

# On the limit of large couplings and weighted averaged dynamics

Wim Wiegerinck <sup>\*1</sup>, Willem Burgers <sup>†1</sup> and Frank Selten <sup>‡2</sup>

<sup>1</sup>Radboud University, Nijmegen, The Netherlands

<sup>2</sup>Royal Netherlands Meteorological Institute, De Bilt, The Netherlands

To appear in: Consensus and Synchronization in Complex Networks, Kocarev, Ljupco (Ed.), Springer, 2013

## Abstract

We consider a network of deterministic non-linear oscillators with non-identical parameters. Interactions between the different oscillators are linear, but the coupling coefficients for each interaction may differ. We consider the case where coupling coefficients are sufficiently large, so that the different oscillators will have their state variables strongly tied together and variables of the different oscillators will rapidly become (almost) synchronized. We will argue that the dynamics of the network is approximated by the dynamics of weighted averages of the vector fields of the different oscillators. Our focus of application will be on so-called supermodeling, a recently proposed model combination approach in which different existing models are dynamically coupled together aiming to improved performance. With large coupling theory, we are able to analyze and better understand earlier reported supermodeling results. Furthermore, we explore the behavior in partially coupled networks, in particular supermodeling with incomplete models, each modeling a different aspect of the truth. Result are illustrated numerically for the Lorenz 63 model.

## 1 Introduction

Synchronization is the phenomenon that coupled oscillating systems fall into the same rhythm. Examples include clocks, singing crickets, firing neurons and applauding audiences [10]. Similar phenomena occur in multi-agent systems where synchronization mechanisms can be used to describe consensus and cooperation [8, 16].

---

\*w.wiegerinck@science.ru.nl

†w.burgers@science.ru.nl

‡selten@knmi.nl

Recently, synchronization mechanisms have been proposed as modeling tools, in particular for data assimilation [15, 4] and for performance improvement by model combination [3]. These proposals were made in the context of atmospheric modeling and climate science. In climate research, there are nowadays about a dozen comprehensive climate models in the world. Despite the improvements in the quality of the model simulations, the models are still far from perfect, and models differ substantially in their climate predictions. The current practice to deal with these different models is to take some form of an ensemble average of the individual outcomes [13]. The proposal using the synchronization mechanism is to dynamically connect the models and so construct one “supermodel”. With sufficient coupling strength, the models in the supermodel will synchronize. The idea is that when coupling strengths are learned from data, the dynamics of the synchronized state may provide a better representation of the climate statistics and dynamics than the conventional ensemble approach would do.

In [1], experiments have been performed with low dimensional systems such as the Lorenz 63 [6], the Rössler [11], and the Lorenz 84 model [7]. As a sort of proxy for climatology, the attractors of the different models and model combinations are assessed. It is found that an optimized supermodel can have an attractor that is a good approximation of the attractor of the assumed ground truth, even when the attractors of the different models are very different. Furthermore, the remarkable result has been found that the supermodel was able to accurately follow the change of the attractor after doubling of one of the system parameters in both the ground truth and models, without the need of a further adaptation of the connections. However, it has also been found that there are many local optima, i.e., about equally good solutions of the connected model with very different connection coefficients. All of these solutions have relatively large connections.

In this chapter, we will further investigate the dynamics and long term behavior of supermodels, i.e. networks of linearly coupled nonlinear oscillators. In the supermodeling context, the couplings differ for each pair of oscillators, and they differ per coupled variable. The behavior is studied in the regime of large supermodel coefficients. Here we find results that generalize on earlier work on synchronization and consensus forming of coupled nonlinear oscillators with large coupling coefficients [5, 12, 8, 16]. We will apply the theoretical results to understand the findings in [1]. The networks considered in [1] are fully connected. In more complex supermodels it is likely that not all variables can be coupled, e.g. due to complexity restrictions, or because different models have variables that describe different aspects of the observable reality. In the latter case, each of the models is an incomplete description. In the supermodel they can complement each other. These are motivations to consider partially connected networks as well.

The chapter is organized as follows. We will first provide a review of supermodeling in section 2). We show in section 3 that in networks of oscillators with large coupling, the individual states synchronize and that the synchronized state will follow a certain weighted averaged dynamics, in which the weights follow from eigenvector analyses of the so-called

Laplacian matrices. In section 4, we will analyze the local optima reported in [1], and show that results can be understood from the weighted averaged dynamics predicted from the theory. In section 5, we generalize on the earlier coupling scheme to partially coupled systems and argue that the system follows a similar partially weighted averaged dynamics. In section 6, we consider a situation where each of imperfect models is incomplete and models different aspects of the ground truth. The incomplete models are coupled by their common variable. In the resulting supermodel the different aspects are now combined in one dynamical system. Finally, we end with a discussion in section 7. All the numerical results in this chapter are in the context of the Lorenz 63 system.

## 2 Supermodels

In this section we review the supermodeling approach and the findings as reported in [1]. The assumption is that there is a ground truth with an observable state  $\vec{x}_{\text{gt}}(t)$  that is driven by a nonlinear chaotic dynamics that is not exactly known. It is further assumed that there are  $M$  good, but imperfect models of this ground truth dynamics. These models are labeled by  $\mu$ . Each of them describes the dynamics of the model state vector  $\vec{x}_\mu$  according to

$$\dot{x}_\mu^i = f_\mu^i(\vec{x}_\mu) \quad (1)$$

in which  $i$  labels the vector components, and dot-notation is used for time derivatives. Here it is assumed that each model is in the same representation as the ground truth, i.e. vector components of each of the models can be compared to each other and with the ground truth. Then, the proposal is to combine the individual models  $\mu$  into one supermodel by inserting nonnegative connections between the model equations,

$$\dot{x}_\mu^i = f_\mu^i(\vec{x}_\mu) + \sum_\nu C_{\mu\nu}^i (x_\nu^i - x_\mu^i). \quad (2)$$

The idea is that with sufficient connectivity, the individual models will synchronize and get a kind of consensus with each other [9]. The solution of the supermodel is defined to be the average of the coupled imperfect models,

$$\vec{x}_{\text{sumo}}(\vec{C}, t) \equiv \frac{1}{M} \sum_\mu \vec{x}_\mu(\vec{C}, t). \quad (3)$$

The connection coefficients  $\vec{C} = \{C_{\mu\nu}^i\}$  are to be inferred from a training set of historical observations. The goal of the supermodel is to do climate simulation, i.e. the supermodel should converge to an attractor that is similar to the attractor of the ground truth. However, since direct attractor learning is difficult, a proxy cost function has been proposed, which basically consists of a sum of short term prediction errors. This cost function is parametrized by a training set, being a time series of observations of the truth  $\{\vec{x}_{\text{gt}}(t)\}$ . To construct the cost function,  $K$  runs of the supermodel are performed. The runs are initialized at times  $t_i$ ,  $i = 1, \dots, K$ , by setting each of the imperfect model states equal to the ground truth state  $\vec{x}_\mu(t_i) = \vec{x}_{\text{gt}}(t_i)$ . Then at each run, the supermodel is

integrated during a (short) period  $\Delta$ . The cost function is now defined by the accumulated error of these runs,

$$E(\vec{C}) = \frac{1}{K\Delta} \sum_{i=1}^K \int_{t_i}^{t_i+\Delta} |\vec{x}_{\text{sumo}}(\vec{C}, t) - \vec{x}_{\text{gt}}(t)|^2 \gamma^t dt, \quad (4)$$

where  $\gamma^t$  is a discount factor, with  $0 < \gamma < 1$  a constant. The idea behind this cost function is that free runs of the supermodel should follow the ground truth trajectories as long as possible. However, the 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 initial divergence between model and truth, but at later times in the short term integrations the internal error growth dominates. The factor  $\gamma^t$  is included to discount the errors at later times to decrease the contribution of internal error growth.

To demonstrate the supermodeling approach, a number of simple chaotic systems has been studied such as the Lorenz 63 model [6]. The equations for the Lorenz 63 model are

$$\dot{x} = \sigma(y - x) \quad (5)$$

$$\dot{y} = x(\rho - z) - y \quad (6)$$

$$\dot{z} = xy - \beta z. \quad (7)$$

This model is used as a metaphor for the atmosphere, because of its regime changes and unstable nature. The model with standard parameter values ( $\sigma = 10$ ,  $\rho = 28$ ,  $\beta = 8/3$ ) is used as ground truth. Imperfect models are assumed to be in the same model class with perturbed parameter values. These imperfect models are connected and combined into a supermodel. Training data is generated from the assumed ground truth and used to tune the connections in the supermodel. By inspecting plots of the attractor, as well as by considering means, variances, covariances and autocorrelations, it is concluded that the supermodeling approach is a promising modeling approach in the case of complex modeling where good, but still imperfect models are available and a machine learning method starting from scratch is infeasible.

One of the issues that the paper addressed is whether the supermodel approach is also able to deal with climate change, for instance the response of the truth to a parameter perturbation. The question is whether the supermodel would generalize well in such a situation. To study this, the parameter  $\rho$  in the true system as well as the corresponding parameters in the imperfect models in the supermodel have been doubled. The doubling of the parameter in the ground truth has been found to yield an increase of size of the ground truth's attractor. It has been remarked that although the connection coefficients in the supermodel have been learned from data generated with ground truth parameter  $\rho = 28$ , the supermodel with

doubled  $\rho$  actually quite accurately reproduces the enlarged attractor of the ground truth with  $\rho = 56$ .

Another issue are local minima. With different starting conditions, solutions with completely different values for  $\vec{C}$  have been found. The supermodels with these different solutions, however, all produce similar attractors that are all close to the ground truth, and all have about the same performance quality.

### 3 Large coupling limit

To understand why this approach worked so well, and also to understand the relation between the different but somehow apparently equivalent solutions, we now further analyze the dynamics of the supermodel. In [1], the values of the  $\vec{C}$  parameters from two optimizations with different initial condition have been reported (see table 1 later in this chapter). The first thing that can be noticed is that for all  $i$ , some of the values  $C_{\mu\nu}^i$  are quite large, which means that component  $i$  of model  $\mu$ 's state is attracted to the component of model  $\nu$ 's state. This is to be expected since the idea behind the coupling is that the models come into a consensus state.

To analyze this further we rewrite the supermodel equations a bit,

$$\dot{x}_\mu^i = f_\mu^i(\vec{x}_\mu) + \sum_\nu L_{\mu\nu}^i x_\nu^i \quad (8)$$

where

$$L_{\mu\nu}^i = C_{\mu\nu}^i - \delta_{\mu\nu} \sum_\kappa C_{\mu\kappa}^i . \quad (9)$$

is the so-called Laplacian matrix, which is known to play an important role in the analysis of synchronization of coupled systems [12]. Now, following e.g. [8] we remark that the matrix  $L^i$  is a "transition rate matrix", known from stochastic processes. That is to say, with such a  $L^i$  the equation  $\dot{P}_\nu = \sum_\mu P_\mu L_{\mu\nu}^i$  is a continuous time Markov process. If  $L^i$  is mixing then  $P$  converges to an equilibrium distribution, which is the normalized left eigenvector  $\vec{w}^i$  of  $L^i$  with eigenvalue 0. Normalization means  $\sum_\mu w_\mu^i = 1$ . Its right eigenvector is the vector with all components equal, i.e., a vector of the form  $x^i(1, \dots, 1)^T$ , which can be interpreted as a fully synchronized state  $x^i$ . Now if  $L^i$  is mixing then the other eigenvalues have a negative real part, which means that the other modes will vanish and states of the different models will synchronize into a joint state,  $x_\mu^i(t) = x_\nu^i(t) = x^i(t)$ . By multiplying the supermodel (8) from the left by  $\vec{w}^i$ , we obtain the synchronized state dynamics

$$\dot{x}^i = \sum_\mu w_\mu^i f_\mu^i(\vec{x}) . \quad (10)$$

This equation states that for large  $\vec{C}$ , the supermodel dynamics is basically described by a vector field of which the components are convex combinations, i.e., weighted averages of the imperfect model components. The weights are given by the left eigenvectors of the Laplacians  $L^i$  of the component-wise coupling matrices  $\vec{C}^i$ .

	Sumo 1	Sumo 2		Sumo 1	Sumo 2
$C_{12}^x$	-0.01	1.52	$C_{12}^y$	7.67	3.53
$C_{13}^x$	4.81	0.03	$C_{13}^y$	18.14	27.36
$C_{21}^x$	5.69	13.28	$C_{21}^y$	3.64	0.00
$C_{23}^x$	13.75	14.99	$C_{23}^y$	10.06	6.50
$C_{31}^x$	17.64	21.51	$C_{31}^y$	2.71	3.89
$C_{32}^x$	-0.01	1.09	$C_{32}^y$	9.79	6.93
				Sumo 1	Sumo 2
			$C_{12}^z$	5.47	3.95
			$C_{13}^z$	4.03	12.24
			$C_{21}^z$	10.72	3.50
			$C_{23}^z$	13.54	2.20
			$C_{31}^z$	8.70	2.89
			$C_{32}^z$	1.50	3.85

Table 1: The connection coefficients of two super-model solutions of the Lorenz 63 system as found in [1].

From this result, one can directly predict when the coupled models are able to reproduce the ground truth. In [1] both the ground truth system and the perturbed imperfect systems are linear in the parameters (i.e.  $\sigma$ ,  $\rho$ , and  $\beta$  in (7), see next section, but also the other models described in [1]). So, if we take the weighted average of the imperfect model equations, we obtain a model that is again in that model class, with parameters that are weighted averages of the perturbed parameters. If there are weights such that the ground truth parameters can be recovered, then it is in principle possible to find a supermodel that reproduces the ground truth.

## 4 Case study: Lorenz 63 supermodel

To what extent does the large coupling theory of the previous section apply to supermodeling and does it help to understand its results, or are other mechanisms more important? In this section, we investigate this by applying the theory of the previous section to analyze the connection coefficients for the Lorenz 63 supermodel that have been reported in [1]. Table 1 shows the results of two independent coupling parameters optimizations with different initializations as reported in [1]. These values are used for analysis in this section, except that small negative values have been set equal to zero. The question here is what these seemingly different solutions have in common. To answer this question, we do the eigenvalue/eigenvector computation of the Laplacians of the two supermodels. For each of the Laplacians, one eigenvalue is  $\lambda_0 = 0$ . The other eigenvalues have negative values. They are listed in table 2. The more negative the eigenvalue, the better the weighted average approximation.

Now assuming that the connections are large enough that the approximations described the previous section hold, the supermodels will effectively follow the dynamics described by the weighted sum of the vector

	$\lambda_1$	$\lambda_2$
$L^x$	-19.4400	-22.4500
$L^y$	-23.9185	-28.0915
$L^z$	-17.2569	-26.7031

	$\lambda_1$	$\lambda_2$
$L^x$	-23.6834	-28.7366
$L^y$	-12.8275	-35.3825
$L^z$	-9.5436	-19.0864

Table 2: Second and third largest eigenvalues of the Laplacian matrices of both supermodels. Left table supermodel 1. Right table: supermodel 2. Note that the largest eigenvalue in both supermodels is  $\lambda_0 = 0$ .

Eigenvector	$w_1$	$w_2$	$w_3$
$\vec{w}^x$	0.7857	0	0.2143
$\vec{w}^y$	0.1083	0.4070	0.4847
$\vec{w}^z$	0.4929	0.1342	0.3729

Eigenvector	$w_1$	$w_2$	$w_3$
$\vec{w}^x$	0.9148	0.0505	0.0347
$\vec{w}^y$	0.0557	0.5019	0.4424
$\vec{w}^z$	0.1644	0.4049	0.4307

Table 3: Left eigenvectors with eigenvalue zero of the Laplacian matrices. Left table supermodel 1. Right table: supermodel 2.

fields of the imperfect models, where the weights are given by the left eigenvectors of the zero eigenvalue. Since the imperfect models in this example are all Lorenz 63 models, and since their vector fields are linear in the parameters  $(\sigma, \rho, \beta)$ , the resulting weighted average will also follow the dynamics of a Lorenz 63 vector field, with the weighted average of the parameters. These parameters could be considered as the effective supermodel parameters. In table 4, the parameters for the assumed ground truth and the imperfect models are listed as well as the effective parameters of the two supermodels. We see that for both supermodels, the effective parameters are closer to the ground truth than any of the imperfect model parameters. However, there is still a considerable discrepancy between the effective parameters and the parameters of the ground truth.

To get an idea of the quality of the approximation, we plot the trajectories of the original supermodels as well as their approximations using the effective supermodel parameters against runs of the ground truth, see figure 1. Here we observe that there is indeed a small discrepancy between the trajectories of the supermodels and the approximations based on their effective parameters. In general, the supermodels seem to be slightly closer to the ground truth. This makes sense because their connections are optimized with respect to the ground truth, whereas the effective supermodel parameters are only derived from the (optimized) connections under a large  $C$  limit approximations. A further remark here is that with finite  $C$  values, the corresponding effective parameters of the optimal model do not have to match the ground truth parameters exactly, since the  $C$ 's are tuned in the presence of the nonlinear imperfect model vectorfields  $\vec{f}_\mu$  in

	$\sigma$	$\rho$	$\beta$
Truth	10	28	2.667
Model 1	13.25	19	3.5
Model 2	7	18	3.7
Model 3	6.5	38	1.7
Supermodel 1	11.8038	27.8024	2.8556
Supermodel 2	12.6998	26.9034	2.8057

Table 4: Ground truth and imperfect model parameters for the Lorenz 63 system as used in [1] as well as the effective supermodel parameters computed by weighting the model parameters according to the eigenvectors for both supermodels.

the supermodel (2). Note finally, that even nearly identical systems will show discrepancies in their trajectories due to chaos.

In [1], the issue of local minima in optimization of the connection coefficients has been addressed. In particular, the shape of the cost function (4) as function of the connections  $\vec{C}$  was studied by taking cross sections of the cost function around the optimized value of  $\vec{C}$ . The cross sections were created by changing one connection coefficient and keeping the others fixed at their values at the minimum. In particular the cross sections of coefficients  $C_{23}^y$  and  $C_{21}^z$  in supermodel 1 were reported as being typical examples. These two connections both have about the same value  $C \approx 10$  (see table 1). However the cross section for coefficient  $C_{23}^y$  showed a clear minimum, whereas the cross section for  $C_{21}^z$  is almost a constant function (see [1] for details).

With the large coupling eigenvector analysis, the differences in the shape of the cross sections can be understood. To illustrate this, we perturbed coefficients  $C_{23}^y$  and  $C_{21}^z$  respectively by multiplying/dividing by 5 while keeping the other constant (see tables 6 and 5). Then the eigenvectors as well as the effective parameters were recomputed. Note that coefficient  $C_{23}^y$  has only influence on eigenvector  $\vec{w}^y$  and effective parameter  $\rho$ , whereas the coefficient  $C_{21}^z$  only influences  $\vec{w}^z$  and  $\beta$ . In table 5, the resulting perturbed eigenvectors are displayed, and in table 6 the resulting perturbed effective parameters. In particular one can see that a factor 5 perturbation in  $C_{23}^y$  results in a relative perturbation of about 20% in  $\rho$ , whereas a factor 5 perturbation in  $C_{21}^z$  results in a relative perturbation of about only 2% in  $\beta$ . In figure 2 trajectories of models with the effective supermodel parameters of supermodel 1, perturbed according to table 6 are plotted, confirming the much greater impact of the 20% perturbation in  $\rho$  than the 2% perturbation in  $\beta$ .

Finally, we remark that the observation that the supermodel was able to reproduce the ground truth's attractor when the parameter  $\rho$  was doubled in both the ground truth and imperfect models can be understood with the large coupling theory: a doubling of all model parameters  $\rho_\mu$ , while  $\vec{w}^x$  is unchanged (because  $\vec{C}^x$  did not change) results in a doubling of the effective parameter  $\rho_{\text{eff}} = \sum_\mu w_\mu^x \rho_\mu$ .



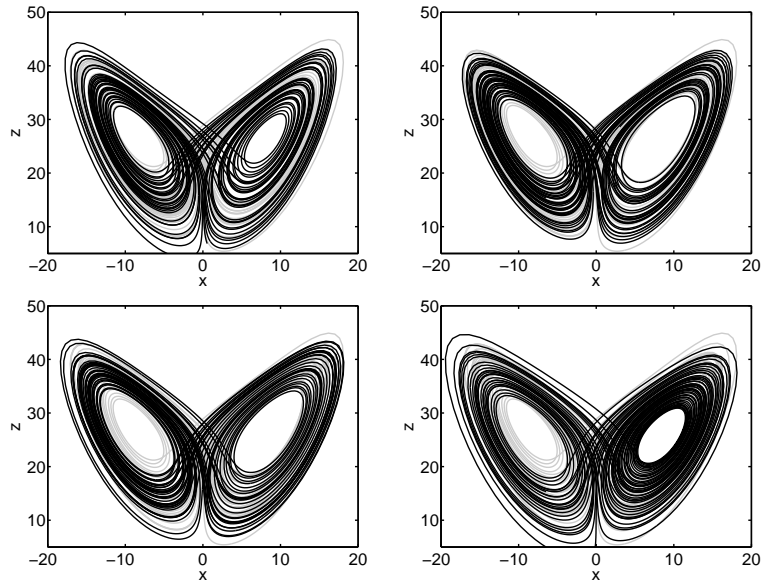


Figure 1: Trajectories for the two supermodels (left figures) and trajectories according to their effective parameters (right figures). Top row: supermodel 1. Bottom row: supermodel 2. Grey: trajectories of the ground truth.

## 5 Partially coupled systems

In this section we consider partially coupled systems, i.e., systems where only subsets of variables are connected. Again we will study the behavior with large connections. We consider for simplicity only the case where for each model variable  $i$ , the graph contains only one connected component and singletons. The set of oscillators  $\mu$  that have their variables  $i$  connected is denoted as  $S(i)$ . The partially connected system dynamics then reads

$$\dot{x}_\mu^i = f_\mu^i(\vec{x}_\mu) + \sum_{\nu \in S(i)} C_{\mu\nu}^i (x_\nu^i - x_\mu^i) \quad \mu \in S(i) \quad (11)$$

$$\dot{x}_\mu^i = f_\mu^i(\vec{x}_\mu) \quad \mu \notin S(i). \quad (12)$$

Now in the limit of large connection coefficients  $\vec{C}$ , the couplings dominate the oscillators within  $S(i)$ , so that we can still apply the earlier theory to the  $i$ 'th component of the oscillators in  $S(i)$ . For all these oscillators we have the left eigenvectors  $\vec{w}^i$  with components  $w_\mu^i$ 's, as well as right eigenvectors with synchronized state components

$$x_\mu^i(t) = x_\nu^i(t) \quad \mu, \nu \in S(i). \quad (13)$$

In the  $i$ -th components of these oscillators we can drop the oscillator index:

$$x_\mu^i(t) = x^i(t) \quad \mu \in S(i). \quad (14)$$

	$w_1^y$	$w_2^y$	$w_3^y$
Supermodel 1	0.108	0.407	0.485
$0.2C_{23}^y$	0.115	0.618	0.267
$5C_{23}^y$	0.100	0.150	0.750
	$w_1^z$	$w_2^z$	$w_3^z$
Supermodel 1	0.493	0.134	0.373
$0.2C_{21}^z$	0.412	0.183	0.405
$5C_{21}^z * 5$	0.620	0.058	0.322

Table 5: Perturbed eigenvectors due to perturbation of supermodel coefficients. Left: perturbed vector  $\vec{w}^y$  due to perturbations in  $C_{23}^y$ . Right: perturbed vector  $\vec{w}^z$  due to perturbations in  $C_{21}^z$ .

	$\rho$	$\beta$
Supermodel 1	27.8024	2.8556
$0.2C_{23}^y$	23.448151	2.8556
$5C_{23}^y$	33.095626	2.8556
$0.2C_{21}^z$	27.8024	2.807157
$5C_{21}^z$	27.8024	2.932322

Table 6: Perturbed effective parameters due to perturbation of supermodel coefficients.

By defining  $S(\mu) = \{i : \mu \in S(i)\}$  and  $S_c(\mu)$  its complement, and using subset notation  $\vec{x}^{(a_1, \dots, a_n)} = (x^{a_1}, \dots, x^{a_n})$ , the state vectors of each oscillator is written as

$$\vec{x}_\mu(t) = (\vec{x}^{S(\mu)}(t), \vec{x}_\mu^{S_c(\mu)}(t)) \quad (15)$$

Multiplying (11), i.e., the connected components of the full system (11) and (12) from the left with their left eigenvectors, the dynamics of the full system in the limit of large couplings is described by the following set of equations, which we call the partially weighted system,

$$\dot{x}^i = \sum_{\mu \in S(i)} w_\mu^i f_\mu^i(\vec{x}^{S(\mu)}, \vec{x}_\mu^{S_c(\mu)}) \quad (16)$$

$$\dot{x}_\mu^i = f_\mu^i(\vec{x}^{S(\mu)}, \vec{x}_\mu^{S_c(\mu)}) \quad \mu \notin S(i). \quad (17)$$

As an illustration, we simulated 10 Lorenz 63 oscillators, each with their own perturbed  $\sigma$ ,  $\rho$  and  $\beta$  parameters. All these parameters were generated by perturbing the standard values by 10% standard zero mean unit variance Gaussian noise  $\xi$ , i.e.,  $\sigma_\mu = \sigma(1 + \xi/10)$ , etc. The oscillators were partially coupled as follows. The  $x$  variables of oscillator 1 to 6 and the  $y$  variables of oscillator 5 to 10 were coupled. The  $z$  variables were not coupled. See figure 3. Couplings were generated randomly according to a uniform distribution between 0 and 100. For the associated partially weighted model the weights were computed using the eigenvector method.

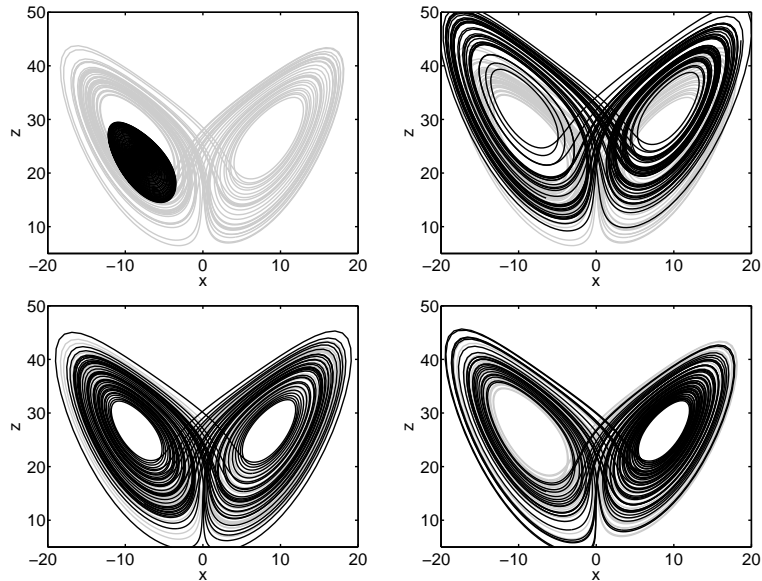


Figure 2: Trajectories for the models with parameters that result from perturbing the connection components of supermodel 1 by a factor 5 . The perturbed parameters are: upper left  $0.2C_{23}^y$ , upper right  $5C_{23}^y$ , lower left  $0.2C_{21}^z$ , lower right  $5C_{21}^z$ . Grey: trajectories of the effective parameter version of the unperturbed supermodel 1.

The variables  $x_1(t), \dots, x_6(t)$  were replaced by a single model variable  $x(t)$ , and  $y_5(t), \dots, y_{10}(t)$  by the single model variable  $y(t)$ . Note that the partially connected is 30 dimensional whereas the partially weighted model is 20 dimensional.

We simulated both the partially connected model and the partially weighted model. All the oscillators were initiated in the same random state. Then the systems were iterated for 100 time units. The first 50 units were discarded from the simulation, the last 50 time units were plotted. For the partially connected model, we plotted the  $(x_1, z_1)$  and the  $(x_1, z_{10})$  component of the partially connected model and the  $(x, z_1)$  and the  $(x, z_{10})$  component of the partially weighted model. See figure 4. The figures suggest a quite strong similarity between the partially connected and the partially weighted model, as expected from the theory.

## 6 Coupling incomplete systems

In the supermodeling paradigm, an example of incomplete subsystems that are to be coupled would occur in the following hypothetical and overly simplified climate modeling case. Assume that the ground truth of the real climate system is governed, among others, by pressure fields,

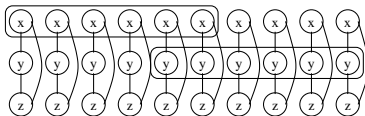


Figure 3: Partially connected network topology. Each vertical set of three interconnected nodes labeled  $x, y$ , and  $z$  represents a Lorenz 63 oscillator. The variables within a box are fully connected via  $C_{\mu\nu}^i$  coefficients. In the partially weighted model, the variables within a box are identified.

clouds and oceans. Suppose we have two imperfect models, model A and model B. Model A has variables for pressure fields and for clouds but not for the oceans. The effect of the oceans and other variables on the pressure fields and the clouds may be parametrized. This is to say that the effect of the ocean is modeled via additional terms in the pressure and cloud dynamics rather than being modeled by its own dynamics. Model B has variables for the pressure fields and for oceans, but not for the clouds. The effect of the clouds may again somehow be parametrized in model B. Now, since model A is modeled as a dynamical system with pressure fields and clouds, it is based on pressure field and cloud data. Data from the oceans is not taken into account. In the same way, model B is based on data of pressure fields and oceans. Now data from clouds is not taken into account. The supermodel should then combine these independently modeled imperfect models. The parameters of the supermodel are then to be estimated based on combined data of pressure fields, clouds and oceans.

To study this further we simulate an assumed ground truth system. We construct two incomplete models, model A and model B. The effect of the missing variable is modeled by a parametrization. Then we combine these incomplete models into a single supermodel. We tune the couplings using the full (training) data and see if this supermodel gives better results on test data. Here we explore this approach in the context of the Lorenz 63 system. The ground truth is given by equations for  $(x, y, z)$ .

$$\dot{x} = \sigma(y - x) \quad (18)$$

$$\dot{y} = x(\rho - z) - y \quad (19)$$

$$\dot{z} = xy - \beta z \quad (20)$$

with standard parameters  $\rho = 28$ ,  $\sigma = 10$ , and  $\beta = 8/3$ .

We assume model A models only dynamics in  $x$  and  $y$ . Assuming some insight from the designers, their dynamics according to model A is given by

$$\dot{x}_a = \sigma(y_a - x_a) \quad (21)$$

$$\dot{y}_a = x_a(\rho - Z(x_a, y_a)) - y_a \quad (22)$$

where we used the label  $a$  to emphasize that these are variables in model A. The function  $Z(x_a, y_a)$  is a parametrization of  $z$ . We model the

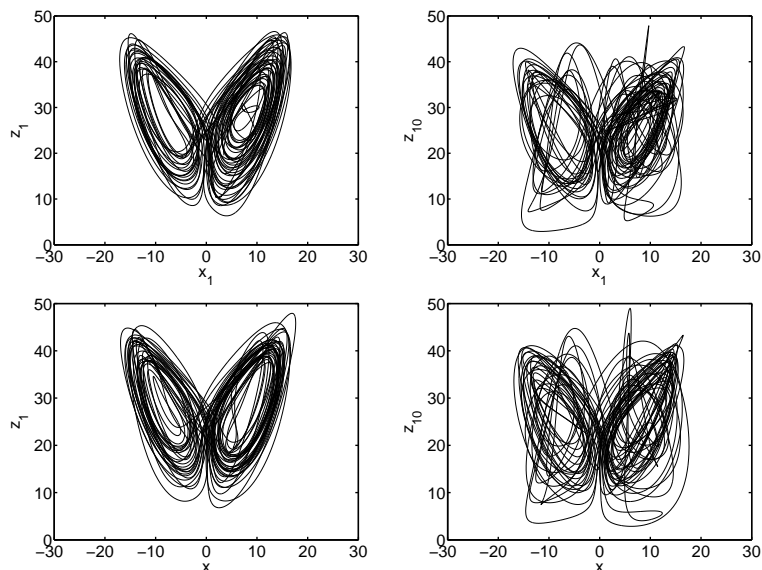


Figure 4: Projections of trajectories from the partially connected models to the  $(x_1, z_1)$  (top left) and  $(x_1, z_{10})$  (top right) plane and from the partially weighted model to the  $(x, z_1)$  (bottom left) and  $(x, z_{10})$  (bottom right) plane.

parametrization  $Z(x, y)$  by a linear combination of radial basis functions [2],

$$Z(x, y) = \sum_i \alpha_i \phi_i(x, y) \quad (23)$$

where the radial basis functions are

$$\phi_i(x, y) = \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_x^2} - \frac{(y - \nu_i)^2}{2\sigma_y^2}\right) \quad (24)$$

and  $\phi_0(x, y) = 1$ . To determine the parameters  $\alpha_i$ , a simulation of the ground truth has been performed, and data has been collected from the  $x, y$  components  $(x(t), y(t))$ . The short term squared prediction error  $E_a$  between the predictions according to model  $A$  and the ground truth data is defined as

$$E_a(\vec{\alpha}) = \sum_t \left( (x(t + dt) - x_a(t + dt|x(t), y(t); \vec{\alpha}))^2 + (y(t + dt) - y_a(t + dt|x(t), y(t); \vec{\alpha}))^2 \right) \quad (25)$$

where  $x_a(t + dt|x(t), y(t); \vec{\alpha})$  represents the prediction of  $x$  at time  $t + dt$  given the state  $(x(t), y(t))$  at time  $t$  and parameter vector  $\vec{\alpha}$ . In the simulations in this section, we used a naive Euler scheme  $x(t + dt) = x(t) + \dot{x}(t)dt$ , so that the short term prediction of model  $A$  depends linearly on  $\vec{\alpha}$ . Therefore the squared error is quadratic in  $\vec{\alpha}$  and can be minimized

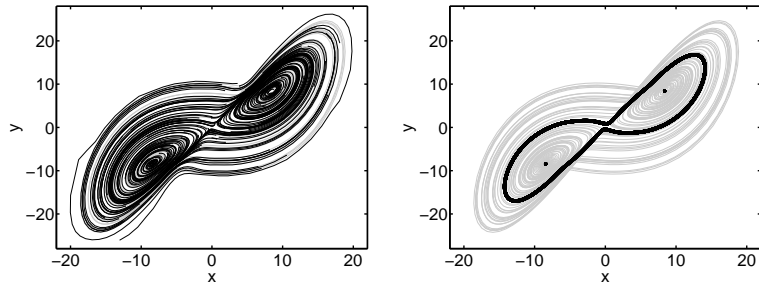


Figure 5: Model A: vector field and asymptotics. Black is model A. Grey is the assumed ground truth, projected on the  $x, y$  plane. The small dots in the left plot are points taken from the ground truth  $(x(t), y(t))$  at some time intervals. They served as starting points for short integrations in model A as a way to illustrate the vector field. In the right plot, model A is run a number of times for 20 time units- starting at different initial conditions, and then the last 2 time units were plotted. Data from runs that diverged are not plotted.

by standard linear algebra methods [2]. The result is displayed in figure 5. The short term predictions, i.e., the vector field, match the ground truth in general quite well. The asymptotic behavior, however, is completely different: model A has two point attractors and a limit cycle. This is to be expected, since two-dimensional continuous time systems cannot exhibit chaotic behavior. Finally, we noted that model A is unstable outside a small basin of attraction, i.e. model trajectories diverge. In a similar way, we assume model B has only dynamics in  $y_b$  and  $z_b$ . The model dynamics is assumed the form

$$\dot{y}_b = X(y_b, z_b)(\rho - z_b) - y_b \quad (26)$$

$$\dot{z}_b = X(y_b, z_b)y_b - \beta z_b \quad (27)$$

The function  $X$  is parametrized by a linear model and optimized for short term predictions in  $y$  and  $z$ , in the same way as  $Z$  in model A. In figure 6 the short term predictions, i.e., the vector field, is displayed, as well as the asymptotics. Again, we conclude that the vector field, match the ground truth in general quite well. The asymptotics seem to consist of many concentric limit cycles. In the plot, orbits seem to cross. This cannot happen in a 2 dimensional ordinary differential equation. Here it is due to the time discretization,  $dt = 0.01$ , combined with the Euler scheme for integration. Outside a basin of attraction, model B is unstable.

So, now we have two models of two different subsystems. In the following we explore the behavior of a supermodel of the full system by combining these two models. We considered a connecting model and a weighted average model. Since the only variable that is common in model A and model B is the variable  $y$ , the coupling is only in this variable. Optimization of coupling parameters, being either the connection coefficients or the weights, has been done by hand, by optimizing the attractor visually.

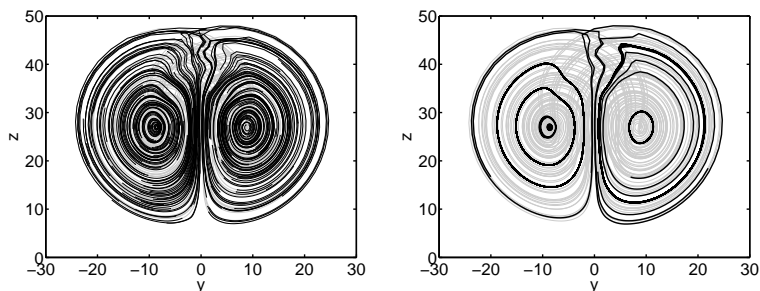


Figure 6: Model B. Vector field and asymptotics. Black is model B. Grey is ground truth, projected on the  $y, z$  plane. See previous figure for other explanations.

The first supermodel is obtained by introducing interacting terms, with connection coefficients  $C_{ab}$  and  $C_{ba}$ ,

$$\dot{x}_a = \sigma(y_a - x_a) \quad (28)$$

$$\dot{y}_a = x_a(\rho - Z(x_a, y_a)) - y_a - C_{ab}(y_b - y_a) \quad (29)$$

$$\dot{y}_b = X(y_b, z_b)(\rho - z_b) - y_b - C_{ba}(y_a - y_b) \quad (30)$$

$$\dot{z}_b = X(y_b, z_b)y_b - \beta z_b \quad (31)$$

the model output  $(x_s(t), y_s(t), z_s(t))$  is then obtained by setting

$$x_s(t) = x_a(t), y_s(t) = \frac{1}{2}(y_a(t) + y_b(t)), z_s(t) = z_b(t) \quad (32)$$

The optimal connection coefficients that visually optimizes the attractor of  $(x_s(t), y_s(t), z_s(t))$  were found to be  $C_{ab} = 6$  and  $C_{ba} = 10$ .

The alternative way of combining is via averaging.

$$\dot{x}_c = \sigma(y_c - x_c) \quad (33)$$

$$\dot{y}_c = w(x_c(\rho - Z(x_c, y_c)) - y_c) + (1 - w)(X(y_c, z_c)(\rho - z_c) - y_c) \quad (34)$$

$$\dot{z}_c = X(y_c, z_c)y_c - \beta z_c \quad (35)$$

The weight  $w$  corresponding with the connection coefficients is  $w = 0.625$ . This turned also out to be about optimal.

Both supermodels have 3-D behavior that looks like the ground truth butterfly (figure 7). Interestingly, both models seem rather stable: after about 20000 iterations, it is still on the attractor (with  $dt = 0.01$ , Euler scheme). However, the connected supermodel looks more similar to the ground truth butterfly dynamics in the way it jumps from one wing to the other. It seems richer, more chaotic than in the weighted supermodel, for example, the plots of the connected supermodel show filaments from the outside of one wing to the inner wheel of the other which seem to be absent in the weighted supermodel.

A possible explanation could be that a weighted supermodel could be understood as a connected supermodel with infinite connections. This

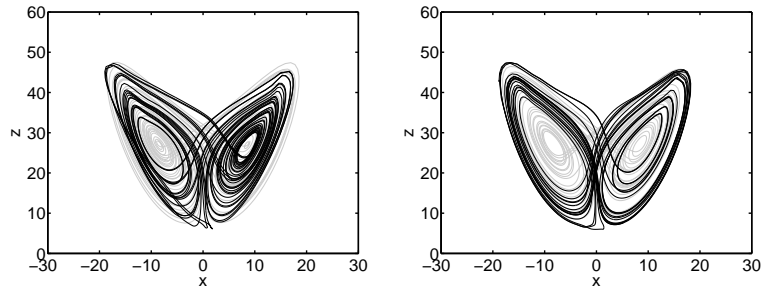


Figure 7: Trajectories of connected supermodel (left) and weighted supermodel (right). Ground truth is plotted in grey.

gives the weighted supermodel less flexibility because the models are in a way instantaneously synchronized and remain so over time. This may hinder transitions between regimes in the attractor. In connected supermodels, where the connections have finite values, synchronization is not immediate, and models can deviate from synchronization. In other words, models are allowed to deviate from the consensus state and follow more their own dynamics for a while. By doing so, the model could make an transition to another other regimes, and by the couplings, other models may follow, which then results in a regime transition of the consensus state.

## 7 Discussion

In this chapter we studied the networks of linearly coupled nonlinear oscillators in the limit of large connections. The motivation for this study were findings in the recently proposed supermodels. Supermodels are dynamically coupled ensembles of models. The connections are optimized, so that the supermodel fits to a data set of observations. Earlier work demonstrated the viability of this approach on low dimensional systems. The connections that were found with this procedure, were typically quite large. This was the motivation to analyze its behaviour theoretically in the large connection limit. Similar to earlier results in coupled systems, it was theoretically argued that the the models in a supermodel synchronizes and that the dynamics of the synchronized state is a weighted average of the imperfect model dynamics. We verified numerically that the supermodel solutions are indeed well approximated by the weighted average approximation. With this analysis, the multiple local optima in connection space that has been found earlier can be better understood. Also the fact that the Lorenz 63 supermodel reported in [1] was able to correctly simulate the the response to parameter change without the need of retraining of the connection coefficients has now a straightforward explanation.

One could consider doing weighted averages of model components from the start. This leads to the weighted supermodel, which is described



in [14] and elsewhere in this book. In practice, weighted supermodels seem to have several advantages. The most important one is the availability of scalable learning schemes. Other advantages are interpretability and transparency, the elimination of equivalent solutions, and possibly performance guarantees (see e.g. ensemble methods in [2]).

In the first approaches to supermodeling, it has been assumed that all models have the same dimension and that each variable in each model is coupled to similar variables in other models. In reality, this may not hold. Partial coupling could be an option if models are too complex for a full coupling. For instance if real-world climate models are to be coupled, the additional overhead to have all variables exchange information will probably be infeasible. A second reason for partial coupling is that different models may have variables that have different interpretations. In this case, one may only want to couple variables with the same interpretation. In this work we argued that also in partially coupled models, large connections leads to averaging of the coupled variables. We verified with simulations in a network of ten partially coupled Lorenz 63 oscillators.

We also simulated supermodels composed of incomplete models. In these simulations, two 2-D models were constructed to model different aspects of the assumed 3-D Lorenz 63 ground truth. By coupling, either through connecting or weighting, the incomplete models were able to complement each other. Both supermodels provided a much better description of the assumed reality than any a posteriori average of the individual models could do. Both supermodels showed a rather complex chaotic butterfly shaped attractor, while the 2-D models only showed simple periodic orbits and point attractors.

In the case where models and ground truth had all the same dimension, there seemed hardly any difference in the performance of connected and weighted supermodels. However, in the incomplete model case, the connected supermodel seemed to be able to better reproduce the butterfly shaped attractor of the assumed Lorenz 63 ground truth than the weighted supermodel. In particular, the dynamics of the weighted supermodel seemed to be a bit less chaotic than the connected supermodel and the ground truth. A possible explanation could be that a weighted supermodel could be understood as a connected supermodel with infinite connections. This gives the weighted supermodel less flexibility because the models are in a way instantaneously synchronized and remain so over time. This may hinder transitions between regimes in the attractor. In connected supermodels, where the connections have finite values, synchronization is not immediate, and models can deviate from synchronization. In other words, models are allowed to deviate from the consensus state and follow more their own dynamics for a while. By doing so, the model could make an transition to another other regimes, and by the couplings, other models may follow, which then results in a regime transition of the consensus state. It would be interesting to find out if this difference occurs in general when incomplete models are coupled. Also it would be interesting to have a more mathematical precise description of this process, which may help to better understand the possible advantages and disadvantages of having finite connections in a supermodel.

There are many subjects for future research. Most of the results pre-

sented in this chapter are explorative and should be made mathematically more precise. The models addressed in this section were all three-dimensional Lorenz 63 systems. For realistic supermodeling application, systems of much higher dimension should be considered and the scalability of the different aspects of the approach should be understood. This chapter considered networks of nonlinear oscillators with large connections with in particular the supermodel application in mind. It would also be interesting to find other applications for the theory.

**acknowledgement** This work has been supported by FP7 FET Open Grant # 266722 (SUMO project).

## References

- [1] van den Berge, L.A., Selten, F.M., Wiegerinck, W., Duane, G.S.: A multi-model ensemble method that combines imperfect models through learning. *Earth System Dynamics* **2**(1), 161–177 (2011)
- [2] Bishop, C.: *Pattern recognition and machine learning*. Springer (2006)
- [3] Duane, G., Tribbia, J., Kirtman, B.: Consensus on Long-Range Prediction by Adaptive Synchronization of Models. In: D. N. Arabelos & C. C. Tscherning (ed.) *EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts*, vol. 11, p. 13324 (2009)
- [4] Duane, G., Tribbia, J., Weiss, J., et al.: Synchronicity in predictive modelling: a new view of data assimilation. *Nonlinear processes in Geophysics* **13**(6), 601–612 (2006)
- [5] Kocarev, L., Shang, A., Chua, L.: Transition in dynamical regimes by driving: A unified method of control and synchronization of chaos. *International Journal of Bifurcation and Chaos* **3**(3), 479–483 (1993)
- [6] Lorenz, E.: Deterministic nonperiodic flow. *Atmos J Sci* **20**, 130–141 (1963)
- [7] Lorenz, E.: Irregularity: a fundamental property of the atmosphere\*. *Tellus A* **36**(2), 98–110 (1984)
- [8] Olfati-Saber, R., Fax, J., Murray, R.: Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE* **95**(1), 215–233 (2007)
- [9] Pecora, L., Carroll, T.: Synchronization in chaotic systems. *Physical review letters* **64**(8), 821–824 (1990)
- [10] Pikovsky, A., Rosenblum, M., Kurths, J.: *Synchronization: A universal concept in nonlinear sciences, Cambridge Nonlinear Science Series*, vol. 12. Cambridge University Press (2003)
- [11] Rössler, O.: An equation for continuous chaos. *Physics Letters A* **57**(5), 397–398 (1976)

- [12] Sun, J., Bollt, E., Nishikawa, T.: Master stability functions for coupled nearly identical dynamical systems. *Europhysics Letters* **85**, 60,011 (2009)
- [13] Tebaldi, C., Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **365**(1857), 2053 (2007)
- [14] Wiegnerinck, W., Selten, F.: Supermodeling: Combining imperfect models through learning. In: *NIPS Workshop on Machine Learning for Sustainability (MLSUST)* (2011). URL <http://people.csail.mit.edu/kolter/mlsust11/lib/exe/fetch.php?media=wiegnerinck-mlsust.pdf>
- [15] Yang, S., Baker, D., Li, H., Cordes, K., Huff, M., Nagpal, G., Okereke, E., Villafañe, J., Kalnay, E., Duane, G.: Data assimilation as synchronization of truth and model: Experiments with the three-variable lorenz system. *Journal of the atmospheric sciences* **63**(9), 2340–2354 (2006)
- [16] Yu, W., Chen, G., Cao, M., Kurths, J.: Second-order consensus for multiagent systems with directed topologies and nonlinear dynamics. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* **40**(3), 881–891 (2010)