1

STOCHASTIC SUPERPARAMETERIZATION IN A ONE-DIMENSIONAL MODEL FOR WAVE TURBULENCE

IAN G. GROOMS [†] AND ANDREW J. MAJDA [‡]

Abstract. Superparameterization is a multiscale numerical method wherein solutions of prognostic equations for small scale processes on local domains embedded within the computational grid of a large scale model are computed and used to force the large scales. It was developed initially in the atmospheric sciences, but stands on its own as a nascent numerical method for the simulation of multiscale phenomena. Here we develop a stochastic version of superparameterization in a difficult one dimensional test problem involving self-similarly collapsing solitons, dispersive waves, and a turbulent inverse cascade of energy from small to large scales. We derive the nonlinear model equations by imposing a formal scale separation between resolved large scales and unresolved small scales; this allows the use of subdomains embedded within the large scale grid to describe the local small scale processes. To decrease the computational cost, we make a systematic quasi-linear stochastic approximation of the nonlinear small scale equations and use the statistical mean of the nonlinear small scale forcing (the covariance) in the large scale equations. The stochastic approximation allows the embedded domains to be formally infinite (unrealistically large scales are suppressed on the embedded domains). Further simplifications allow us to precompute the small scale forcing terms in the large scale equations as functions of the large scale variables only, which results in significant computational savings. The results are positive. The method increases the energy in overdamped simulations, decreases the energy in underdamped simulations, and improves the spatial distribution and frequency of collapsing solitons.

Key words. Intermittent inverse cascade, no scale separation, stochastic modeling

AMS subject classifications. 76F55

Complex interactions across a wide range of spatial and temporal scales represent a major challenge for computational physics in a variety of settings. Direct resolution of all relevant scales is often impossible even with the current generation of supercomputers; even if it were possible, it would in many cases generate more information than desired. Lacking the ability to resolve all dynamically relevant scales, or interested only in the largest ones, one is led to model the interaction of resolved and unresolved scales.

Models of unresolved small, fast scales proceed by separating the dynamics into 'mean' and 'fluctuation' (or 'eddy') components, where the former is directly resolved and the latter is not. Separation into mean and eddies can proceed in a variety of ways, for example by low- and high-pass filtering or by ensemble averaging (although the latter is not guaranteed to result in a 'mean' with resolvably large scales). The equations governing the mean component generally involve contributions from the eddy component, which must be modeled in such a way as to achieve a faithful representation of the evolution of the mean.

This difficulty of parameterizing unresolved processes has been a particular concern of the atmospheric modeling community. To illustrate the specific example of unresolved cloud processes we quote [1],

> "The representation of cloud processes in global atmospheric models has been recognized for decades (e.g. Arakawa 1975; Charney 1979; Houghton et al. 2001) as the source of much uncertainty surrounding predictions of climate variability. Despite the best efforts of our community, and notwithstanding the achievement of significant advances ...the problem remains largely unsolved. ...A new and different strategy is needed"

The "new and different strategy" referred to above was first proposed by [2]. They embedded high resolution simulations of the unresolved cloud processes within each vertical column of grid cells in a low resolution, large scale atmospheric model. These embedded high resolution simulations are horizontally periodic and not directly connected to each other which allows them to be solved in parallel, making good use of the massively parallel architecture of contemporary supercomputers. The strategy of using local, embedded, high resolution simulations to model the interaction of large and small scales is termed 'superparameterization.' It has been tested in a variety of atmospheric and oceanic contexts, e.g. [2–8], with good results.

Despite the high degree of parallelism afforded by making the embedded high resolution simulations horizontally periodic and disconnected from each other, the cost of using fully three dimensional embedded domains was deemed excessive by [2], who made use of two dimensional embedded domains instead; their use of two dimensional embedded domains has been imitated by many subsequent authors [3–9]. Another way to reduce the cost and potentially improve the efficiency of superparameterization is to run the embedded simulations on domains that do not fill the spatial and/or temporal grid of the large scale model [2, 6, 9]. The success of embedded models with reduced dimensionality (2D versus 3D) and reduced spatial and temporal extent suggests that in many cases the exact details of the eddies are not needed to correctly capture the impact of the eddies on the mean. The process of simplifying the embedded eddy model was taken in a new direction by [10], who proposed to use a linear stochastic model of the eddy dynamics; their ideas were further expanded and generalized in [11].

In this article we expand upon the work of [10, 11] by implementing a stochastic superparameterization in the idealized one dimensional model of wave turbulence. The parameterization is spectacularly successful, resulting in a good approximation of the mean dynamics in a difficult setting with breaking solitons and an inverse cascade of energy from unresolved scales [12–17]. The parameterization is extremely efficient and robust, retaining high performance even with embedded domains much smaller than the large scale grid; it increases the energy of overdamped solutions, decreases the energy of underdamped solutions, and improves the spatial and temporal distribution of breaking solitons in low resolution simulations.

In section 1 we introduce the one dimensional test model, and we set up the basis for a stochastic superparameterization in section 2. In section 3 we describe numerical experiments testing the accuracy, efficiency, and robustness of our stochastic superparameterization. We conclude in section 4.

1. MMT: A one-dimensional test model   The test model used in this paper is a member of the family of dynamical equations introduced in [12]

$$i\partial_t\psi = |\partial_x|^\alpha\psi + \lambda|\partial_x|^{-\beta/4}\left(||\partial_x|^{-\beta/4}\psi|^2|\partial_x|^{-\beta/4}\psi\right), \lambda = \pm 1, \qquad (1.1)$$

where $\psi(x,t)$ is a complex scalar. In a periodic or infinite domain, the fractional derivative may be simply defined by $\widehat{|\partial_x|^\alpha\psi} = |k|^\alpha\hat{\psi}_k$ where $\widehat{(\cdot)}$ denotes the Fourier

transform and k is a wavenumber. The system is referred to as 'focusing' for $\lambda = -1$ and 'defocusing' for $\lambda = 1$. The motivation for the introduction of the MMT[1] equation in [12] was to provide a setting simple enough to test thoroughly the predictions of weak turbulence theory; for our purposes, the model provides a chaotic, turbulent dynamical system with coherent structures, waves, and forward and inverse energy cascades over a very broad range of scales while remaining accessible to high resolution simulation [12–17].

The inviscid, unforced MMT equation is Hamiltonian with

$$H = H_L + H_{NL} = \int \left| |\partial_x|^{\alpha/2} \psi \right|^2 \mathrm{d}x + \frac{\lambda}{2} \int \left| |\partial_x|^{-\beta/4} \psi \right|^4 \mathrm{d}x. \tag{1.2}$$

In addition to the Hamiltonian, it also conserves wave action N and momentum P

$$N = \int |\hat{\psi}_k|^2 \mathrm{d}k = \int |\psi|^2 \mathrm{d}x \text{ and } P = \int \left(\psi(\partial_x \psi)^* - (\partial_x \psi)\psi^* \right) \mathrm{d}x \tag{1.3}$$

where $*$ denotes the complex conjugate. We refer to $n_k = |\hat{\psi}_k|^2$ as the 'energy spectrum.' The ratio H_L/H_{NL} is a useful measure of the nonlinearity of the system [16,17].

The focusing equation has unstable, non-translating soliton solutions and self-similar focusing events, 'collapses' which can rapidly transport wave action to a large saturation wavenumber k_s [13–17]. Previous investigations [14, 15] reported simulations where resonant wave interactions carry the wave action transferred to k_s by focusing events partly upscale in an inverse cascade with the spectrum predicted by weak turbulence theory $n_k \propto k^{-5/6}$, and partly downscale in a forward cascade with the spectrum predicted by weak turbulence theory $n_k \propto k^{-1}$; remarkably, the inverse cascade proceeds even when k_s lies within the high wavenumber dissipation range [14, 15]. The defocusing equation does not admit soliton solutions, but admits quasisolitons, or 'envelope solitons,' which are semi-localized structures with a finite lifetime [16, 17].

The test cases considered here use $\alpha = 1/2$, $\beta = 0$, and $\lambda = -1$ (focusing) with additional forcing at large scales and damping at both large and small scales. Specifically, we solve

$$i\partial_t \psi = |\partial_x|^{1/2} \psi - |\psi|^2 \psi + iF + iD\psi \tag{1.4}$$

in a periodic domain of length $L = 400$ where $F = F_0 \sin(4\pi x/L)$ is a steady, large scale forcing and the damping $D\psi$ is defined by

$$\widehat{D\psi}_k = \begin{cases} -\hat{\psi}_k \text{ for } |k| = 2\pi/L, \\ -\hat{\psi}_k(|k| - 2600(2\pi/L))^2 \text{ for } |k| > 2600(2\pi/L), \\ 0 \text{ otherwise.} \end{cases} \tag{1.5}$$

From an initial condition (which is largely irrelevant to our results and included here only for completeness)

$$\psi(x, t = 0) = e^{2i\pi x/L} \mathrm{sech}(100(x/L - 0.5)) \tag{1.6}$$

the high resolution reference solutions are integrated to at least $t = 2 \times 10^5$, by which point the total wave action has saturated; statistics, including the time averaged spectrum, are calculated over the last 1000 time units of each simulation. We compute the reference solutions using a standard Fourier pseudospectral spatial discretization with $2^{13} = 8192$ points in tandem with the fourth order Runge-Kutta exponential integrator of [18] with step size $\Delta t = 0.005$.

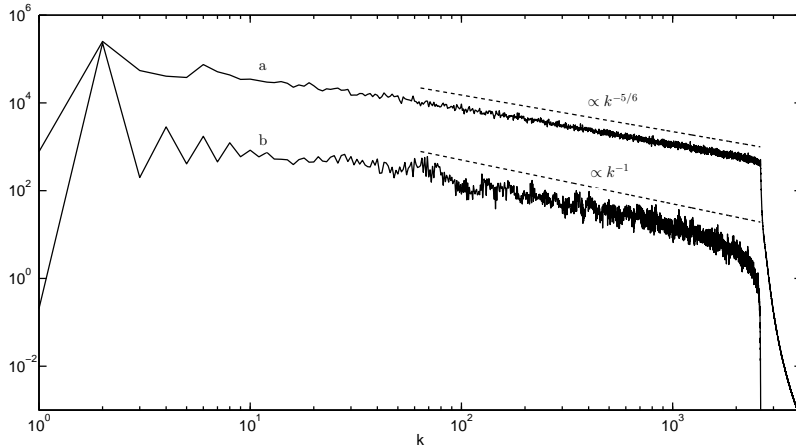[1] MMT stands for Majda-McLaughlin-Tabak, after the authors of [12].

FIG. 1.1. **Spectral bifurcation in focusing MMT** *The energy spectra $n_k = |\hat{\psi}_k|^2$ averaged over the last 1000 time units of the reference simulations of equation (1.4) with (a) $F_0 = 0.0163$ and (b) $F_0 = 0.01625$. Lines proportional to $k^{-5/6}$ and k^{-1} are included for reference. The abscissa has been normalized by $2\pi/L$. The level of nonlinearity in solution (a) is $|H_{NL}/H_L| = 0.09$ and in solution (b) is $|H_{NL}/H_L| = 0.06$, averaged over the last 1000 time units of simulation (see (1.2).*

Figure 1.1 shows the time averaged spectrum of two reference solutions in the focusing case using $F_0 = 0.01625$ and 0.0163. A clear bifurcation is evident in the spectra (spectra at $F = 0.0162$ and 0.0164, not shown, are indistinguishable from those at 0.01625 and 0.0163); although the energy in both solutions is similar at the forcing wavenumber $k = 4\pi/L$, there is a gap of approximately two orders of magnitude between the spectra at higher k. Previous investigations [13,15] found that the spectra of solutions could exhibit significant changes as the solutions relaxed slowly towards a statistically steady state. To rule out the possibility that our spectra were not saturated we extended the simulations at $F_0 = 0.01625$ and 0.0163 to $t = 4 \times 10^5$ and no change was evident in the spectra; this does not contradict the results of [13, 15] because the details of the forcing and dissipation are different.

The spectrum of the high energy simulation in figure 1.1 is proportional to $k^{-5/6}$ over a wide range of scales, which is consistent with the inverse cascade spectrum of weak turbulence theory. Following previous work [14, 15] we interpret this spectrum to be the signature of an energetic loop where collapsing singularities rapidly transport energy (wave action) to a saturation wavenumber k_s that lies in the dissipation range, and weak turbulence transports energy upscale from the dissipation range in an inverse cascade. The spectra from the low energy simulation is very shallow at large scales except for the prominent peak at the forcing wavenumber $k = 4\pi/L$; at higher wavenumbers the spectra are proportional to k^{-1}, which is consistent with the weak turbulence prediction of a downscale energy cascade [12].

Figure 1.2 compares grayscale renderings of $|\psi(x,t)|$ over the last 1000 time units of the simulations at $F_0 = 0.01625$ and $F_0 = 0.0163$. The low energy solution is clearly dominated by the pattern of the large scale forcing, with weak dispersive wave turbulence superposed; in contrast, the high energy solution shows almost no sign of the forcing pattern, displaying instead a sea of dispersive wave turbulence punctuated by localized collapsing singularities.
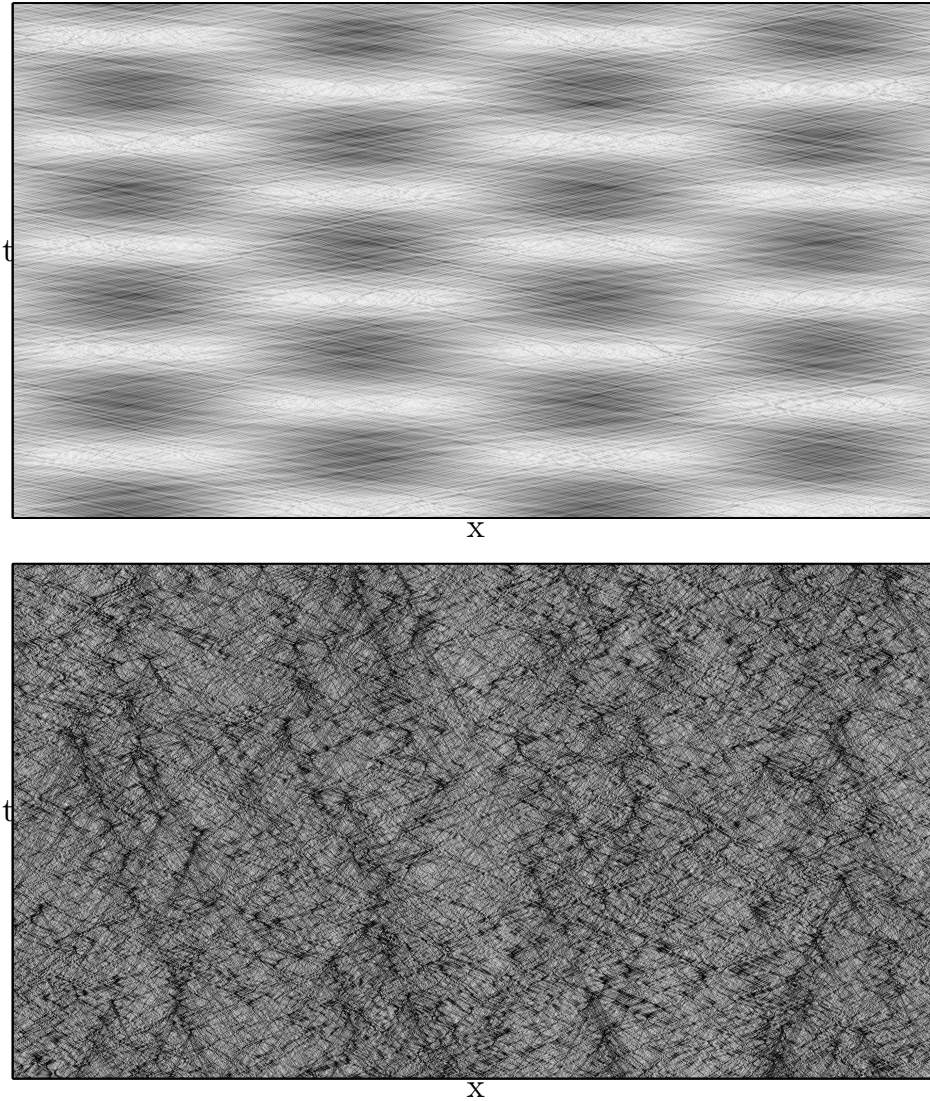
FIG. 1.2. **Low and high energy solutions in focusing MMT** *Grayscale images of $|\psi(x,t)|$ over the last* 1000 *time units of reference simulations of equation (1.4) with $F_0 = 0.01625$ (left) and $F_0 = 0.0163$ (right); darker colors indicate higher values. The grayscale is different, reflecting the much higher energy level of the solution on the right; time increases downwards along the vertical axis of each figure.*

The high energy solution is a difficult test case for superparameterization, with large-scale structures evolving self-similarly toward small scales and energy cascading upscale from the dissipation range. In the following sections we formulate and test a stochastic superparameterization for this difficult test problem.

2. Stochastic superparameterization in MMT: Theory   The mean equation

$$i\partial_t\overline{\psi}=|\partial_x|^{1/2}\overline{\psi}+\lambda|\overline{\psi}|^2\overline{\psi}+\lambda\left[2\overline{|\psi'|^2}\,\overline{\psi}+\overline{(\psi')^2}\,\overline{\psi}^*+\overline{|\psi'|^2\psi'}\right]+i(D\overline{\psi}+\overline{F}).\qquad(2.1)$$

is derived by applying a statistical average and low-pass filter (which can conveniently be defined as a Fourier truncation to low wavenumbers) to the MMT equation (1.4). The use of a statistical average in addition to the low-pass filter allows immediate cancellation of terms which are linear in eddy variables.

The nonlinear eddy equation

$$i\partial_t\psi'=|\partial_x|^{1/2}\psi'+\lambda\left[2|\overline{\psi}|^2\psi'+\overline{\psi}^2\psi'^*+2\lambda(|\psi'|^2)'\overline{\psi}+\lambda((\psi')^2)'\overline{\psi}^*+\lambda(|\psi'|^2\psi')'\right]$$
$$+i(D\psi'+F')\quad(2.2)$$

is derived by subtracting the mean equation from the full equation (1.4). We next make the quasi-linear gaussian closure approximation [11] of the eddy equation wherein F' and all terms which are nonlinear in the eddy variables are replaced by stochastic forcing $i\sigma\dot{W}$, which is spatially correlated and white in time, and deterministic dissipation $i\Gamma\psi'$

$$i\partial_t\psi'=|\partial_x|^{1/2}\psi'+\lambda\left[2|\overline{\psi}|^2\psi'+\overline{\psi}^2\psi'^*\right]+i\left[(\Gamma+D)\psi'+\sigma\dot{W}\right].\qquad(2.3)$$

Note that we have not yet required the use of embedded domains. The gaussian closure approximation implies $\overline{|\psi'|^2\psi'}=0$, so the mean equation becomes

$$i\partial_t\overline{\psi}=|\partial_x|^{1/2}\overline{\psi}+\lambda|\overline{\psi}|^2\overline{\psi}+\lambda\left[2\overline{|\psi'|^2}\,\overline{\psi}+\overline{(\psi')^2}\,\overline{\psi}^*\right]+i(D\overline{\psi}+\overline{F}).\qquad(2.4)$$

In order to signify the superparameterization approximation, wherein the eddy equation is interpreted to apply on local domains embedded in the large scale domain, we introduce a new spatial coordinate y; the use of embedded domains imposes an artificial spatial scale separation between the mean and eddies regardless of the size of the embedded domains. We additionally approximate the mean variables as being time-independent in the eddy equation, which imposes a temporal scale separation; this is signified by the introduction of a new temporal coordinate τ for the embedded domains. With these changes the eddy equation on an embedded domain, parameterized by the local value value of $\overline{\psi}$, is

$$i\partial_\tau\psi'=|\partial_y|^{1/2}\psi'+\lambda\left[2|\overline{\psi}|^2\psi'+\overline{\psi}^2\psi'^*\right]+i\left[(\Gamma+D)\psi'+\sigma\dot{W}\right],\qquad(2.5)$$

and the mean equation (2.4) remains unchanged.

The introduction of embedded domains requires re-interpretation of the low-pass filter and statistical average denoted by the overbar. In addition to the statistical average, the overbar is now interpreted an average over the new coordinates of the embedded domains, which is consistent with standard practice in conventional superparameterization. This interpretation applies regardless of the details of the low-pass filter used to separate the full dynamics into a low-pass mean and a high-pass eddy component.

Because it involves ψ'^*, equation (2.5) is not linear in ψ', strictly speaking. We write it as a linear system for the real and imaginary parts of $\psi'=\psi'_r+i\psi'_i$

$$\partial_t\boldsymbol{u}=\mathcal{L}\boldsymbol{u}+\sigma\dot{W}\qquad(2.6)$$

where $\boldsymbol{u} = (\psi'_r, \psi'_i)^T$ and the meaning of $\sigma \dot{W}$ has changed (but not the notation), and

$$\mathcal{L} = \begin{pmatrix} \lambda \mathcal{I}\{\overline{\psi}^2\} + D + \Gamma & |\partial_y|^{1/2} + 2\lambda |\overline{\psi}|^2 - \lambda \mathcal{R}\{\overline{\psi}^2\} \\ -|\partial_y|^{1/2} - 2\lambda |\overline{\psi}|^2 - \lambda \mathcal{R}\{\overline{\psi}^2\} & -\lambda \mathcal{I}\{\overline{\psi}^2\} + D + \Gamma \end{pmatrix}. \tag{2.7}$$

Here $\mathcal{R}\{\cdot\}$ and $\mathcal{I}\{\cdot\}$ denote the real and imaginary parts of a complex number.

As in [10] we let the embedded domains be of infinite extent, and represent $\boldsymbol{u}'$ as a homogeneous random function (see also [19])

$$\boldsymbol{u}(x,t,y,\tau) = \int \hat{\boldsymbol{u}}_k(x,t,\tau) e^{iky} \mathrm{d}W_k. \tag{2.8}$$

The Fourier coefficients satisfy the autonomous Itō stochastic differential equation

$$\mathrm{d}\hat{\boldsymbol{u}}_k = \mathcal{L}_k \hat{\boldsymbol{u}}_k \mathrm{d}\tau + \sigma_k \mathrm{d}W_k \tag{2.9}$$

where σ_k is a complex 2×2 matrix and W_k is a 2-vector of independent complex white noise. Also

$$\mathcal{L}_k = \begin{pmatrix} \lambda \mathcal{I}\{\overline{\psi}^2\} - d_k - \gamma_k & |k|^{1/2} + 2\lambda |\overline{\psi}|^2 - \lambda \mathcal{R}\{\overline{\psi}^2\} \\ -|k|^{1/2} - 2\lambda |\overline{\psi}|^2 - \lambda \mathcal{R}\{\overline{\psi}^2\} & -\lambda \mathcal{I}\{\overline{\psi}^2\} - d_k - \gamma_k \end{pmatrix} \tag{2.10}$$

where the d_k and γ_k are nonnegative real numbers; specifying these specifies the form of the damping.

We define the 2×2 Fourier covariance matrix

$$C_k = \mathbb{E}[\hat{\boldsymbol{u}}_k \hat{\boldsymbol{u}}_k^*] \tag{2.11}$$

where $*$ is both a complex conjugate and transpose. Note that the eddy terms in the mean equation (2.4)

$$\overline{|\psi'|^2} = \overline{(\psi'_r)^2} + \overline{(\psi'_i)^2}; \quad \overline{(\psi')^2} = \overline{(\psi'_r)^2} - \overline{(\psi'_i)^2} + 2i\overline{\psi'_r \psi'_i} \tag{2.12}$$

may be written in terms of the Fourier covariance matrix using

$$\overline{\boldsymbol{u}\boldsymbol{u}^T} = \begin{bmatrix} \overline{(\psi'_r)^2} & \overline{\psi'_r \psi'_i} \\ \overline{\psi'_r \psi'_i} & \overline{(\psi'_i)^2} \end{bmatrix} = \epsilon \int_0^{\epsilon^{-1}} \int_{-\infty}^{\infty} C_k \mathrm{d}k \mathrm{d}\tau, \tag{2.13}$$

where we have introduced a scale ϵ^{-1} for fast-time averaging. The Fourier covariance matrix is Hermitian by definition, $C_k = C_k^*$; because $\boldsymbol{u}$ is real valued ($\hat{\boldsymbol{u}}_k^* = \hat{\boldsymbol{u}}_{-k}$) we also have $C_{-k} = C_k^*$ which implies $C_k = C_{-k}$. The integral above can therefore be carried out over only half of k.

$$\overline{\boldsymbol{u}\boldsymbol{u}^T} = \begin{bmatrix} \overline{(\psi'_r)^2} & \overline{\psi'_r \psi'_i} \\ \overline{\psi'_r \psi'_i} & \overline{(\psi'_i)^2} \end{bmatrix} = 2\epsilon \int_0^{\epsilon^{-1}} \int_0^{\infty} C_k \mathrm{d}k \mathrm{d}\tau. \tag{2.14}$$

The Fourier covariance obeys the ordinary differential equation

$$\frac{d}{d\tau} C_k = \mathcal{L}_k C_k + C_k \mathcal{L}_k^T + \sigma_k \sigma_k^* \tag{2.15}$$

where the superscript ^{T} denotes a simple transpose to emphasize the fact that $\mathcal{L}_k$ is a real matrix (the derivation of (2.15) amounts to a simple application of Itō's lemma, see also [10,11]). This differential equation can be written as a 4×4 real system where the real and imaginary parts of the off-diagonal element of C_k are treated separately.

The system for the Fourier covariances is

$$\frac{d}{dt}\boldsymbol{c}_k = \mathbf{M}_k\boldsymbol{c}_k + \Sigma_k \tag{2.16}$$

where

$$\mathbf{M}_k = \left[\begin{array}{c|c} \tilde{\mathbf{M}}_k & 0 \\ \hline 0 & 0 \end{array}\right] - 2(\gamma_k + d_k)\mathbf{I}, \tag{2.17}$$

$$\tilde{\mathbf{M}}_k = \begin{bmatrix} 2\lambda\mathcal{I}\{\overline{\psi}^2\} & 2\left(m_k - \lambda\mathcal{R}\{\overline{\psi}^2\}\right) & 0 \\ -\left(m_k + \lambda\mathcal{R}\{\overline{\psi}^2\}\right) & 0 & m_k - \lambda\mathcal{R}\{\overline{\psi}^2\} \\ 0 & -2\left(m_k + \lambda\mathcal{R}\{\overline{\psi}^2\}\right) & -2\lambda\mathcal{I}\{\overline{\psi}^2\} \end{bmatrix}, \tag{2.18}$$

and

$$m_k = |k|^{1/2} + 2\lambda|\overline{\psi}|^2, \tag{2.19}$$

$$\boldsymbol{c}_k = (\{C_k\}_{1,1}, \mathcal{R}\{\{C_k\}_{1,2}\}, \{C_k\}_{2,2}, \mathcal{I}\{\{C_k\}_{1,2}\})^T, \tag{2.20}$$

$$\Sigma_k = (\{\sigma_k\sigma_k^*\}_{1,1}, \mathcal{R}\{\{\sigma_k\sigma_k^*\}_{1,2}\}, \{\sigma_k\sigma_k^*\}_{2,2}, \mathcal{I}\{\{\sigma_k\sigma_k^*\}_{1,2}\})^T. \tag{2.21}$$

Some interesting and useful facts about this system:
- The imaginary part of the off-diagonal element of C_k (the fourth element of $\boldsymbol{c}_k$) decouples from the rest of the system. The remaining comments refer to the reduced subsystem which excludes the imaginary part of the off-diagonal element of C_k.
- It is non-normal for most parameters (exceptions noted below). Nevertheless, it has distinct eigenvalues and is therefore diagonalizable for most parameters. The eigenvalues are

$$-2(\gamma_k + d_k), \ -2(\gamma_k + d_k) \pm 2[-(|k|^{1/2} + \lambda|\overline{\psi}|^2)(|k|^{1/2} + 3\lambda|\overline{\psi}|^2)]^{1/2} \tag{2.22}$$

- For $\lambda = 1$ (defocusing) it is uniformly stable, i.e. its eigenvalues have purely negative real parts equal to $-2(\gamma_k + d_k)$. Nevertheless, due to the non-orthogonality of its eigenvalues solutions may exhibit short-time growth. With $\lambda = 1$ the system is non-normal but diagonalizable for all k.
- For $\lambda = -1$ (focusing) interaction with the mean can lead to exponential growth of modes in the range $|\overline{\psi}|^4 < |k| < 9|\overline{\psi}|^4$; the exact range of unstable modes is smaller than this and depends on the damping $\gamma_k + d_k$. The system is diagonalizable for all k except $|k| = |\overline{\psi}|^4, 9|\overline{\psi}|^4$, where it degenerates.

Provided $\mathbf{M}_k$ is nonsingular, the solution to (2.16) for the elements of the Fourier covariance matrix C_k is

$$\boldsymbol{c}_k(\tau) = e^{\mathbf{M}_k\tau}\boldsymbol{c}_{k,0} + \mathbf{M}_k^{-1}\left[e^{\mathbf{M}_k\tau} - \mathbf{I}\right]\Sigma_k \tag{2.23}$$

where

$$\boldsymbol{c}_{k,0} = \boldsymbol{c}_k(\tau = 0). \tag{2.24}$$

The coefficient matrix $\mathbf{M}_k$ can only be singular in the focusing case. The solution for c_k at any values of k for which $\mathbf{M}_k$ is singular needs to be treated specially, or one can prevent $\mathbf{M}_k$ from being singular by adjusting the value of γ_k.

To calculate the eddy terms in the mean equation (2.4), one needs to compute the time integral of the above exact solution, and then the integral over k. The time average can be easily calculated by integrating the original ODE (2.16) for c_k:

$$\epsilon \int_0^{\epsilon^{-1}} c_k(\tau)\mathrm{d}\tau = \epsilon \mathbf{M}_k^{-1} \left[c_k(\epsilon^{-1}) - c_{k,0} - \epsilon^{-1}\Sigma_k \right]$$

$$= \epsilon \mathbf{M}_k^{-1} \left[e^{\mathbf{M}_k/\epsilon} - \mathbf{I} \right] \left(c_{k,0} + \mathbf{M}_k^{-1}\Sigma_k \right) - \mathbf{M}_k^{-1}\Sigma_k. \qquad (2.25)$$

3. Stochastic superparameterization in MMT: Tests

3.1. Implementation details The stochastic superparameterization computes the eddy terms in the mean equation (2.4) at each time step using the covariance integral (2.14) and the analytical form of the integrand (2.25). To do this one must choose a forcing Σ_k and damping γ_k, an initial condition $c_{k,0}$ for (2.25), a length of averaging ϵ^{-1}, and a quadrature rule for (2.14). In principle, each of these choices could vary across the large scale spatio-temporal domain. In the following we detail our choices for each of these tunable parameters.

Because the eigenvalues of $\mathbf{M}_k$ have negative definite real part when $\overline{\psi} = 0$, (see (2.22)) the system relaxes to a steady balance between forcing and dissipation given by (see (2.23))

$$\lim_{\tau \to \infty} c_k = - [\mathbf{M}_{k,0}]^{-1}\Sigma_k, \qquad \mathbf{M}_{k,0} = \mathbf{M}_k(\overline{\psi} = 0). \qquad (3.1)$$

We choose the stochastic forcing matrix Σ_k and total damping $\gamma_k + d_k$ so that the eddy spectrum coincides with observations when $\overline{\psi} = 0$. Analysis of the reference simulation at $F = 0.0163$ indicates that $|\hat{\psi}_{r,k}|^2 \sim |\hat{\psi}_{i,k}|^2 \sim |k|^{-5/6}$. We therefore choose

$$\Sigma_k = -n_k \mathbf{M}_{k,0}(1,0,1,0)^T \qquad (3.2)$$

where

$$n_k = \begin{cases} 0, & |k| < k_0 \\ \frac{A}{|k|^{5/6} + e^{|k| - 2600(2\pi/L)}}, & k_0 \le |k| \le 4096(2\pi/L). \end{cases} \qquad (3.3)$$

The cutoff k_0 is equal to the largest resolved wavenumber on the large scale grid used for the solution of the mean equation, and the 'eddy amplitude' A is a tunable parameter. The forcing is still dependent on the damping, which we set to

$$\gamma_k + d_k = \begin{cases} 10^{-5}, & |k| < 2600(2\pi/L) \\ 10^{-5} + (|k| - 2600(2\pi/L))^2, & 2600(2\pi/L) \le |k| \le 4096(2\pi/L). \end{cases} \qquad (3.4)$$

Recall that the largest resolved wavenumber in our reference simulations is $k = 4096(2\pi/L)$ and its dynamics are undamped for $1 < Lk/(2\pi) < 2600$.

In a conventional superparameterization with periodic embedded domains the initial condition for each eddy simulation is finite-dimensional, consisting of the value of the eddy variables at gridpoints in the embedded domain. In contrast, the initial condition $c_{k,0}$ is a function of k; to avoid keeping track of finite dimensional approximations of the initial condition at each grid point of the large scale domain we simply set the initial condition to the 'equilibrium' value

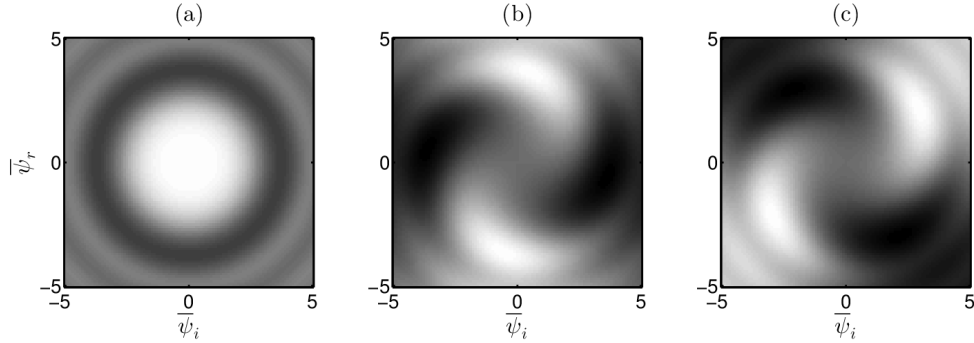$$c_{k,0} = n_k(1,0,1,0)^T. \qquad (3.5)$$

FIG. 3.1. **Precomputed eddy terms as functions of the mean** *The eddy terms in the mean equation (2.4) are precomputed as functions of $\overline{\psi}$ using a time interval $\epsilon^{-1} = 1/10$ and a large-scale cutoff $k_0 = 8\pi/25$. (a) The value of $\overline{|\psi'|^2}$, (b) the real part of $\overline{(\psi')^2}$, and (c) the imaginary part of $\overline{(\psi')^2}$. Higher values are darker, and the exact magnitudes are dependent on the eddy amplitude A.*

If the state of the eddies were tracked from one time step of the mean equations to the next, it would be natural to set the length of the eddy simulations, ϵ^{-1}, equal to the time step of the mean equations. Because our initial condition for the eddies contains no information about their previous state, we leave ϵ as a free parameter to allow the eddies a little more time to respond to the local values of the mean variables.

The integral over k in (2.14) is approximated by a quadrature which is truncated at $k = 4096(2\pi/L)$, the largest wavenumber resolved in the reference simulation. Conventional superparameterization using periodic embedded domains is analogous to using the trapezoid rule to calculate the integral over k in (2.14). Following [10] we compare adaptive Simpson's quadrature and trapezoid rule calculations; the details of the trapezoid rule integrations are discussed further below. The matrix function

$$\epsilon \mathbf{M}_k^{-1} \left[e^{\mathbf{M}_k/\epsilon} - \mathbf{I} \right] \tag{3.6}$$

in equation (2.25) can be difficult to evaluate numerically; we evaluate it using a [7,7] Padé approximant [20].

There are values of k for which $\mathbf{M}_k$ is singular and (2.25) is incorrect. In principle these can be dealt with by perturbing the damping $\gamma_k + d_k$ slightly at those k for which $\mathbf{M}_k$ is singular; this makes the damping a function of the mean $\gamma_k = \gamma_k(\overline{\psi})$. This is not necessary in practice since the quadrature rules almost never evaluate the integrand at those k for which $\mathbf{M}_k$ is singular. Following [10], if any entry of the integrand (2.25) of the covariance integral (2.14) becomes larger than 1000, the integrand is scaled back to a maximum of 1000, which improves the stability of the quadrature. This is only necessary for long averaging times (small ϵ) and has minimal impact on the results.

Because our choices of forcing Σ_k and damping γ_k, length of averaging ϵ^{-1}, quadrature rule, and in particular of initial condition $\boldsymbol{c}_{k,0}$ are independent of the large scale and long time variables x and t, the eddy terms in the mean equation depends only on $\overline{\psi}$. This allows us to precompute the values of the eddy terms over a range of $\overline{\psi}$, which significantly increases the efficiency of the code. Figure 3.1 shows the eddy terms in the mean equations ($\overline{|\psi'|^2}$ and the real and imaginary parts of $\overline{(\psi')^2}$) as precomputed functions of $\overline{\psi}$ using $\epsilon = 10$, $k_0 = 8\pi/25$ (the same as using a

large scale grid with 128 points).

3.2. Stochastic superparameterization in infinite embedded domains

To test the stochastic superparameterization, we solve the mean equation (2.4) using the same numerical algorithm as the reference simulations and identical large scale forcing ($F_0 = 0.0163$) and damping but fewer gridpoints, either 128 or 512. In all simulations we precompute the terms $\overline{|\psi'|^2}$ and $\overline{(\psi')^2}$ as functions of $\overline{\psi}$ on a grid of 101×101 points with $\mathcal{R}\{\overline{\psi}\}, \mathcal{I}\{\overline{\psi}\} \in [-5,5]$ which is sufficient to cover the range of $\overline{\psi}$ encountered in the simulations. Once the eddy terms are precomputed, they are evaluated at each time step by querying a linear interpolant.

At both low resolutions we test three different values of $\epsilon = 2, 10, 50$. At each resolution for each value of ϵ we add three different forms of additional damping to the mean equation to remove energy carried to high wavenumbers by aliasing rather than by intrinsic dynamics; thus, in (2.4) we interpret

$$\widehat{D\overline{\psi}}_k = \begin{cases} -\hat{\overline{\psi}}_k \text{ for } |k| = 2\pi/L, \\ -\hat{\overline{\psi}}_k(|k| - k_d(2\pi/L))^2 \text{ for } |k| > k_d(2\pi/L), \\ 0 \text{ otherwise} \end{cases} \tag{3.7}$$

for selective Laplacian damping,

$$\widehat{D\overline{\psi}}_k = \begin{cases} -\hat{\overline{\psi}}_k \text{ for } |k| = 2\pi/L, \\ -0.5\hat{\overline{\psi}}_k \text{ for } |k| > k_d(2\pi/L), \\ 0 \text{ otherwise} \end{cases} \tag{3.8}$$

for 'strong' damping, or

$$\widehat{D\overline{\psi}}_k = \begin{cases} -\hat{\overline{\psi}}_k \text{ for } |k| = 2\pi/L, \\ -0.1\hat{\overline{\psi}}_k \text{ for } |k| > k_d(2\pi/L), \\ 0 \text{ otherwise} \end{cases} \tag{3.9}$$

for 'weak' damping. In simulations with 128 points $k_d = 42$ and in simulations with 512 points $k_d = 170$. In each simulation we test a range of eddy amplitudes, as defined by A in equation (3.3). All simulations are run to $t = 10^5$ since they equilibrate faster than the reference simulations, using a time step of $\Delta t = 0.02$.

Figure 3.2 shows prototypical spectra from the stochastic superparameterization simulations, compared to the reference simulation and to simulations of the mean equation (2.4) in the absence of eddy forcing. The spectra from simulations with 128 points shown in the upper panel of figure 3.2 use strong damping; without the eddy terms the solution has low energy similar to the low energy reference simulation with $F_0 = 0.01625$ in figure 1.1. The addition of the eddy terms increases the energy in the high wavenumber components of the solution, making it a close approximation of the reference solution. This demonstrates the ability of the eddy terms to correctly represent the effects of an upscale cascade from unresolved small wavenumbers, which is even more striking considering the fact that the eddy terms in the mean equation are nonlinear deterministic functions of the mean variables rather than additive stochastic forcing terms. This example is typical of all simulations with strong damping: the addition of eddy terms lifts the overdamped high wavenumber energy spectrum by nearly two orders of magnitude, making it comparable to the reference simulation even though the coarse grids, 128 and 512 points, radically under-resolve the reference solution with 8192 points.
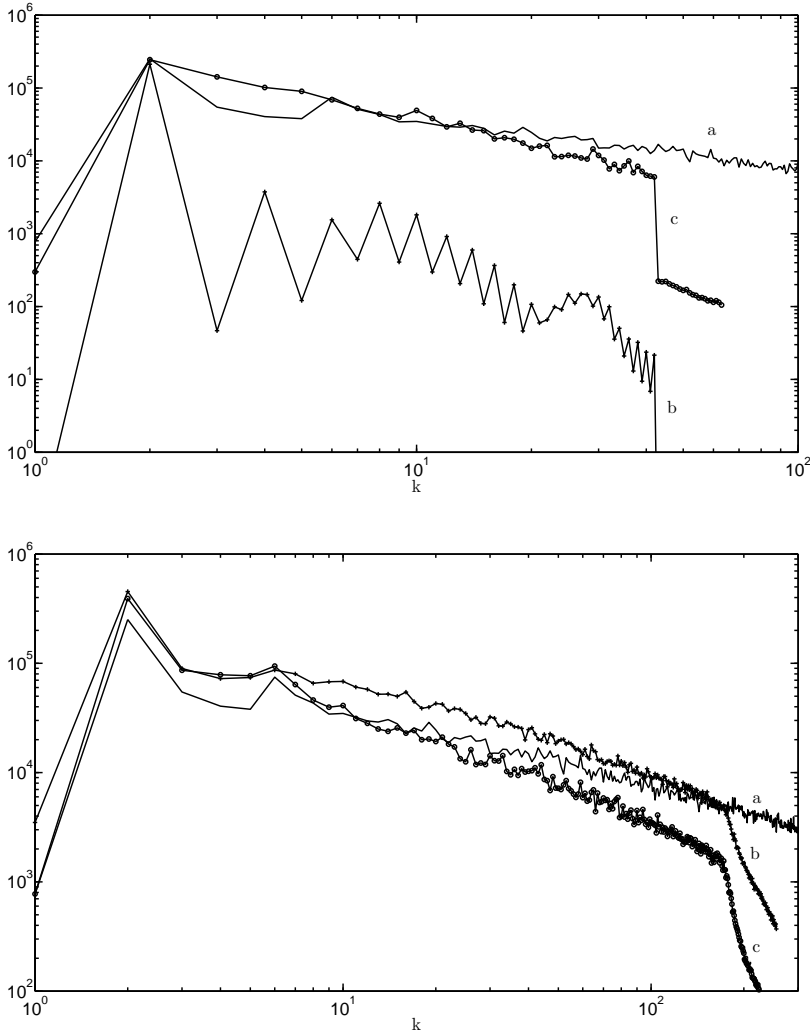
FIG. 3.2. **Low resolution spectra with and without eddy terms** *The energy spectrum $n_k = |\hat{\psi}_k|^2$ from the reference solution to equation (1.4) with $F_0 = 0.0163$ is marked 'a' in both panels. The upper panel shows the spectra from solutions of the mean equation (2.4) with 128 points and strong damping at $k > 42(2\pi/L)$ with the eddy terms (marked 'c' and with open circles) and without the eddy terms (marked 'b' and with plus signs). The lower panel shows the spectra from solutions of the mean equation (2.4) with 512 points and selective Laplacian damping at $k > 170(2\pi/L)$ with the eddy terms (marked 'c' and with open circles) and without the eddy terms (marked 'b' and with plus signs). The eddy terms in both sets of simulations are calculated for $\epsilon = 10$. The abscissae have been normalized by $2\pi/L$.*

The spectra from simulations with 512 points shown in the lower panel of figure 3.2 use selective Laplacian damping; without the eddy terms the spectrum is already similar to that of the reference simulation with $F_0 = 0.0163$ in figure 1.1, although it has slightly too much energy at all scales outside the dissipation range. The addition of the eddy terms decreases the energy of the solution and makes the spectrum

slightly steeper at high wavenumbers. The eddy terms effect only a slight improvement in the low resolution spectrum, mainly by decreasing the total energy; this demonstrates that the stochastic superparameterization can both increase the energy of an overly-damped solution and decrease the energy of an overly energetic solution. This example is typical of all simulations with weak or selective Laplacian damping: the addition of eddy terms lowers the overly energetic energy spectrum, primarily at high wavenumbers, making it comparable to the reference solution.

Figure 3.3 shows grayscale images of $|\overline{\psi}(x,t)|$ over the last 1000 time units of simulations with 512 points and weak damping (compare to the high energy solution in figure 1.2). The collapsing singularities without eddies (upper panel) are much more clustered around the peaks of the large scale forcing than in the reference solution shown in figure 1.2; there are also too many collapses, reflecting an excess of energy similar to that displayed in the spectrum shown in figure 3.2. By contrast, the simulation with eddy terms (lower panel) is much more similar to the reference solution both in terms of the number of collapses and their spatio-temporal distribution.

A more quantitative measure of this improvement is given by the fraction of the domain occupied by collapsing solitons, averaged over time. The time averaged root-mean-square $|\psi|$ for the reference solution is 0.39, and it attains a maximum of 3.5; we consider any location with $|\psi| > 1$ to be occupied by a collapsing soliton. The time averaged area occupied by collapsing solitons in the reference solution is 0.0056; the time averaged area occupied by collapsing solitons in the solution without eddy terms (upper panel of figure 3.3) is 0.0188 and in the solution with eddy terms (lower panel) is 0.0063. If we consider collapsing solitons to be defined by $|\psi| > 1.25$ the results are similar: 0.0016 for the reference solution, 0.0095 for the simulation without the eddy terms and 0.0027 for the simulation with the eddy terms. This example is typical of all simulations with weak or selective Laplacian damping: the addition of eddy terms improves the physical character of the solutions by improving the spatial distribution and reducing the frequency of collapsing singularities.

The effects of the eddy terms shown in figures 3.2 and 3.3 are generic across all of the simulations in our suite of tests. The strong damping simulations with 128 and 512 points both have far too little energy at high wavenumbers; yet at every ϵ there is a range of eddy amplitudes A (see equation (3.3)) for which the addition of eddy terms brings the spectrum of the solution back into close approximation of the reference solution. The physical characteristics of the field, i.e. the number and locations of collapses, are also improved by the addition of the eddy terms in the strong damping simulations. The weak damping simulations with 128 and 512 points produce spectra that are overly energetic at all scales, and the solutions have too many collapses which are clustered at the peaks of the large scale forcing function. The addition of the eddy terms slightly decreases the total energy and improves the physical characteristics of the field. Furthermore, for a given resolution and ϵ, any eddy amplitude A that improves an overdamped solution will also improve an underdamped solution. In general, as the value of A (the magnitude of the eddy terms) is increased from zero, the approximation improves; for values of A greater than some threshold which depends on ϵ and the resolution, the eddy terms heavily damp the mean solution producing a poor approximation.

3.3. Stochastic superparameterization in periodic embedded domains
As noted above and developed in [10], use of the trapezoid rule to approximate the Fourier covariance integral (2.14) is analogous to the use of periodic embedded domains in conventional superparameterization. In conventional superparameteriza-
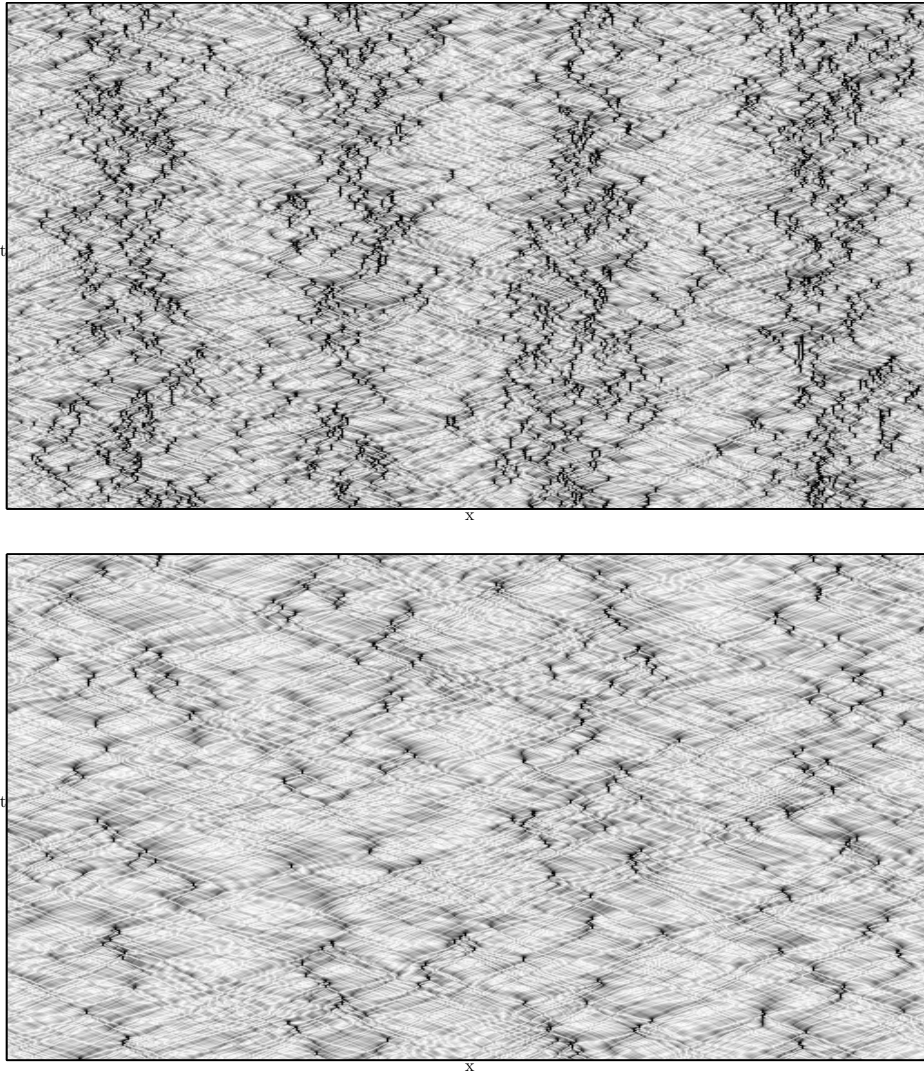
Fig. 3.3. **Improvement in the pattern of low resolution simulations with eddy terms** *Grayscale images of* $|\overline{\psi}(x,t)|$ *over the last* 1000 *time units of* 512-*point solutions to the mean equation (2.4) with weak damping. In the solution without the eddy terms (upper) the collapses are clustered on the peaks of the large scale forcing; the eddy terms improve the solution by decreasing the number of collapses and spreading them more evenly over the domain. Compare to figure 1.2. Time increases downwards along the vertical axis of each figure, and darker colors indicate higher values.*

tions, where individual solutions of nonlinear eddy equations are computed on the embedded domains, one may attempt to decrease computational cost by reducing the size and/or spatial discretization of the embedded domains [2, 6, 9]. Such a reduction limits the range of wavenumbers accessible on the embedded domains, which can exclude important unstable dynamics if these occur on a range of scales not captured on the reduced, discretized periodic embedded domains. Because we precompute the eddy terms the cost savings afforded by using periodic embedded domains is negli-
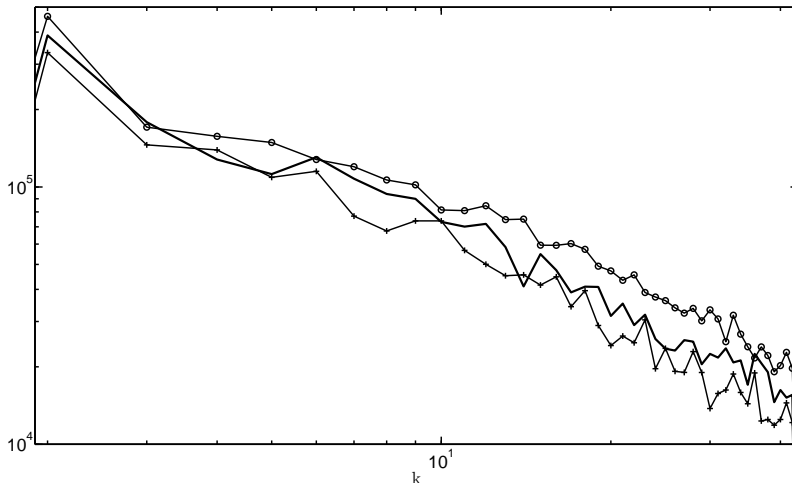
Fig. 3.4. **Superparameterization spectra using periodic embedded domains** *The energy spectrum $n_k = |\hat{\psi}_k|^2$ from solutions of the mean equation (2.4) with 128 points and strong damping at $k > 42(2\pi/L)$. The eddy terms are computed using adaptive quadrature (solid with no markers), corresponding to a homogeneous random function in an infinite domain, and using the trapezoid rule, corresponding to periodic embedded domains (i) twice the size of the large scale grid and with 130 points (marked with plus signs), and (ii) one eighth the size of the large scale grid with 10 points (marked with open circles). All simulations use the same eddy amplitude $A = 0.0044$ (see equation (3.3)). The abscissa has been normalized by $2\pi/L$.*

gible; as in [10] we test stochastic superparameterization using periodic embedded domains to assess the impact of reducing the set of wavenumbers accessible to the modeled eddies.

The size of the embedded domains is $2\pi/k_1$ where k_1 is the smallest nonzero wavenumber used in the trapezoid rule calculation. Taking k_1 equal to the largest resolved wavenumber on the large scale grid makes the embedded domain equal to two large scale grid cells; conventional superparameterizations use embedded domains equal to or smaller than the large scale grid in order to save computational cost [2,6,9]. We repeat the tests discussed above using strong damping, 128 points, and $\epsilon = 2, 10, 50$, but using the trapezoid rule to calculate the eddy terms instead of adaptive Simpson's quadrature. At each value of ϵ we tested six discretizations of the trapezoid rule corresponding to embedded domains (i) twice the size of the large scale grid with 130 gridpoints, (ii) 66 gridpoints, and (iii) 34 gridpoints, and (iv) equal to the large scale grid with 66 gridpoints and (v) 34 gridpoints, and (vi) half the large scale grid with 34 gridpoints. Neither decreasing the size of the embedded domains nor coarsening their grid had significant detrimental impact on the ability of the method to correctly reproduce the spectrum and physical characteristics of the large scale field, though we found that the eddy amplitude A (see equation (3.3)) which produces the best agreement with the reference solution differs between the calculations with the trapezoid rule and those with adaptive Simpson's quadrature.

Figure 3.4 compares the spectra of solutions to the mean equation (2.4) with 128 points and strong damping where the eddy terms are computed using (i) adaptive quadrature, and discretizations of the trapezoid rule corresponding to (ii) an embed-

ded domain one eighth the size of the large scale grid with 10 points per embedded domain, and (iii) an embedded domain twice the size of the large scale grid with 130 points per embedded domain. All three solutions use the same eddy amplitude A. The results exhibit the typical variations displayed by solutions using different embedded domains; the solution with the small, low resolution embedded domains has too much energy at all scales, while the solution with the large, high resolution embedded domain has slightly too little energy at all scales. These energetic disparities can be corrected by changing the eddy amplitude A, which is true of all our results with embedded domains, but the direction in which A must be changed to improve a given solution does not appear to have any obvious dependence on the size or resolution of the embedded domains.

The success of stochastic superparameterization in periodic embedded domains can be related to the typical range of unstable wavenumbers in the eddy domains. For an embedded domain where the local mean has amplitude $|\overline{\psi}| = 1$, the range of unstable eddy wavenumbers is very nearly $1 \leq |k| \leq 9$, as given by (2.22) with $\gamma_k + d_k = 10^{-5}$. Every test performed includes at least one wavenumber in this range, except the test using an embedded domain one eighth the size of the large scale grid, which includes unstable modes for $|\overline{\psi}| \gtrsim 4/3$. Furthermore, even for simulations that lack unstable modes the transient response of the eddies can include non-normal growth of linearly stable modes. Clearly, for this problem the use of periodic embedded domains that are smaller than the large scale grid and have low spatial resolution is warranted.

4. Concluding discussion and future directions   Superparameterization, which was initially developed in the context of atmospheric dynamics [2–4], is a promising method of capturing the impact of unresolved small scale processes on resolved large scales in simulations of multiscale phenomena. In superparameterization, solutions of prognostic equations for small scale processes on local, periodic domains embedded within the computational grid of a large scale model are computed and used to feed back onto the large scales. The computational cost of a superparameterization with periodic embedded domains is less than the cost of a high resolution simulation because simulations on the embedded domains can be run in parallel with fewer points than a fully resolved simulation (see, e.g. [6, 8, 9]).

It is often useful and sometimes necessary to further decrease the computational cost; this is done by using embedded domains with lower dimensionality (e.g. using 2D embedded domains in a 3D large scale grid as in [3–9]), by making the embedded domains smaller than the size of the large scale grid cells, and/or by integrating the embedded simulations over less time than the large scale model time steps [3, 6, 9]. The success of these approximations of the small scale dynamics suggests that even more radical approximations may produce comparably accurate large scale solutions. This line of reasoning led [10] to propose the use of quasi-linear stochastic models of the small scale processes; a systematic theory of self-consistent quasi-linear Gaussian stochastic models was introduced by [11].

We extend the work of [10] and [11] by implementing and testing a stochastic superparameterization in a nonlinear one-dimensional model of wave turbulence. This idealized model, introduced in [12] to test the validity of weak turbulence theory, includes many realistic features including unstable solitons which collapse self-similarly from large to small scales and a turbulent inverse energy cascade from small to large scales [12, 14–17]. In short, the test case for stochastic superparameterization is extremely difficult.

In contrast to many superparameterizations [2–6, 8, 9], where the equations gov-

erning the large and small scale processes are posited directly based on knowledge of the relevant physics, we derive our model equations from a single equation governing the dynamics at all scales. The small scale dynamics are radically simplified by imposing an artificial scale separation through the use of embedded domains and by a quasi-linear stochastic approximation, the 'Gaussian closure' of [11], but no stochasticity is retained in the equations governing the large scale: only the expected value of the eddy forcing, defined by the eddy covariance, is used in the large scale equations.

Additionally, we make a 'memoryless' approximation of the small scale dynamics, where the transient response of the small scales to the local large scale variables is calculated at every time step of the large scale equations from an initial state given by a prescribed 'climatological' equilibrium rather than from the state of the small scale variables at the end of the previous time step of the large scale equations. This transient response need not equilibrate (and generally does not equilibrate), but is rather truncated after a specified time interval comparable to the length of the large scale time step. The 'memoryless' approximation allows us to precompute the response of the small scales to a range of large scale conditions, and the small scale forcing terms in the large scale equations are evaluated by querying the precomputed interpolant in a look-up table.

The stochastic approximation allows us to treat the embedded domains as being formally infinite, the small scale variables being spatially homogeneous random functions on the embedded domains. Linear instabilities in the small scale dynamics sometimes occur over a limited range of wavenumbers (not just in the test model, but generally). The use of periodic embedded domains restricts the allowable small scale wavenumbers to a finite set which can miss the most important modes in the small scale response to the large scale conditions [10]. Our use of infinite domains alleviates this problem by allowing a continuum of wavenumbers in the small scale dynamics. Our particular test problem does not offer clear evidence of the superiority of infinite embedded domains since the range of unstable wavenumbers in the small scale equations is wide enough to be captured by even extremely small or coarsely grained periodic embedded domains: we do not find the use of periodic embedded domains to be clearly inferior or superior to the use of infinite domains in this particular problem.

Despite the many and severe approximations of the small scales, the resulting stochastic superparameterization performs admirably. It increases the energy in overdamped simulations and decreases the energy in underdamped simulations (figure 3.2), and improves the spatial distribution and frequency of collapsing singularities (figures 1.2 and 3.3). Because of the ability to precompute the small scale forcing terms in the mean equations the computational cost is similar to the cost of not using any parameterization at all.

The resounding success of stochastic superparameterization in the current test problem warrants further investigation of the method in more physically relevant settings. In particular, the use of stochastic superparameterization for filtering turbulent signals from sparse observations in order to overcome the curse of ensemble size is an enticing future topic [21, 22].

REFERENCES

[1] D. Randall, M. Khairoutdinov, A. Arakawa, and W. Grabowski, *Breaking the cloud parameterization deadlock*, Bull. Amer. Meteor. Soc., 84, 1547–1564, 2003.

[2] W. W. Grabowski and P. K. Smolarkiewicz, *CRCP: A cloud resolving convection parameterization for modeling the tropical convective atmosphere*, Physica D, 133, 171–178, 1999.

[3] W. W. Grabowski, *Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization (CRCP)*, J. Atmos. Sci., 58, 978–997, 2001.

[4] M. Khairoutdinov and D. A. Randall, *A cloud resolving model as a cloud parameterization in the NCAR community climate system model: Preliminary results*, Geophys. Res. Lett., 28, 3627–3620, 2001.

[5] M. Khairoutdinov, D. A. Randall, and C. DeMott, *Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes*, J. Atmos. Sci., 62, 2136–2154, 2005.

[6] Y. Xing, A. J. Majda, and W. W. Grabowski, *New efficient sparse space-time algorithms for superparameterization on mesoscales*, Mon. Wea. Rev., 137, 4307–4324, 2009.

[7] A. J. Majda and Y. Xing, *New multi-scale models on mesoscales and squall lines*, Commun. Math. Sci., 8, 113–144, 2010.

[8] J.-M. Campin, C. Hill, H. Jones, and J. Marshall, *Super-parameterization in ocean modeling: Application to deep convection*, Ocean Mod., 36, 90–101, 2011.

[9] W. W. Grabowski, *Large-scale organization of moist convection in idealized aquaplanet simulations*, Int. J. Numer. Meth. Fl., 39, 843–853, 2002.

[10] A. J. Majda and M. J. Grote, *Mathematical test models for superparameterization in anisotropic turbulence*, Proc. Natl. Acad. Sci. U.S.A., 106, 5470–5474, 2009.

[11] A. J Majda, *Challenges in climate science and contemporary applied mathematics*, Commun. Pur. Appl. Math., 65, 920–948, 2012.

[12] A. J. Majda, D. W. McLaughlin, and E. G. Tabak, *A one-dimensional model for dispersive wave turbulence*, J. Nonlinear Sci., 7, 9–44, 1997.

[13] D. Cai, A. J. Majda, D. W. McLaughlin, and E. G. Tabak, *Spectral bifurcations in dispersive wave turbulence*, Proc. Natl. Acad. Sci. U.S.A., 96, 14216–14221, 1999.

[14] D. Cai and D. W. McLaughlin, *Chaotic and turbulent behavior of unstable one-dimensional nonlinear dispersive waves*, J. Math. Phys., 41, 4125–4153, 2000.

[15] D. Cai, A. J. Majda, D. W. McLaughlin, and E. G. Tabak, *Dispersive wave turbulence in one dimension*, Physica D, 152-153, 551–572, 2001.

[16] V. E. Zakharov, P. Guyenne, A. N. Pushkarev, and F. Dias, *Wave turbulence in one-dimensional models*, Physica D, 152-153, 573–619, 2001.

[17] V. E. Zakharov, F. Dias, and A. N. Pushkarev, *One-dimensional wave turbulence*, Phys. Rep., 398, 1–65, 2004.

[18] S. M. Cox and P. C. Matthews, *Exponential time differencing for stiff systems*, J. Comput. Phys., 176, 430–455, 2002.

[19] A. M. Yaglom, *An Introduction to the Theory of Stationary Random Functions*, Prentice-Hall, 1962.

[20] H. Berland, B. Skaflestad, and W. M. Wright, *EXPINT—A MATLAB package for exponential integrators*, ACM Trans. Math. Softw., 33, 1–17, 2007.

[21] A.J. Majda, J. Harlim, and B. Gershgorin, *Mathematical strategies for filtering turbulent dynamical systems*, Discrete Cont. Dyn. S., 27, 441–486, 2010.

[22] A. J. Majda and J. Harlim, *Filtering Complex Turbulent Systems*, Cambridge University Press, 2012.