SUMO - Supermodeling by combining imperfect models

Workpackage 3: Year 1

Gregory S. Duane, Lasko Basnarkov

November 7, 2011
In this WP we report on supermodeling of systems modeled with partial differential equations with low complexity. Different learning strategies and algorithms were developed and tested on the Kuramoto-Sivashinsky system. The research was generally performed as was stated in the tasks 3.1 and 3.2:

Task 3.1 Development of learning strategies for PDE supermodels: Both learning algorithms introduced in WP2 will be applied to various simple systems of PDEs, with accent set on Kuramoto-Sivashinsky system (KS). KS is a natural choice for preliminary investigation on the subject, due to the following reasons: in the past the system was already used as a natural bridge between ODEs and PDEs; it is used to represent turbulent fluid dynamics, therefore is relevant to geophysical situations. Galerkin projections of different orders will give rise to hierarchy of systems in which supermodeling approach can be explored. Both the number of connections as well as the number of interdependently trained connections can be varied as the order increases. Previous results obtained on simple parameter estimation models will be further analyzed here in regards to possible lessons about performance vs. density of required connections for analyzed algorithms.

Task 3.2 Comparison of learning algorithms for connections in PDE supermodels: The conclusions emerging from WP2 about the relative merits of different learning approaches will be re-examined in PDEs context. A comparison of the spatial density of connections required in each approach will be examined. Special attention will be paid to the issue of locally optimal supermodel in the PDE context, with accent on the system dimensionality. The software frame for supermodeling that will later be applied to climate models, will be constructed at this stage.

Although we have generally followed the plan, the investigation of supermodeling of PDE’s using Galerkin projections is ongoing research.
1 Introduction

Before extending the preliminary demonstrations of supermodeling with ODE’s to climate models of varying complexity, it is useful to extend them generally to systems defined by partial differential equations (PDE’s). In combining different PDE systems to form a supermodel, the main differences from the simpler ODE supermodels lie in the far greater number of possible connections and the possible impact on training requirements. On the other hand, the basic translational symmetry of such systems is expected to facilitate the task, greatly reducing the number of connections that need to be independently trained. The issue of local vs. global optimality in the connection scheme is expected to arise in PDE supermodels, as with the simpler systems, if the connection schemes are too severely restricted.

As with ODE systems, the phenomenon of synchronization between a pair of PDE systems is key to their use in the larger collection of systems that forms the supermodel. Because the models are different from each other and all are even more different from reality, the relevant notion is that of generalized synchronization in which the correspondence between states of different systems, while fixed, is not the identity (Rulkov et al., 1995). But for the supermodel to be useful, such correspondences must either be known explicitly or be reasonably close to the identity. Then the main question that arises in structuring a supermodel is about the required spatial density of connections between systems, a question that can be addressed by studying the generalized synchronization of a single pair of systems.

Investigations in WP3 have examined synchronization and supermodeling with the Kuramoto-Sivashinsky (KS) model. As a well-studied PDE in one space and one time dimension, KS provides a testbed in which issues pertaining to more complex models can be readily investigated. In particular, we have focused on generalized synchronization, on the density of connections required for synchronization, and on learning strategies to establish optimal connection weights. Fourier-space representations of KS (Galerkin projections) define a useful relationship with ODE systems and ODE supermodels.

2 Structures for Synchronization of KS systems

2.1 Generalized synchronization of KS systems

2.1.1 Model

We considered a generalized version (Kudryashov, 1990) of the Kuramoto-Sivashinsky PDE (Kuramoto and Tsuzuki, 1976; Sivashinsky, 1977) given by

\[ u_t = -u_{xxxx} - \alpha_u u_{xx} - u_{xx} - 2uu_x, \]  

(1)

where \( \alpha_u \) is a parameter and subscripts \( x \) and \( t \) denote partial derivatives in space and time respectively. Its solution \( u(x,t) \) evolves in time \( t \) confined in the interval \( [0,L] \) and subjected to periodic boundary conditions \( u(0,t) = u(L,t) \). We use it as a master that drives another system defined by a PDE with a different parameter value \( \alpha_v \).
Previous studies of synchronization of spatially extended chaotic systems have shown that synchronization of two identical systems ($\alpha_u = \alpha_v$) is possible even when the coupling between them is sparse, or more precisely, when the coupling points can be taken at some distance lower than some threshold value (Kocarev et al., 1997a,b). We chose to take the coupling points at equal distance because that choice was found to be nearly optimal (Junge et al., 1999). The dynamics of the driven system $v(x,t)$ thus reads

$$v_t = -v_{xxxx} - \alpha_v v_{xxx} - v_{xx} - 2vv_x + K[u(x) - v(x)]f(x),$$

where $K$ denotes the coupling strength, and the coupling selection function $f(x)$ is

$$f(x) = \begin{cases} 1, & x = nd - d/2; \\ 0, & \text{otherwise.} \end{cases}$$

In the last equation $d$ is the distance between the coupling points that is a multiple of the integration step $\Delta x$ and $n$ is a positive integer. A measure of the quality of synchronization is the average squared distance between corresponding points or the so-called synchronization error

$$e_{u,v}(t) = \frac{1}{L} \int_0^L dx \left[ u(x,t) - v(x,t) \right]^2.$$  

In this work we use its discretized version

$$e_{u,v}(t) = \frac{1}{N} \sum_{k=0}^{N-1} \left[ u(k\Delta x,t) - v(k\Delta x,t) \right]^2,$$

where the averaging is performed for all $N$ discretization points. It was previously obtained that two identical KS models (with $\alpha_u = \alpha_v = 0$) synchronize identically $e_{u,v}(t) \to 0$ as $t \to \infty$ (Junge et al., 1999). We have verified the same type of synchronization in the case of generalized KS, i.e. when the parameters $\alpha_u = \alpha_v > 0$. However, in reality one should expect that the two systems have at least minor differences which can be exemplified by a difference of the parameter values $\alpha_u \neq \alpha_v$. Then the quality of the synchronization weakens and generalized synchronization replaces the complete form. The synchronization error (5) is non-zero and is small if the parameter mismatch $|\alpha_u - \alpha_v|$ is small and coupling strength $K$ large enough. With increase of the difference $|\alpha_u - \alpha_v|$ or by weakening the coupling strength $K$, the synchronization error becomes comparable with the values of $u$ and $v$.

Another estimator of generalized synchronization is based on the auxiliary systems method (Abarbanel et al., 1996). In that case the master drives two identical slaves $v$ and $w$ with $\alpha_v = \alpha_w$ and one calculates the distance (5) between the slaves. The master is considered to be synchronized with the slave in the generalized sense if $e_{v,w} \to 0$ when $t \to \infty$. We have found that to be the case for some parameter values with $\alpha_u \neq \alpha_v = \alpha_w$. It appeared even in the extreme case when $\alpha_v = \alpha_w = 0$. That means that the master is synchronizing even a reduced model. The reduction can sometimes be interpreted to mean the absence of some physical mechanism.
2.1.2 Results

We have integrated the coupled PDE’s numerically with the explicit Euler method. In all simulations the time step was taken $\Delta t = 0.0002$ and in most of them the spatial resolution was set to $\Delta x = 0.25$. The initial values $u(x, 0)$, $v(x, 0)$ and $w(x, 0)$ were taken randomly and independently from the uniform distribution in the interval $[-0.5, 0.5]$. The coupling was turned on after transient period $t = 200$. The synchronization error between the master and the slave (5) was calculated as an average value of interval with length 500, after adaptation period of 500.

We first analyzed the influence of the value of the parameter $\alpha$ on the dynamics of the Kuramoto-Sivashinsky PDE. It appeared that larger values of the parameter lead the system to state of regular dynamics with traveling waves which are shown in fig. 1c. For smaller $\alpha$ the system is in a chaotic regime. Its existence persists for rather long time–much longer than the time scale needed for generalized synchronization to settle.

That the behavior is chaotic can be verified by finding a positive value of the largest Lyapunov exponent $\Lambda$. In this case $\Lambda$ is the mean rate of exponential growth of the error function (5) between two identical uncoupled systems with close initial conditions $e_{u,v}(0) < \epsilon$

$$\Lambda = \lim_{t \to \infty} \frac{1}{t} \ln \frac{e_{u,v}(t)}{e_{u,v}(0)}$$  \hspace{1cm} (6)

Figure 1 shows the time evolution of the solution $u(x, t)$ after a transition period of 5000 time units. For the upper two panels the Lyapunov exponent $\Lambda$ has a positive value confirming chaotic dynamics, while for the bottom panel it is zero as is expected for regular behavior.

We have examined the dependence of the synchronization error between the driver and its simpler slave (with $\alpha_v = 0$) on the coupling strength $K$. The distance between the coupling points $d$ was taken as a parameter. As expected, stronger coupling forces the slave closer to the master as can be seen from the decreasing synchronization error shown in fig. 2. It was found that when the parameter of the master $\alpha_u$ has a larger value the synchronization is weaker because the difference between the two systems is greater. Also, the sparser the coupling the harder is to synchronize the systems, as seen when the distance $d = 2$, as compared to $d = 1$ in the same figure. For large distance between the coupling points, $d = 5$, the systems behave as if they are uncoupled. The tendency of the synchronization error is to stabilize at some non-zero value as $K$ increases.

To verify the occurrence of generalized synchronization (GS) we coupled two systems with $\alpha_v = \alpha_w = 0$ with the driver. GS was assumed to be established if the error obtained was vanishingly small. Figure 3 shows the synchronization error between the slaves as a function of time for three particular cases. In the first case, with $\alpha_u = 0.001$, it can be seen that even for small coupling strength $K = 1$, when the distance between the coupling points is small enough, the synchronization error is decreasing exponentially. Then it settles at a value near the computer precision and the fluctuations are small. Clearly, the GS exists even when one cannot infer it from observation of the distance between the master and the slave (see fig. 2). In the second case the distance $d = 5$ is
Figure 1: Snapshot of the evolution of the generalized Kuramoto-Sivashinsky equation for various values of the parameter $\alpha_u$. From top to bottom panel the values are 0.001, 0.01 and 0.1 respectively and $L = 50$. 
Figure 2: Synchronization error between the master and the slave [Eq. (5)] as a function of the coupling strength $K$, for various distances between the coupling points.

Table 1: Type of generalized synchronization as a function of the number of coupling points for $L = 50$ and $\alpha_u = 0.01$.

<table>
<thead>
<tr>
<th>coupling strength</th>
<th>sync</th>
<th>intermittency</th>
<th>no-sync</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K \to \infty$</td>
<td>6 or more</td>
<td>4, 5</td>
<td>3 or less</td>
</tr>
<tr>
<td>$K = 10$</td>
<td>6 or more</td>
<td>5</td>
<td>4 or less</td>
</tr>
<tr>
<td>$K = 5$</td>
<td>8 or more, 6</td>
<td>5, 7</td>
<td>4 or less</td>
</tr>
</tbody>
</table>

larger, as is the coupling, $K = 4$, but the coupling is not strong enough to give GS. The parameter value in this and in the third case, $\alpha = 0.01$, gives larger synchronization error. Then the existence of GS cannot be inferred with the auxiliary system method because the fluctuations are large (although the fluctuating error remains small, suggesting some form of intermittent generalized synchronization – see below). The third case is similar to the second but with stronger coupling $K = 5$. The GS seems to be present although the fluctuations are more pronounced than in the first case.

An important issue in coupling of spatially extended systems is determination of the density of connections needed for synchronization. We have found numerically the minimal number of equally spaced coupling points that still permits generalized synchronization (small nonuniformities in the spacing had to be introduced so that each coupling distance is a multiple of the integration step $\Delta x$). The coupling strength $K$ was allowed to vary. The case $K \to \infty$ corresponds to complete replacement (cf. (Pecora and Carroll, 1990)) of the values of the slave solution with corresponding values of the master $w(x_c, t) = v(x_c, t) = u(x_c, t)$ (see fig. 4).
Figure 3: Synchronization error between both slaves as a function of the time. The coupling is switched on at time zero.

Figure 4: Synchronization error between both slaves for the case of complete replacement $K \to \infty$. The number of connection points is as indicated in the legend. The other parameters are $L = 50$ and $\alpha = 0.01$. 
We found that the break-up of generalized synchronization is not sharp. Instead there exists an interval in the number of coupling points where intermittency appears. Intermittent synchronization is a situation when the periods of synchronization are interrupted with periods of desynchronization. Such periods are seen in Fig. 4 for $d = 4$ (blue) and $d = 5$ (green). The intermittency window is itself broken by cases of synchronization for some number of connections. For example for $L = 50$, when $K = 5$ for 6 coupling points there is synchronization, although there is intermittency for 5 or 7 coupling points. The appearance of intermittency was previously found in studies of complete synchronization of identical PDE’s (Junge et al., 1999). The results for one particular example ($L = 50$ and $\alpha_u = 0.01$) are summarized in table 1. From visual inspection of the profile of the solution shown in fig. 1 and from the results in table 1 one can conclude that the distance between coupling points needed for synchronization is approximately equal to the wavelength. We have verified that relation also for larger system size $L = 100$.

Central to strategies for optimal coupling will be coherent structures found both in the model and in reality. It was hypothesized in (Duane, 2009) that the existence of such internal coherence is both necessary and sufficient for synchronizability with an external system. The combinations of variables that must be coupled to give synchronization are informed by the patterns of internal coherence. Here, the “fingers” found in solutions of the ordinary KS equation form traveling waves when a third-derivative term is added. That one needs to sample reality only at their wavelength, which corresponds to the width of a finger, is in accord with the general hypothesis - the entire structure is synchronized with its counterpart when only one variable within it is coupled because of internal synchronization within the structure.

2.2 Coupling in Fourier space

As a one-dimensional problem, the KS PDE can be appropriately solved by representing the solution with Fourier series. Then solving the PDE is replaced by the task of solving a system of (infinitely many) ODE’s of the Fourier components. For practical purposes the series is truncated to some reasonable number of components. For the KS PDE, it is known from the literature that the solution of the truncated Fourier series can depend on the number of Fourier components kept. For example, for truncation at 15 vs. 16 Fourier components the solution can go to different asymptotic states (Cvitanović et al., 200). Coupling of the Fourier components instead of the solution $u(x,t)$ at certain points in physical space is another alternative to achieve synchronization. Nudging the corresponding low-order components toward each other leads to similar large-scale structures of the models. To couple spectral components while avoiding errors because of the truncation of the number of components one should perform integration in the physical space. At every time step the Fourier spectrum is calculated, the components are adjusted accordingly and then the new values are fed back in the physical space integrator. Precisely, for master-slave synchronization of two identical KS systems the procedure goes as follows: At every time step the lowest part of the Fourier spectrum is calculated for both KS systems. Next, the slave Fourier components $f_k^s$ are nudged
toward the master’s $f_k^m$

$$f_k^m(t + \Delta t) = f_k^m(t) + C[f_k^m(t) - f_k^s(t)].$$

(7)

where $C$ is the coupling strength. (For simplicity, we assume the nudging and the dynamical evolution occur at different, alternating time steps.) The role of coupling is to adjust or correct the dynamics. We have tested the extreme coupling case $C \to \infty$ which is identical to replacing the lowest Fourier components (up to order $K$) with their corrected values. One can introduce real valued functions that represent the inverse Fourier transform of the lowest Fourier components

$$u_s^F(k\Delta x, t) = \frac{1}{M} \sum_{j=-K}^{K} f_s^k(t)e^{ij\Delta x};$$

$$u_s^F(k\Delta x, t + \Delta t) = \frac{1}{M} \sum_{j=-K}^{K} f_s^k(t + \Delta t)e^{ij\Delta x}.$$  

The normalizing factor $M$ is the number of discretization points (samples in the physical space). Then in the RHS of the slave’s equation (2), at a nudging time step, one uses the following values of the solution

$$u(t) + u_s^F(k\Delta x, t + \Delta t) - u_s^F(k\Delta x, t),$$

(9)

which simply replaces the lower part of Fourier representation with the updated value. By applying the procedure described above one can obtain complete synchronization of two identical KS systems by coupling only Fourier components up to order $K = 4$. Omitting the average $f_0$, the systems can not synchronize completely. This can be seen by using Fourier expansion in the defining equation of KS (1)–without the coupling–and obtaining that the average is constant $\dot{f}_0 = 0$. So, the lowest component is also uncoupled from the higher components and removing the direct communication between the two averages leaves them independent. Synchronization of the KS systems by coupling in Fourier space allows an interpretation in terms of master and slave subspaces of a single KS system, defined in terms of lower and higher Fourier components, resp., and perhaps in terms of the theory of inertial manifolds (see Appendix).

3 KS Supermodels and Learning Strategies

3.1 Learning algorithms for supermodeling KS systems

To form a KS supermodel, we couple three ordinary KS models with different values of $\nu$. The choice $\nu = 1$ is taken as truth that we will try to model with $N$ individual models coupled in a supermodel. The viscosities of the individual models are assumed different from each other $\nu_i \neq \nu_j$, and from the truth’s $\nu_i \neq \nu$. The individual models are coupled to each other and driven by the truth, which means they influence mutually
and are subjected to forcing by the truth which evolves freely. Their dynamics assumes the form

$$\frac{\partial u_i}{\partial t} = -\nu_i \frac{\partial^4 u_i}{\partial x^4} - \frac{\partial^2 u_i}{\partial x^2} - 2u_i \frac{\partial u_i}{\partial x} + f(x) \left[ \sum_{j=1}^{N} C_{j,i}(u_j - u_i) + K(u - u_i) \right].$$

(10)

In the equation above $C_{j,i}$ is the coefficient expressing the strength of influence of the model $j$ to $i$, and $K$ is the strength of the coupling of the truth to the supermodel. The function $f(x)$ is determining the points in space $x$ where the coupling is applied and is given by

$$f(x) = \begin{cases} 1, & x = md - d/2 \\ 0, & x \neq md - d/2 \end{cases},$$

(11)

where $d$ is the distance between the coupling points. We considered uniform spacing of coupling points across $[0,L]$, because such choice was found to be nearly optimal in synchronization of identical PDE’s(Junge et al., 1999).

The quality of synchronization is estimated by the synchronization error which is simply the temporal average of the squared distance between the truth and the simple average of the individual models

$$e(t) = \frac{1}{L} \int_{0}^{L} \int_{0}^{L} dx \left[ u(x,t) - \frac{1}{N} \sum_{i=1}^{N} u_i(x,t) \right]^2. \tag{12}$$

Because we know the solution only at discrete set of points we use discrete version of the integral above

$$e(t) = \frac{1}{M} \sum_{k=0}^{M-1} \left[ u(k\Delta, t) - \frac{1}{N} \sum_{i=1}^{N} u_i(k\Delta, t) \right]^2, \tag{13}$$

where $\Delta = L/M$ is the integration step.

Integration of the PDEs is done with the simple Euler scheme which assumes discretization of the physical space. Then the PDE integration problem is reduced to integration of a system of ODEs. For example at one particular discretization point
located at x, the corresponding ODE is:

\[
\frac{du_i(x)}{dt} = -\nu_i\left(u_i(x + 2\Delta) - 4u_i(x + \Delta) + 6u_i(x) - 4u_i(x - \Delta) + u_i(x - 2\Delta)\right) \Delta^4 - \frac{u_i(x + \Delta) - 2u_i(x) + u_i(x - \Delta)}{\Delta^2} \nonumber \\
-2u_i(x) \frac{u_i(x + \Delta) - u_i(x - \Delta)}{2\Delta} + \sum_{j=1}^{N} C_{j,i}(u_j(x) - u_i(x)) + K(u(x) - u_i(x)) \tag{14}
\]

The key issue in supermodeling is determination of the connection coefficients \(C_{j,i}\). Here we have applied a few different strategies. Two of them are deterministic and two are stochastic. The first one of the deterministic algorithms considers equal strength \(C_{j,i} = C\), while the second assumes an adaptation law given by

\[
\frac{dC_{j,i}}{dt} = \alpha \frac{1}{L} \int_0^L dx \left[ u(x) - \frac{1}{N} \sum_{i=1}^{N} u_i(x) \right] [u_j(x) - u_i(x)], \tag{15}
\]

where \(\alpha\) is a parameter determining the adaptation rate (Duane, 2011). The law (15) is based on previous work about determination of the unknown parameters of ODE by means of synchronization (Parlitz, 1996; Duane et al., 2007). It has also been applied to a rudimentary PDE supermodel of geophysical interest, as described in the Appendix. Intuitively, a time integral of (15) states that connections between models are strengthened or weakened in proportion to the correlation between the tendencies induced by those connections and error in the variables thus computed.

Here we present supermodeling with three individual models \((N = 3)\). The viscosity parameter of the truth is 1 as stated above and those of the individual models are \(\nu_1 = 0.75\), \(\nu_2 = 1.1\) and \(\nu_3 = 1.2\). The coupling between the individual models was performed at every integration point while the influence from the master is applied at different densities ranging from every integration point up to every tenth. The general conclusion is that when coupling coefficients \(C_{j,i}\) have large values compared to that of the master \(K\), the supermodel performs better than the average of the individual models—when \(C_{j,i} = 0\). Also, by using adaptation law (15) the synchronization error (13) is even smaller but the adaptation does not converge which means that the coefficients \(C_{j,i}\) continuously change their values even for rather long simulation time (see figs. 5 and 6). However, the coupling coefficients quickly settle to some values and then their drift is very small in comparison to their values. The experiments with the adaptation law (15) were repeated ten times for different random initializations of the individual models \(u_i(x,0)\). The values of the connection coefficients \(C_{i,j}(200)\) obtained after 200 time units of adaptation from the initial identical values \(C_{i,j}(0) = 10\) were found to
be nearly identical. That suggests that possibly those values correspond to the global minimum.

At an extreme value of coupling strength (not shown) the values of $u$ of the individual models are replaced by corresponding values of the truth. Then the synchronization error is smaller compared to that when finite $K$ is assumed. Also, the synchronization break up happens at larger coupling distance when complete replacement is made instead of nudging.

The stochastic algorithms wander around the vector space of the coupling coefficients randomly in search of combination of coupling strengths that corresponds to smallest synchronization error. They are based on the following idea: at every step a perturbation of the coupling coefficients is made $C_{i,j} + \delta C_{i,j}$. Then starting with the same initial conditions, the truth and the supermodel are integrated forward in time for some transition time $t_{tr}$, and then synchronization error is calculated for some other integration time $t_{sync}$. We found that both times can have values as small as 10. If the synchronization error is smaller than for the unperturbed case the connection coefficients are updated with the perturbation; otherwise the perturbation is discarded and new one is made. In one of the algorithms the perturbation had constant magnitude $\| \delta C_{i,j} \| = \Delta = \text{const}$, while in the second at every step its magnitude was chosen randomly from exponential distribution $p(\lambda) = \lambda e^{-\lambda \Delta}$. This was done to get out of possibly local minima in which the algorithm can get stuck. Also the algorithms were implemented with a Metropolis-like modification which accepted the perturbation of the coefficients–although it gives larger synchronization error–but with some probability. The stochastic algorithms did not find any better combination in the sense of smaller synchronization error. However the connection coefficients had different values from those found by the synchronization-based algorithm. We can conclude that at least for the supermodeling of Kuramoto-Sivashinsky model the adaptive law is at least approximately as good as the searching algorithms. But the adaptive law is computationally more effective because it calculates the changes of the coefficients at every integration step. The searching algorithms need finite integration of the equations of motion to test whether the synchronization error is smaller or not.

3.2 Computational costs of the learning algorithms

The larger computational cost of the searching algorithms arises as a result of the impossibility of determining the form of the quality-of-solution function \textit{a priori}. In this case it is the synchronization error (12). Its calculation at every point in the space of the connection coefficients requires running a simulation of the KS systems. It is obtained as a temporal and spatial average of the distance between the truth and the mean of the models, or time average of errors defined by (4). The spatial average has complexity of order of the number of discretization points $N$. The temporal average is of the order of the product of the number of time steps in a time unit $1/\Delta t$ and integration time $t_{sync}$. When the adaptation rule (15) is implemented, the change of the connection coefficients depends only on sums which have computational complexity similar to that of the synchronization error (12) which is of order $N$. Thus the simple rule (15) is much more
Figure 5: Evolution of coupling coefficients according to the law (15). The coupling is implemented at every integration point. The viscosity parameters are $\nu_1 = 0.75; \nu_2 = 1.1; \nu_3 = 1.2$ and learning rate has value $a = 50$. The system length is $L = 50$ in all cases considered in this text.

Figure 6: Time averaged synchronization error (12) versus master coupling strength $K$. The transition period is 500 time units and $e$ is obtained as time average for period 200. The other parameters are as in fig. 5.
cost-effective than the other algorithms. We have verified numerically that the quality of the results obtained with (15) are close to those based on a more global search with regard to synchronization. One should check whether such a result is obtained for other measures of the quality of supermodeling: comparison of averages or autocorrelations to determine whether the supermodel has captured the shape of the true attractor, even in the absence of nudging, as would be relevant for climate projection (van den Berge et al., 2010). If such a generalization is possible, then that result could save a lot of computational power, especially for the supermodeling of the real climate. The time integrations for that application should be performed for at least a few years in order to smooth the seasonal effects and year-to-year fluctuations.

4 Software framework for coupling of PDEs and Climate Models

We have secured the involvement of the National Center for Atmospheric Research (NCAR) in modifying their Data Assimilation Research Testbed (DART) as one track toward a generic software framework for combining alternative PDE models of the same objective process to form a supermodel. DART allows “identical twin” experiments with a hierarchy of models, ranging from 3-variable Lorenz systems to quasigeostrophic models to full climate models. In each case, one model plays the role of reality, feeding data to a copy of the model, or modified version thereof. The software needs to be modified so that models can assimilate data from one another, as well as from reality. NCAR personnel are providing assistance with software modifications to 1) replace ensemble Kalman filtering with the simple nudging form used in supermodeling; 2) replace the use of the ensemble as a way to determine background error with a new role as an interactive suite of different models; and 3) incorporate the various machine learning approaches to determine the nudging coefficients. We will retain the flexibility of DART in being able to easily substitute more complex models for simpler ones.

5 Summary

The work completed thus far in WP3 has contributed to a foundation for supermodeling with climate models in several ways. The relationships among the individual models, as well as the relationship of the supermodel to reality in the training stage, will have the form of near-identical generalized synchronization (Fenat and Solis-Perales, 2002). That is also the usual form of the relationship between a weather-prediction model and reality (Duane et al., 2006; Yang et al., 2006), allowing the state of one system to be inferred from the state of another, to a reasonable degree of approximation. (Generalized synchronization defined by a correspondence that is far from the identity (So et al., 2002; Yang and Hu, 2007) is typically not useful (Yang et al., 2006).) The required density of connections between models will be governed by the coherent structures (e.g. hurricanes, fronts, Rossby waves) that exist within the models separately. Questions that may arise about detailed methods to design a connection scheme based on such structures will lead
us to investigate refinements to the coupling of KS systems. Our use of the Fourier-space representation of KS leads to a formulation in terms of master- and slave subspaces that should be transferrable to climate models where such decompositions are familiar.

In regard to learning, we have demonstrated that the synchronization-based approach gives satisfactory results for PDEs with little or no degradation as compared to the extended approach based on costs of trajectories. The large difference in computation cost may make the former approach preferable. But the simple one-parameter problem we have examined does not lend itself readily to the study of non-global local optima and methods to escape them. The KS study will be re-visited as information about local optima becomes available from the study of more complex systems.

Finally, the collaboration around the use of the Data Assimilation Research Testbed (DART) will allow the ready construction of supermodels from both simple PDEs and from larger climate models.

**A Appendix: Application of Synchronization-Based Learning to a Geophysical PDE Supermodel**

A synchronization scheme has been used in previous work to fuse two models of the mid-latitude flow, one with forcing in the Atlantic sector and the other with forcing in the Pacific sector, to form a two-sector model (Duane and Tribbia, 2001). While a two-sector model is of course easily formulated, the construction from single-sector models illustrates the utility of the synchronization paradigm.

The uncoupled single-system model is given by the quasigeostrophic equation for potential vorticity $q$ in a two-layer reentrant channel on a $\beta$-plane

$$\frac{Dq_i}{Dt} \equiv \frac{\partial q_i}{\partial t} + J(\psi, q_i) = F_i + D_i$$  \hspace{1cm} (16)

where the layer $i=1, 2$, $\psi$ is streamfunction, and the Jacobian $J(\psi, \cdot) = \frac{\partial \psi}{\partial x} \frac{\partial \cdot}{\partial y} - \frac{\partial \psi}{\partial y} \frac{\partial \cdot}{\partial x}$ gives the advective contribution to the Lagrangian derivative $D/Dt$. The forcing $F$ is a relaxation term designed to induce a jet-like flow near the beginning of the channel: $F_i = \mu_0(q^*_i - q_i)$ for $q^*_i$ corresponding to the choice of $\psi^*$ shown in Fig. 7a. The dissipation terms $D$, boundary conditions, and other parameter values are given in (Duane and Tribbia, 2004).

We couple two models of the form (16), with $F_i = \mu_0(q^*_i - q_i)$, using a different $q^*_i$ in each model. The coupling here is in the advection terms of the two models. The configuration is given by

$$\frac{Dq^A}{Dt} + cJ(\psi^A, q^B - q^A) = F^A + D^A$$
$$\frac{Dq^B}{Dt} + cJ(\psi^B, q^A - q^B) = F^B + D^B$$  \hspace{1cm} (17)

with forcing terms defined in terms of their spectral components

$$F^A,B_k = (\mu_0 - \mu_c^k) [q^A,B^* - q^A,B]$$  \hspace{1cm} (18)
Figure 7: The forcing $\psi^*$ (a,b), and the evolving flow $\psi$ (c-f), in the parallel channel model with advective coupling (given by (17)), with $c = 1/2$, $\mu_0 = 8.8 \times 10^{-4} s^{-1}$, and other parameters as in (Duane and Tribbia, 2004), for the indicated numbers $n$ of time steps in a numerical integration. One time step is $\approx 0.4$ hr. Near-identical synchronization occurs by the last time step shown (e,f). The synchronized flows closely approximate the flow in a model with forcing $\tilde{q}^*$ (corresponding to $\tilde{q}^* = (qA^* + qB^*)/2$) describing jets in both sectors (g). (Boxed regions were used in (Duane and Tribbia, 2004) to study the zonal/blocked index cycle.)
Figure 8: Estimation of the inter-model coupling according to (19), finding \( c \to 0.5 \), its “true” value as the two models synchronize with each other and with “truth” (not shown in Fig. 7).

where \( q_k \) is the wave number \( k \) spectral component of \( q \) (suppressing the layer index \( i \)). The coefficients \( \mu_k \) are chosen so as to couple only the medium scales as described in detail in (Duane and Tribbia, 2004). The flow fields in the coupled channels governed by (17) are found to synchronize, regardless of differences in initial conditions, as seen in Fig. 7. The correspondence is not exactly the identity, because of the difference in forcings, but is an instance of generalized synchronization. Likewise here, at \( c = 1/2 \) we have \( \psi^A \approx \psi^B \). That the smallest scales need not be coupled is thought to reflect the existence of an inertial manifold (Temam, 1988) on which they are slaved.

The existence of (approximate) inertial manifolds for forced-dissipative systems is the theoretical underpinning of the synchronization-based approach.

It is readily seen that the average \( \hat{q} = (q^A + q^B)/2 \) of the solutions of (17a) and (17b), for strong coupling \( c = 1/2 \), is the solution of a model with the average forcing, that is, of a model with \( \hat{q}^* = (q^A + q^B^*)/2 \), describing two jets, as in Fig. 7g. The flow in either channel \( \psi^A \approx \psi^B \) approximates the flow \( \hat{\psi} \) in a channel with two jets. As the coupling between the channels is increased from \( c = 0 \) to \( c = 1/2 \), the dynamics of each channel changes so as to incorporate an approximate, “virtual” counterpart of the dynamics of the sector that is forced in the other channel.

The desired value \( c = 1/2 \) can be obtained by adapting to a real (two-jet) data set \( q_{obs}(x, t) \) according to

\[
\frac{dc}{dt} = a \int d\mathbf{x}[J(\psi^A, q^B - q^A)(q^A - q_{obs}) + J(\psi^B, q^A - q^B)(q^B - q_{obs})] \tag{19}
\]

as was indeed tested, with results shown in Fig. 8. The slowness of the convergence and the instability of the resulting state suggest the need for an improved learning algorithm, although the synchronization-based algorithm is simple and versatile.


