

## Synchronization of Alternative Models in a Supermodel and the Learning of Critical Behavior

GREGORY S. DUANE<sup>a</sup> AND MAO-LIN SHEN<sup>b</sup>

<sup>a</sup> University of Colorado Boulder, Boulder, Colorado

<sup>b</sup> Geophysical Institute, University of Bergen, Bergen, Norway

(Manuscript received 20 May 2022, in final form 17 February 2023, accepted 27 February 2023)

**ABSTRACT:** “Supermodeling” climate by allowing different models to assimilate data from one another in run time has been shown to give results superior to those of any one model and superior to any weighted average of model outputs. The only free parameters, connection strengths between corresponding variables in each pair of models, are determined using some form of machine learning. It is demonstrated that supermodeling succeeds because near critical states, interscale interactions are important but unresolved processes cannot be effectively represented diagnostically in any single parameterization scheme. In two examples, a pair of toy quasigeostrophic (QG) channel models of the midlatitudes and a pair of ECHAM5 models of the tropical Pacific atmosphere with a common ocean, supermodels dynamically combine parameterization schemes so as to capture criticality, associated critical structures, and the supporting scale interactions. The QG supermodeling scheme extends a previous configuration in which two such models synchronize with intermodel connections only between medium-scale components of the flow; here the connections are trained against a third “real” model. Intermittent blocking patterns characterize the critical behavior thus obtained, even where such patterns are missing in the constituent models. In the ECHAM-based climate supermodel, the corresponding critical structure is the single ITCZ pattern, a pattern that occurs in neither of the constituent models. For supermodels of both types, power spectra indicate enhanced interscale interactions in frequency or energy ranges of physical interest, in agreement with observed data, and supporting a generalized form of the self-organized criticality hypothesis.

**SIGNIFICANCE STATEMENT:** In a “supermodel” of Earth’s climate, alternative models (climate simulations), which differ in the way they represent processes on the smallest scales, are trained to exchange information as they run, adjusting to one another much as weather prediction models adjust to new observations. They form a consensus, capturing atmospheric behaviors that have eluded all the separate models. We demonstrate that simplified supermodels succeed, where no single approach can, by correctly representing *critical phenomena* involving sudden qualitative transitions, such as occur in El Niño events, that depend on interactions among atmospheric processes on many different scales in space and time. The correct reproduction of critical phenomena is vital both for predicting weather and for projecting the effects of climate change.

**KEYWORDS:** Blocking; ENSO; Adaptive models; Data assimilation; Model comparison; Subgrid-scale processes

### 1. Introduction

The dynamics of Earth’s climate system is to a large degree a story of interactions between processes on very different scales. It is commonly believed that there are fewer real degrees of freedom in the climate system than there are in a typical physical description because the smallest scales are somehow slaved. Weather prediction on the basis of severely limited observations is indeed possible because of such low dimensionality. If processes on the smallest physical scales in time or space were entirely independent, the butterfly effect would preclude prediction on any useful time scale.

Yet the dimensionality is also clearly greater than the number of independent observations used in weather prediction. The dissipative dynamics are such that assimilating a relatively small number of observations causes the remaining degrees of freedom, while dynamically independent, to partially

synchronize with their “real” counterparts. Data assimilation is thus an instance of the dynamical systems paradigm of chaos synchronization (Duane et al. 2006; Yang et al. 2006), wherein two chaotic systems, though sensitively dependent on initial conditions, can be made to synchronize through the exchange, unidirectionally or bidirectionally, of only one or of a few dynamical variables (Fujisaka and Yamada 1983; Afraimovich et al. 1986; Pecora and Carroll 1990; Pecora et al. 1997). Standard data assimilation algorithms prescribe couplings between the systems, here “reality” and “model,” that can also be proven optimal for synchronization and thus for prediction (Duane et al. 2006).

Models can similarly be made to synchronize with *one another* when coupled only loosely, a phenomenon that can also be put to advantage. Unobserved subgrid-scale processes, some of them slaved and others that are dynamically important, are parameterized in different ways in different climate models. If such models are allowed to assimilate data from one another in run time, the models will partially synchronize. The “supermodeling” agenda that relies on the synchronization of alternative models of the same objective process is one of the ultimate applications of the chaos synchronization

Corresponding author: Gregory S. Duane, gregory.duane@colorado.edu

phenomenon to large real-world systems (Duane et al. 2017). The program is to tune the strengths of the intermodel assimilation for the different variables so that the synchronized variables simulate their counterparts in observations, in the belief that results superior to those of any one model, or to any *ex post facto* combination of model outputs, can thus be obtained. That hypothesis has been confirmed for supermodels formed from systems of low-order ODEs (van den Berge et al. 2011; Duane 2015; Mirchev et al. 2012; Du and Smith 2017<sup>1</sup>), quasigeostrophic models, as described herein, versions of the intermediate-complexity SPEEDO model (Selten et al. 2017), and full climate models connected in a rudimentary way, only at the ocean–atmosphere interface (Shen et al. 2016, 2017).

Despite the proliferation of climate models and of subgrid-scale parameterization schemes, the partially synchronized large scales in the trained supermodel afford a superior description of the real climate. The improvement is related to the better known advantage gained in *ex post facto* averaging of outputs of different IPCC-class climate models, which almost always gives an improved representation of the real climate as compared to that of any single model (Reichler and Kim 2008; T. Reichler 2013, personal communication). It is as though the error in the separate models can be treated as random model error, which is reduced upon averaging. Supermodeling employs this assumption at every intermodel assimilation cycle.

The success of the supermodeling agenda, and of the new effective parameterization scheme on which it is based, should tell us something about the small-scale behavior of the real climate, and about relationships among processes on different scales. Here we provide evidence that supermodeling succeeds because it captures the interscale interactions that are essential to maintaining *critical behavior* that is missed in each of the separate models. Further, because different models commonly err in the same way in missing such behavior, no *ex post facto* average of the outputs of those models can reproduce it.

It is precisely when models are in critical states that they are most sensitive to the poorly represented small scales. According to the self-organized criticality (SOC) hypothesis of Bak et al. (1987), systems naturally tend toward such states. They are in any case commonly encountered in the climate system, and SOC has been applied to tropical sea surface temperatures previously (Andrade et al. 1995). What the training of supermodels achieves, in this view, is criticality. This is possible because prior modeling experience, essentially performing optimization in various subspaces of the space of all possible models, has been used to narrow the training task to the dimensions in model space along which ambiguity remains. That ambiguity arises because optimization of a model within the confines of any given algorithmic scheme chosen for subgrid-scale parameterization tends to miss the full

<sup>1</sup> The cross-pollination in time II scheme of Du and Smith (2017) is based on intermodel data assimilation and thus is regarded here as a form of supermodeling, though it was conceived and developed independently.

scope of interscale interactions that come into play near criticality.

The plan of this paper is as follows. In the next section we review the concept and history of supermodeling. Then in section 3 we discuss supermodeling in the case of a toy model with no unresolved processes, but with intermodel coupling in a limited range of scales. It is seen that critical behavior can be achieved where there is none in the constituent models, with an emergent power spectrum that reflects the expected interscale cascade. In section 4, we examine the Shen et al. (2016, 2017) full climate supermodel, with intermodel connections only through a common ocean, from the same point of view. We discuss the possibility of monolithic single-model descriptions of small-scale processes in section 5, and conclude with a reassessment of the self-organized criticality hypothesis in the final section.

## 2. Background: Synchronization of competing models in a supermodel

### a. Single-scale ODE supermodels

The supermodel concept is first illustrated with three Lorenz (1963) systems, each with a different setting of the standard parameters, that we combine to mimic the behavior of a “true” system, defined by a fourth choice of parameters.

The resulting suite is  $\dot{x} = \sigma(y - z)$ ,  $\dot{y} = \rho x - y - xz$ ,  $\dot{z} = -\beta z + xy$ ,

$$\begin{aligned}\dot{x}_i &= \sigma_i(y_i - z_i) + \sum_{j \neq i} C_{ij}^x(x_j - x_i) + K_x(x - x_i), \\ \dot{y}_i &= \rho x_i - y_i - x_i z_i + \mu_i + \sum_{j \neq i} C_{ij}^y(y_j - y_i) + K_y(y - y_i), \\ \dot{z}_i &= -\beta_i z_i + x_i y_i + \sum_{j \neq i} C_{ij}^z(z_j - z_i) + K_z(z - z_i),\end{aligned}\quad (1)$$

where  $(x, y, z)$  is the real Lorenz system and  $(x_i, y_i, z_i)$   $i = 1, 2, 3$  are the three models. An extra term  $\mu$  is present in the models but not in the real system. Because of the relatively small number of variables available in this toy system, all possible directional couplings among corresponding variables in the three Lorenz systems were considered, giving 18 connection coefficients  $C_{ij}^A$ ,  $A = x, y, z$ ,  $i, j = 1, 2, 3$ , and  $i \neq j$ . The constants  $K_A$ ,  $A = x, y, z$  are chosen arbitrarily so as to effect “data assimilation” from the “real” Lorenz system into the three coupled “model” systems. The corresponding configuration for a general set of models, in any given assimilation context represented by a gain matrix  $\mathbf{K}$ , is schematized in Fig. 1.

If we add an equation to adjust the connections, treated as additional dynamical variables,

$$\dot{C}_{i,j}^x = a(x_j - x_i) \left( x - \frac{1}{3} \sum_k x_k \right), \quad (2)$$

with analogous equations for  $\dot{C}_{i,j}^y$  and  $\dot{C}_{i,j}^z$ , where the adaptation rate  $a$  is an arbitrary constant, we find that the three systems synchronize with each other as well as with the “true” system, as seen in Fig. 2a, and that the connection coefficients rapidly asymptote to optimal values. The arrangement is analogous to an idealized form of weather prediction with continuous

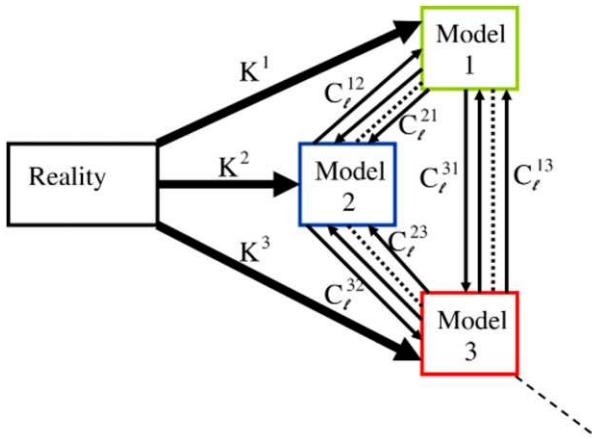


FIG. 1. In a supermodel, “model” systems are linked to each other, generally in both directions, and to “reality” in one direction. Separate links between models  $i, j$ , with distinct values of the connection coefficients  $C_{ij}^{\ell}$ , are introduced for different variables  $\ell$  and for each direction of possible influence.

assimilation of data from “truth” into three alternative forecast models, none of which track truth (Figs. 2b–d) as well as the supermodel.

That Eq. (2) is the correct way to extend synchronization of states to estimation of connection coefficients follows from a result on extending synchronization of states to synchronization of parameters generally (Duane et al. 2007), as explained in detail by Selten et al. (2017). As a heuristic justification, if one considers a time integral of (2), the prescribed change in any intermodel nudging coefficient  $C_{ij}^A$ , where  $A = x, y$ , or  $z$ , is seen to be proportional to the correlation of the intermodel nudging term and the overall signed error between supermodel and truth. With care to the correct choice of sign in the latter factor, model  $j$  is then given more weight, through nudging, if such nudging of model  $i$  to model  $j$  tends to reduce truth–supermodel synchronization error, and conversely, as desired. For models of real processes, the learning rate  $a$  can be made to depend on the amount of random observational error, as in ordinary data assimilation, so that learning proceeds more slowly when the data are noisy.

If we turn off the connection to truth and run the trained supermodel freely, so as to represent long-range climate projection rather than weather prediction, we would expect to lose synchronization, but to still have a supermodel attractor that matches the true attractor. Results of such a comparison are shown in Figs. 3a–e. A fair degree of attractor matching, still far from perfect, has been achieved. Further, if we modify an ancillary parameter of the models, such as  $\rho$ , to mimic the change in radiative forcing of all models in a changed climate, the modified supermodel attractor (Fig. 3e) captures the new properties—a shift in  $Z$  and expansion in the  $Y$  dimension—of the modified true attractor (Fig. 3d), even where the connections defining the new supermodel were obtained by training with the original value of  $\rho$  in all models.

Better attractor matching is achieved with algorithms that consider finite trajectory segments, instead of instantaneous states, in setting values of the connection coefficients for models with very different attractors (van den Berg et al. 2011). In the general case of highly imperfect models, algorithms that minimize short-term prediction error are not adequate for matching attractors (Wiegerinck and Selten 2017). The task of optimizing connections in a supermodel such as (1) thus defines a problem in machine learning that can be addressed with a variety of traditional and less traditional methods (Wiegerinck and Selten 2017; Schevenhoven and Carrassini 2022).

### b. Multiscale supermodels: The quasigeostrophic example

Supermodeling with multiscale models, the main subject of this paper, is first illustrated by extending an example that was previously used to study synchronization-based teleconnections—weak ones of mainly theoretical interest—between the Atlantic and Pacific sectors of the midlatitude circulation in a quasigeostrophic channel model (Duane and Tribbia 2001, 2004). It was previously established that two such models, one forced by a jet in the Atlantic and the other by a jet in the Pacific, would synchronize if connected through only a limited range of Fourier components of the total flow. In the present context, we recast the partially synchronized configuration of connected models as a toy supermodel, engaged in three-way synchronization with a “true” system that has a jet in both sectors.

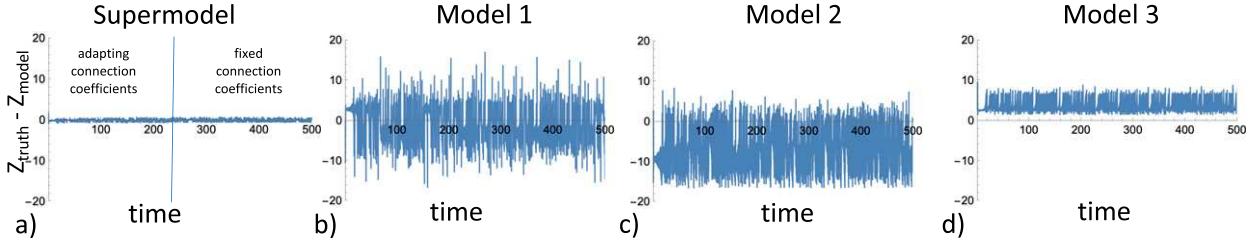


FIG. 2. (a) Synchronization error in the  $z$  variable between a “true” Lorenz system, with parameters  $\rho = 28$ ,  $\beta = 8/3$ ,  $\sigma = 10.0$ ,  $\mu = 0$ , and a supermodel with  $z_{\text{model}} \equiv (1/3) \sum_k z_k$  connected to truth as in (1) with nudging coefficients  $K_x = 10$ ,  $K_y = 10$ ,  $K_z = 0$ , and intermodel nudging coefficients  $C_{ij}^{x,y,z}$  that adapt on the fly according to (2) until  $t = 250$ , and are thenceforth frozen at values that continue to give nearly perfect synchronization. (b)–(d) The corresponding synchronization error for the three models of which the supermodel is comprised, each nudged to truth as in (a), but with no connections among themselves. The three models are defined by parameter values  $\rho_{1,2,3} = \rho$ ,  $\beta_1 = \beta$ ,  $\sigma_1 = 15.0$ ,  $\mu_1 = 30.0$ ,  $\beta_2 = 1.0$ ,  $\sigma_2 = \sigma$ ,  $\mu_2 = -30.0$ ,  $\beta_3 = 4.0$ ,  $\sigma_3 = 5.0$ , and  $\mu_3 = 0$ .

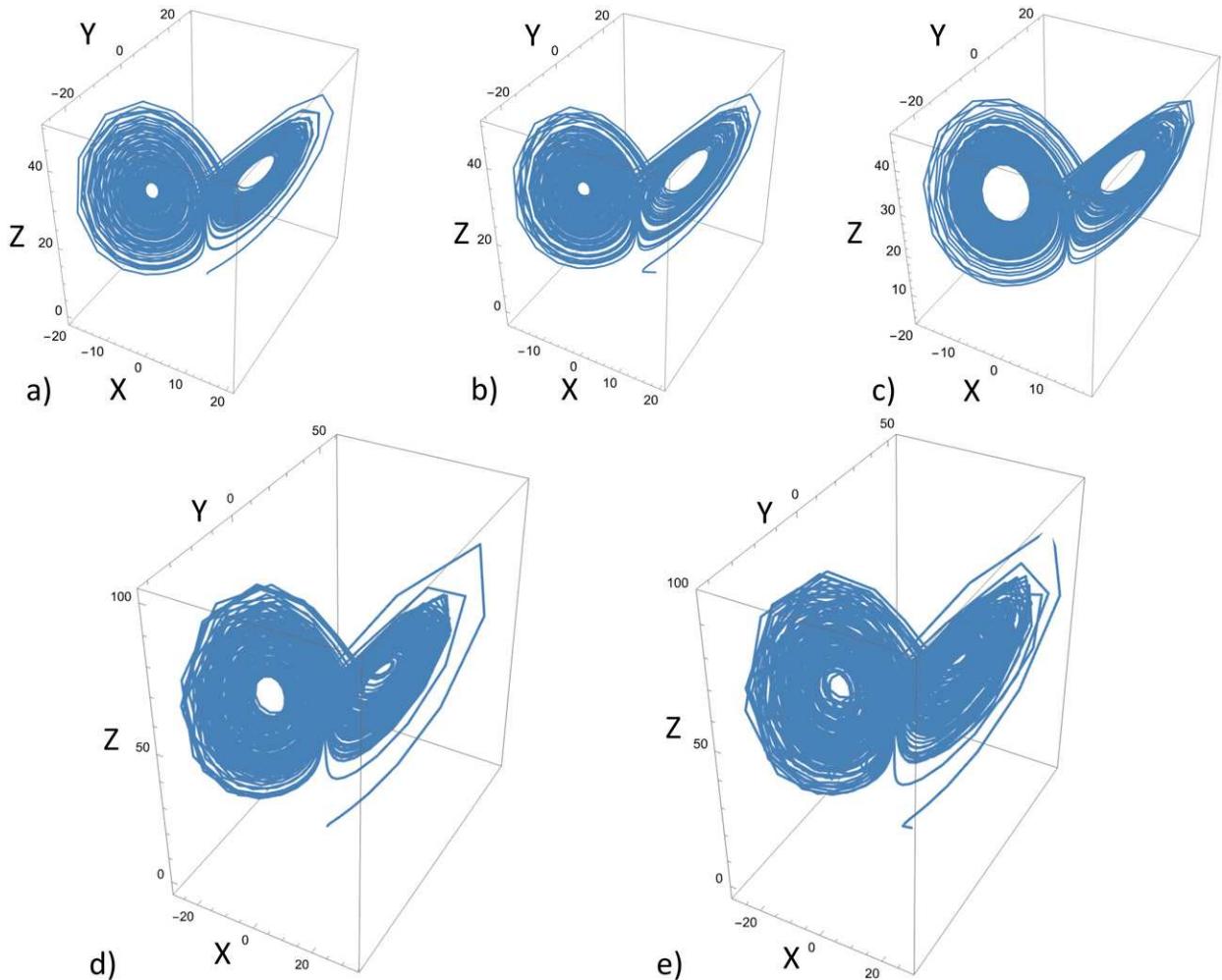


FIG. 3. Attractors of (a) the “true” Lorenz system with parameters as in Fig. 2; (b) the supermodel connected to truth as in Fig. 2a; and (c) the same supermodel run freely, i.e., with  $K_{x,y,z} = 0$ . Also shown are the attractors for all models with a modified  $\rho = 56$ : (d) the modified “true” system and (e) the supermodel formed from the modified models with the original connection coefficients. [Supermodel variables are defined as averages  $(x, y, z) \equiv (1/3)\sum_k (x_k, y_k, z_k)$ .]

The two imperfect models are based on a channel model of zonal/blocked flow vacillation due to Vautard et al. (Vautard et al. 1988; Vautard and Legras 1988), who attempted to capture the index cycle, known as a prominent mode of variability in the midlatitudes at least since the early work of Namias (1951). Each model is given by a quasigeostrophic potential vorticity equation for potential vorticity  $q$  in a two-layer reentrant channel on a  $\beta$  plane:

$$\frac{Dq_i}{Dt} \equiv \frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = F_i + D_i, \quad (3)$$

where the layer  $i = 1, 2$ ,  $\psi$  is streamfunction, and the Jacobian  $J(\psi, \cdot) = (\partial\psi/\partial x)(\partial\cdot/\partial y) - (\partial\psi/\partial y)(\partial\cdot/\partial x)$  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. 4a. The dissipation terms  $D$ , boundary conditions, and other parameter values are given in Duane and Tribbia (2004).

We couple two models of the form (3), 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

$$\begin{aligned} \frac{Dq^A}{Dt} + cJ(\psi^A, q^B - q^A) &= F^A + D^A, \\ \frac{Dq^B}{Dt} + cJ(\psi^B, q^A - q^B) &= F^B + D^B, \end{aligned} \quad (4)$$

(suppressing the layer index  $i$ ) with forcing terms defined in terms of their spectral components:

$$F_k^{A,B} = (\mu_0 - \mu_k^c)[q_k^{A^*,B^*} - q_k^{A,B}], \quad (5)$$

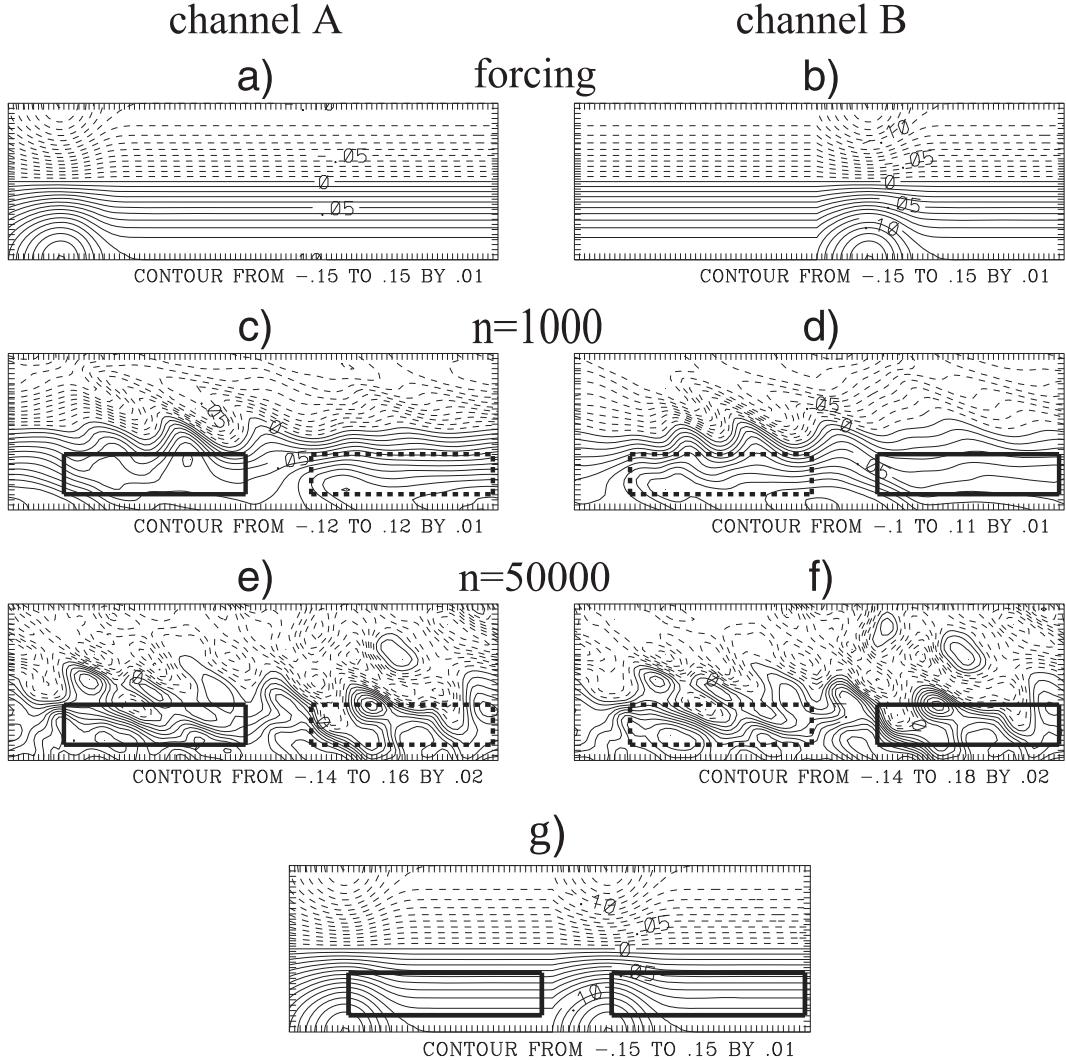


FIG. 4. (a),(b) The forcing  $\psi^*$  and (c)–(f) the evolving flow  $\psi$ , in the parallel channel model with advective coupling as given by (4) with external forcing  $F_k^{A,B} \neq 0$  only in the large scales (6), with  $c = 1/2$ , for the indicated numbers  $n$  of time steps in a numerical integration. In a dimensionalized interpretation of the toy model,  $\mu_0 = 8.8 \times 10^{-4} \text{ s}^{-1}$  and one time step is  $\approx 0.4 \text{ h}$ . An average streamfunction for the two vertical layers  $i = 1, 2$  is shown, for the bottom half of the full domain of the model (see footnote 2). Other parameters are as defined in Duane and Tribbia (2004). In each channel, a blocked-zonal index cycle is induced in the solid-boxed region by the jet in that channel, and in the dashed-boxed region by the jet in the other channel. Near-identical synchronization occurs by the last time step shown in (e) and (f). The synchronized flows closely approximate the flow in a model with forcing  $\hat{\psi}^*$  [corresponding to  $\hat{q}^* \equiv (q^{A*} + q^{B*})/2$ ] shown in (g), describing jets in both sectors. [Reprinted from Duane and Tribbia (2004).]

where  $q_{\mathbf{k}}$  is the wavenumber  $\mathbf{k}$  spectral component of  $q$  in three dimensions.<sup>2</sup> The adjustments  $\mu_{\mathbf{k}}^c$  in (5) were chosen to

allow only the large scales to respond to external forcing, yielding a form that defines a slightly smoothed step function of wavenumber:

$$F_k^{A,B} = \begin{cases} \mu_0 [q_k^{A*,B*} - q_k^{A,B}] & \text{if } |k_x| \leq k_{x0} \text{ and } |k_y| \leq k_{y0}, \\ \mu_0 (k_0/|k^{\perp}|)^4 [q_k^{A*,B*} - q_k^{A,B}] & \text{otherwise} \end{cases} \quad (6)$$

with  $k_0 = |\mathbf{k}_0|$ ,  $\mathbf{k}_0^{\perp} = (k_{x0}, k_{y0})$ , and  $\mathbf{k}^{\perp} = (k_x, k_y)$ . The dynamical variables on all scales are coupled to their counterparts in the opposite channel through the Jacobian terms in (4).

<sup>2</sup> Following Duane and Tribbia (2001, 2004), periodic boundary conditions are used in the  $y$  dimension as well as the  $x$  dimension. Their use in  $y$  implies that there are *two* flows, in opposite directions, displaced from one another in latitude. Only half of the full domain is shown in the figures in this paper, so as to focus on a single flow. But here the wavenumber  $\mathbf{k}$  refers to the number of waves per domain width or domain length. So here  $k_y$  is the number of waves across the width of the full domain, that is, twice the number of waves across the width of the displayed channel.

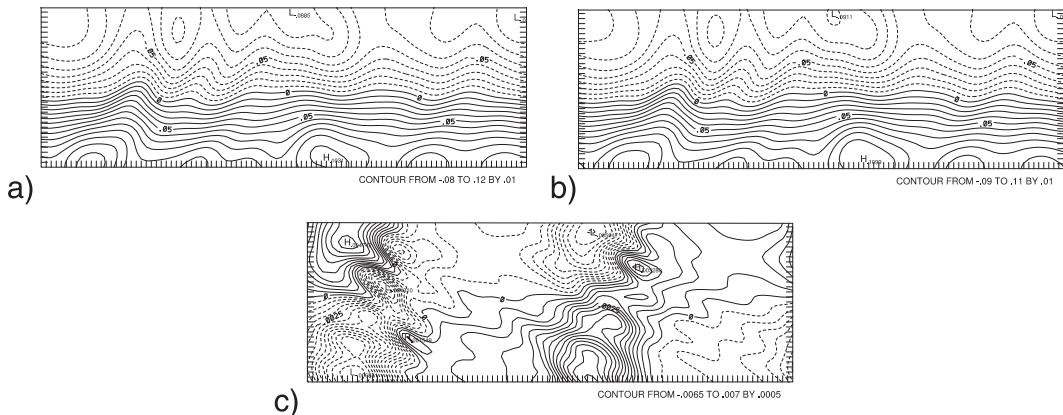


FIG. 5. (a),(b) At a typical instant of time in the evolution of two coupled quasigeostrophic channel models, the streamfunction is projected onto the components describing the large and medium scales, displayed as in Fig. 4, and here seen to be approximately synchronized. The intermodel coupling and parameters are as in Fig. 4, but with external forcing strength given by  $\mu_0 = 3.5 \times 10^{-2} \text{ s}^{-1}$ , and with added intermodel nudging in the medium scales of strength equal to  $0.1 \times \mu_0$ . Medium scales are those for which  $|\mathbf{k}^\perp| \leq k_{\text{noise}}$  and either  $|k_x| > k_{x0}$  or  $|k_y| > k_{y0}$  [see Eq. (13)], where  $k_{\text{noise}} = 15$ ,  $k_{x0} = 3$ , and  $k_{y0} = 4$ . (c) The difference between the streamfunction fields in the two channels, as shown, is mostly in the high-spatial-frequency components.

The flow fields in the coupled channels governed by (4) are found to synchronize, regardless of differences in initial conditions, as seen in Fig. 4. Synchronization was also found when only the medium scales were coupled, by replacing the advective coupling terms  $cJ(\psi^{A,B}, q^{B,A} - q^{A,B})$  in (4) with  $cJ(\psi^{A,B}, q_{\text{medium}}^{B,A} - q_{\text{medium}}^{A,B})$ , where  $q_{\text{medium}}^{A,B}$  is constructed from the medium-scale Fourier components of  $q^{A,B}$ . For faster convergence, intermodel nudging terms can be added, in various configurations as described by Duane and Tribbia (2004).

The correspondence between the channels is not exactly the identity, because of the difference in forcings, but is an instance of *generalized synchronization*, as known to occur in pairs of systems of ordinary differential equations (Rulkov et al. 1995) when small differences in parameters are introduced. Likewise here, at  $c = 1/2$  we have  $\psi^A \approx \psi^B$ , with the difference between the flows in the two channels given mostly by a difference in the high-spatial-frequency components, as illustrated in Fig. 5. Control of the small scales by the assimilation of large-scale data agrees with previous studies of such behavior QG models, e.g., Tanguay et al. (1995). 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 (4a) and (4b), for strong coupling  $c = 1/2$ , is the solution of a model with the average forcing, that is, of a model with  $\hat{q}^* \equiv (q^{A*} + q^{B*})/2$ , describing *two* jets, as in Fig. 4g. (If intermodel nudging terms are included in the forcing, they cancel in the average.) 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  $1/2$ , the dynamics of each channel changes so as to incorporate a “virtual” counterpart of the dynamics of the sector that is

forced in the other channel, including blocking behavior in that sector.

The value  $c = 1/2$  that defines a useful supermodel can be obtained by training against a “real” (two-jet) dataset  $q_{\text{obs}}(\mathbf{x}, t)$  to which the model is nudged, according to the training rule:

$$\frac{dc}{dt} = a \int d\mathbf{x} [J(\psi^A, q^B - q^A)(q^A - q_{\text{obs}}) + J(\psi^B, q^A - q^B)(q^B - q_{\text{obs}})], \quad (7)$$

the connection adaptation rule for coupling in the advective term that is analogous to the rule (2) that was used when the coupling was in the nudging term. The rule (7), not discussed in previous work, tends to minimize the RMS difference between the supermodel and observations (see appendix) and suffices to give the correct coupling, as shown in Fig. 6.

### c. Climate supermodels

At an intermediate level of complexity in the hierarchy of models, two SPEEDO models (Severijns and Hazeleger 2010) sharing a common ocean, but with distinct atmospheric components, each with its own parameter settings, were connected by nudging at all atmospheric grid points (Seltten et al. 2017). The resulting supermodel exhibited a climatology and a climate response to a CO<sub>2</sub> increase that was closer to those of a reference “true” model than were those of a standard multimodel average. That result suggests that the toy supermodel results described above will carry over to full climate models.

The manner in which supermodels surpass multimodel averages is, however, more clearly illustrated by the application to full climate models connected to one another only through a common ocean, without interatmosphere connections. We use the COSMOS configuration (Giorgi et al. 2013) composed of an ECHAM5 atmosphere and an MPIOM ocean. Two versions of the ECHAM5 atmosphere model, “Nordeng” and “Tiedtke,”

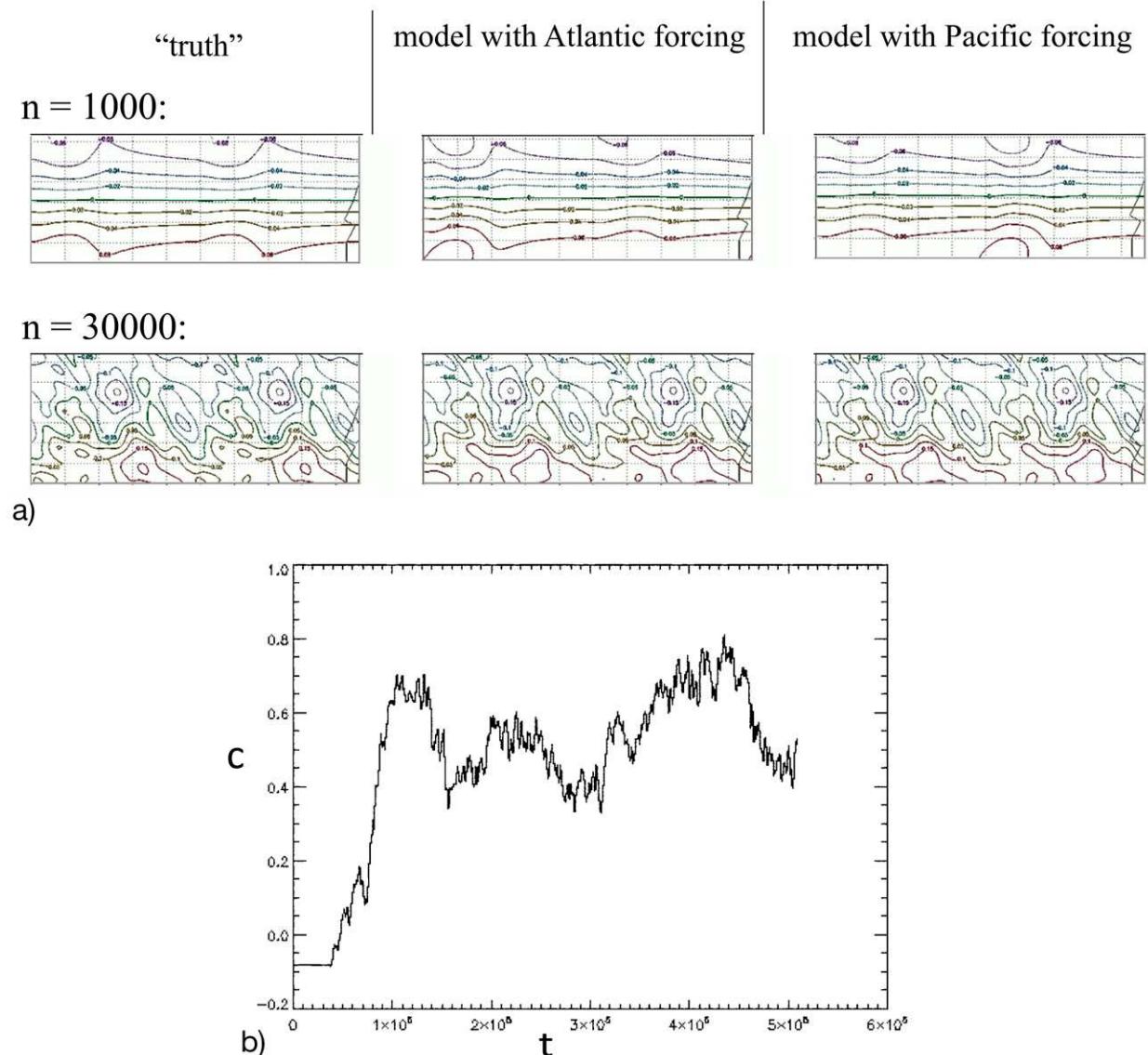


FIG. 6. (a) As two QG models synchronize with each other and also with the truth, to which each channel model is nudged, (b) the intermodel coupling is estimated according to (7), finding  $c \rightarrow 0.5$ , its “true” value. The intermodel coupling, external forcing strength  $\mu_0$ , and other parameters are as in Fig. 5, but with medium-scale intermodel nudging coefficients equal to  $0.025 \times \mu_0$ , a narrower range of medium-scale nudging bounded above by  $k_{\text{noise}} = 10$ , and coefficients of nudging to truth equal to  $0.5 \times \mu_0$ .

each with a different parameterization of subgrid-scale convective processes, were coupled to a single MPIOM ocean model (Shen et al. 2016, 2017): Both atmosphere models feel air-sea fluxes based on the same SST field, while the ocean receives a weighted average of the air-sea fluxes, as depicted schematically in Fig. 7. We refer the interested reader to the original sources (Tiedtke 1989; Nordeng 1994) for details of the differences between the two models, which are rather involved.

Different weights, labeled  $\alpha$ ,  $\beta$ , and  $\gamma$  are used for each of the air-sea fluxes felt by the common ocean—heat, momentum, and freshwater, respectively. That is, the two COSMOS models are given by

$$\dot{\mathbf{A}} = f_N[\mathbf{A}, \mathbf{Q}_N(\mathbf{A}, \mathbf{O}), \boldsymbol{\tau}_N(\mathbf{A}, \mathbf{O}), \mathbf{q}_N(\mathbf{A}, \mathbf{O})], \quad (8a)$$

$$\dot{\mathbf{O}} = g[\mathbf{O}, \mathbf{Q}_N(\mathbf{A}, \mathbf{O}), \boldsymbol{\tau}_N(\mathbf{A}, \mathbf{O}), \mathbf{q}_N(\mathbf{A}, \mathbf{O})], \quad (8b)$$

and

$$\dot{\mathbf{A}} = f_T[\mathbf{A}, \mathbf{Q}_T(\mathbf{A}, \mathbf{O}), \boldsymbol{\tau}_T(\mathbf{A}, \mathbf{O}), \mathbf{q}_T(\mathbf{A}, \mathbf{O})], \quad (9a)$$

$$\dot{\mathbf{O}} = g[\mathbf{O}, \mathbf{Q}_T(\mathbf{A}, \mathbf{O}), \boldsymbol{\tau}_T(\mathbf{A}, \mathbf{O}), \mathbf{q}_T(\mathbf{A}, \mathbf{O})], \quad (9b)$$

where  $f_N$  and  $f_T$  are the dynamics describing the evolution of the atmospheric state vector  $\mathbf{A}$  in the Nordeng and

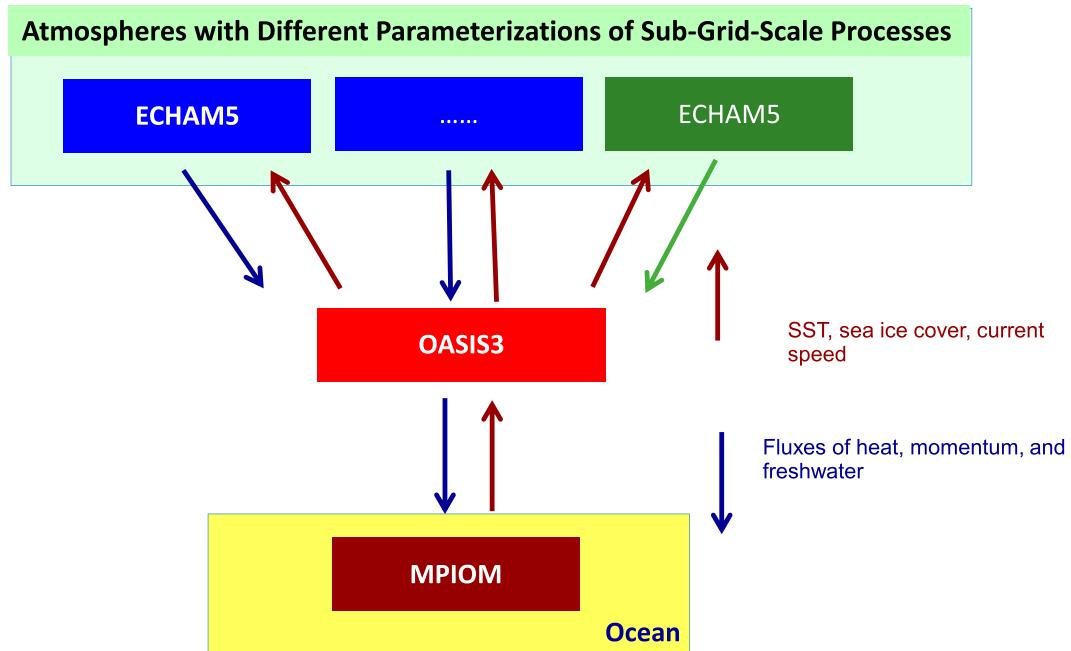


FIG. 7. The general form of the COSMOS supermodel. Two or more atmospheric models impact a common ocean model via heat, momentum, and mass (water) fluxes, which are weighted combinations of the fluxes from the separate models. Each atmosphere is impacted only by its own fluxes.

Tiedtke models, respectively, which include a dependence on the atmosphere–ocean fluxes  $\mathbf{Q}$ ,  $\boldsymbol{\tau}$ , and  $\mathbf{q}$ , for heat, momentum (wind stress), and water (precipitation minus evaporation), respectively. Each flux depends on the states of atmosphere and ocean in a different way for each model. The ocean state  $\mathbf{O}$  evolves according to the dynamics  $g$  which is the same in both models. The supermodel state is the conjoined state vector  $(\mathbf{A}_N, \mathbf{A}_T, \mathbf{O})$ . The state evolves according to

$$\dot{\mathbf{A}}_N = f_N[\mathbf{A}_N, \mathbf{Q}_N(\mathbf{A}_N, \mathbf{O}), \boldsymbol{\tau}_N(\mathbf{A}_N, \mathbf{O}), \mathbf{q}_N(\mathbf{A}_N, \mathbf{O})], \quad (10a)$$

$$\dot{\mathbf{A}}_T = f_T[\mathbf{A}_T, \mathbf{Q}_T(\mathbf{A}_T, \mathbf{O}), \boldsymbol{\tau}_T(\mathbf{A}_T, \mathbf{O}), \mathbf{q}_T(\mathbf{A}_T, \mathbf{O})], \quad (10b)$$

$$\begin{aligned} \dot{\mathbf{O}} = & g[\mathbf{O}, \alpha \mathbf{Q}_N(\mathbf{A}_N, \mathbf{O}) + (1 - \alpha) \mathbf{Q}_T(\mathbf{A}_T, \mathbf{O}), \\ & \beta \boldsymbol{\tau}_N(\mathbf{A}_N, \mathbf{O}) + (1 - \beta) \boldsymbol{\tau}_T(\mathbf{A}_T, \mathbf{O}), \\ & \gamma \mathbf{q}_N(\mathbf{A}_N, \mathbf{O}) + (1 - \gamma) \mathbf{q}_T(\mathbf{A}_T, \mathbf{O})]. \end{aligned} \quad (10c)$$

The weights  $\alpha$ ,  $\beta$ , and  $\gamma$  are free parameters that are trained using a gradient descent scheme, so that the simulated monthly climatology of SST over the tropical Pacific region is closest to the observed monthly climatology, with respect to RMS error in SST in the tropical Pacific ( $10^{\circ}\text{S}$ – $10^{\circ}\text{N}$ ,  $160^{\circ}\text{E}$ – $90^{\circ}\text{W}$ ) (Shen et al. 2016)

The climatological SST and precipitation fields for the two models, the supermodel, and reference observations are shown in Fig. 8. It is seen that while the two models each exhibit the error of a double intertropical convergence zone (ITCZ), that is also found in a large variety of other GCMs (Mechoso et al. 1995; Zhang et al. 2007), the trained supermodel exhibits the single ITCZ that is found in observed tropical Pacific behavior

surrounding ENSO.<sup>3</sup> Importantly, any ex post facto weighted average of the outputs of the two models would also exhibit the double ITCZ error.

To be consistent with our definition of supermodeling as intermodel data assimilation, we note that the configuration described above, given by Eqs. (8)–(10), approximates a configuration of two coupled ocean–atmosphere models—one with a Nordeng atmosphere and an MPIOM ocean, the other with a Tiedtke atmosphere and a separate MPIOM ocean—connected by intermodel nudging between the ocean components.<sup>4</sup> To see this, imagine that the connections between the two oceans are strong enough to cause them to synchronize, while the interocean connections for different prognostic variables are chosen judiciously. For example, if the atmosphere–ocean heat flux only enters the prognostic equation for ocean temperature, and the momentum flux

<sup>3</sup> It is important to distinguish the trained supermodel behavior from that of the *interactive ensemble* of Kirtman and Shukla (2002), who also obtained improvement in the representation of ENSO by coupling several atmospheres, there with equal weights, to the same ocean. In that work, the atmospheres were different realizations of the *same* model, so the improvement was due to a reduction in weather noise. Here, training is essential, since the double ITCZ remains when equal weights are used (Fig. 8c), and the improvement cannot be attributed to noise reduction. Further, no improvement in SST pattern similar to that obtained here was reported in that previous work. The *multimodel* interactive ensemble of Kirtman et al. (2003), on the other hand, involved swapping of variables between different models and was a forerunner of supermodeling.

<sup>4</sup> As this article goes to press, it is noted that such an ocean-connected supermodel has recently been constructed by Counillon et al. (2023).

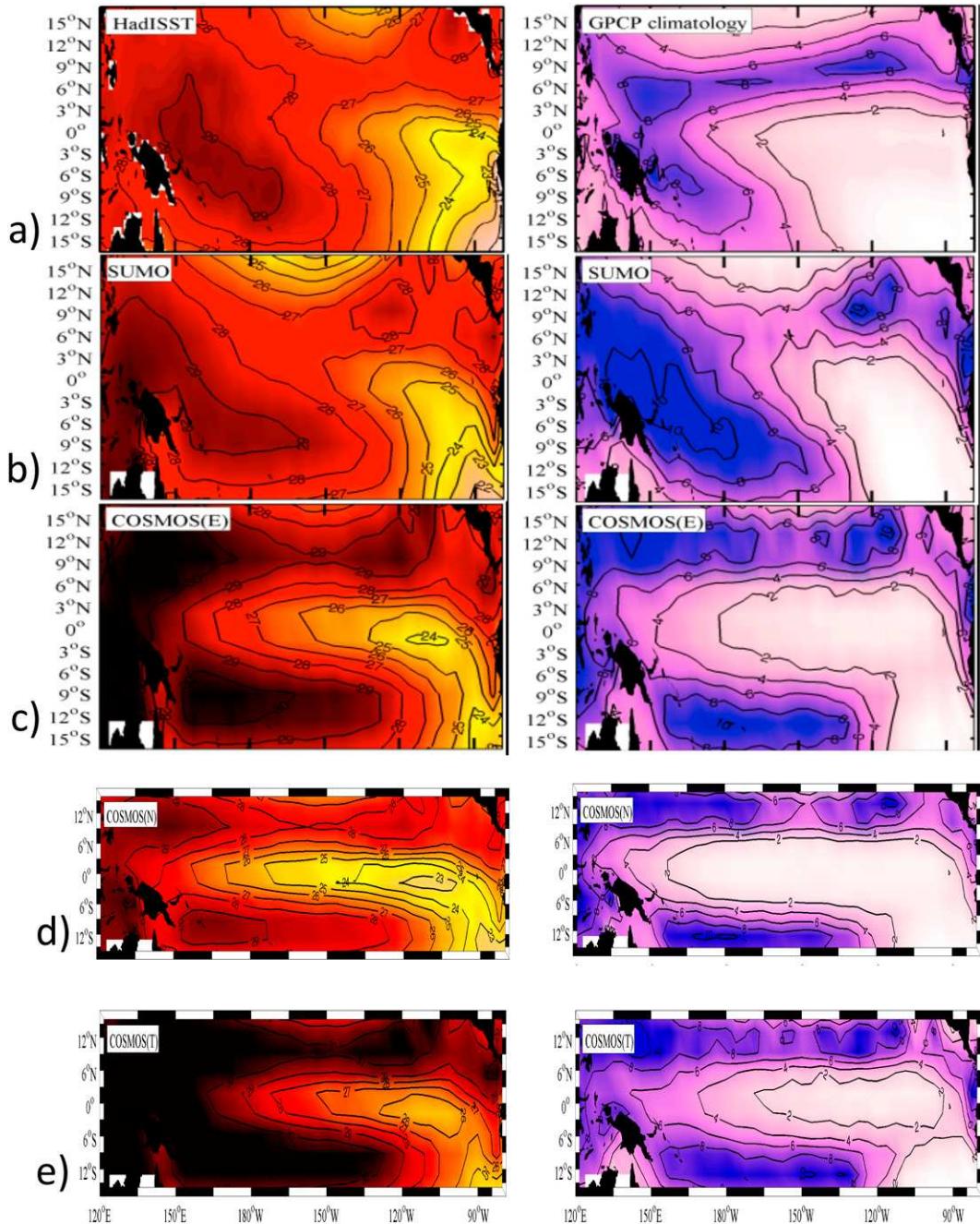


FIG. 8. The (left) climatological sea surface temperature (SST) (units:  $^{\circ}\text{C}$ ) and (right) precipitation (units:  $\text{mm day}^{-1}$ ) in the tropical Pacific from (a) observations, (b) the trained supermodel (SUMO), (c) the untrained, equal-weighted supermodel [COSMOS(E)], and the two constituent models, (d) COSMOS(N) and (e) COSMOS(T). The three independent weights in the supermodel are trained by optimizing the match between model climatological monthly means and observed climatology. Observed SST is from the HadISST dataset ([Met Office Hadley Centre 2013](#)) for the period 1948–79, the period also used as a training set; observed precipitation is from the GPCP dataset ([Physical Sciences Laboratory 2013](#)) for the period 1979–2012. There is a double ITCZ in both of the two separate models and the equal-weighted supermodel, while there is a single ITCZ in observations and in the trained supermodel.

only enters the prognostic equations for ocean velocity, then different coefficients for the nudging of temperature and velocity between the two oceans would give different weights  $\alpha$  and  $\beta$  for the heat and momentum fluxes, resp., in the limit of infinite nudging,  $C_l^{12} \rightarrow \infty$  ( $l = T, u, v, \dots$ ) while  $C_T^{12}/C_u^{12}$  and  $C_T^{12}/C_v^{12}$  are held fixed. Thus, it is thought that the behavior of the configuration described above, with a single ocean, would be obtained in a supermodel comprised of models with distinct oceans. Further, it is asserted that such behavior, involving elusive structures such as the single ITCZ, is typical of supermodels generally, describing the way in which they surpass their constituent models and also surpass any weighted average of the outputs thereof.

### 3. Relationships between processes on different scales in a quasigeostrophic supermodel

To further explore the hypothesis that supermodeling succeeds because of the importance of interscale interactions near critical states, such as those associated with the ENSO cycle, we consider the simpler case of another quasigeostrophic supermodel, formed from a different configuration of models than the one described above. We connect two quasigeostrophic models, one strongly forced by a jet, and the other weakly forced or unforced. The models are here connected both by simple nudging and by advective coupling as in the two-sector example:

$$\begin{aligned} \frac{Dq^A}{Dt} + c^A J(\psi^A, q^B - q^A) &= F^A + D^A \\ \frac{Dq^B}{Dt} + c^B J(\psi^B, q^A - q^B) &= F^B + D^B \end{aligned} \quad (11)$$

The forcing terms of the two models include external and intermodel nudging terms, as in the first quasigeostrophic supermodel discussed in the last section, but with the same external jet pattern  $q^*$  used to force both models:

$$\begin{aligned} F_k^A &= \mu_k^{\text{ext } A}(q_k^* - q_k^A) + c'^A \mu_k^{\text{coup } A}(q_k^B - q_k^A) \\ F_k^B &= \mu_k^{\text{ext } B}(q_k^* - q_k^B) + c'^B \mu_k^{\text{coup } B}(q_k^A - q_k^B) \end{aligned} \quad (12)$$

and with scale-dependent nudging coefficients chosen so as to nudge only medium-wavenumber components of the flow to the other model:

$$\mu_k^{\text{coup } A,B} = \begin{cases} 0 & \text{if } |k_x| \leq k_{x0} \text{ and } |k_y| \leq k_{y0} \\ \mu_0^{A,B}[(k_{\text{noise}}/|\mathbf{k}^\perp|)^4] & \text{if } |\mathbf{k}^\perp| > k_{\text{noise}} \\ \mu_0^{A,B}[1 - (k_0/|\mathbf{k}^\perp|)^4] & \text{otherwise} \end{cases}, \quad (13)$$

for horizontal wavenumbers defined as  $\mathbf{k}^\perp = (k_x, k_y)$ ,  $k_0 = |\mathbf{k}_0^\perp|$ , and  $\mathbf{k}_0^\perp = (k_{x0}, k_{y0})$ , and with

$$\mu_k^{\text{ext } A,B} = \begin{cases} \mu_0^{A,B} - \mu_k^{\text{coup } A,B} & \text{if } |\mathbf{k}| \leq k_{\text{noise}} \\ 0 & \text{if } |\mathbf{k}| > k_{\text{noise}} \end{cases} \quad (14)$$

so that the smallest-scale “noise” components are neither coupled to the other channel nor constrained by the external forcing. (We could dispense with the advective coupling if we are

only interested in the medium and large scales, since they can be made to synchronize through nudging. We retain the advective coupling so that we can define the small-scale variables of the supermodel as the average of small-scale variables in the two constituent models that are not synchronized.)

The supermodel behavior, for appropriately chosen connections, is exactly the same as that of a single quasigeostrophic channel model with moderate forcing strength. To see this, consider a weighted average  $\hat{q}$  of the potential vorticity fields for the two channels separately, that is,  $\hat{q} \equiv w^A \mathbf{q}^A + w^B \mathbf{q}^B$ , for some weights  $w^A, w^B \in [0, 1]$ , with  $w^B = 1 - w^A$ . Define the coefficients of the advective coupling terms in (11) differently for the two channels, specifically let  $c^A = w^B$  and  $c^B = w^A$ . Then, combining the two equations in (11) and expanding the advective derivatives  $D/Dt$ , the averaged potential vorticity  $\hat{q}$  is seen to satisfy

$$\begin{aligned} \frac{\partial \hat{q}}{\partial t} + w_A J(\psi^A, w^A q^A + w^B q^B) + w_B J(\psi^B, w^A q^A + w^B q^B) \\ + w^A w^B J(\psi^A, q^B - q^A) + w^B w^A J(\psi^B, q^A - q^B) \\ = w^A F^A + w^B F^B + w^A D^A + w^B D^B \end{aligned} \quad (15)$$

It is then seen that the four Jacobian terms in (15), two from the advective derivatives and two from the advective couplings, sum to  $J(w^A \psi^A + w^B \psi^B, w^A q^A + w^B q^B) = J(\hat{\psi}, \hat{q})$ , which is the usual Lagrangian contribution to the advective derivative of  $\hat{q}$ , so that (15) can be written:

$$\frac{D\hat{q}}{Dt} = (w^A F^A + w^B F^B) + (w^A D^A + w^B D^B). \quad (16)$$

For  $\mathbf{k}$  in any range of scales for which the models are synchronized, we naturally define the corresponding component of the supermodel  $q^{\text{sm}}$  as  $q^{\text{sm}} \equiv q_k^A = q_k^B = \hat{q}_k$ . This range includes all  $\mathbf{k}$  for which the external forcing coefficients  $\mu_k^{\text{ext } A,B} \neq 0$ . If the models or Fourier components thereof are not synchronized, we can arbitrarily define the supermodel potential vorticity as the weighted average of the values for the two models  $q^{\text{sm}} \equiv \hat{q}$ . As in Duane and Tribbia (2004) we assume the dissipation terms  $D^{A,B}$  are of the same form and combine linearly. So the supermodel field  $\mathbf{q}^{\text{sm}}$  also satisfies (16), the usual potential vorticity equation with the forcing terms  $F^{A,B}$  averaged. The contribution of any intermodel nudging terms to the averaged forcing can also be made to vanish for an appropriate ratio of  $c'^A$  to  $c'^B$  in (12), leaving only an averaged value of external forcing by the single jet, as was to be shown.

The weights can be obtained dynamically, in an initial learning stage, from an adaptation rule of the same origin as (2) and (7), that augments a given dynamical system with a dynamical equation for a parameter whose value is to be adjusted based on observed data. The general rule is that, for any parameter that appears linearly in one of the dynamical equations, one adds an equation for the parameter, regarded as a new dynamical variable, that sets the time derivative of the parameter to be proportional to the product of the cofactor of that parameter in the given equation and the partial derivative of some error function of the state variables with respect to the variable that is prognosed by the given equation. It can be shown that if the model is nudged to another

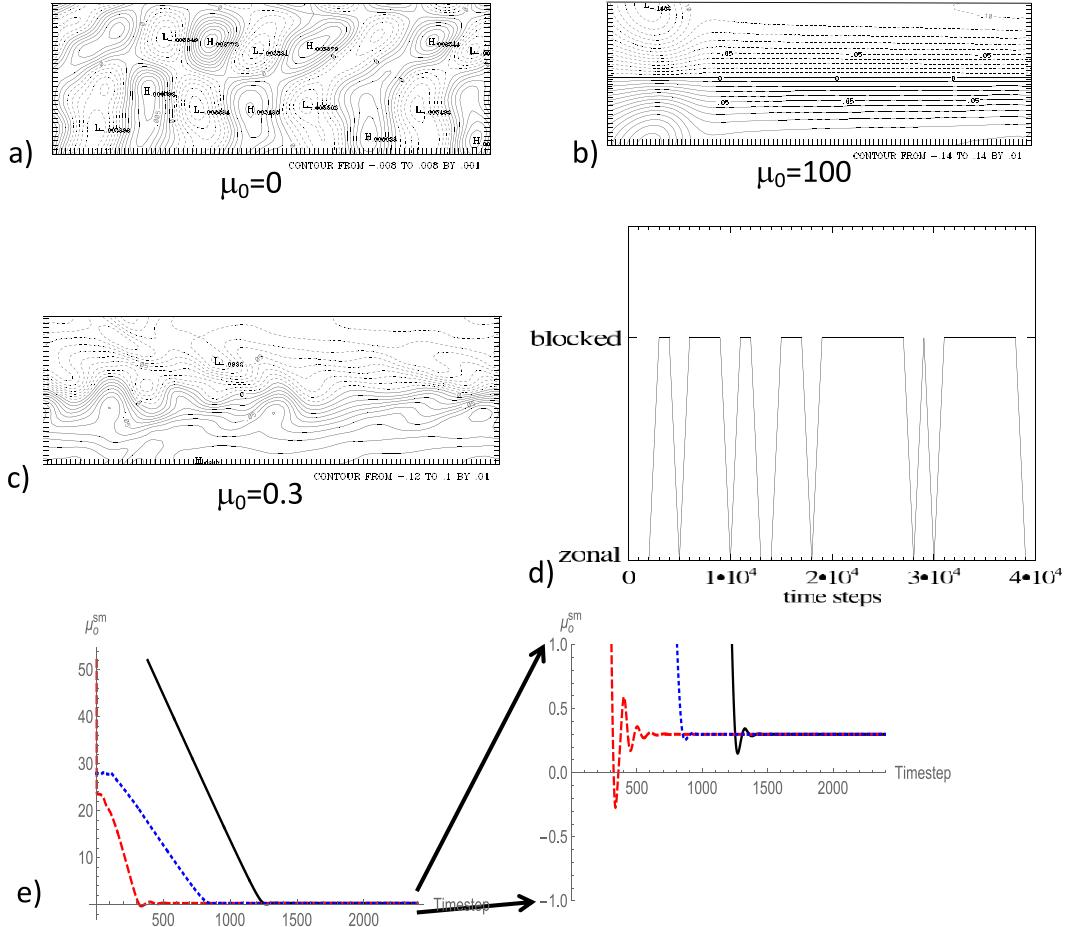


FIG. 9. Typical flows in the uncoupled QG channel model with (a) no forcing ( $\mu_0 = 0$ ), (b) strong forcing ( $\mu_0 = 100$  in nondimensional units), and (c) “realistic” forcing ( $\mu_0 = 0.3$ ); the last case as represented also by a supermodel. For no forcing and large forcing, there is no blocking. [The values in (a), from  $-0.008$  to  $0.008$ , indicate low-amplitude turbulence, not blocking.] (d) For “realistic” forcing, the history of vacillation of the flow between zonal and blocked regimes, sampled at low temporal resolution over the course of a simulation is plotted, using the blocking diagnostic defined in Duane and Tribbia (2004). (e) Synchronization-based optimization of the connections in a supermodel composed of  $\mu_0^A = 100$  and  $\mu_0^B = 0$  models, nudged to truth in the medium scales, produces a model with forcing  $\mu_0^{\text{sm}} = w^A \mu_0^A + (1 - w^A) \mu_0^B$  that asymptotically approaches the correct value  $\mu_0^{\text{sm}} = 0.3$ , starting from  $\mu_0^{\text{sm}} = 100$ , regardless of the strength of nudging to truth  $\nu \mu_0^{\text{sm}}$ , which can be the same as the strength of external forcing  $\mu_0^{\text{sm}}$  (solid black line), decreased to  $0.5 \mu_0^{\text{sm}}$  (dashed red line), or increased to  $1.5 \mu_0^{\text{sm}}$  (dotted blue line); all cases are also displayed with an expanded vertical axis.

version of itself with correct parameter values, and if the configuration is such as to make the states converge when the parameters match, than the stated adaptation rule will correct the parameter mismatch. For an appropriate choice of error function, the RMS error in states and parameters is minimized if we adapt the weights according to

$$\begin{aligned} \frac{dw_A}{dt} = a \sum_{|k_x| \leq k_{x0}, |k_y| \leq k_{y0}} (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}})(q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}}) \\ + a \sum_{\substack{\|\mathbf{k}\| < k_{\text{noise}}, \\ (|k_x| > k_{x0} \text{ or } |k_y| > k_{y0})}} [(k_0/|\mathbf{k}^\perp|)^4 (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}})(q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}}) \\ + \nu [1 - (k_0/|\mathbf{k}^\perp|)^4] (q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}})^2], \end{aligned} \quad (17)$$

where  $a$  is a learning rate, with  $w_B \equiv 1 - w_A$ , and  $\nu$  is the ratio, held constant, between the strength of nudging to “observations” during training and the strength of external forcing. This adaptation rule is the analog of (2) and (7) for the single-jet two-forcing-strength QG supermodel (see appendix).

Two models that are unforced and strongly forced, respectively, exhibit no blocking, as seen in Fig. 9, while a model with moderate forcing strength exhibits a blocked-zonal index cycle, arguably similar to that found in the real atmosphere, as does also the supermodel with the two extreme cases as constituents, after training the connections (Fig. 9e) to simulate “realistic” forcing. With only medium scales connected, the constituent models synchronize with each other and exhibit the usual index cycle. Note that the models cannot

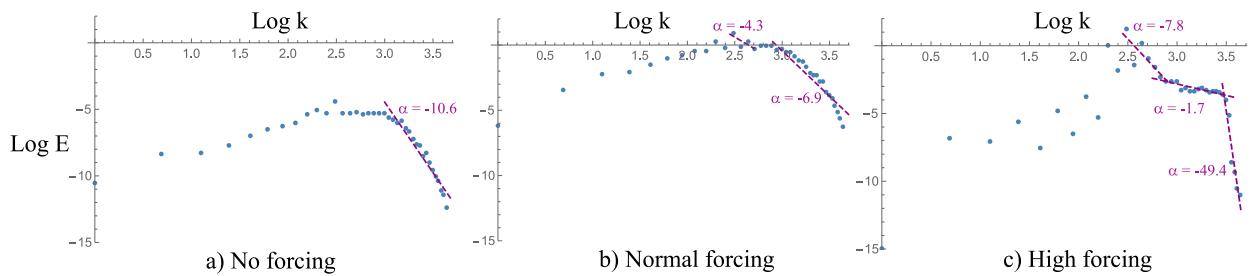


FIG. 10. Energy  $E(k) = \sum_{k < |k| < k+1} [u^2(k) + v^2(k)]/2$  vs total wavenumber  $k$ , plotted logarithmically in arbitrary units, for the three cases of the quasigeostrophic channel model defined by forcing strength (a)  $\mu_0 = 0$ , (b)  $\mu_0 = 0.3$ , and (c)  $\mu_0 = 100$ , which are illustrated in Fig. 9. Dashed lines are best linear fits, each with slopes  $\alpha$  equal to the estimated critical exponent, in wavenumber ranges where the energy falloff is approximately exponential,  $E \approx k^\alpha$ ,  $\alpha < 0$ .

possibly synchronize exactly, because they would satisfy different equations if the nudging terms vanish. The differences are taken up in the small-scale components that are not connected between models, as shown for the two-jet supermodel in Fig. 5. That some dynamical components of the constituent models are *not* connected, so that the models remain semiautonomous, is key to the supermodeling program.

With strong large-scale forcing, small-scale dynamics are ineffective in disrupting the jet-like flow (Fig. 9a). With no forcing, no interesting interaction between large and small scales occurs, and only low-amplitude turbulence (Fig. 9b) ensues. The interesting behavior occurs in the intermediate dynamical regime (Fig. 9c). We take the blocked-zonal flow vacillation to be one example of *critical behavior*, of which there are indeed many in the climate system, each characterized by chaotically intermittent vacillation of a climate subsystem between qualitatively different dynamical regimes.

In the SOC picture of Bak et al. (1987), as applied to the model studied, blocked and zonal flows are “minimally stable” states which break down in response to small fluctuations. One can indeed cast blocked and zonal flows as two equilibria that exist simultaneously in a simpler, barotropic system over a narrow range of parameters (Ghil and Childress 1987). Vacillation among minimally stable states, in the Bak et al. view, occurs in a wide variety of open systems that are far from thermodynamic equilibrium, such as the famous example of a sandpile onto which sand is continually poured, giving rise to a fractal pattern of avalanching structures on the surface. The fractal form arises, in the idealized SOC case, because information cascades freely from small scales to large scales when a system is at or near a critical point, implying that there is no preferred scale in the range of scales over which this cascade occurs. In the nonideal case studied here, no fractal form is discerned, but there is interscale interaction across a wide range of scales (e.g., Branstator 1995), so something of the SOC picture will apply.

Rather than further describe the specific interscale interactions that give rise to criticality, we investigate the applicability of the power-law form of the fluctuations that was conjectured by Bak et al. (1987) to characterize critical states generally. If  $P(s)$  represents the distribution of scales  $s$  in a power spectrum, then the lack of a well-defined value for  $\int ds sP(s)$ , which would define an average or typical value of  $s$ ,

suggests that  $P(s)$  is of the form  $Ns^\alpha$  for some constants  $\alpha$  and  $N$ , so that  $\log P(s)$  is linear in  $\log s$ , with slope  $\alpha$ . Logarithmic plots of the energy spectrum of the flow in the QG channel model are shown in Fig. 10. The energy falloff with wavenumber in the situation of blocked-zonal flow vacillation, as occurs in “reality” and in the supermodel, has a form closer to that given by a single log-linear range than does the spectrum of either of the two constituent models defined by extreme values of the forcing strength, except for a range of energies and of length scales both too small to be of interest, in the case of zero forcing. The behavior supports the notion that the cascade of information from the small scales is important in the case of vacillation, and that it is achieved by the supermodel. In the toy model there are no scales that are unresolved by design, but there are relevant processes on many scales that are collectively parameterized by the forcing strength  $\mu_0$ . We suggest that the demonstrated log-linear form is a vestige of the form that would be found in a more complete model that exhibits the same vacillation.

It is suggested that the representation of such spatiotemporal structures as blocking patterns is at the root of a trained supermodel’s ability to avoid errors present in its constituent models, even when those errors are qualitatively similar in all constituents. Such structures, like the fractal patterns on avalanching sandpiles, appear only near criticality, and thus are absent in the noncritical constituent models. They are also absent in any *ex post facto* average of the constituent model outputs. In the following sections, further evidence is presented that supermodels can generally surpass the common method of output averaging for the same reason—the unique character of the spatiotemporal patterns that emerge only near a critical state and are missed by all constituent models.

#### 4. Trained criticality in the full climate supermodel

We argue that the COSMOS supermodel that was described in section 2, in reaching a critical state, qualitatively different from that of either of its constituent models, exhibits behavior comparable to that of the quasigeostrophic supermodel discussed above. Here the critical state, imagined to coincide with climatology, as characterized by a single ITCZ and a reduced cold tongue, is supported by suppressed ocean upwelling at the equator under conditions where the equatorial and off-equatorial ocean circulations combine in just the

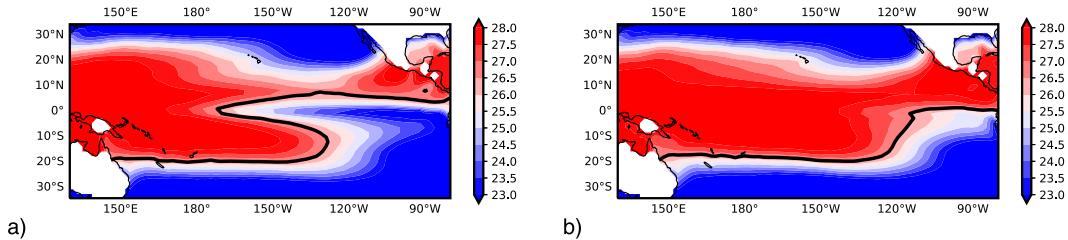


FIG. 11. A composite of observed SST patterns in the (a) cold phase and (b) warm phase of the ENSO cycle, showing an extended cold tongue, a usual corollary of the double ITCZ, in the cold phase. The boundary of the cold tongue is taken to be the  $26.5^{\circ}\text{C}$  contour (bold black line). Observations are from the period 1982–2021 as given by the NOAA High Resolution SST dataset (Physical Sciences Laboratory 2016; Reynolds et al. 2007; Huang et al. 2021).

right way to suppress the excessive cold tongue found in either model separately, in the manner previously described by Shen et al. (2017).

Briefly, the Shen et al. mechanism relies on an interplay between Ekman pumping of subsurface cold water, upwelled at the equator due to local wind stress on the one hand, and off-equatorial downwelling driven by wind stress gradients at  $5^{\circ}$ – $10^{\circ}\text{N}$  and  $5^{\circ}$ – $10^{\circ}\text{S}$  on the other. The off-equatorial downwelling induces equatorial upwelling via the meridional tropical cells in the ocean, in either hemisphere. The equatorial wind stress and consequent upwelling is stronger in the Nordeng model, while the off-equatorial downwelling and consequent equatorial upwelling is stronger in the Tiedtke model. The two effects are combined in the supermodel, with relative strengths determined by the value of the momentum-flux weight  $\beta$  in (10), which determines wind stress. The effects do not themselves combine linearly, due to some of the same nonlinearities that are essential to the ENSO cycle (e.g., Jin 1997, 1998; Zebiak and Cane 1987). That is, the cooling due to upwelling at the equator is not itself a weighted average of the cooling in Nordeng and Tiedtke. Nature has in effect selected a special value of  $\beta$  that gives a singular pattern in the tropical Pacific SST, different from what is observed at both higher and lower values.

Following the self-organized criticality analysis of Bak et al. (1987), we examine the power spectra for observations and the different models, as in section 3 for the QG model, but here using temporal rather than spatial frequency. The spectra for both the Nordeng and Tiedtke models, in simulations of the twentieth century, exhibit a falloff with frequency that is too steep at low frequencies (Figs. 12a,b), as compared to observations (Fig. 12d). This behavior agrees with the fact that the ENSO cycle in both models is too rapid, with too much energy concentrated around a period of 2 years, an error commonly encountered in models of ENSO (e.g., Kirtman 1997). The single-ITCZ pattern is a promising indication that the ocean–atmosphere dynamics in the supermodel is a more faithful representation of the true dynamics.

The chaotic vacillation here that is analogous to the vacillation between blocked and zonal flow regimes in the midlatitudes, depicted in Fig. 9d, is that between the warm phases and cold phases of the ENSO cycle, illustrated in Fig. 11. While the climatological precipitation patterns in the two phases are less distinct than the temperature patterns shown

in the figure, the single ITCZ in climatology is indeed an average of the two phases. Nordeng, Tiedtke, and other climate models of similar resolution overrepresent the cold (La Niña) phase in which the ITCZ is divided. That the single ITCZ behavior in the supermodel is the result of vacillation, and does not represent a uniform climatological shift, was demonstrated in Shen et al. (2016), where it was shown that the supermodel better reproduces the El Niño phase. The system is in a critical state in regard to the two flow regimes. As with blocking, El Niño events interrupt the background cold phase pattern that is so easily reproduced in climate models.

Analogously to the blocking patterns in the quasigeostrophic model, here it is the spatial structure associated with ENSO that is characteristic of the critical behavior. The extension of the Pacific warm pool, defining an El Niño event, is an extreme form of the suppression of the cold tongue in the supermodel in a climatological average. The vacillation between El Niño and La Niña regimes reduces the size of the cold tongue on average.

The extended warm pool in El Niño events, in this view, is a dissipative structure, as are the blocking patterns in the QG model, of the sort that arises in nonequilibrium forced-dissipative systems. And as with the fractal structure on the surface of Bak et al.’s sandpiles that gives rise to avalanches, the El Niño structure has a unique form, though not fractal, that disappears as one ventures away from the critical state.

Interscale interactions are thought to play a key role in the ENSO cycle and in the form of El Niño events, as they do in blocked/zonal vacillation and in blocking patterns. All models exhibit cascades (Figs. 12a–c), down to the scale of the  $4$ – $12\text{ yr}^{-1}$  Madden–Julian oscillation (MJO), which has indeed been hypothesized to trigger El Niño events (Moore and Kleeman 1999). The cascades actually continue to higher frequencies (not shown in the figure) that would include westerly wind bursts with periods of 5–20 days, which have also been thought to play such a triggering role (Harrison and Vecchi 1997). Neither of these phenomena are reproduced well either by the separate models or by the supermodel. The supermodel, however, while giving a falloff rate that is still too steep, slightly steeper than in the separate models at high frequencies, avoids the low-frequency error of the separate models, in each of which the region of rapid falloff extends to below-annual frequencies. In this regard the supermodel spectrum (Fig. 12c) is more like that of observations (Fig. 12d).

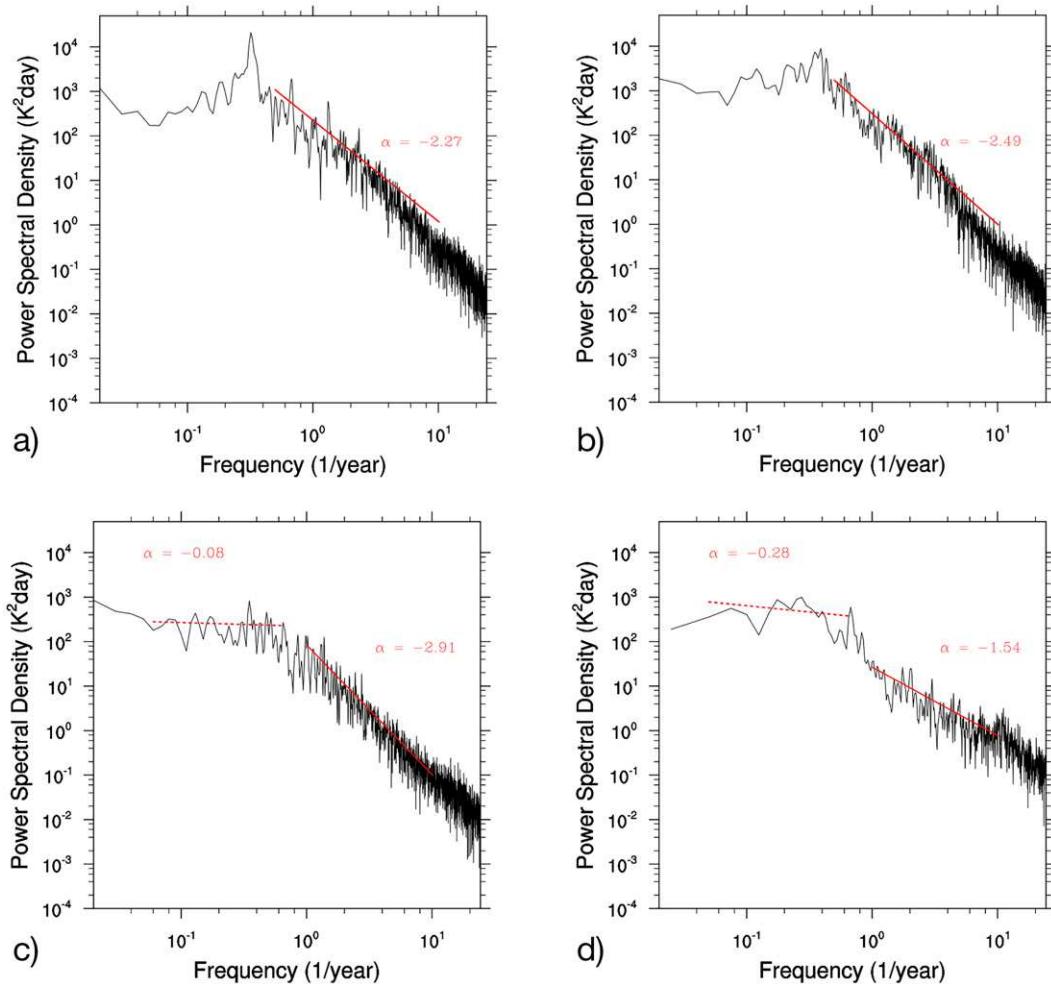


FIG. 12. Power spectral density of Niño 3.4 daily SST anomaly vs frequency, plotted logarithmically, in (a) the Nordeng model, (b) the Tiedtke model, (c) the supermodel formed from Nordeng and Tiedtke, and (d) observations for the period 1982–2021 (as in Fig. 11). Solid and dashed lines are best linear fits, each with slopes  $\alpha$  equal to the estimated critical exponent, in wavenumber ranges where the falloff of power spectral density  $P$  with frequency  $f$  is approximately exponential,  $P \approx f^\alpha$ ,  $\alpha < 0$ .

Thus, it is the form of the criticality in this case that is better represented by the supermodel than it is by either of the constituent models separately, although the latter qualitatively agree with each other in this regard, as they do in their SST patterns. The improvement in the spectrum is thought to accompany a more realistic transfer of energy between scales that underlies the phenomena producing the observed SST pattern.

## 5. Discussion

### a. Generality of supermodeling

The use of a handful of models to create a useful supermodel rests on the assumption that the set of combinations of the constituent model dynamics includes the true dynamics, or comes closer to it, than does any one constituent model. It is preferred to consider only combinations formed with non-negative coefficients or weights, since the use of negative

coefficients can lead to instability. If we envision the models as defined in a common, large parameter space, then the constituent models should form a convex hull surrounding the “true” model (e.g., Schevenhoven et al. 2019). A desirable set of constituent models can be created by design, as with the Lorenz system supermodel described in section 1, and with both of the two quasigeostrophic supermodels described in sections 2b and 3, respectively. Alternatively, a desirable model set can be constructed using models that have each been locally optimized in their respective parameter spaces, as with the Nordeng and Tiedtke versions of the COSMOS model, relying on the empirical result that model error in such situations can be viewed as random error in the large model space. That empirical result seems reasonable, as it does for the set of IPCC-class models discussed in the introduction, because the local optimization starts from choices made by different groups of modelers that arguably are

equally realistic or nearly so. Some combination of such models will likely come closer to the true dynamics.

For state-of-the-art climate models, the main differences are in the subgrid-scale parameterization schemes. The improvement obtained by supermodeling is therefore expected to be in the representation of phenomena on these scales. Further, realistic anthropogenic changes in radiative forcing are not expected to have a large impact on the combination of subgrid-scale parameterizations best suited to represent the true dynamics. So, the effective parameterizations in supermodels trained on twentieth century data are expected to be robust against climate change, requiring only the usual change in forcings and not in intermodel connections, as with the Lorenz system example in Fig. 3. Models are most sensitive to behavior on the smallest scales when they are near criticality, irrespective of the specific theory of criticality put forward by Bak et al. and their predicted power spectra. Thus, the improvement due to supermodeling is expected to be most pronounced when the system is near criticality, presently or in the projected future.

It remains to argue that similar improvement could not be obtained by averaging the outputs of the separate models, in the general case. We assert, without proof, that the occurrence of such structures as blocking patterns and a single ITCZ near and only near critical points is common. The structures are then absent in any *ex post facto* average of states in models that do not represent the critical behavior well.

#### *b. Effective parameterization of the subgrid scales in a climate supermodel*

It is natural to ask whether there is some parameterization of the subgrid-scale dynamics, intermediate between the Nordeng and Tiedtke parameterizations, that would yield the same results as the supermodel but with a single atmosphere. Indeed, it is expected that better results in the upper atmosphere and in the extratropics will be obtained by introducing connections between the atmospheres. If the atmospheres themselves then nearly synchronize, it would thus be expected that the desynchronized small scales could be used to define such an intermediate parameterization.

We argue that the lesson of the supermodeling study is that such parameterizations must be more complex, dynamical entities. By contrast, the quasigeostrophic supermodel formed from unforced and strongly forced models (11)–(14), is equivalent to a single model with a parameterization defined simply by averaging the forcing strengths, as shown in section 3. This will generalize to any supermodel formed from models that differ only in parameters that appear linearly in the prognostic equations, perhaps as multipliers of entire tendencies, provided that the nonlinear terms are only the familiar Jacobian terms for advection.

Averaging of tendencies (e.g., Wiegerinck and Selten 2017) will generally not suffice in the case of climate models with multiple nonlinearities, such as the ECHAM models with different convective parameterizations, coupled to one another by nudging with finite coefficients. With climate models that are fully connected in such standard ways, among which the large scales synchronize, there would be additional degrees of freedom at the smallest scales and, in general, no definition of

supermodel variables, known *a priori*, that would satisfy a modified set of prognostic equations. We recall that not all variables can synchronize in systems that satisfy different equations given by different subgrid-scale parameterizations, if coefficients remain finite. If the large- and medium-scale variables synchronize, the smallest-scale variables cannot. The only recourse might be a limit of infinite nudging coefficients in which all scales are nudged and synchronize, with nudging terms  $C_{ij}(\mathbf{x}_i - \mathbf{x}_j)$  that remain finite as  $C_{ij} \rightarrow \infty$  and  $\mathbf{x}_i - \mathbf{x}_j \rightarrow 0$ , to compensate for the model differences. In the limiting case, such finite terms can be taken as dynamical variables that satisfy their own prognostic equations. Thus, a combined parameterization is only possible, in general, at the cost of introducing extra degrees of freedom, whether we introduce desynchronized variables or use  $C_{ij} \rightarrow \infty$ .

It is interesting to compare our method with that of stochastic parameterization, in which random variables are added (Palmer et al. 2009; Berner et al. 2017). Specifically, one could consider random variables that are sums of multiplicative noise terms, each random term based on one of the models' instantaneous parameterization of unresolved processes. Here the new variables are deterministically related to the entire history of the models' states.

Extra degrees of freedom are needed, of course, if we assume a need to represent the physical processes occurring at subgrid scales explicitly. The usual approach would be to increase the spatial resolution of the grid. But we note that we have achieved a single ITCZ by at most doubling the number of degrees of freedom through the use of a separate model. A similar increase in dynamical dimension in the increased resolution approach would require that resolution be increased by  $\sqrt{2}$ , at most, in each horizontal dimension. That such a small increase would suffice to qualitatively change the tropical SST pattern seems doubtful.

Thus, supermodeling gives an economy of extra degrees of freedom. If we assume, for finite nudging, that synchronized and desynchronized variables segregate according to scale, as is common, we can have unique variables to describe the large and medium scales, corresponding to the putatively synchronized variables in our supermodel. This subsystem is analogous to the single ocean in the COSMOS supermodel. But there would be two or more models of the smallest resolved scales, which need not be connected to one another, but would be connected to the larger scales with variable weights. The proposed architecture, schematized in Fig. 13, emerges as the form of effective parameterization that directly extends the supermodel approach to combining models.

Near criticality, the two-way interaction between the smallest scales and the larger scales is especially important. The supermodel approach, in our construction, is useful for two related reasons: 1) it uses prior modeling experience to effectively isolate the dimensions in a high-dimensional parameter space along which there is ambiguity in how to parameterize the subgrid scales, so that optimization can be restricted to those dimensions—by training intermodel connections, as in conventional supermodeling; 2) it uses those dimensions to define extra degrees of freedom, in a dynamical parameterization, that capture the part of the subgrid-scale dynamics that is relevant to interscale interactions in a critical state.

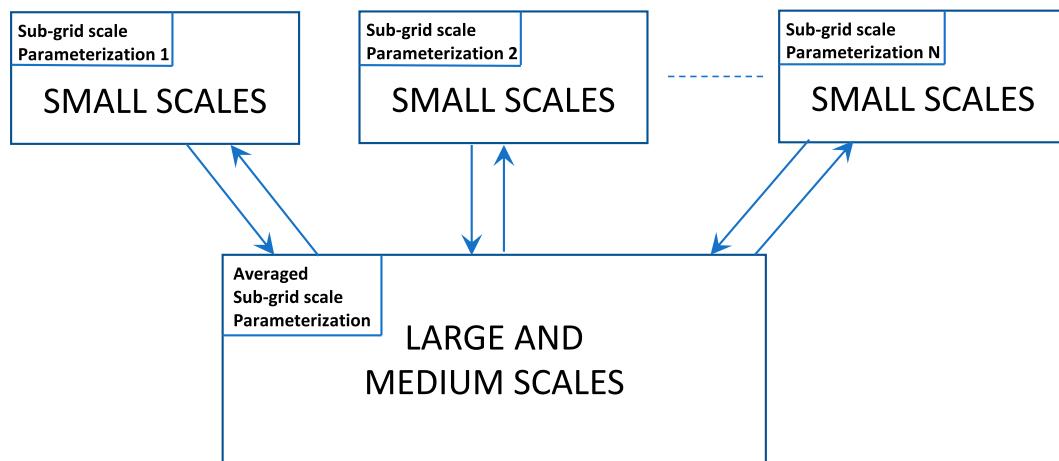


FIG. 13. The form of the effective parameterization of subgrid-scale processes suggested by supermodeling.  $N$  copies of the smallest scales are connected to the larger scales in the usual way, except impacting them through weighted averages of their own state variables in the larger-scale prognostic equations. The small-scale models each use a different parameterization of the subgrid-scale processes but are otherwise identical. Where the equations for the larger-scale variables refer to the subgrid-scale processes directly, an averaged parameterization is used.

## 6. Concluding remarks—Beyond self-organized criticality

We have used the Bak et al. (1987) analysis as a reference point for our investigation of criticality in the two systems studied here. However, we are not wedded to the specific log-linear form of the spectrum that is expected in the original concept of self-organized criticality. That form has been criticized (Jaeger et al. 1989) as not even descriptive of the behavior of real sandpiles, a primary example in the original proposal, and is too specific. A multifractal form applies in the more general case discussed by Kadanoff et al. (1989), which may be piecewise linear or even given by a continuously varying exponent in the putative power law. Correspondingly, the spatial structures that arise intermittently near a critical state have a unique form, but one more general than that implied by the fractal structure on the surface of Bak et al.'s sandpiles—a generalization that has not heretofore received attention. Blocking patterns and extended warm pools can be viewed as examples of the “dissipative structures” that appear in Prigogine's (Nicolis and Prigogine 1977; Prigogine and Stengers 1979) qualitative approach to the nonequilibrium thermodynamics of open forced-dissipative systems, in which fluctuations play a large role near critical points, as here. But general rules about the form of the structures and their intermittency near a critical point have not been elucidated. In the absence of a general theory, our point about the scaling is simply that however the true power spectrum deviates from the idealized SOC form, the supermodel spectrum is closer to it than is the spectrum of any of the separate models from which the supermodel is constructed.

Specifically, for the toy quasigeostrophic model, the spectrum (Fig. 10) has a form suggestive of the expected log-linear behavior over a significant range of frequencies and energies. For the constituent models, the behavior appears piecewise multifractal (for strong forcing) at best, or has the expected power-law form over energies and length scales that are too small to be of interest (for no forcing). The multifractal form is in fact common in

geophysical turbulence (Schertzer and Lovejoy 2011). For the tropical Pacific SST and its detailed dynamical models, the spectrum (Fig. 12) of the real system exhibits log-linear behavior over a range that is more restricted than in simple models. In reality, and in the supermodel, specific dynamical processes appear to limit the power-law form at the low end of that range—processes that are indeed key to the ENSO phenomenon as known. Thus, while any claim to universality of the power-law form is exaggerated, that form provides a benchmark for the comparison of observed and modeled behavior.

What we seek to preserve of self-organized criticality is the qualitative assertion that natural systems tend toward critical states, and toward unique, temporally intermittent spatial structures associated with those states. For the two examples given, that would mean that processes that are not represented in the models or are not fully resolved effectively set the values of the forcing strength or the flux weights, respectively, to give criticality and associated structures. The power spectra confirm this tendency, in view of the expected scale relationships near such states. Beyond the scope of this paper is a possible connection with the idea that such critical states arise so as to maximize heat transport from the equator to the poles and the associated principle of maximum entropy production (Paltridge 1975), another qualitative principle in climate science that has yet to yield verifiable predictions. As the proponents of SOC commented in response to criticism, their claim was that SOC could explain power-law behavior when it was observed, not that it universally implied power-law behavior (Bak et al. 1989).

In any case, phenomena of interest in weather and climate tend to involve critical states, and that does explain the need for the methods of this paper. It is to be remembered that subgrid-scale processes to which critical states are sensitive are dynamical processes that are only approximately, not completely slaved to the larger scales. Especially near criticality, they cannot be correctly represented by any diagnostic parameterization. Such

situations are common if not universal. Criticality was achieved in the present work not by design, but by crudely training a small number of coefficients combining different models, each of them arguably physical, against observed data. It appears that nature has chosen criticality over simplicity in regard to approximate parameterization schemes. Thus, where limited computational resources necessitate a compact representation of subgrid-scale processes, either supermodeling or other use of empirical methods (i.e., learning) to combine alternative parameterizations is in order.

**Acknowledgments.** Michael Ghil is thanked for useful discussions regarding self-organized criticality and synchronization generally. Andrew Mai is thanked for assistance with software modifications to adapt the connection coefficient in the quasi-geostrophic supermodel. The original version of the code used herein for a single QG channel model, and the original suggestion for its use, are due to Joe Tribbia, to whom we are grateful. We thank the three anonymous reviewers for their constructive contributions and the editor, Peter Bartello, for his diligent consideration of all points of view. This work was supported by the European Commission (EU FP7 Grant 266722, ERC Grant 648982 and Marie Skłodowska-Curie Actions—Individual Fellowship Grant 658602), the U.S. Department of Energy (Grant DE-SC0005238), and the National Science Foundation (Grant 2015618).

**Data availability statement.** Software and data relevant to this publication are archived at <https://zenodo.org/communities/duaneshenjas2022data>.

## APPENDIX

### Synchronization-Based Optimization of Connections in a PDE Supermodel

The method used in this work to optimize connections in a supermodel is a variant of the approach, well known in data assimilation, of adjoining parameters to state vectors so as to estimate parameter values based on observed data. We consider data assimilation to be an instance of the synchronization of two dynamical systems. Here one dynamical system, the model, is synchronized with another dynamical system, “truth,” which influences the first system via nudging or other form of connection through a reduced set of corresponding dynamical variables. The evolution of the true system is given by

$$\frac{dx_l}{dt} = f_l^0(\mathbf{x}) + pf_l^p(\mathbf{x}), \quad (\text{A1})$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  is the  $n$ -dimensional state vector. The evolution of the model system,  $\mathbf{x}^M(t)$  is described similarly:

$$\frac{dx_l^M}{dt} = f_l^0(\mathbf{x}^M) + p^M f_l^p(\mathbf{x}^M) + c_l(\mathbf{x}, \mathbf{x}^M), \quad (\text{A2})$$

where the dynamics of the true system is assumed known, except for the value of the parameter  $p$ , and an extra term  $c_l(\mathbf{x}, \mathbf{x}^M)$  has been added to represent nudging or other form of data assimilation.

Let the Lyapunov function  $L(\mathbf{x}, \mathbf{x}^M; t)$  represent the error in the model’s state at time  $t$ . Commonly,  $L = \sum_l [x_l(t) - x_l^M(t)]^2$ .

If the model system (A2) synchronizes with the true system (A1), i.e.,  $L \rightarrow 0$ , regardless of initial states, for a correct choice of parameter  $p^M = p$ , we add an equation for adaptation of the parameter along with states:

$$\frac{dp^M}{dt} = -a \sum_l \frac{\partial L}{\partial e_l} f_l^p, \quad (\text{A3})$$

where  $e_l \equiv x_l^M - x_l$ , the sum is over all  $l$  such that the prognostic equation for the state variable  $x_l$  contains the parameter  $p$ , and  $a$  is an arbitrary learning rate. The rule (A3) can be rigorously shown to give asymptotic values of the model parameter that allows the states to synchronize (Duane et al. 2007; Selten et al. 2017).<sup>5</sup>

For a supermodel of the form depicted in Fig. 1, the parameters to be estimated are the connection coefficients  $C_{ij}^l$ , while the parameters of the constituent models remain fixed. For this to be possible, according to the above scheme, it is only necessary that for some choice of connection coefficients, the supermodel will synchronize with “reality” for some choice of nudging coefficients contained in the matrix  $\mathbf{K}$ , the one we will use to train the supermodel. In general, there is more than one way to define the state error  $L$ . For a supermodel composed of  $m$  models, two broad categories are to define the error for each constituent model and then average, i.e.,  $L \equiv (1/m) \sum_l \sum_{k=1}^m (x_l - x_l^k)^2$ , or alternatively to first define supermodel state variables, e.g., as the average of corresponding state variables in the constituent models (equal to the variables in any one model for variables that synchronize), so that  $L \equiv \sum_l [x_l - (1/m) \sum_{k=1}^m x_l^k]^2$ . Using the latter definition, we find that to estimate  $p \equiv C_{ij}^l$ , we set coefficient  $f_l^p = x_l^j - x_l^i$  and find  $\partial L / \partial e_l = [(1/m) \sum_{k=1}^m x_l^k] - x_l$ , so that the general connection adaptation rule is

$$\dot{C}_{ij}^l = -a \left( \frac{1}{m} \sum_{k=1}^m x_l^k - x_l \right) (x_l^j - x_l^i), \quad (\text{A4})$$

the generalization of the training rule (2) used for the Lorenz system supermodel.

For the quasigeostrophic supermodels, one needs to generalize (A3) to PDEs. This is done by regarding a PDE as a collection of ODE’s, either one for each point in space or one for each spectral component. For the supermodel formed from models with forcing at different locations, given by (4)–(6), the parameter to be estimated is  $c$ , the coefficient of the advective coupling term. The same parameter appears in the two Eqs. (4) for every point in space. So there are two terms in the sum in (A3) for each point in space and an integral over space. The Lyapunov function we choose is a sum of error relative to truth over the two models separately, giving factors of  $q^A - q_{\text{obs}}$  and  $q^B - q_{\text{obs}}$ .

<sup>5</sup> The rule that appears in both references is stated for dynamical equations with any form of dependence on the parameters, but a linear dependence, as in (A1) and (A2), was assumed in the derivation, so we can express the rule as (A3) without loss of generality. Selten et al. (2017) address the restricted case where the Lyapunov function  $L$  has the usual RMS form.

The cofactors  $f^p$  are  $J(\psi^A, q^B - q^A)$  and  $J(\psi^B, q^A - q^B)$ . The resulting parameter adaptation rule is

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

which is Eq. (7).

For the supermodel composed of QG models with different forcing strengths, the different spectral components are treated differently. One can simplify the problem of optimizing the supermodel given by (11)–(14) by using the demonstrated equivalence to a simple model with averaged forcing, given by (16). We rewrite (16) as

$$\begin{aligned} \frac{D\hat{q}}{Dt} \Big|_{\mathbf{k}} = & [w^A \mu_{\mathbf{k}}^{\text{ext } A} + (1 - w^A) \mu_{\mathbf{k}}^{\text{ext } B}] (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}}) \\ & + [w^A D_{\mathbf{k}}^A (q^A) + (1 - w^A) D_{\mathbf{k}}^B (q^B)]. \end{aligned} \quad (\text{A5})$$

---


$$\frac{D\hat{q}}{Dt} \Big|_{\mathbf{k}} = \begin{cases} [w^A \mu_o^A + (1 - w^A) \mu_o^B] (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}}) + D_{\mathbf{k}}(\hat{q}) & \text{if } |k_x| \leq k_{x0} \text{ and } |k_y| \leq k_{y0} \\ \nu [w^A \mu_o^A + (1 - w^A) \mu_o^B] [(k_{\text{noise}}/|\mathbf{k}^\perp|)^4] (q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}}) + D_{\mathbf{k}}(\hat{q}) & \text{if } |\mathbf{k}^\perp| > k_{\text{noise}} \\ \nu [w^A \mu_o^A + (1 - w^A) \mu_o^B] [1 - (k_0/|\mathbf{k}^\perp|)^4] (q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}}) & \text{otherwise} \\ + [w^A \mu_o^A + (1 - w^A) \mu_o^B] (k_0/|\mathbf{k}^\perp|)^4 (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}}) + D_{\mathbf{k}}(\hat{q}) \end{cases}, \quad (\text{A6})$$


---

so the cofactors of  $w^A$  are modified as follows:

---


$$f_{\mathbf{k}}^{w^A} = \begin{cases} (\mu_o^A - \mu_o^B) (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}}) & \text{if } |k_x| \leq k_{x0} \text{ and } |k_x| \leq k_{x0} \\ \nu (\mu_o^A - \mu_o^B) (q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}}) & \text{if } |\mathbf{k}^\perp| > k_{\text{noise}} \\ \nu (\mu_o^A - \mu_o^B) [1 - (k_0/|\mathbf{k}^\perp|)^4] (q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}}) & \text{otherwise} \\ + (\mu_o^A - \mu_o^B) (k_0/|\mathbf{k}^\perp|)^4 (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}}) \end{cases}, \quad (\text{A7})$$


---

An adequate error metric is based on differences only in the large and medium scales:

$$L_o \equiv \sum_{\|\mathbf{k}\| < k_{\text{noise}}} (\hat{q}_{\mathbf{k}} - q_{\mathbf{k}}^{\text{obs}})^2, \quad (\text{A8})$$

---


$$\frac{dw_A}{dt} = a \sum_{|k_x| \leq k_{x0}, |k_y| \leq k_{y0}} (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}}) (q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}}) + a \sum_{\|\mathbf{k}\| < k_{\text{noise}}, (|k_x| > k_{x0} \text{ or } |k_y| > k_{y0})} \{ (k_0/|\mathbf{k}^\perp|)^4 (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}}) (q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}}) + \nu [1 - (k_0/|\mathbf{k}^\perp|)^4] (q_{\mathbf{k}}^{\text{obs}} - \hat{q}_{\mathbf{k}})^2 \},$$

(A9)

which is Eq. (17), where we have absorbed other constant factors into the general learning rate factor  $a$ .

## REFERENCES

Afraimovich, V. S., N. N. Verichev, and M. I. Rabinovich, 1986: Stochastic synchronization of oscillation in dissipative systems.

There is only one free parameter  $p = w_A$  to be determined, having taken  $w_B = 1 - w_A$ . We assume, as in Duane and Tribbia (2004), that the dissipative terms  $D^A$  and  $D^B$  are linear and are of the same form, so we can simplify  $w^A D^A (q^A) + (1 - w^A) D^B (q^B) = D(w^A q^A + w^B q^B) = D(\hat{q})$  and there is no net dependence in these terms on the parameter  $w^A$ . The cofactor of  $w_A$  in (A5), in each large-scale equation where  $w_A$  appears, is  $(\mu_o^A - \mu_o^B) (q_{\mathbf{k}}^* - \hat{q}_{\mathbf{k}})$ . However, in coupling the model to a “true” system to optimize  $w^A$ , we need to include a nudging term in (A5). While such a term would usually be independent of the parameter to be estimated, as in previous examples, in the present case the nudging strength (in the medium scales) must be kept proportional to the external forcing (in the large scales) for dynamical consistency. With nudging to “truth” included in the forcing for the medium scales with a proportionality factor  $\nu$ , (A5) becomes

so only the large- and medium-scale cofactors contribute to the product in (A3), yielding the adaptation rule for the weights in the strong/weak forcing QG supermodel:

*Radiophys. Quantum Electron.*, **29**, 795–803, <https://doi.org/10.1007/BF01034476>.

Andrade, J. S., Jr., I. Wainer, J. M. Filho, and J. E. Moreira, 1995: Self-organized criticality in the El Niño Southern Oscillation. *Physica A*, **215**, 331–338, [https://doi.org/10.1016/0378-4371\(95\)00004-Q](https://doi.org/10.1016/0378-4371(95)00004-Q).

Bak, P., C. Tang, and K. Wiesenfeld, 1987: Self-organized criticality: An explanation of 1/f noise. *Phys. Rev. Lett.*, **59**, 381–384. <https://doi.org/10.1103/PhysRevLett.59.381>.

—, —, and —, 1989: Comment on “Relaxation at the angle of repose.” *Phys. Rev. Lett.*, **62**, 110, <https://doi.org/10.1103/PhysRevLett.62.110>.

Berner, J., and Coauthors, 2017: Stochastic parameterization: Toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–588, <https://doi.org/10.1175/BAMS-D-15-00268.1>.

Branstator, G., 1995: Organization of storm-track anomalies by recurring low-frequency circulation anomalies. *J. Atmos. Sci.*, **52**, 207–226, [https://doi.org/10.1175/1520-0469\(1995\)052<0207:OOSTAB>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<0207:OOSTAB>2.0.CO;2).

Counillon, F., N. Keenlyside, S. Wang, M. Devilliers, A. Gupta, S. Koseki, and M.-L. Shen, 2023: Framework for an ocean-connected supermodel of the Earth system. *J. Adv. Model. Earth Syst.*, **15**, e2022MS003310, <https://doi.org/10.1029/2022MS003310>.

Du, H., and L. A. Smith, 2017: Multi-model cross-pollination in time. *Physica D*, **353–354**, 31–38, <https://doi.org/10.1016/j.physd.2017.06.001>.

Duane, G. S., 2015: Synchronicity from synchronized chaos. *Entropy*, **17**, 1701–1733, <https://doi.org/10.3390/e17041701>.

—, and J. J. Tribbia, 2001: Synchronized chaos in geophysical fluid dynamics. *Phys. Rev. Lett.*, **86**, 4298–4301, <https://doi.org/10.1103/PhysRevLett.86.4298>.

—, and —, 2004: Weak Atlantic–Pacific teleconnections as synchronized chaos. *J. Atmos. Sci.*, **61**, 2149–2168, [https://doi.org/10.1175/1520-0469\(2004\)061<2149:WATASC>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<2149:WATASC>2.0.CO;2).

—, —, and J. B. Weiss, 2006: Synchronicity in predictive modelling: A new view of data assimilation. *Nonlinear Processes Geophys.*, **13**, 601–612, <https://doi.org/10.5194/npg-13-601-2006>.

—, D.-C. Yu, and L. Kocarev, 2007: Identical synchronization, with translation invariance, implies parameter estimation. *Phys. Lett.*, **371A**, 416–420, <https://doi.org/10.1016/j.physleta.2007.06.059>.

—, C. Grabow, F. Selten, and M. Ghil, 2017: Introduction to focus issue: Synchronization in large networks and continuous media—Data, models, and supermodels. *Chaos*, **27**, 126601, <https://doi.org/10.1063/1.5018728>.

Fujisaka, H., and T. Yamada, 1983: Stability theory of synchronized motion in coupled-oscillator systems. *Prog. Theor. Phys.*, **69**, 32–47, <https://doi.org/10.1143/PTP.69.32>.

Ghil, M., and S. Childress, 1987: *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics*. Applied Mathematical Sciences, Vol. 60, Springer-Verlag, 485 pp.

Giorgetta, M. A., and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572–597, <https://doi.org/10.1002/jame.20038>.

Harrison, D. E., and G. A. Vecchi, 1997: Westerly wind events in the tropical Pacific, 1986–95. *J. Climate*, **10**, 3131–3156, [https://doi.org/10.1175/1520-0442\(1997\)010<3131:WWEITT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<3131:WWEITT>2.0.CO;2).

Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H.-M. Zhang, 2021: Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) version 2.1. *J. Climate*, **34**, 2923–2939, <https://doi.org/10.1175/JCLI-D-20-0166.1>.

Jaeger, H. M., C.-H. Liu, and S. R. Nagel, 1989: Relaxation at the angle of repose. *Phys. Rev. Lett.*, **62**, 40–43, <https://doi.org/10.1103/PhysRevLett.62.40>.

Jin, F.-F., 1997: An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *J. Atmos. Sci.*, **54**, 811–829, [https://doi.org/10.1175/1520-0469\(1997\)054<0811:AEORPF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2).

—, 1998: A simple model for the Pacific cold tongue and ENSO. *J. Atmos. Sci.*, **54**, 2458–2469, [https://doi.org/10.1175/1520-0469\(1998\)055<2458:ASMFTP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<2458:ASMFTP>2.0.CO;2).

Kadanoff, L. P., S. R. Nagel, C.-H. Liu, and S. M. Zhou, 1989: Scaling and universality in avalanches. *Phys. Rev. A*, **39**, 6524–6537, <https://doi.org/10.1103/PhysRevA.39.6524>.

Kirtman, B. P., 1997: Oceanic Rossby wave dynamics and the ENSO period in a coupled model. *J. Climate*, **10**, 1690–1704, [https://doi.org/10.1175/1520-0442\(1997\)010<1690:ORWDAT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1690:ORWDAT>2.0.CO;2).

—, and J. Shukla, 2002: Interactive coupled ensemble: A new coupling strategy for CGCMs. *J. Geophys. Res. Lett.*, **29**, 1367, <https://doi.org/10.1029/2002GL014834>.

—, D. Min, P. S. Schopf, and E. K. Schneider, 2003: A new approach for CGCM sensitivity studies. COLA Tech. Rep. 154, 50 pp., <http://www.m.monsoondata.org/pubs/tech.html>.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).

Mechoso, C. R., and Coauthors, 1995: The seasonal cycle over the tropical Pacific in coupled ocean–atmosphere general circulation models. *Mon. Wea. Rev.*, **123**, 2825–2838, [https://doi.org/10.1175/1520-0493\(1995\)123<2825:TSCOTT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<2825:TSCOTT>2.0.CO;2).

Met Office Hadley Centre, 2013: HadISST1 data. Accessed 5 February 2013, <https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>.

Mirchev, M., G. S. Duane, W. K. Tang, and L. Kocarev, 2012: Improved modeling by coupling imperfect models. *Commun. Nonlinear Sci. Numer. Simul.*, **17**, 2741–2751, <https://doi.org/10.1016/j.cnsns.2011.11.003>.

Moore, A. M., and R. Kleeman, 1999: Stochastic forcing of ENSO by the intraseasonal oscillation. *J. Climate*, **12**, 1199–1220, [https://doi.org/10.1175/1520-0442\(1999\)012<1199:SFOEBT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1199:SFOEBT>2.0.CO;2).

Namias, J., 1951: The index cycle and its role in the general circulation. *J. Meteor.*, **7**, 130–139, [https://doi.org/10.1175/1520-0469\(1950\)007<0130:TICAIR>2.0.CO;2](https://doi.org/10.1175/1520-0469(1950)007<0130:TICAIR>2.0.CO;2).

Nicolis, G., and I. Prigogine, 1977: *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*. Wiley, 491 pp.

Physical Sciences Laboratory, 2013: GPCP version 2.3 combined precipitation data set. NOAA/OAR/ESRL, accessed 5 February 2013, <https://psl.noaa.gov/data/gridded/data.gpcp.html>.

—, 2016: NOAA OI SST V2 high resolution dataset. NOAA/OAR/ESRL, accessed 21 April 2022, <https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html>.

Nordeng, T. E., 1994: Extended versions of the convective parametrization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics. ECMWF Tech. Memo. 206, 42 pp., <https://www.ecmwf.int/en/elibrary/75843-extended-versions-convective-parametrization-scheme-ecmwf-and-their-impact-mean>.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 44 pp., [www.ecmwf.int/sites/default/files/elibrary/2009/11577-stochastic-parametrization-and-model-uncertainty.pdf](http://www.ecmwf.int/sites/default/files/elibrary/2009/11577-stochastic-parametrization-and-model-uncertainty.pdf).

Paltridge, G. W., 1975: Global dynamics and climate—A system of minimum entropy exchange. *Quart. J. Roy. Meteor. Soc.*, **101**, 475–484, <https://doi.org/10.1002/qj.49710142906>.

Pecora, L. M., and T. L. Carroll, 1990: Synchronization in chaotic systems. *Phys. Rev. Lett.*, **64**, 821–824, <https://doi.org/10.1103/PhysRevLett.64.821>.

—, —, G. A. Johnson, D. J. Mar, and J. F. Heagy, 1997: Fundamentals of synchronization in chaotic systems, concepts, and applications. *Chaos*, **7**, 520–543, <https://doi.org/10.1063/1.166278>.

Prigogine, I., and I. Stengers, 1979: *The New Alliance*. Gallimard, 312 pp.

Reichler, T., and J. Kim, 2008: How well do coupled models simulate today's climate? *Bull. Amer. Meteor. Soc.*, **89**, 303–311, <https://doi.org/10.1175/BAMS-89-3-303>.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. *J. Climate*, **20**, 5473–5496, <https://doi.org/10.1175/2007JCLI1824.1>.

Rulkov, N. F., M. M. Sushchik, and L. S. Tsimring, 1995: Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E*, **51**, 980–994, <https://doi.org/10.1103/PhysRevE.51.980>.

Schertzer, D., and S. Lovejoy, 2011: Multifractals, generalized scale invariance, and complexity in geophysics. *Int. J. Bifurcation Chaos*, **21**, 3417–3456, <https://doi.org/10.1142/S0218127411030647>.

Schevenhoven, F., and A. Carrassi, 2022: Training a supermodel with noisy and sparse observations: A case study with CPT and the synch rule on SPEEDO. *Geosci. Model Dev.*, **15**, 3831–3844, <https://doi.org/10.5194/gmd-15-3831-2022>.

—, F. Selten, A. Carrassi, and N. Keenlyside, 2019: Improving weather and climate predictions by training of supermodels. *Earth Syst. Dyn.*, **10**, 789–807, <https://doi.org/10.5194/esd-10-789-2019>.

Selten, F. M., F. J. Schevenhoven, and G. S. Duane, 2017: Simulating climate with a synchronization-based supermodel. *Chaos*, **27**, 126903, <https://doi.org/10.1063/1.4990721>.

Severijns, C. A., and W. Hazleger, 2010: The efficient global primitive equation climate model SPEEDO V2.0. *Geosci. Model Dev.*, **3**, 105–122, <https://doi.org/10.5194/gmd-3-105-2010>.

Shen, M.-L., N. Keenlyside, F. M. Selten, W. Wiegerinck, and G. S. Duane, 2016: Dynamically combining climate models to “supermodel” the tropical Pacific. *Geophys. Res. Lett.*, **43**, 359–366, <https://doi.org/10.1002/2015GL066562>.

—, —, B. C. Bhatt, and G. S. Duane, 2017: Role of atmosphere-ocean interactions in supermodeling the tropical Pacific climate. *Chaos*, **27**, 126704, <https://doi.org/10.1063/1.4990713>.

Tanguay, M., P. Bartello, and P. Gauthier, 1995: Four-dimensional data assimilation with a wide range of scales. *Tellus*, **47A**, 974–997, <https://doi.org/10.3402/tellusa.v47i5.11967>.

Temam, R., 1988: *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*. Springer-Verlag, 650 pp.

Tiedtke, M., 1989: A comprehensive mass-flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2).

van den Berge, L. A., F. M. Selten, W. Wiegerinck, and G. S. Duane, 2011: A multi-model ensemble method that combines imperfect models through learning. *Earth Syst. Dyn.*, **2**, 161–177, <https://doi.org/10.5194/esd-2-161-2011>.

Vautard, R., and B. Legras, 1988: On the source of midlatitude low-frequency variability. Part II: Nonlinear equilibration of weather regimes. *J. Atmos. Sci.*, **45**, 2845–2867, [https://doi.org/10.1175/1520-0469\(1988\)045<2845:OTSOML>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<2845:OTSOML>2.0.CO;2).

—, —, and M. Déqué, 1988: On the source of midlatitude low-frequency variability. Part I: A statistical approach to persistence. *J. Atmos. Sci.*, **45**, 2811–2843, [https://doi.org/10.1175/1520-0469\(1988\)045<2811:OTSOML>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<2811:OTSOML>2.0.CO;2).

Wiegerinck, W., and F. M. Selten, 2017: Attractor learning in synchronized chaotic systems in the presence of unresolved scales. *Chaos*, **27**, 126901, <https://doi.org/10.1063/1.4990660>.

Yang, S.-C., and Coauthors, 2006: Data assimilation as synchronization of truth and model: Experiments with the three-variable Lorenz system. *J. Atmos. Sci.*, **63**, 2340–2354, <https://doi.org/10.1175/JAS3739.1>.

Zebiak, S. E., and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.*, **115**, 2262–2278, [https://doi.org/10.1175/1520-0493\(1987\)115<2262:AMENO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<2262:AMENO>2.0.CO;2).

Zhang, X., W. Lin, and M. Zhang, 2007: Toward understanding the double intertropical convergence zone pathology in coupled ocean-atmosphere general circulation models. *J. Geophys. Res.*, **112**, D12102, <https://doi.org/10.1029/2006JD007878>.