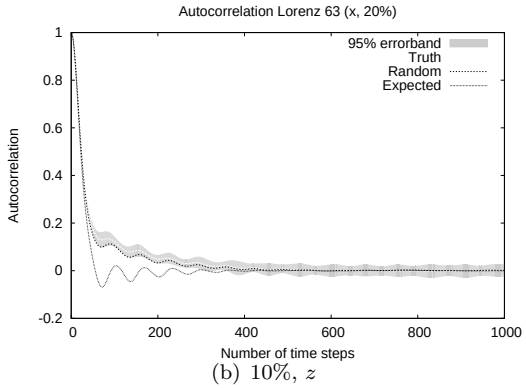


(f) Expected (30%)

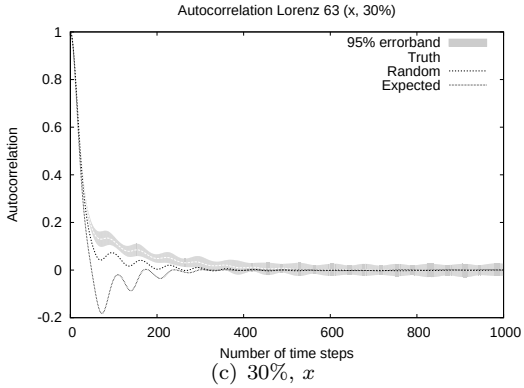
Figure 8.11: Solution for the three perturbed models without connection (black) and the truth (grey) for the Lorenz 63 system.



(a) 10%, x



(b) 10%, z



(c) 30%, x

Figure 8.12: The autocorrelation functions for the two super-models and the truth, including a 95% error band around the truth in grey, for all three perturbations for the x variable. Similar results were found for y and z.

close to that of the truth, but the values for the first super-models are closer than that of the expected super-models. Also for the autocorrelation, shown for x in figure 8.12, shows that the expected models give solutions that are much further away from the truth than that of the randomly found super-models.

For synchronization, however we find completely different results. The randomly found super-models need a nudging strength of 10, 13 and 15 for 10, 20 and 30% perturbations respectively, whereas the expected super-models need strengths of 12, 8 and 9. The tolerances were chosen equal in each of the experiments ($tol = 0.2$ and $tol_2 = 0.4$), as well as the initial conditions. Therefore it seems that for synchronization the expected values do work well. Still notice that only for the larger perturbations this is the case. Why the results for these attractors are so different is not clear, but it might just be that, while the attractors do not match, the trajectories are not that far off. Still synchronization in most cases agrees with the other measures and this is just an exception.

It is clear from the other measures that the expected values for the connection coefficients do not result in a super-model that has comparable skills to the super-models found with random initial connection coefficients. It may be possible to find different expected values for this that do result in a better super-model, but this experiment shows that we ought to be carefull with expected values. It also indicates that the learning process to find a good super-model in an objective way, plays an important role.

(a) Lorenz 63

	σ	ρ	β
Truth	10	28	$\frac{8}{3}$
Model 1	11 (+10%)	25.48 (-9%)	2.88 (+8%)
Model 2	10.9 (+9%)	25.76 (-8%)	2.93 (+10%)
Model 3	10.8 (+8%)	25.2 (-10%)	2.91 (+9%)

(b) Lorenz 84

	a	b	F	G
Truth	0.25	4	8	1
Model 1	0.23 (-8%)	3.64 (-9%)	7.2 (-10%)	0.89 (-11%)
Model 2	0.2275 (-9%)	3.6 (-10%)	7.12 (-11%)	0.92 (-8%)
Model 3	0.225 (-10%)	3.56 (-11%)	7.36 (-8%)	0.91 (-9%)

(c) Rössler

	a	b	c
Truth	0.2	0.2	5.7
Model 1	0.216 (+8%)	0.218 (+9%)	6.27 (+10%)
Model 2	0.218 (+9%)	0.22 (+10%)	6.156 (+8%)
Model 3	0.22 (+10%)	0.216 (+8%)	6.213 (+9%)

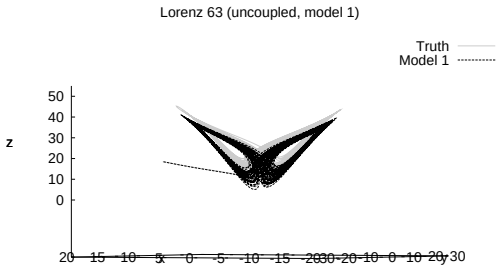
Table 8.8: The sets of perturbed parameters for the Lorenz 63, Lorenz 84 and Rössler systems. The models all have perturbations with the same sign and are perturbed up to 11%.

8.3 Three similarly wrong imperfect models

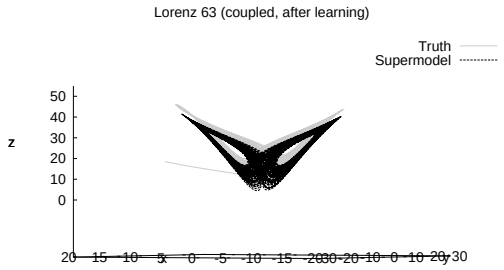
One of the advantages of using multiple imperfect models to find a solution, is that the strengths of the different models can be exploited to find a better solution. But there are models that have similar flaws and strengths. In that case the exchange of information might not add value, as all models already have (a large part of) the information that can be provided by the other models. To see if this is true we created similarly wrong imperfect models by perturbing the parameters with similar size and equal sign and tried to find a good combination for a super-model. This experiment is done for all three systems, for which the parameter values are given in table 8.8.

As the perturbations for each parameter have the same sign, the imperfect models are approximately equal too. In figure 8.13 one of the models without connections to the other models and the super-model are shown for each of the three systems. The attractors of super-model and the unconnected model are very similar, indicating that there has not been much improvement. Also with different initial connection coefficients, the solution did not improve.

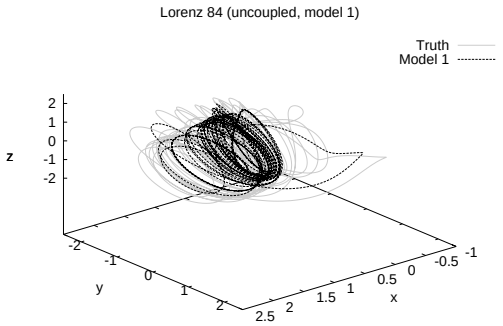
These results show that our hypothesis was right. If the models that are combined have the same strengths, this can not produce a much better super-model. This clearly indicates that choosing a set of models may be more important than choosing for instance the parameters for the systems. This also shows that the strengths of the different models can be exploited by combining imperfect models by information exchange, as with models with different strengths we are able to improve the simulations. It even suggests that including a model that we know to be unrealistic in some sense, but that is different in its imperfection from the other models, can improve the performance of the super-model.



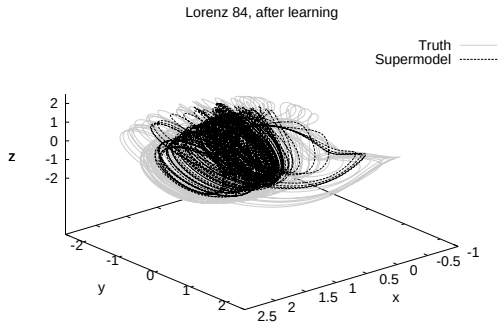
(a) Lorenz 63, model 1, 10%, unconnected



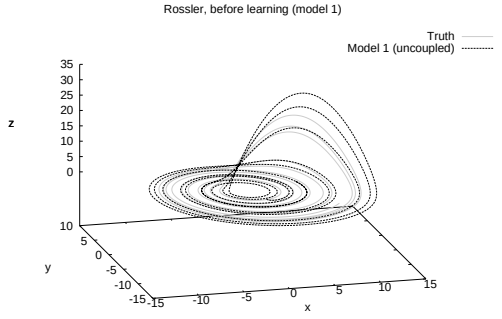
(b) Lorenz 63, super-model



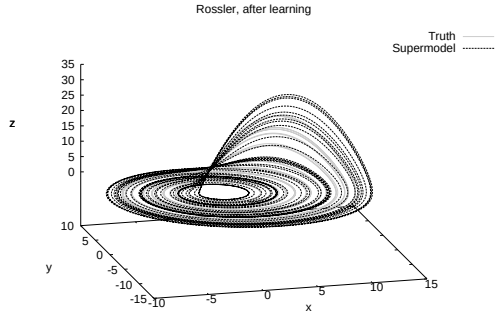
(c) Lorenz 84, model 1, unconnected



(d) Lorenz 84, super-model



(e) Rössler, model 1, unconnected



(f) Rössler, super-model

Figure 8.13: The solutions for the unconnected imperfect model 1 on the left and the resulting super-models on the right. For each of the three systems one imperfect model is shown, since the other models look (almost) identical.

Chapter 9

Discussion

9.1 Multiple solutions

We have found that we can find multiple good minima by just starting from different initial connection coefficients. This is good news as finding a combination will not be hard, but on the other hand it does make it harder to find a deeper minimum as the method is very likely to find a local instead of a global minimum. Trying a method like the amoeba method, that explores the landscape of the cost function, might help to find such a global minimum, but still the method depends on the initial conditions, indicating that it is not able to find the global minimum.

9.2 Measures

There are six measures that can be used for larger models as well, namely: mean, standard deviation covariance, autocorrelation, the value of the cost function and synchronization. Especially the cost function is an important measure, as it is used in the learning process. The cost function is dependent on the different parameters γ, Δ and K, that are on their turn dependent on the system, as described in section 8.1. Therefore these parameters have to be chosen well, which will be harder since the truth will in practice only consist of measurements. Due to the larger time to calculate solutions, a trial and error approach will not work either. Therefore it is encouraging to know that the learning process is not too dependent on the parameters in all cases.

The statistical measures can definitely be calculated in practice. There is enough knowledge of the truth to calculate the values for the truth as well and compare the super-models to the truth. Still this will be based on historical values and the values might be much different in the future, but this problem will exist for all climate models available.

Also for synchronization measurements of the truth are needed to nudge the model towards the truth. In the study for data assimilation [Yang et al., 2006] the nudging didn't necessarily take place at each time step. Synchronization could also take place if the system was nudged every 8 or even 25 time steps. This did result in higher nudging strength, but it does mean that not that many observations in time are needed for the nudging. From the daily practice of data assimilation for weather forecasting we know that enough observations are available to obtain a realistic estimate of the complete state of the atmosphere, which suggests that the observational network is dense enough for synchronization to occur.

Allthough synchronization might have potential, there are also some difficulties with it, which were also mentioned briefly in section 5.4. Synchronization is strongly dependent on the tolerances used in the definition. Especially changing the ϵ has a rather drastic effect. When a certain treshold for ϵ is reached, the nudging strength needed for synchronization will be much smaller. For different supermodels this

treshold is different, which causes super-models that are comparable in quality to need very different nudging strengths. We even saw in section 8.2.3 that sometimes the synchronization does not agree with the other measures and although this is not very common we cannot check how good synchronization works for climate models and it may therefore be a complicated measure to use.

It seems that it may be necessary to introduce different measures or to adapt the existing measures to the larger climate models to be able to say anything about the quality of the solutions. Especially in synchronization it might be necessary to do some testing. Furthermore it seems to be important to use multiple measures to compare the quality of super-models.

9.3 Application to climate models

As discussed in the previous two sections as well as in section 6.5 the application of the approach to larger climate models may be difficult. This is mainly caused by the fact that the information about the truth will consist of measurements and it is not clear if enough measurements are available for the learning process. As shown in section 8.2 it may be hard to find combinations that work well based on what we expect to work. Still the low dependence of the learning process on the spread of the initial conditions and the low number of initial conditions that is actually needed (section 8.1.3) suggest that not much information of the truth is needed.

Another important issue is whether all variables need to be connected. As the climate models have a much larger number of variables it may be necessary to choose a couple of them to connect. As synchronization can work with only one variable connected this can work as well. In that case it should be studied which variables should be connected to get a good solution. Also it is important to use climate models that do not have the same strengths, but that can contribute to a better solution (section 8.3).

Climate models make use of grids to do simulations. These grids consist of multiple points that are situated at a certain distance from each other. This distance and the spread of the grid points can be different for various models. Regridding should be applied to the models before the information exchange can take place. Regridding is a standard procedure in climate modeling, so this will probably not pose a problem.

The problem of predicting the future is that we cannot be sure how well models will make simulations of the future in advance. As mentioned in the introduction climate models are trained on historical measurements, but as the climate is changing it is not true that the models will perform well on future climate if they do well on historical climate simulations. This is a general problem in climate modeling. This problem can be addressed in the setting of the small chaotic dynamical systems for which the truth is already known.

Chapter 10

Conclusion

The goal of this study was to find a way of connecting imperfect models and combining them into a super-model such that the super-model solutions can outperform the separate imperfect models. We were successful in finding these connections in all three small chaotic dynamical systems for many different imperfect models by using a learning process to objectively determine these connections. Many different combinations of connection coefficients were capable of improving the simulations, which makes it easy to find a good super-model.

The measures that were introduced showed that the super-models were indeed good approximations of the truth and they can be used to compare different super-models. The measures agree with each other and with the image of the attractor that could be used to monitor the measures. Also the cost function seems to be a good quality measure to use in the learning process as it was generally true that super-models with low values for the cost function were a good approximation of the truth.

The method needs models with a variety of strengths to improve the simulations, which was shown in section 8.3 so the choice of which models to connect is important. It is also of some importance to choose the parameters needed in the learning process, but the dependence on them for the small chaotic dynamical systems is not very large in all cases (section 8.1).

The approach as we have implemented it is able to find good super-model solutions for almost all initial connection coefficients. Finding a global minimum, however, is hard, even for the amoeba method, due to the large 18-dimensional space of connection coefficients. On the other hand most super-models found have a high enough quality to use and it may therefore not be necessary to look for a global minimum at all.

Chapter 11

Commonly used terms

Term	Definition	Section
Amoeba minimization	A minimization method that explores the landscape of a function to find a global minimum	3.5.3
Attractor	A bounded set that attracts the solutions of a dynamical system	2.1
Autocorrelation function	The similarity of a function to itself at two different points in time	5.3
Bifurcation theory	The study of studying the behaviour of a dynamical system when the values of the parameters are changed.	2.1
Chaos	Sensitive dependence on initial conditions	2
Conjugate gradients	A minimization method	3.5.1
Connection coefficients	Determine the information exchange between models	3.2
cost function	A function that indicates the difference between truth and model	3.4
Covariance	Similarity between two variables in a system	5.2
Dynamical system	The flow of an initial value problem (in this study)	2
Error estimation	An indication of how large the errors in the data are	5.2
Floating point exception	An errror that arises when the numbers in a computer program get too large (tend to infinity)	7.1
Global minimum	The lowest minimum of a function	3.5.1
Hopf bifurcation	A transition for periodic solutions to fixed points	2.1
IPCC	Intergovernmental Panel on Climate Change	1
Local minimum	A minimum of a function that is not the lowest	3.5.1
Lorenz 63	Small chaotic dynamical system used for testing	2.1
Lorenz 84	Small chaotic dynamical system used for testing	2.3
Lyapunov stability	A stability criterium for fixed points of maps	5.4

Term	Definition	Section
Monte Carlo approach	Repeating an experiment many times for random initial conditions	3.5.2
Nudging	Trying to push a model in the direction of another model	3.2
Robustness	The dependence of the approach on which initial conditions are used and the ability to find a minimum from any initial condition	6.3.1, 7
Rössler	Small chaotic dynamical system used for testing	2.2
Simulated annealing	A strategy of slowly coming to a global minimum	3.5.2
Standard deviation	Spread of a data set with respect to the mean value	5.2
Super-model	Combination of multiple imperfect models	3.3
Synchronization	Linking chaotic systems can lead to synchronization.	1.1, 5.4
Training set	The set of initial conditions used in the learning process	6.3.1
Truth	A small chaotic dynamical system with certain parameter values that is used as reality	3.1

Bibliography

V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer, 1989.

R. Barrio, F. Blesa, and S. Serrano. Qualitative analysis of the Rössler equations: Bifurcations of limit cycles and chaotic attractors. *Physica D: Nonlinear Phenomena*, 238(13):1087–1100, 2009.

V. Cerny. Thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm. *Journal of optimization therory and applications*, 45(1):41–51, 1985.

G.P. Compo, J.S. Whitaker, and P.D. Sardeshmukh. Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, 87:175–189, 2006.

E.J. Doedel, B. Krauskopf, and H.M. Osinga. Global bifurcations of the lorenz manifold. *Nonlinearity*, 19(12):2947–2972, 2006.

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

G.S. Duane and J. Tribbia. Weak atlantic-pacific teleconnections as synchronized chaos. *J. Atmos. Sci.*, 61:2149–2168, 2004.

G.S. Duane, J. Tribbia, and J. Weiss. Synchronicity in predictive modelling: a new view of data assimilation. *Nonlin. Proc. in Geophys.*, 13:601–612, 2006.

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

G.S. Duane, J. Tribbia, and B.P. Kirtman. Consensus on long-range prediction by adaptive synchronization of models. Vienna, Austria, 2009. EGU General Assembly.

R. Fletcher and C.M. Reeves. Function minimization by conjugate gradients. *The Computer Journal*, 6: 149–154, 1963.

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

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

E.N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20:130–140, 1963.

E.N. Lorenz. Irregularity, a fundamental property of the atmosphere. *Tellus*, 36 A(2):98–110, 1984.

N. Nakicenovic. Greenhouse gas emissions scenarios. *Technological Forecasting and Social Change*, 65 (2):149–166, 2000.

J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7: 308–313, 1965.

L.M. Pecora and T.L. Carroll. Synchronization in chaotic systems. *Physical Review Letters*, 64:821–824, 1990.

M.J. Rodwell and T. Jung. Understanding the local and global impacts of model physics changes: an aerosol example. *Quarterly Journal of the Royal Meteorological Society*, 134:1479–1497, 2008.

O.E. Rössler. An equation for continuous chaos. *Physics Letters*, 75 A(5):397–398, 1976.

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

L.A. van den Berge, F.M. Selten, W. Wiegerinck, and G.S. Duane. A multi-model ensemble method that combines imperfect models through learning. Submitted to Earth System Dynamics, chapter 6 of this thesis, 2010.

L. van Veen. *Time scale interaction in low-order climate models*. PhD thesis, Utrecht University, 2002.

F. Verhulst. *Nonlinear Differential Equations and Dynamical Systems*. Springer, 2000.

S. Yang, D. Baker, H. Li, K. Cordes, M. Huff, G. Nagpal, E. Okereke, J. Villafañe, E. Kalnay, and G.S. Duane. Data assimilation as synchronization of truth and model: Experiments with the three-variable lorenz system. *Journal of the Atmospheric Sciences*, 63:2340–2354, 2006.