Large-Scale Dynamical Response to Subgrid-Scale Organization Provided by Cellular Automata

LISA BENGTSSON
Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

HEINER KÖRNICH
Department of Meteorology, Stockholm University, Stockholm, Sweden

ERLAND KÄLLEN
European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

GUNILLA SVENSSON
Department of Meteorology, Stockholm University, Stockholm, Sweden

(Manuscript received 21 December 2010, in final form 18 July 2011)

ABSTRACT
Because of the limited resolution of numerical weather prediction (NWP) models, subgrid-scale physical processes are parameterized and represented by gridbox means. However, some physical processes are better represented by a mean and its variance; a typical example is deep convection, with scales varying from individual updrafts to organized mesoscale systems. This study investigates, in an idealized setting, whether a cellular automaton (CA) can be used to enhance subgrid-scale organization by forming clusters representative of the convective scales and thus yield a stochastic representation of subgrid-scale variability. The authors study the transfer of energy from the convective to the larger atmospheric scales through nonlinear wave interactions. This is done using a shallow water (SW) model initialized with equatorial wave modes. By letting a CA act on a finer resolution than that of the SW model, it can be expected to mimic the effect of, for instance, gravity wave propagation on convective organization. Employing the CA scheme permits the reproduction of the observed behavior of slowing down equatorial Kelvin modes in convectively active regions, while random perturbations fail to feed back on the large-scale flow. The analysis of kinetic energy spectra demonstrates that the CA subgrid scheme introduces energy backscatter from the smallest model scales to medium scales. However, the amount of energy backscattered depends almost solely on the memory time scale introduced to the subgrid scheme, whereas any variation in spatial scales generated does not influence the energy spectra markedly.

1. Introduction
Physical processes that are not explicitly resolved by the model grid in numerical weather prediction (NWP) models are parameterized. To address model errors associated with parameterization schemes and subgrid-scale variability, development of stochastic representations of atmospheric processes is becoming more frequent (e.g., Lin and Neelin 2002; Shutts 2005; Teixeira and Reynolds 2008; Plant and Craig 2008; Berner et al. 2008). In the context of ensemble prediction, in particular within the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS), stochastic perturbations of the physical tendencies have been applied in order to account for uncertainties arising from subgrid variability (Buizza et al. 1999). Such a stochastic parameterization is expected to improve the mean state as well as the spread around the mean in an EPS. However, Bengtsson et al. (2008) found that even though stochastic physics is included in the ECMWF EPS, the ensemble forecast was underdispersive in comparison with the characteristic variability of the atmosphere. Furthermore, Magnusson (2009) showed that in order to

Corresponding author address: Lisa Bengtsson, Swedish Meteorological and Hydrological Institute, SMHI, SE-601 76, Norrköping, Sweden.
E-mail: lisa.bengtsson@smhi.se

DOI: 10.1175/JAS-D-10-05028.1

© 2011 American Meteorological Society
obtain a good ensemble system using breeding vectors, ensemble transform perturbations, or ensemble transform Kalman filters, it is crucial for the skill of the EPS to have a model with a sufficient level of variance.

A typical example of subgrid variability arises from organized deep convection in the atmosphere, which in turn is inherently associated with small-scale motions. Such processes are not well represented by gridbox means in NWP models (Lin and Neelin 2002; Palmer 2001). In light of the above, it is desired to address uncertainties that arise from subgrid variability, such as deep convection, both within the deterministic forecast model and within its accompanying EPS.

Following an idea by Palmer (2001), Shutts (2005) implemented a stochastic physics scheme to the ECMWF EPS in which stochastic noise is generated at the smallest resolved spatial and temporal scales using a cellular automaton (CA). This resulted in backscatter of energy that compensates for some of the energy dissipated by numerical advection error and horizontal diffusion. Berner et al. (2008) extended the use of a CA stochastic scheme to the seasonal time scales and found that the scheme resulted in a reduction of systematic model errors. Here, the CA scheme acted as a source of energy on the mesoscale that backscattered to the planetary-scale components of the flow.

Another usage of a CA could be to represent subgrid variability of, for instance, organized deep convection by letting information from the dynamical model serve as input to the CA. Numerous processes in the atmosphere serve as organizing mechanisms of deep convection. Examples of such processes are vertical wind shear, underlying sea surface temperature (SST) gradients, cold pool dynamics, and water vapor feedbacks (Tompkins 2001). Also, ducted gravity waves, initiated from deep convection, act to organize convective clusters and mesoscale convective systems (Huang 1988). Fast-moving gravity waves are either damped or not resolved in time in most NWP models. The suggestion of including a CA as part of a deep convection parameterization scheme is intriguing, since a CA encompasses several components that are of interest for deep convection organization, such as lateral communication, temporal memory, and stochasticity.

In this study, we investigate, in an idealized setting, whether a CA can be used in order to enhance subgrid-scale organization and form clusters representative of the convective scales. We study the transfer of energy from the convective to the larger atmospheric scales through wave interaction. Ensemble calculations are used to explore the inherent stochasticity of the scheme that results from the initialization of the CA. The idealized model environment that we employ is a shallow water (SW) model. By letting a CA act on a finer resolution than that of the SW model, it can be expected to mimic the effect of gravity wave propagation on convective organization. Introducing such organization, subgrid cells may combine and form clusters with spatial scales larger than that of the truncation scale of the model. In addition, and in contrast to what is done in conventional parameterizations, the option of advecting information by the wind field on the subgrid level is explored.

The SW model is initialized with equatorial wave modes in order to study the interaction between the subgrid parameterization and the large-scale flow as an idealized analogy to convectively coupled equatorial waves. Although the dynamics of these equatorial waves are very complex, observed waves display dispersive properties derived from simple two-dimensional SW theory (Kiladis et al. 2009). Thus, a SW model provides an idealized setting that can be used to investigate the interaction between subgrid scales formed by the CA and large-scale dynamics of equatorial waves.

In contrast to previous studies, where the CA is used as a pattern generator in order to perturb physical tendencies, we introduce a two-way interaction between the resolved scales of the SW model and the subgrid scales from the CA. The SW horizontal divergence field is given as input to the CA, letting the scheme serve as a crude representation of convective divergence in a SW model. In turn, information from the CA is given back to the large-scale dynamics through a mass source term, mimicking latent heat release. Thus, the regions in which the subgrid scheme is active are determined by deterministic information from the large-scale dynamical field of the model, while the subgrid-scale variability is determined by the CA, initialized with a white spectrum in space and time. The rules that govern the CA are designed to achieve a statistical representation of the subgrid-scale motions, in particular the horizontally propagating gravity waves. As a tunable parameter, the CA used in this study contains a memory in time. By construction, the CA scheme contains both a spatial and temporal autocorrelation, unlike a purely random scheme. In the present study we investigate how the choice of scheme, CA or random, affects the scale interaction. The paper is organized as follows. In section 2, we describe the numerical configuration of the shallow water model, as well as the subgrid schemes. The experimental setup is defined in section 3, followed by main results in section 4, and concluding remarks are presented in section 5.

2. Model description

a. The shallow water model

The numerical model framework consists of a system of SW equations that govern the vertically independent
motion of a single thin layer of incompressible and homogeneous fluid on a rotating sphere. The equations consider the flow for one particular vertical mode only, for which an appropriate choice of an equivalent depth $h$ must be made (Kiladis et al. 2009). The SW equations are

$$\frac{du}{dt} + \mathbf{V} \cdot \nabla u - f v = -\frac{g}{\rho} \frac{\partial h}{\partial x} + K_u, \quad (1)$$

$$\frac{dv}{dt} + \mathbf{V} \cdot \nabla v + fu = -\frac{g}{\rho} \frac{\partial h}{\partial y} + K_v, \quad (2)$$

$$\frac{dh}{dt} + \mathbf{V} \cdot \nabla h + hf \cdot \nabla \mathbf{V} = \left( Q - \bar{Q} \right) + K_h. \quad (3)$$

Here, we use the equivalent depth, the horizontal wind vector $\mathbf{V} = (u, v)$, containing the zonal and meridional velocities, the gravitational acceleration $g$ (m s$^{-2}$), and the Coriolis parameter $f$ (s$^{-1}$). Also, $K_x$ is the tendency of any prognostic variable $x$ due to horizontal diffusion. The mass source term $Q$ (m s$^{-1}$) is both space and time dependent. To ensure mass conservation, a mass correction is implemented by removing the global mass source term $\bar{Q}$. A two-dimensional spectral transform formulation using Fourier series is utilized to find a numerical solution of the system (1)–(3).

Semi-implicit time-stepping is done using a leapfrog scheme. The boundary conditions are periodic on the eastern and western boundaries, whereas a relaxation zone is imposed in the north–south direction in order to damp waves at the boundaries. The model framework and spectral solutions to Eqs. (1)–(3) follow closely the methods utilized for the spectral High Resolution Limited Area Model (HIRLAM) model developed by Gustafsson (1998).

b. The cellular automaton subgrid scheme

An elementary CA is a dynamical system with a state vector that takes on a number of discrete states determined by a given rule (Wolfram 1983). This rule relates the state at one point in space and time to the state of the neighboring CA grid cells at the previous time step. Since given rules can generate self-organization of cells, complex patterns emerge from the implementation of very simple rules.

In this study, rules based on the CA “game of life” are used, which describes the evolution of a given initial condition, governing a self-organizing system (Chopard and Droz 1998). With proper parameter selection these rules are able to generate continuous patterns that appear to be close to the spatial scales of organized deep convection, as suggested by Shutt’ (2005). In the following, a subgrid variability scheme, based on the CA game of life, serves here as a crude representation of convective organization in a SW model. The CA yields a statistical representation of the subgrid-scale variability, with the possibility of organizing clusters larger than the truncation scales of the SW model. The CA is acting on a grid finer than that of the SW model. From here on, we separate between cells, when speaking of subgrid cells on the CA grid, and grid boxes when referring to the SW model grid. If active cells form on the finer CA grid, these cells can act to spread information across model grid boxes as a way of representing communication between grid boxes, analogous to communication via gravity waves that propagate radially outward from a convective cell (Huang 1988). We also explore the option of advecting active subgrid cells on the CA grid, allowing subgrid information to propagate between subgrid cells, something a conventional parameterization scheme does not readily do.

The proposed subgrid scheme consists of four components. 1) The game-of-life CA provides the background field (BF), which serves as potentially active subgrid cells yet uncoupled to the dynamics. 2) Information from the SW model is combined with the current state of the BF in order to determine whether a subgrid cell becomes active. An active cell is assigned a prescribed lifetime. 3) In case of advection, an active cell with prescribed lifetime is advected in a quasi-Lagrangian fashion at the subgrid level. 4) Finally, the active subgrid cells are used to calculate the mass source term in the height tendency equation, which couples the representation of the convective scales with the large-scale dynamics. More details of the cellular automaton model used in this study can be seen in Fig. 1.

The BF is initialized randomly and, in step 1 the following CA rules are applied according to the game of life:

- Each grid cell can take on the state of either 0 or 1, and the state of the BF cell at the next time step is dependent on its own current state and on the state of its surrounding eight neighbors at the current time step.
- If the current state of the BF cell is 1, and it has exactly two or three neighboring cells with the state 1, it will remain at the state 1 at the next time step.
- If the current state of the BF cell is 1, but it has less than two neighboring cells with the state 1, it will become 0 at the next time step due to “loneliness.”
- If the current BF cell state is 1 and is surrounded by more than three neighbors with state 1 it will become 0 at the next time step due to “overcrowding.”
- If the current state is 0 and it is surrounded by exactly three neighbors with the BF cell state of 1, it will take on the value 1 at the next time step; otherwise, it will remain at 0.

To reach an equilibrium state of the game of life, the BF is initialized by integrating 2000 time steps. These rules alone yield a stochastic representation of the subgrid variability, as the BF is initialized randomly. The impact
of this stochasticity on the ensemble spread will be explored in section 4d.

Figure 2a shows a snapshot of the BF, where black cells represent a cell state of 1 and white cells indicate value 0. In this example the grid of the BF is $3 \times 3$ times as fine as the SW model grid. It can be seen that with the CA rules applied, clusters form in quasi-stochastic patterns.

As stated above, the BF only gives information where a subgrid cell may potentially be active at the next time step. The motivation for this is twofold: 1) the rules governing the BF, based on the game of life, yield a clustering effect of subgrid cells; and 2) given a large enough initial sample of cells with state 1, the CA constructed with such rules is continuous in time. Next, the divergence field from the SW model is used in combination with the BF to create “active” subgrid cells. Active cells are assigned a lifetime that introduces a memory time scale to the parameterization. The lifetime of an active cell is reduced by 1 for each model time step until it reaches 0. Upon reaching 0, a new active subgrid cell can be generated if the neighbors in the BF are either 2 or 3 and the SW model divergence is less than a threshold value (e.g., in areas of strong large-scale convergence). The threshold is chosen to be $0.35 \times \max(\text{div} U^\ast)$ at each time step. This is an important property of the CA model since this information provides a physical large-scale input to where a new active CA cell can be generated. The maximum lifetime $L$ is defined in terms of the model time step and is updated for each newly activated cell. An example of the lifetime field with $L = 20$ is shown in Fig. 2b, where the cells can only be active in the region of large-scale convergence within a Kelvin wave generated in the SW model.

When exploring the effect of advection, we choose to advect active subgrid cells with an assigned lifetime (see Fig. 2b) by quasi-Lagrangian advection on the subgrid level. It is not desirable to advect CA cells within the BF, as the automaton is sensitive to the rules governing the self-organization of the system. Advecting cells at this level will destroy the evolution of the CA. However, the advection of the lifetime corresponds to the advection of convective cells with the mean flow.

The subgrid-scale information provided by the scheme is averaged onto the SW model grid in order to yield a “CA fraction” $\sigma$ describing the fraction of active cells for a grid box of the SW model. This fraction, scaled with a factor $\alpha$, provides the spatially and temporally varying $Q$ on the right-hand side of Eq. (3) and represents the coupling back to the large-scale field, mimicking latent heat release:
\[ Q(x, y, t) = a \sigma. \]  

The mass source strength \( \alpha \) is chosen such that the SW model remains stable, \( \alpha = 1 \times 10^{-4} \text{ m s}^{-1} \). As noted above, a global mass correction is applied in the SW model in order to conserve the global mass. The maximum of \( Q \) occurs if \( \sigma \) equals 1. By choosing the same \( \alpha \) in all of the experiments, we ensure that the local mass source maxima are comparable. The fraction of the CA field from the example above can be seen in Fig. 2c.

c. The random subgrid scheme

To examine the impact of larger space and time scales, generated by the CA, on the large-scale dynamics, the scheme is compared with a fully random representation of the subgrid motions. Convective scales generated completely randomly are not expected to perform well with respect to energy backscatter as they create only small structures in space and time that are not able to force the large-scale flow effectively. For the sake of comparison, a few choices need to be made in order to set up the random scheme. The random scheme acts on the same finer mesh as the CA scheme and is only active in regions of sufficient horizontal convergence, using a threshold in the same way as the CA scheme.

The percentage of active subgrid cells in the random scheme is on average the same as for the active values in the BF of the CA scheme after the spinup time of 2000 time steps (i.e., 10% of all subgrid cells are active). The cells are determined randomly in space, with the same percentage of active cells at each time step. The CA scheme will generate larger spatial and temporal scales of \( Q \) because of neighborhood rules and the additional lifetime, which are not present in the random scheme. Thus, for the same \( \alpha \), \( \overline{Q} \) will be larger within the CA scheme than within the random scheme.

3. Experimental setup

The maximum wavenumbers in the zonal and meridional direction of the SW model are given by \( N_K = 89 \) and \( N_L = 59 \), respectively (corresponding to about 0.6° horizontal resolution), and a Fourier elliptic truncation

\[ \left( \frac{k}{N_K} \right)^2 + \left( \frac{l}{N_L} \right)^2 \leq 1 \]  

ensures a homogeneous and isotropic resolution over the whole model domain (Gustafsson 1998). The smallest wave to be resolved by the model is 3Δx, corresponding to about 200 km. The time step is chosen to be 600 s. Fourth-order horizontal diffusion is applied, where the damping time scale of the smallest wave is about 25 min.

<table>
<thead>
<tr>
<th>( L )</th>
<th>( G )</th>
</tr>
</thead>
<tbody>
<tr>
<td>L10G3</td>
<td>10</td>
</tr>
<tr>
<td>L20G3</td>
<td>20</td>
</tr>
<tr>
<td>L30G3</td>
<td>30</td>
</tr>
<tr>
<td>L20G7</td>
<td>20</td>
</tr>
<tr>
<td>ADVL20G3</td>
<td>20</td>
</tr>
<tr>
<td>RANDOM</td>
<td>—</td>
</tr>
<tr>
<td>RANDOM, ( L = 20 )</td>
<td>20</td>
</tr>
</tbody>
</table>

In the atmosphere, the convective scales are twofold: the scales of individual convective, nonhydrostatic cells, and the scales of mesoscale organization. Obviously, the SW model has a resolution far beyond both scales and the proposed CA scheme tries to represent the subgrid variability of both processes.

The SW model is initialized with two different initial conditions: either a broad selection of equatorial waves or a single Kelvin wave. For the former, the choice of eigenmodes is based on a cutoff at a critical frequency of 6 h\(^{-1}\). The spectral variance distribution of the waves follows the red spectral behavior of Zagar et al. (2004). The total variance is split among the different waves such that the equatorial Rossby waves contains about 50%, Kelvin modes and westward mixed Rossby gravity modes about 15% each, and eastward mixed Rossby-gravity and westward equatorial inertio-gravity modes receive about 10% of the total variance as chosen in Zagar et al. (2004), estimated from outgoing longwave radiation observations by Wheeler and Kiladis (1999).

Simulations of the SW model for the two initial conditions are compared with simulations using either the CA or the random scheme. Since the spatial and temporal scales generated by the CA scheme depend on the prescribed lifetime \( L \) and the horizontal resolution of the CA grid, four sensitivity experiments of these tuning parameters are constructed (see Table 1). The names of the experiments are in the form \( LXY \), where \( X \) denotes the number of time steps for the maximum lifetime and \( Y \) denotes how much finer the CA mesh is than the SW grid. For example, in experiment L10G3, the maximum lifetime is 10, and the CA mesh is 3 x 3 times finer than the resolution of the SW model. The experiments L10G3, L20G3, and L30G3 enable us to examine the feedback between the choice of maximum lifetime and the SW dynamics. One model time step is 600 s; thus, the lifetimes correspond to 1 h 40 min, 3 h 20 min, and 5 h, respectively. The different lifetimes span the range for tropical deep convection or mesoscale convective systems that can last from a couple of hours to over a day (Hodges and Thorncroft 1997). With the experiments L20G3 and L20G7 the impact of the CA grid resolution
on the small-scale structures formed by the CA scheme can be examined. Both experiments use a prescribed maximum “convective” lifetime $L$ of 20 time steps, but different CA grid resolutions that are either $3 \times 3$ ($G_3$) or $7 \times 7$ ($G_7$) times as fine as the SW model grid. The effect of advection is explored using the single Kelvin wave setup. The experiment is done using $L = 20$ and $G = 3 \times 3$, with the experiment name ADVL20G3.

To achieve smoother kinetic energy (KE) spectra, four runs were created for each experiment of Table 1, using different initial conditions, achieved by shifting the equatorial waves with a random phase. All of the results from here on are presented as an average of these four samples.

An extension of the random scheme was constructed in order to investigate the relative impact of the memory time scale on the large-scale dynamics, using a lifetime of 20 (RANDOM, $L = 20$; Table 1). A new subgrid cell is generated only if the lifetime has been reduced to zero. Similarly to the CA scheme, a fraction of active subgrid cells, $\sigma$ is generated and multiplied by the scaling parameter $a$ to yield a value for $Q$.

All of the model experiments, using the CA and random scheme, are constructed such that the maximum value of $Q$ is the same for the same CA grid horizontal resolution. However, $Q$ will differ since the different approaches will generate clusters and cells of different size and frequency. The impact of these subgrid properties on the resolved SW dynamics is examined in the next section.

As noted above, the CA scheme holds several tuning parameters: the scaling parameter for the horizontal resolution of the CA scheme, the “convective” lifetime, the strength of the mass source term, and the threshold of large-scale divergence. The sensitivity to the spatial and temporal components of the CA scheme (i.e., the scaling factor of the horizontal resolution of the CA scheme $G$ and the “convective” lifetime $L$) is investigated. Changes in $a$ will yield locally larger values of $Q$, generating a strong local feedback on the SW dynamics. As we are interested in the large-scale response, $a$ is chosen to be sufficiently small. The value of $a$ is held fixed in all of the experiments for a clear comparison. Furthermore, the threshold criterion for the divergence, which determines the region where active subgrid cells are allowed (strong convergence), is kept constant in all experiments.

4. Results

a. The Kelvin wave

Figure 3a shows the geopotential height anomaly and wind vectors of a single Kelvin wave. The overall shape and location of this single wave is compared with the wave coupled with convective scale structures, generated from the subgrid scheme using the random scheme (RANDOM; Fig. 3b) and the CA with different CA-grid resolution (L20G7, L20G3; Figs. 3c,d). The interaction between the subgrid scheme and the large-scale flow depends on the parameters in the scheme. It can be seen that in the regions where convergence is large, the wave travels with a slower phase speed compared to regions of divergence, where convection is suppressed. In those regions the wave continues to travel with a phase speed of an uncoupled wave. These two effects result in an asymmetric wave field, since it is stretched out where the geopotential height anomaly is negative and more compact where the geopotential height anomaly is positive. This signal is stronger in the case in which subgrid-scale motions are organized using the CA scheme, and most prominent in the L20G3 experiment. This behavior of our simple model suggests that the well-known reduced phase speed of convectively coupled waves applies only to the region of the wave that enhances convection. Observations of convectively coupled Kelvin waves confirm
that the convective region of the wave travels with a slower speed than the dynamical signal (usually lower tropospheric easterlies to the east of the convection). For instance, in Fig. 4 of Straub and Kiladis (2002), as well as Fig. 5 of Wheeler et al. (2000) and Fig. 7 of Kiladis et al. (2009), it appears that the observed anomalies ahead of the convection get progressively wider, while those regions coupled to the convection maintain their more compact size. Figure 4a shows the Hovmöller diagram of the experiment L20G3. The region of maximum convergence, where the interaction with the subgrid scheme occurs, is marked by a dashed white line. The region moves from 100°W to 28°E over a period of 6 days, giving an average phase speed of about 27 m s\(^{-1}\). The equivalent region for the uncoupled run is represented by a black dashed line, and follows the analytical phase speed of \(\sqrt{gh} = 31 \text{ m s}^{-1}\), where \(h\) is 100 m. The reduction of phase speed after 10 days is only 4 m s\(^{-1}\), which is a small reduction of phase speed compared with observations. For example, convectively coupled Kelvin waves are observed to propagate eastward at 15–20 m s\(^{-1}\), whereas dry Kelvin waves in the lower stratosphere propagate at speeds of 30–40 m s\(^{-1}\) (Wheeler and Webster 2000). To obtain a more substantial reduction in wind speed, we would also need to model wind shear, which is not represented in the SW model. However, it is interesting to observe that the introduction of subgrid organization using a CA reduces the phase speed notably. Such a reduction in phase speed is not seen using the random subgrid scheme (not shown). Furthermore, looking at the Hovmöller diagram of the fraction \(\sigma\) of active subgrid cells (Fig. 4b), it can be seen that subgrid cells can organize and propagate against the mean flow of the model, as a consequence of the neighborhood rules of the CA. This is a unique feature of the CA scheme, and such propagation is not seen in conventional deep convection parameterizations. Straub and Kiladis (2002) demonstrate that an eastward-propagating convectively coupled Kelvin wave envelope consists primarily of smaller-scale, westward-moving convective features. Thus, this is an interesting result, demonstrating the capabilities of a cellular automaton to be used within a deep convection parameterization.

From Figs. 3a–d it is evident that the different horizontal resolutions of the CA grid yield different sizes and numbers of the structures formed in the CA scheme, as well as different magnitudes of \(Q\). In the L20G7 experiment (Fig. 3d), where the CA grid resolution is 7 × 7 times as fine as for the SW model, the probability of having a nonzero \(\sigma\) on the model grid increases, even though this fraction may be small. Thus, L20G7 yields more frequent cells representing convection but on average less intense magnitude of \(Q\), compared with L20G3 (Fig. 3c). Similarly, in RANDOM (Fig. 3b) \(\sigma\) is generally smaller than in the different CA experiments, since the fraction of the grid box seeded is prescribed each time step and there is no interaction with neighboring cells, nor any memory time scale in the scheme that can contribute to form structures on the convective scales. The impact of the different tuning parameters, such as the lifetime and horizontal resolution of the CA grid, on the

---

**FIG. 4.** (a) A Hovmöller diagram of a U vector (m s\(^{-1}\)) of a single Kelvin wave from experiment L20G3. The region of maximum convergence of the L20G3 experiment is illustrated by a dashed white line. The equivalent region for the uncoupled run is represented by a black dashed line and follows \(\sqrt{gh} = 31 \text{ m s}^{-1}\). (b) Fraction of active subgrid cells from experiment L20G3, zoomed in over the last five forecast days.
spatial and temporal scales of the structures formed is discussed more thoroughly in the next subsection.

The option of advecting active subgrid cells by the large-scale wind is explored. Such advection of subgrid features is not readily done in conventional parameterizations, and it is of interest to see how it affects the interaction with a single Kelvin wave. The impact of the experiment ADVL20G3 can be seen in Fig. 3e, using the same parameters as in L20G3 (Fig. 3d). In the case of advection, the active subgrid cells are confined to the leading edge of the positive height anomaly for the eastward-propagating Kelvin wave. This is causing the wave to become more asymmetric, as the subgrid scales are allowed to interact for a longer time period with the leading edge. Without advection the active subgrid cells remain longer in one place, affecting a broader part of the wave as it travels over the region of active subgrid cells. Although the advection has a clear impact on the region over which the interaction occurs, it does not affect the phase speed of the wave significantly (not shown). The option of advection is not used in the following experiments with the broader spectrum of equatorial waves.

b. Spatial and temporal scales of CA structures

Although a prescribed $L$ is present, the rules of neighboring cells of the background state vector also influence how long an active subgrid cell can live (or when it dissipates). Similarly, the size of the clusters forming is also dependent on the prescribed lifetime. As noted above, the coupling between the subgrid-scale organization and the equatorial wave is sensitive to the spatial scales of the cells and clusters formed by the subgrid scheme. Thus, it is of interest to assess how the different parameters of the CA scheme determine the spatial extension and the actual lifetime of the clusters formed by the scheme. To do so, the model is initialized with a mixture of equatorial wave solutions providing realistic energy spectra.

In Fig. 5 the domain $\mathcal{Q}$ as a function of time for the different experiments is presented. Since we “turn on” the subgrid scheme in a model that is at rest, an initial spinup can be seen in all of the experiments lasting about 1.5–2 days, after which a quasi-steady state is reached. As the maximum values of $\mathcal{Q}$ within the domain are the same for all of the experiments with the same subgrid resolution, the difference in $\mathcal{Q}$ is a result of the frequency and size of the structures generated by the separate subgrid schemes. Thus, by construction, $\mathcal{Q}$ is the same for the two experiments using the random scheme (RANDOM and RANDOM, $L = 20$), and it is higher for the experiments using the CA scheme. When $L$ is increased in the CA scheme, clusters are allowed to survive longer, hence $\mathcal{Q}$ becomes larger with increasing $L$. No large difference is seen between L20G3 and L20G7 as far as the global mean of $\mathcal{Q}$ is concerned, although, as seen in Fig. 3, the structures generated from the two schemes look very different.

We now examine how the size of active grid clusters depends on the parameters of the subgrid parameterizations. The size of an active cluster is defined by the number of connected SW model grid boxes with a nonzero CA fraction. These clusters are identified with an algorithm which finds all clusters of connected cells with nonzero $\sigma$. The frequency of the number of elements within the identified clusters, for each time step throughout the model integration, is shown in Fig. 6a. Only clusters...
with size larger than two grid boxes are shown. The figure shows that for larger prescribed values of $L$, the size and number frequency of the cluster size increases. However, the shape of the frequency distribution is similar for different values of $L$. In all of the experiments, the frequency is largest for clusters of one. The experiment with higher horizontal resolution of the CA grid $L20G7$ yields a flatter frequency distribution with more large clusters than with the $L20G3$ experiment. Additionally, there are also more clusters identified as 1s and 2s in the $L20G7$ experiment. This behavior can be understood, since the probability is larger for having a nonzero fraction of active subgrid cells in the $7 \times 7$ than in the $3 \times 3$ experiment. A CA grid of $7 \times 7$ provides 49 possible active cells, while a mesh of $3 \times 3$ contains only 9. Similarly, there is a higher probability that several grid points on the model grid with active cells will appear next to each other in $L20G7$ compared with $L20G3$. When using the same $\alpha$, the smaller fraction reduces $Q$ averaged over the cluster. This is shown in Fig. 6b where $Q$ is summed over each identified cluster, multiplied by the number frequency, and plotted as a function of cluster size. It can be seen that $L20G7$ falls below the curves of $L20G3$ and $L20G3$ for convective clusters of 6–13 model grid points, although the frequency of these clusters was higher in the $L20G7$ experiment. This confirms that the larger clusters contain only a relatively weak mass source term. Within the two random experiments, almost all of the contribution to $Q$ comes from clusters of one SW model grid box. In summary, the maximum lifetime parameter increases the size frequency but keeps its shape, while the increase in resolution of the CA mesh leads to larger clusters but generally with a smaller mass source term.

To investigate how the prescribed maximum lifetime within the different experiments influences the subgrid structures formed, the total lifetimes of the individual active SW grid boxes were counted throughout the entire forecast period. Figure 7 shows the lifetimes that were longer than $L$ in the CA experiments. Each experiment is centered around its prescribed $L$; however, given the rules of the BF a subgrid cell may stay active longer than its prescribed maximum lifetime. In particular, experiment $L20G3$ displays nonzero cells, which can extend the prescribed maximum lifetime by more than a factor of 2. Furthermore, the experiment $L20G7$ with higher CA grid resolution yields more cells with a longer total lifetime than the experiment $L20G3$ having the same prescribed $L = 20$. The extension in lifetime occurs because of the interaction with its surrounding neighbors. The total lifetime therefore also depends on the horizontal resolution of the CA grid. However, it should be noted here that 90% of all active grid boxes achieve only the prescribed maximum lifetime. The impact of the CA rules on any additional lifetime is relatively small.

c. Kinetic energy spectra

The one-dimensional KE spectra are computed separately for each latitude band, between $30^\circ S$ and $30^\circ N$, at each time step of the last 5 days, and for each ensemble.
member. The KE spectra averaged over latitudes, time, and ensembles are plotted in Fig. 8 for the different experiments. The KE spectra of the initial conditions are determined by the prescribed variance distribution among the different equatorial wave solutions, as described in section 3. When the subgrid scheme is switched off, no kinetic energy is added to the system, although it can be redistributed between the different atmospheric scales through nonlinear interactions among the different wave modes. All experiments with the CA scheme display increased kinetic energy throughout the spectrum compared with simulations uncoupled to a subgrid scheme. An important contribution to this extra energy results from the memory time scale added to the scheme through the additional lifetime. The RANDOM case show very little added energy. With increased \( L \) in the CA simulations, most of the added kinetic energy is on the smallest and intermediate scales, and also visible for low wavenumber when \( L = 20 \) and \( L = 30 \). This behavior demonstrates that the CA scheme acts as a source of energy on the convective space scales, which can backscatter energy to the larger scales through organization on to scales larger than the truncation scales of the SW model. Interestingly, the experiments L20G3 and L20G7, which behave very differently in terms of spatial scales and intensity of \( Q \) but have similar \( \overline{Q} \), yield almost identical KE spectra, suggesting that the temporal scales (i.e., the additional lifetime of the subgrid-scale motion) is the decisive factor in terms of the amount of kinetic energy backscatter.

The KE spectrum of the RANDOM experiment lines up with the KE spectrum of the uncoupled case, hence the space and time scales are not large enough in the RANDOM experiment to contribute to any added energy in the system. However, when a lifetime is added to kinetic energy...
the random experiment as in RANDOM, \( L = 20 \), it can be seen that the KE spectrum is lifted slightly from the uncoupled case. Again, the temporal component of the subgrid motions plays a crucial role in the interaction with the dynamics. It should be noted that the comparably small global mean convective mass source in the RANDOM experiments makes the memory time-scale effect smaller than in the CA experiments. Figure 9 shows the KE spectrum difference between the experiments and the initial condition, scaled with the KE spectrum of the initial condition. There is no source of energy inserted into the system at the larger atmospheric length scales (greater than 500 km) in any of the experiments, and within the case coupled to a subgrid scheme there is only a sink of energy through horizontal diffusion. Therefore, with time, a loss of energy on the larger atmospheric scales (between wavenumbers 10 and 35, corresponding to scales between 500 and 1800 km) can be witnessed in the uncoupled case. The gain of energy within scales smaller than 500 km in the uncoupled experiment can be explained by a cascade of energy from larger to smaller atmospheric scales. In the random experiments, energy is inserted at the truncation scale of the model (~200 km for 3D), and without an additional lifetime, such a scheme does not illustrate a behavior that is markedly different from the results obtained in the uncoupled case. Only when the maximum lifetime is large enough to allow for organization of larger clusters to form, yielding an interaction with the equatorial waves, can a gain of energy on the larger scales be seen. Such behavior is most prominent in the L20G3 experiment, illustrating that there is a kinetic energy backscatter from the subgrid-scale motions, which is organized into clusters larger than the truncation scales of the model, to the large scales (greater than 500 km). Increasing \( L \) from 10 to 30 clearly increases the backscatter (Fig. 9; L20G3, L20G3, and L20G3), while the CA grid resolution apparently has no impact on the backscatter (cf. L20G3 and L20G7).

d. Stochasticity

A stochastic representation of processes that are too small or too fast to be explicitly modeled can lead to improvements in climate simulation and prediction (Palmer and Williams 2009). In the CA scheme, the stochasticity of the model comes from the random initial condition of the CA field. The interaction between the subgrid scheme and large-scale dynamics will lead to a spread in the wind and height field around an ensemble mean. Such a spread is not achieved by conventional deterministic parameterizations. Figure 10 shows the spread of the wind speed as a function of time (Fig. 10a) and the geographical distribution at one time instant (Fig. 10b). The nonlinearity of the CA scheme gives rise to an exponential growth of the spread, although the amplitude of the spread remains modest at about 1% of the large-scale winds. A large part of the spread in wind speed is found at the larger scales, even though the randomness of the CA parameterization is introduced initially at the subgrid scales. However, it can be seen that the spread is mostly confined to the region in which the subgrid scheme is active. In a full NWP model, we expect that nonlinear interaction and convective instability would lead to a larger spread of the wind speed compared with our simple model since 1) because of the scaling of the equivalent depth in the SW model, the wind speeds are substantially lower than in reality; 2) there is no wave

![Figure 9](https://example.com/figure9.png)

**Fig. 9.** As in Fig. 8, but for the difference between the KE spectra of the experiments and the KE spectra of the initial state, scaled with the KE spectra of the initial state (kg m\(^{-2}\) s\(^{-1}\)).
instability in the SW model, and thus no exponential or linear growth of the internal variance of the wind field is expected; and 3) the model, although initialized with a mixture of equatorial waves including different phases and amplitudes, has very limited nonlinear interactions with a slow advection of the equatorial waves.

Furthermore, as mentioned in section 2, the region over which the subgrid cells are allowed to be active is controlled by a threshold, $0.35 \times \max[\text{div}(\mathbf{U})]$, at each time step. This time step–dependent control of the “convective” activity ensures that subgrid cells do not organize and spread over a large region. This was chosen in order to stabilize the model simulations. However, such a condition also damps exponentially growing perturbations, and in order to achieve a larger ensemble spread, a threshold based on the divergence of the wind field at the initial time only could be explored.

5. Conclusions

We have, in an idealized setting, used a cellular automaton (CA) to study how a subgrid scheme with properties of lateral organization and a memory in time interacts with large-scale dynamical waves. Unlike a purely random equivalent, the present CA scheme yields an autocorrelation in both space and time.

Compared to a more conventional parameterization, the CA scheme can organize subgrid information across model grid boxes and propagate this information against the mean flow, mimicking organization by gravity waves initiated from deep convection. It has the option of advecting subgrid-scale information with the resolved flow, which is not implemented in a conventional parameterization.

Furthermore, the scheme is inherently stochastic, which generates an ensemble spread in the large scale.

It is demonstrated that for a single Kelvin mode, the proposed CA scheme allows for interaction between the simulated subgrid scales and the large-scale flow known from convectively coupled equatorial waves (Straub and Kiladis 2002). However, the modeled reduction in phase speed of the equatorial Kelvin mode occurs only in those regions where the wave enhances convection by strong low-level convergence. This effect leads to an asymmetry in the Kelvin wave where the “convectively” active area slows down, while the convectively suppressed area propagates with the phase speed of the dry Kelvin wave.

The kinetic energy spectra reveal a backscatter of energy from the smallest to the largest atmospheric scales, using the CA scheme. The time scales over which the subgrid-scale motions are allowed to interact with the dynamics appears to be the determining factor for the amount of energy backscatter. Random perturbations do not have the right properties to feed back on the large-scale flow (i.e., are not very efficient in mimicking the effects of organized, unresolved convection on the resolved, large-scale flow).

It is also demonstrated that the spatial and temporal scales, and number of clusters formed in the shallow-water model, depend strongly on a number of parameters given in the CA scheme, the horizontal resolution of the CA grid, and a prescribed memory time scale. The scales generated by the CA scheme determine the interaction between the subgrid scales and the dynamics. While the size of the active gridbox clusters is clearly affected by the resolution of the finer CA mesh, it showed no impact on the energy backscatter. For that process, the temporal scales of the CA are most important.
Bechtold et al. (2008) has shown that a change in entrainment rates and the adjustment time scale of a convection parameterization scheme has a clear effect on the large-scale variability of a forecast model. In addition, Tennant et al. (2011) have demonstrated that the introduction of a stochastic physics forcing to parameterize the effects of unresolved physical processes on the resolved scales has to have a significant fraction of its energy directly forcing large scales well above the resolution limit. Both studies point to the importance of the forcing acting on the large scales directly and that the stochastic process representation has a sufficiently long memory time scale. This is in agreement with findings in this study.

A natural extension to this work is to use a CA as part of a deep convection parameterization scheme in a state-of-the-art NWP model. Such work is ongoing, where ideas include using the CA as an additional term in the closure assumption of the description of the prognostic updraft mesh fraction in order to introduce additional lateral organization. In the study presented here, the rules of the CA were chosen such that a clustering of subgrid cells is achieved; also, it is important that the CA does not simply die out. One of the important aspects of the “game of life” set of rules is that the CA is active throughout the model integration. Of course, the rules of the CA are not set in stone, and one can think of many possible sets of rules, based on physical parameters of the model. For instance, one could use the topography of the model for CA initialization or set the rules to be probabilistic as a function of model parameters, rather than deterministic. However, the temporal continuity of a CA is very sensitive to the rules of the system, and if we introduce probabilistic rules we may need to seed new CA cells throughout the forecast period in order to keep the CA active. Such seeding could also be done on a physical basis. However, this could jeopardize the self-organization aspects introduced by the CA in the first place and therefore defeat the original purpose of the CA scheme.

Acknowledgments. We thank two anonymous reviewers for very helpful suggestions. Part of this work was performed within the framework of the COST ES0905 action.

REFERENCES