A simple dynamical model linking radiative-convective instability, convective aggregation and large-scale dynamics

Matthew Davison[1,*] and Peter Haynes[1,*]

[1]Department of Applied Mathematics and Theoretical Physics, University of Cambridge
[*]These authors contributed equally to this work

Correspondence: Matt Davison (md752@cam.ac.uk)

Abstract. A simple model is presented which is designed to analyse the relation between the phenomenon of convective aggregation at small scales and larger scale variability that results from coupling between dynamics and moisture in the tropical atmosphere. The model is based on the single-layer dynamical equations coupled to a moisture equation to represent the dynamical effects of latent heating and radiative heating. The moisture variable q evolves through the effect of horizontal convergence, nonlinear horizontal advection and diffusion. Following previous work, the coupling between moisture and dynamics is included in such a way that a horizontally homogeneous state may be unstable to inhomogeneous disturbances and, as a result, localised regions evolve towards either dry or moist states, with respectively divergence or convergence in the horizontal flow. The time evolution of the spatial structure of the dry and moist regions is investigated using a combination of theory and numerical simulation. One aspect of the evolution is a spatial coarsening that, if moist regions and dry regions are interpreted respectively as convecting and non-convecting, represents a form of convective aggregation. When the weak temperature gradient (WTG) approximation (i.e. a local balance between heating and convergence) applies and horizontal advection is neglected the system reduces to a nonlinear reaction-diffusion equation for q and the coarsening is a well-know aspect of such systems. When nonlinear advection of moisture is included the large-scale flow that arises from the spatial pattern of divergence and convergence leads to a distinctly different coarsening process. When thermal and frictional damping and f-plane rotation are included in the dynamics, there is a dynamical length scale L_{dyn} that sets an upper limit for the spatial coarsening of the moist and dry regions. The f-plane results provide a basis for interpreting the behaviour of the system on an equatorial β-plane, where the dynamics implies a displacement in the zonal direction of the divergence relative to q and hence to coherent equatorially confined zonally propagating disturbances, comprising separate moist and dry regions. In many cases the propagation speed and direction depend on the equatorial wave response to the moist heating, with the relative strength of the Rossby wave response to the Kelvin wave response determining whether the propagation is eastward or westward. Within this model, the key overall properties of the propagating disturbances, the spatial scale and the phase speed, depend on nonlinearity in the coupling between moisture and dynamics and any linear theory for such disturbances therefore has limited usefulness. The model described here, in which the moisture and dynamical fields vary in two spatial dimensions and important aspects of nonlinearity are captured, provides an intermediate model between theoretical models based on linearisation and one spatial dimension and GCMs or convection-resolving models.

1 Introduction

Much theoretical and modelling work over the past few decades has focused on the coupling between dynamics and moisture in the tropical atmosphere, which it has been argued should be taken into account at leading-order to explain many tropical phenomena. Two topics that continue to attract significant attention are on the one hand convective aggregation, identified as a behaviour in numerical simulations in convection-representing models, and on the other the Madden-Julian Oscillation (hereafter MJO), identified in observations as a dominant mode of intraseasonal variability of the real tropical atmosphere. Convective aggregation (e.g Wing et al., 2017; Muller et al., 2022) has been identified as a pattern of behaviour of a hypothetical tropical atmosphere in which there is no imposed spatial inhomogeneity, but which exhibits spontaneous organisation of the circulation and convection into regions of two types, one type with active convection and large-scale ascent, and the other with convection suppressed by large-scale subsidence. The relevance of convective aggregation to the behaviour of the real atmosphere remains a topic of debate, but the study of aggregation in convection-representing numerical models (hereafter CRMs) has provided a great deal of insight into the physics of the tropical atmosphere, particularly the interactions between convecting and non-convecting regions, and the way in which this physics is represented in models. It has been suggested that the same physics that is responsible for convective aggregation in numerical simulations is also part of the mechanism for the MJO (Bretherton et al., 2005; Arnold and Randall, 2015). Investigating this possibility in numerical simulations has been challenging because the numerical resolution for CRM simulations is a few km or less and the spatial structure of the MJO has scales of several thousand km. A small number of papers (e.g Arnold and Randall, 2015; Khairoutdinov and Emanuel, 2018) have described simulations that bridge this gap and have given insight into the relation between convective aggregation and the spontaneous generation of large-scale MJO-like disturbances. Other papers (e.g Carstens and Wing, 2022, 2023) have used CRM simulations to explore more generally the relation between convective aggregation and larger scale dynamics. Since such simulations are at the very edge of current computational capacity and the scope for thorough examination of parameter space is limited, it is desirable to find a simpler theoretical and modelling framework within which the link between processes such as aggregration on the mesoscale and larger scale organisation of dynamics can be investigated further. The focus of this paper is the formulation and study of a simple model for such a purpose.

Several different physical processes have been proposed as important for aggregation, often supported by the results of mechanism-denial experiments in CRMs in which the effects of particular processes have been altered or omitted altogether. As emphasised by Muller et al. (2022), there remains considerable uncertainty over which of these various descriptions of the aggregration process is most relevant to convective aggregation and, within each of them, the relative importance of different physical mechanisms. Examples of suggested mechanisms include the spatio-temporal propagation of triggering of convection through different effects, either through the formation and propagation of cold pools (e.g. Hirt et al., 2020), or through the propagation of gravity waves within the boundary layer (Yang, 2021). The study described in this paper is prompted in particular by the proposal that aggregation occurs via an instability of a spatially homogeneous radiative-convective equilibrium

state resulting from feedbacks between moisture and radiation (Raymond, 2000; Emanuel et al., 2014), primarily in the free troposphere, though the boundary layer (Yang, 2018) may play a crucial role in the dynamics of these feedbacks. Instability, generally described as radiative instability or radiative-convective instability, may result if these feedbacks can overcome the dynamical stability of the convecting atmosphere. The competition between these processes is typically represented by the gross moist stability (e.g Raymond et al., 2009), though changes in convective moisture and heat transport should be taken properly into account (Beucler et al., 2018). The growing instability is expected ultimately to saturate at finite amplitude with the result that locally the system tends towards a moist or a dry state. Such behaviour has been demonstrated using single-column radiation-convection calculations using various standard radiative and convective parametrizations and adopting the weak temperature gradient (WTG) assumption where the environmental temperature is specified and the vertical mass flux is allowed to be non-zero (Sobel et al., 2007; Emanuel et al., 2014). Under suitable conditions the radiative-convective (RCE) state is unstable and the system evolves towards a moist state or a dry state, depending on the imposed initial perturbation.

A more generic and fundamental approach to the study of convective aggregation, starting with instability as the initial mechanism for the growth of moisture inhomogeneities on the homogeneous state, but then seeking a description of the spatial evolution of the resulting moist and dry regions, has been undertaken by Craig and Mack (2013) and Windmiller and Craig (2019) (hereafter CMWC). The model system considered in this work is an evolution equation for a time-evolving two-dimensional moisture concentration field q , incorporating a source-sink term that is a nonlinear function of q, $G(q)$ say, with three zeros, each corresponding to possible steady states. The form of the function $G(q)$ is such that the large-q (moist) and small-q (dry) states are stable and the intermediate-q state is unstable. Transport of moisture is assumed to be diffusive. This system is equivalent to a reaction-diffusion equation with bistable reaction, sometimes known as the Allen-Cahn equation, which has been much studied using theoretical and numerical approaches. Such systems exhibit *coarsening* where, after the initial separation into high-q and low-q regions, typically on small scales, the scale of these regions increases monotonically with time, until constrained by the large-scale geometry imposed on the system. This is presented by CMWC as a mechanism for aggregation. The two essential ingredients required for this mechanism to operate are the q-dependence, and hence the 'bistable' nature, of the source-sink term, which CMWC argue results from the dependence of subsidence drying and convective moistening on free-tropospheric moisture, and the diffusive transport. Windmiller and Craig (2019) argue that diffusive transport can be justified on the basis of a simple model in which a stochastic convective cloud moistens the environment of a moist region and provide an estimate for the resulting diffusivity as $4 \times 10^2 \mathrm{m}^2\mathrm{s}^{-1}$. A larger value of the diffusivity, $10^5 \mathrm{m}^2\mathrm{s}^{-1}$, is used in Craig and Mack (2013) envisaged as based on the typical horizontal velocity and length scales of convective systems. The latter justify this as an eddy diffusivity based on the typical horizontal velocity and length scales of convective motions. The appropriate value of the diffusivity therefore depends significantly on what the diffusivity is intended to represent. This reaction-diffusion model for aggregation is interesting, and it is discussed further in the Muller et al. (2022) review, but, since large-scale dynamics is omitted from the model, it cannot be used to investigate the link between convective aggregation and large-scale dynamics.

The model presented in this paper combines aspects of the CMWC model with a simple description of large-scale dynamics. The structure of this paper is as follows. In Sect. 2 we define the mathematical model to be studied, following previous

approaches in using the shallow-water equations augmented by a prognostic equation for moisture, with the moisture coupling to the shallow-water dynamics. We then set out the behaviour expected of the model on the basis of previous work together with various scaling arguments, explaining the relation to CMWC. In Sect. 3 we present results from numerical simulations of the system on a doubly periodic domain. In particular we verify that aggregation occurs and that the mechanism for aggregation can be dominated by horizonal diffusion of moisture or by horizontal advection of moisture, depending on the external parameters defining the system. In Sect. 4 we then show that rotation, thermal damping and frictional damping can each, or in combination, lead to a finite upper limit on the aggregation scale. This is both with theoretical arguments and results from numerical simulations on a doubly periodic f-plane (including the zero-rotation case $f = 0$). Then in Sect. 5 we exploit the f-plane results derived previously to consider the system on an equatorial β-plane and show that the process of aggregation is then confined to a low-latitude region with the result of aggregation being the formation of coherent propagating disturbances. The scale and propagation speed of these disturbances depend on the external parameters, in particular, in some regimes, the relative strength of the equatorial Kelvin and Rossby wave responses to the moist heating. In Sect. 6 we discuss the results and present overall conclusions.

Whilst we have mentioned above the MJO as a key motivation for further investigation of tropical dynamics, the model as presented in this paper does not reach the stage of direct relevance to the MJO or other aspects of tropical intraseasonal variability. That requires further development to be reported in a future paper. We simply note that the model presented here can be regarded as belonging to the class of 'moisture-mode' models studied by Sugiyama (2009a, b); Sobel and Maloney (2012, 2013); Adames and Kim (2016) and others, but that this class is one several different classes of theoretical models for the MJO that continue to be studied, as recently reviewed by Jiang et al. (2020); Zhang et al. (2020).

2 Model system to be studied

2.1 Model equations

The model to be studied in the remainder of the paper follows much previous work in being based on the dynamical equations for a rotating shallow-water system, describing the evolution of horizontal velocity $\boldsymbol{u}$ and free-surface displacement h, augmented by a moisture variable q which is transported by the horizontal velocity. Such single-layer equations can be derived from the primitive equations and the corresponding moisture equation for a stratified 3-D atmosphere following the systematic procedure set out by Neelin and Zeng (2000) based on a vertical mode decomposition for dynamical quantities and for moisture. Specific further simplifications, following e.g. Sugiyama (2009b), over the set of equations presented by Neelin and Zeng (2000) (their 5.1-5.4) are that the barotropic flow and the nonlinear terms in the dynamical equations are neglected.

To emphasise the overall structure of the equations, rather than the details determined by the modal decomposition, we simply write the dynamical equations in the standard form for the motion, linearised about a state of rest, of a shallow layer of fluid with undisturbed depth H:

$$\boldsymbol{u}_t = -f\boldsymbol{k} \times \boldsymbol{u} - g\nabla h - \alpha \boldsymbol{u} \tag{1}$$

and

$$h_t + H\nabla.\boldsymbol{u} = F_h(q) - \lambda h, \tag{2}$$

where g is the gravitational acceleration and f is the Coriolis parameter which will either be taken to be constant $f = f_0$ corresponding the the f-plane, or to be linearly dependent on the y coordinate, $f = \beta y$, corresponding to the equatorial β-plane. $\boldsymbol{k}$ is the unit vector in the vertical. α is a linear friction coefficient and λ a thermal damping rate. The corresponding equation for the moisture variable q is

$$q_t + Q\nabla \cdot \boldsymbol{u} + \nabla \cdot (q\boldsymbol{u}) - \kappa\nabla^2 q = F_q(q), \tag{3}$$

including both horizontal advective transport and horizonal diffusive transport, the latter with diffusivity κ, assumed constant. In these equations $\boldsymbol{u}$ is to be interpreted as representing the lower tropospheric horizontal velocity field and the displacement h is a surrogate for mid-troposphere temperature, with temperature increasing as h decreases. Given typical choices for the vertical basis function for moisture (Zeng et al., 2000), q is to be interpreted as the lower tropospheric moisture concentration. The values of the constants H, α and λ could be chosen to match the values implied by the modal decomposition and by corresponding approximations to detailed parametrizations of physical processes, as set out in Neelin and Zeng (2000); Sugiyama (2009b), but for the purposes of the work reported here we will simply chose values typical of those chosen in previous work on simple modelling of the tropical atmosphere.

The term $F_h(q)$ included on the right-hand side of Eq. (2) represents a moisture-dependent cooling term (cooling because of the relation between h and temperature), potentially including both latent heating and radiative heating. If cooling decreases with moisture, as is physically plausible, then $F_h(q)$ will be a decreasing function of q. Correspondingly the term $F_q(q)$ on the right-hand side of Eq. (3) represents the combined effects of evaporation and precipitation. The part of $F_q(q)$ representing precipitation (note that $F_q(q)$ is a negative contribution to precipitation) is also expected, on the basis of the observed correlation between precipitation and moisture in the free troposphere (e.g. Holloway and Neelin, 2009), to be a decreasing function of q. Q is a suitably chosen constant, so that q is the perturbation away from a background state where the 'total' moisture variable is Q, and it is convenient to choose Q such that $F_q(0) = 0$, i.e. such that the in the background state there is a balance between evaporation and precipitation. Previous papers developing and exploiting the simplified set of equations, Eq. (1), Eq. (2) and Eq. (3), such as Neelin and Zeng (2000); Zeng et al. (2000); Sugiyama (2009b) have given detailed arguments for how these moisture-dependent terms $F_h(q)$ and $F_q(q)$ might be constructed from specific parametrization schemes for radiation and convection. Here we will choose simplified ad hoc forms for these terms. A specific simplification made is that both terms F_q and F_h are assumed to be independent of h, i.e. the evaporation-precipitation and the moisture dependent part are assumed to be independent of temperature. Whilst temperature (or h) dependence of these moisture coupling terms is neglected in many several previous papers (Adames and Kim, 2016; Sobel and Maloney, 2012, 2013) it is included by Sugiyama (2009a, b). We have briefly examined the effect of such dependence, with results reported in Appendix E and will comment further in in Sect. 6.

Within the constraints of the very simple model specified above, we may identify the possibility of choosing H and Q such that $h = q = 0$ corresponds to a spatially homogeneous radiative-convective equilibrium (RCE) state with $\boldsymbol{u} = \boldsymbol{0}$, and

F_h and F_q satisfying the conditions $F_h(0) = F_q(0) = 0$. We may restrict the forms of $F_h(q)$ and $F_q(q)$ to those for which the system is stable to spatially homogeneous perturbations, which holds if $\partial F_q/\partial q < 0$, with the partial derivative being evaluated at $q = 0$. Then a key question is whether, within this restriction, the radiative-convective equilibrium (RCE) state is unstable to spatially inhomogenous disturbances. We will demonstrate below that such instability is possible for relatively simple choices of the functions $F_h(q)$ and $F_q(q)$ and will interpret this instability as the analogue, within this simple model, of 'radiative-convective instability' that has previously been discussed in several papers and demonstrated in suitable single-column radiation-convection calculations (e.g Sobel et al., 2007; Emanuel et al., 2014).

The key dimensional quantities that define the above system include g and H, which determine the dry gravity-wave speed $c = \sqrt{gH}$, the horizonal diffusivity κ, the Coriolis parameter f, the thermal and frictional damping rates, respectively λ and α, a typical background value of moisture, Q, say, and μ, an inverse timescale for the moist processes represented by F_h and F_q. It is convenient to take the dimensions of the moisture Q to be the same as those of the thickness H, and indeed this corresponds to a simple re-scaling of the parameters in F_h. To assess the importance of the advective term in Eq. (3) an additional dimensional quantity is needed which sets the magnitude of the spatially inhomogeneous part of the moisture field and, hence, via the leading-order balances operating in Eq. (2) and Eq. (3), the magnitude of the corresponding horizontal flow. These magnitudes are set by the nonlinear dependence of F_q and F_h on q, but it is convenient to choose the magnitude D, say, of the divergence $\nabla.\boldsymbol{u}$, as the relevant dimensional quantity. The reason for this will become clear from the discussion in Sect. 2.2 below.

The fact that the magnitude of the advective nonlinearity arising in the equations, Eq. (1), Eq. (2) and Eq. (3), as written above, depends on the choice of $F_h(q)$ and $F_q(q)$, potentially complicates interpretation of the behaviour of the system and there are advantages in allowing the advective nonlinearity to be varied independently of this choice. We therefore introduce the parameter ϵ into Eq. (3) to give

$$q_t + Q\nabla \cdot \boldsymbol{u} + \epsilon\nabla \cdot (\boldsymbol{u}q) - \kappa\nabla^2 q = F_q(q). \tag{4}$$

The linear instability analysis of the RCE state, for example, takes $\epsilon = 0$ in the above equation. Note that in the form of the equations considered by Sugiyama (2009b), derived from the Neelin and Zeng (2000) QTCM equations, there is a distinct constant multiplying the nonlinear advective term. This constant is determined in principle in the derivation of the single-layer equations by the projection of a horizontal moisture advection term that is varying in height on to the single basis function used to represent the moisture field, but can also be conveniently be varied as a independent parameter, and the ϵ being introduced here plays the same role as that parameter. We later illustrate the role of advective nonlinearity by comparing $\epsilon = 1$ behaviour with $\epsilon = 0$ behaviour, but note that, for the reasons just given, $\epsilon = 1$ cannot be regarded as the only 'correct' choice for including advective nonlinearity.

2.2 WTG and the relation to the CMWC reaction-diffusion system

A standard approach, particularly at low latitudes where f is small, to analysing the system defined by Eq. (1)–(3) is to make the *weak temperature gradient approximation* (WTG) (e.g. Sobel et al., 2001). This neglects horizontal variation of h and can

be justified provided that the horizontal length scale L satisfies $L/c \ll T_q$ i.e. that the time scale for gravity-wave propagation through L is much less than the time scale T_q for moist processes. T_q could either be a timescale μ^{-1} set by an appropriate combination of F_q and F_h (see below) or an emergent property of the system. Additionally, when damping and rotation are included, it must be the case that $L \ll L_{\text{dyn}}$, where L_{dyn} is a dynamical length scale that is typically determined by c together with some combination of f, α and λ. We will focus on the zero damping case in this section and return to the dynamical effects of damping and rotation in Sect. 4.

Whilst under WTG h is constant in space, it may not be constant in time. Taking the spatial average of Eq. (2), using the notation $\overline{}$ to denote the spatial average. It follows that

$$\frac{d\overline{h}}{dt} = \overline{F_h(q)} - \lambda\overline{h}. \tag{5}$$

The spatially varying part of Eq. (2) then has the form

$$H\nabla.\boldsymbol{u} = F_h(q) - \overline{F_h(q)}. \tag{6}$$

implying that $\nabla \cdot \boldsymbol{u}$ and hence the irrotational part of the velocity field, is determined instantaneously by the moisture field q. Under the assumption $f = \alpha = 0$ in this section, the rotational part of $\boldsymbol{u}$ is constant in time. When provided with this initial rotational part of the flow, assumed to be zero for the purposes of this section, Eq. (4) becomes a self-contained equation for the evolution of the q field with the form:

$$q_t + \epsilon\nabla \cdot (\boldsymbol{u}[q]q) - \kappa\nabla^2 q = F_q(q) - \frac{Q}{H}(F_h(q) - \overline{F_h(q)}) = G_{hq}(q; \overline{F_h(q)}). \tag{7}$$

where the second equality defines the function G_{hq}. Note that whilst evaluation of $\overline{F_h(q)}$ requires knowledge of the q field, for the purposes of expressing the right-hand side of the equation as a function of q, $\overline{F_h(q)}$ is simply a parameter that appears in the definition of that function. The notation $\boldsymbol{u}[q]$ simply expresses the fact that at each instant $\boldsymbol{u}$ is determined completely, but non-locally, by the q field, through Eq. (6).

Neglecting for the moment the advection term $\epsilon\boldsymbol{u}[q].\nabla q$, this may be recognised as a reaction-diffusion equation of the type studied by CMWC. The difference is that, whereas the nonlinear 'reaction' term on the right-hand side of Eq. (7) was in CMWC's case entirely motivated by the q-dependence of precipitation and evaporation, in this case the reaction term is a combination of the moisture driven heating/cooling $F_h(q)$ and the moisture source/sink $F_q(q)$. A further structural difference from the system considered by CMWC is the evolving quantity $\overline{h}(t)$. The effect of this is felt by the system through the corresponding $\overline{F_h(q)}$ appearing in the definition of the reaction term. The reaction term is therefore not completely specified in advance as a function of q but contains the spatially constant term $\overline{F_h(q)}$ which also drives changes in $\overline{h}$, as specified by Eq. (5). The CMWC model, on the other hand, defines the reaction term as $G(q) - \overline{G(q)}$, where $G(q)$, as $F_h(q)$ and $F_q(q)$ in the model being presented in this study, is a function that is specified in advance. Again this means that the complete reaction term requires knowledge of the spatial distribution of q.

Simple theory of the reaction-diffusion system with specified reaction term $G(q)$ is that (i) homogenous steady states are possible with q equal to the constant value q_s, if $G(q_s) = 0$ and (ii) those homogeneous states are stable if $G'(q_s) < 0$

and unstable if $G'(q_s) > 0$. CMWC consider a 'bistable' system with three possible values for q_s, $q_- < q_0 < q_+$, such that $G(q_-) = G(q_0) = G(q_+) = 0$, and $G(q_-) < 0$, $G'(q_0) > 0$, $G'(q_+) < 0$. $G'(q_0)$ provides a useful definition of a reaction inverse timescale μ. The generic behaviour for a non-linear reaction diffusion equation of this type is that locally q tends to one of the stable values, partitioning the domain into two regions one with $q = q_+$ and the other with $q = q_-$, separated by interfaces of thickness $(\kappa/\mu)^{1/2}$. In the absence of rotation and damping WTG will break down on length scales of order cT_q, so we require $cT_q \gg (\kappa/\mu)^{1/2}$, i.e., if the reaction timescale μ^{-1} is such that $\kappa\mu \ll c^2$. The initial geometry of these two regions is set by the initial conditions. A useful simple solution is a 1-dimensional propagating reactive-diffusive wave solution with $q = q_+$ on one side of the wave and $q = q_-$ on the other. The speed of propagation of the wave, $c_{\mathrm{RD}} \sim (\kappa\mu)^{1/2}$ is determined by the form of the reaction function $G(q)$. Defining $V(q)$ by $dV/dq = G$, so that $V(q)$ has turning points where $G(q)$ has zeros, then if $V(q_+) > V(q_-)$ the region with $q = q_+$ propagates into the region with $q = q_-$. The corresponding result for the initial value problem, in one or more space dimensions, is that q tends everywhere to q_+. Similarly if $V(q_+) < V(q_-)$ then q eventually tends everywhere to q_-. Only in the case $V(q_+) = V(q_-)$, which applies in particular to the Allen-Cahn equation, do both regions $q = q_+$ and $q = q_-$ persist. Note that $V(q)$ represents the area under the graph of $G(q)$. To be precise, if the choice $V(q_0) = 0$ is made, then $V(q_+)$ is the area under the graph of $V(q)$ in the interval $[q_0, q_+]$ and $V(q_-)$ is the corresponding (positive) area in the interval $[q_-, q_0]$.

An important effect in two dimensions is that the reaction-diffusion velocity c_{RD} becomes a local property of each point on each interface, depending not only on the form of $G(q)$ but also on curvature of the interface. The reaction velocity c_{RD} should be replaced by $c_{\mathrm{RD}} + \kappa/R$, where R is the (signed) radius of curvature of the interface (such that that the propagation is towards the interior of the curve). If $|c_{\mathrm{RD}}| < \kappa/|R|$ the velocity speed of the boundary may even change sign. Therefore, c_{RD} decreases as the curvature of the interface increases (R decreases) (Rubinstein et al., 1989). This tends to smooth out the boundary between moist and dry regions, as small -scale irregularities or indeed small-scale regions will tend to disappear. Larger moist regions can therefore expand while smaller moist regions shrink. This is the standard coarsening behaviour, i.e. the geometric simplification of the geometry between the two regions through an increase in spatial scales, observed in reaction-diffusion systems (Bray et al., 2003).

CMWC's inclusion of the $\overline{G(q)}$ term, so that the total reaction term becomes $G_{CMWC}(q,t) = G(q) - \overline{G(q)}$, is important because this ensures that even if $V(q_+) \neq V(q_-)$ (with $V(q)$ defined as above), the system does not simply evolve to $q = q_-$ or $q = q_+$ everywhere. Both values of q persist as coarsening proceeds. Indeed if this sort of constraint is not applied then in most cases the reaction-diffusion system with a bistable reaction evolves everywhere towards one of the stable states. It will be demonstrated below that the same property holds for the model system being considered in this study, i.e. for a reaction-diffusion system with the reaction term as specified by the right-hand side of Eq. (7).

Following the arguments presented above, the stability of the spatially homogeneous RCE state will be determined by the derivative with respect to q of $G_{hq}(q,0) = F_q(q) - (Q/H)F_h(q)$ at $q = 0$, with instability if the derivative is positive, provided that the domain size is large enough that diffusion does not stabilise the system through the action of the $\kappa\nabla^2 q$ term. (This derivative will later be identified as proportional to the negative of the gross moist stability, i.e. the RCE state will be unstable if the gross moist stability is negative.) It will be assumed that the derivative is indeed positive and furthermore that $G_{hq}(q;0)$ is

bistable in the sense that there are $q_+(0)$ and $q_-(0)$ such that $q_-(0) < 0 < q_+(0)$, with $G_{hq}(q_-(0);0) = G_{hq}(q_+(0);0) = 0$ and $G'_{hq}(q_-(0);0) < -0$, $G'_{hq}(q_+(0);0) < 0$. Note that this property of $G_{hq}(q;0)$ implies a similar property, with corresponding $q_-(\overline{F_h(q)})$, $q_0(\overline{F_h(q)})$ and $q_+(\overline{F_h(q)})$, for the more general right-hand side of Eq. (7) $G_{hq}(q;\overline{F_h(q)})$ provided that $|\overline{F_h(q)}|$ is not too large, For notational convenience the explicit dependence of e.g. $q_-(\overline{F_h(q)})$ on $\overline{F_h(q)}$ will not be displayed unless essential.

Numerical solutions below will show that if G_{hq} is bistable in the sense defined then the system indeed evolves towards two values of q and that coarsening occurs. However some further insight can be obtained by assuming that after the initial adjustment the region $q = q_+$ has fills an area fraction A_+ and the region $q = q_-$ fills an area fraction A_-, with $A_+ + A_- = 1$. The area of the interfaces between the regions is assumed negligible. The configuration is therefore determined by the three unknowns q_-, q_+ and A_+ (or A_-).

Then the above equations imply

$$G_{hq}(q_+, \overline{F_h(q)}) = F_q(q_+) - (Q/H)(F_h(q_+) - \overline{F_h(q)}) = 0 \tag{8}$$

$$G_{hq}(q_-, \overline{F_h(q)}) = F_q(q_-) - (Q/H)(F_h(q_-) - \overline{F_h(q)}) = 0 \tag{9}$$

$$\overline{F_h(q)} = A_+ F_h(q_+) + A_- F_h(q_-). \tag{10}$$

These determine any two of the three variables q_-, q_+ and A_+ in terms of the third. For the system to be at a steady state an additional constraint, obtained by integrating Eq. (3) over the domain, might seem to be $A_+ F_q(q_+) + A_- F_q(q_-) = 0$, however this can be deduced from Eq. (8) and Eq. (9) above and provides no extra information. Therefore it has to be accepted that one piece of further information in addition to the above is required for a unique solution. In general these equations determine a state that is quasi-steady rather than exactly steady. After the initial adjustment, when the required extra information will be determined by the initial conditions (which might, for example, set A_+), we expect a further slow time evolution of the three variables and correspondingly of the geometry of the dry and moist regions.

Assume that the overall effect of this slow time evolution can be captured by the classical theory for 1-D reactive diffusive waves, describing the propagation of the thin interfaces between regions of piecewise constant q. This suggests the (slow) time evolution equation

$$\frac{dA_+}{dt} = L_{\text{interface}} c_{\text{RD}}(q_+, q_-, \overline{F_h(q)}) \tag{11}$$

where $L_{\text{interface}}$ is the length of the interface between the regions and c_{RD} is the reaction diffusion velocity, with the convention that this is positive if the region with $q = q_+$ propagates into the region with $q = q_-$. $L_{\text{interface}}$ will vary in time but is certainly positive. This equation allows a steady state when $c_{\text{RD}}(q_+, q_-, \overline{F_h(q)}) = 0$.

That such a steady state exists can be deduced by considering the graphs of relevant functions of q. As noted previously, for given $\overline{F_h(q)}$ the reaction function is $G_{hq}(q, \overline{F_h(q)}) = F_q(q) - (Q/H)F_h(q) + (Q/H)\overline{F_h(q)})$. Consider first the graph of $G_{hq}(q, 0) = F_q(q) - (Q/H)F_h(q)$ as shown by the curve in Fig. 1, which intersects the q-axis at $q_-(0)$, $q_0(0)$ and $q_+(0)$. The value of $G_{hq}(q, \overline{F_h(q)})$ is represented by the vertical distance between this curve and the horizontal line $-(Q/H)\overline{F_h(q)}$, shown on the figure for various values of $\overline{F_h(q)}$, varying between $C_{\text{min}} < 0$ and $C_{\text{max}} > 0$. For C outside this range then

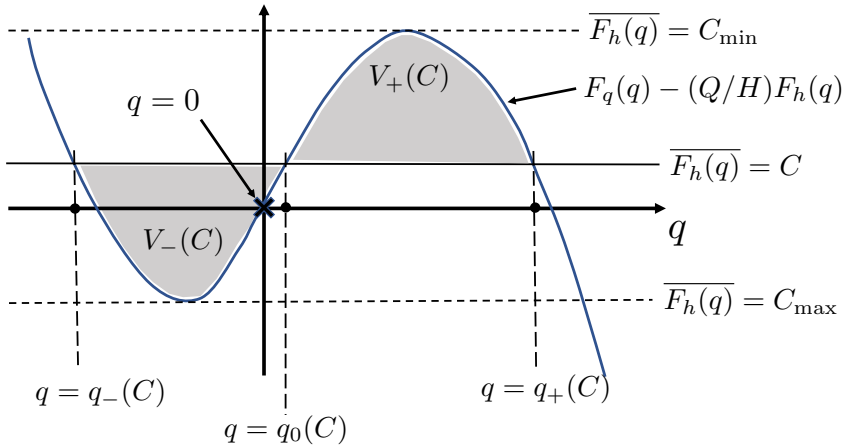

Figure 1. Functions of q controlling the behaviour. The curve shows the function $G_{hq}(q, \overline{F_h(q)}) = F_q(q) - (Q/H)F_h(q) + (Q/H)\overline{F_h(q)}$ for the value $\overline{F_h(q)} = 0$. $q = 0$ marked by the cross, corresponds to the homogeneous RCE state. Various horizontal straight lines are shown corresponding to the values $-(Q/H)\overline{F_h(q)}$ for different values of $\overline{F_h(q)}$. The value of the function $G_{hq}(q, \overline{F_h(q)}) = F_q(q) - (Q/H)F_h(q) + (Q/H)\overline{F_h(q)}$ corresponds to the vertical difference between the curve and the relevant line. For $C_{\min} < \overline{F_h(q)}) < C_{\max}$ the curve and the straight line intersect at three values of q, denoted by $q_-(\overline{F_h(q)})$, $q_0(\overline{F_h(q)})$ and $q_+(\overline{F_h(q)})$. These are indicated in the diagram for $q_-(\overline{F_h(q)}) = C$). The areas between the curve and the straight line in the intervals $[q_-(C), q_0(C)]$ and $[q_0(C), q_+(C)]$ are denoted by $V_-(C)$ and $V_+(C)$ respectively. Note that it is clear that there is a choice of C, with $C_{\min} < C < C_{\max}$ such that $V_-(C) = V_+(C)$.

$G_{hq}(q; C)$ no longer has three roots. Areas V_+ and V_- are marked on the figure for a particular value of $\overline{F_h(q)}$. The condition $c_{\text{RD}}(q_+, q_-, \overline{F_h(q)}) = 0$ is satisfied if and only if $V_+ = V_-$. It is clear from the Figure that there is one value of $\overline{F_h(q)}$, C_s say, for which this holds, lying in the range $(C_{\min}, C_{\max})$. Substituting this value into Eq. (8)-(10) gives the corresponding values of q_- and q_+, the dry and moist values of q, and A_+ the fractional area occupied by the moist region.

A further question concerns the stability of this steady state. It is clear from Fig. 1 that c_{RD} is an increasing function of C (the area V_+ increases and the area V_- decreases as C increases. Suppose that $C > C_s$, so that $V_+(C) > V_-(C)$ and c_{RD} is positive, i.e. regions of q_+ will propagate into regions of q_-. The consequence will be that the relative area occupied by q_+ will

increase resulting in a decrease in $C = \overline{F_h(q)}$, if $F_h(q)$ is a decreasing function of q. Similarly if $C < C_s$ then C will increase, indicating that the steady state $C = C_s$ is stable. Note that the above arguments do not describe the process of coarsening, but indicate that the two values of q persist, just as they do for the special case of the Allen-Cahn equation and for the system considered by CMWC, suggesting that coarsening is relevant.

We will show examples in Sect. 3 to demonstrate that the our model system, in regimes where the approximations leading to Eq. (5) and Eq. (7) can be justified, naturally evolves to a piecewise constant configuration with both of the values of q, q_- and q_+, consistent with the relevant form of $G_{hq}(q, \overline{F_h(q)})$.

The strict WTG form of the evolution equation, Eq. (7), suggests that for the system Eq. (1), Eq. (2) and Eq. (4) locally q will tend to one of two values q_- or q_+ as was the case for the CMWC pure reaction-diffusion system. The distinct additional feature of the system being considered here is that there is an associated pattern of convergence and divergence. The WTG balance in Eq. (2) implies that the divergence $\nabla.\boldsymbol{u}$ will also tend to one of two values respectively $D_- = -H^{-1}F_h(q_-, \overline{h}) = Q^{-1}F_q(q_-, \overline{h}) > 0$ or $D_+ = -H^{-1}F_h(q_+, \overline{h}) = Q^{-1}F_q(q_+, \overline{h}) < 0$. (Note that these estimates of convergence in moist regions and divergence in dry regions are the basis for the discussion above concerning the introduction of the parameter ϵ to control the magnitude of advective nonlinearity.) Assume that the corresponding values of $\nabla \cdot \boldsymbol{u}$ are $-D_- < 0 < D_+$. Since area integrated $\nabla \cdot \boldsymbol{u}$ is zero we expect that the areas A_- and A_+ filled respectively by dry regions and moist regions satisfy $A_- D_- \sim A_+ D_+$.

This non-zero divergence has no consequence for the evolution of the system under the strict WTG approximation, but if this approximation is relaxed, as is likely to be required at large horizontal scales, then the coupling between moisture and divergence may lead to distinctly different behaviour from that predicted by reaction-diffusion alone. Furthermore, even if WTG can be justified in Eq. (2), the non-zero divergence will potentially be important if the nonlinear advection term $\epsilon \boldsymbol{u}[q].\nabla q$ is included, since Eq. (2) implies an advecting velocity field that depends on the distribution of q. A typical velocity at length scale L will be $U \sim DL$. The advective velocity becomes comparable with the reaction-diffusion velocity when $\epsilon DL \sim (\kappa\mu)^{1/2}$, i.e. $L = L_{adv} \sim (\kappa\mu/D^2\epsilon^2)^{1/2}$. Since moist regions are associated with convergence and dry regions with divergence, the effect of advection will be to reduce the area of moist regions relative to those of dry regions. This suggests that in a steady state the reaction-diffusion velocity c_{RD} has to be positive rather than zero, i.e. that $V_+ > V_-$ rather than $V_+ = V_-$. It also suggests that the values of q in both moist and dry regions are increased relative to their values without the advective nonlinearity. On scales larger than those of a single aggregated region, the combination of regions of divergence of opposite signs will generate a large scale velocity field. Nonlinear advection will therefore cause aggregation when there is a high density of distinct convergent regions, on a timescale of ϵ/D. The advective regime of aggregation becomes dominant for length scales $L > L_{adv}$. The consequence of this will be illustrated in Sect. 3.2.

We can also now give a better estimate for T_q, and hence the scale at which WTG will arrest. In the linear instability phase a possible estimate is $T_q \sim \mu^{-1}$. However in the nonlinear aggregation phase a potentially more relevant estimate is $T_q \sim L/c_{\text{RD}}$, i.e. the timescale increases as the length scale increases. In this case where $\alpha = \lambda = f = 0$, then for WTG the first estimate would require $L \ll c\mu^{-1}$. However the second estimate would require $L \ll cL/c_{\text{RD}}$, suggesting that if $c_{\text{RD}} \ll c$ then in the nonlinear aggregation phase the WTG description remains valid at all scales, i.e. aggregation simply proceeds until the length scale is the largest allowed by the geometry.

3 Numerical simulations in the WTG regime

3.1 Model details

The model equations defined above in equations Eq. (1)–(3) can be integrated numerically, and in this section we will use this to confirm the previous results and further investigate model behaviour. Numerical details are given in Appendix A. Recall that the thickness variable h and the moisture variable q represent the departure from the spatially homogeneous RCE state. In all simulations reported below the initial condition is taken to be $\boldsymbol{u} = \mathbf{0}$, $h = 0$ and q small with $|q| \ll |q_\pm|$. The values of q are chosen randomly at each grid point.

The theoretical discussion in the previous section did not depend on the precise form of the functions F_h and F_q appearing respectively in Eq. (2) and Eq. (3), requiring only that they together lead to a bistable moisture equation, Eq. (7). However, in order to investigate the behaviour numerically we will need to define these forms explicitly. For illustrative processes we will use a piecewise linear construction, with

$$F_q(q) = -\mu_1 q \tag{12}$$

$$F_h(q) = \begin{cases} -\mu_2 q_p - \mu_1(q - q_p), & q > q_p \\ -\mu_2 q, & q_m < q < q_p \\ -\mu_2 q_m - \mu_1(q - q_m), & q < q_m. \end{cases} \tag{13}$$

With $\mu_2 > \mu_1 > 0$ this represents larger effective latent heat release of precipitation near to the RCE state. Note that the key quantity $\mu = (d/dq)G_{hq}(q;0)|_{q=0}$ is equal to $-\mu_1 + \mu_2 Q/H$ which we write as $-\mu_1(1 - \mu_2 Q/H\mu_1) = -\mu_1 M$ where M is a normalised gross moist stability of the RCE state. Throughout the paper we choose μ_1 and μ_2 such that $M < 0$, implying that the RCE state is unstable. This simple formulation of F_h and F_q has the advantage that we can easily tune the locations of the fixed points $q_\pm$ using the parameters q_p and q_m. Note that the $q_+ > q_p$ and $q_- < q_m$. Q and H are chosen such that $1 - Q/H > 0$, implying that the fixed points q_p and q_m are stable according to the analysis in Sect. 2.2. We will choose $|q_p| > |q_m|$, corresponding to a more extreme values of moisture in moist regions than in dry regions. Together with the constraint of zero net heating in steady state this implies small moist regions with strong upwelling and large dry regions weak downwelling, as typically observed in convective aggregation (e.g Muller and Bony, 2015).

Now that the equations are fully defined, the most basic starting point is to consider the system with no rotation or damping, $f = \lambda = \alpha = 0$. The computational domain is taken to be square with sides of length $10^7\mathrm{m}$. We will initially take the other parameters in the system as $g = 10\mathrm{ms}^{-2}$, $H = 30\mathrm{m}$, $\mu_1^{-1} = 36000\mathrm{s}$, $\mu_2 = 3\mu_1$, $\kappa = 10^5\mathrm{m}^2\mathrm{s}^{-1}$, $Q = 15\mathrm{m}$ and $q_p/Q = 0.1$, $q_m/Q = -0.025$. The precise values of these parameters are unimportant as we are aiming here to understand the general behaviour of the system, and indeed the parameters will be varied throughout the paper, however the values are chosen to give similar moisture time scales to those deduced from the system studied by Sugiyama (2009b). We use the Craig and Mack (2013) value of $\kappa = 10^5\mathrm{m}^2\mathrm{s}^{-1}$ rather than the $\kappa = 1.5 \times 10^5\mathrm{m}^2\mathrm{s}^{-1}$ used by Sugiyama (2009b). The parameter values used in

	Figures	f-plane regime(s)	β-plane regime	steady state	f [s^{-1}]	α [s^{-1}]	λ [s^{-1}]	κ [m^{2}s^{-1}]	ϵ	Q [m]
1	2,3,4a	I	-	Y	0	0	0	10^5	0	15
2	4b	I	-	Y	0	0	0	10^5	0.5	15
3	4c	I	-	Y	0	0	0	10^5	1	15
4	6A	I	-	Y	10^{-5}	4×10^{-6}	4×10^{-6}	10^5	0	15
5	6B	IIa	-	N	10^{-5}	10^{-7}	10^{-6}	10^5	0	15
6	6C	IIb	-	N	10^{-5}	10^{-6}	10^{-4}	10^5	0	15
7	9a	I	-	Y	10^{-6}	10^{-6}	10^{-6}	10^5	1	15
8	9b	I	-	Y	10^{-5}	10^{-5}	10^{-5}	10^5	0	15
9	9c	I	-	Y	0	10^{-5}	10^{-5}	10^5	0	15
10	11a	I,III	B	N	$2\times10^{-11}y$	10^{-5}	10^{-5}	10^5	0	15
11	11b	I,III	B	N	$2\times10^{-11}y$	10^{-5}	3×10^{-5}	10^5	0	15
12	11c	I,III	A	N	$2\times10^{-11}y$	3×10^{-5}	3×10^{-5}	10^5	0	15
13	11d	I,II,III	C	N	$2\times10^{-11}y$	10^{-6}	10^{-5}	10^5	0	15
14	11e	I,II,III	C	N	$2\times10^{-11}y$	10^{-6}	3×10^{-6}	10^5	0	15
15	11f, 13a,b	I,III	C	Y	$2\times10^{-11}y$	10^{-5}	10^{-5}	4×10^5	0	15
16	12, 13c,d	I,III	D	Y	$2\times10^{-11}y$	10^{-6}	10^{-6}	1.5×10^5	0	10.5
17	14a, 15, 16	I,III	C	N	$2\times10^{-11}y$	10^{-5}	10^{-5}	4×10^5	1	15
18	14b	I,III	B	N	$2\times10^{-11}y$	10^{-5}	10^{-5}	10^5	1	15
19	14c	I,II,III	B	N	$2\times10^{-11}y$	3×10^{-6}	3×10^{-5}	10^5	1	15
20	14d	I,III	D	N	$2\times10^{-11}y$	10^{-6}	10^{-6}	1.5×10^5	1	10.5

Table 1. Parameters for all examples of two-dimensional simulations shown in figures. Parameters which remain constant are $\mu_1 = 1/36000$ s^{-1}, $\mu_2 = 1/12000$ s^{-1}, $c^2 = 300$ m^{2}s^{-2}, $q_p = 0.1Q$ and $q_m = -0.025Q$. For 1–9 the f-plane regime is specified on the basis of Fig. 5. For 10–20 (equatorial β-plane simulations) the set of f-plane regimes encountered as f increases from zero is given. For 10–20 the β-plane regime is specified on the basis of Fig. 10. The 'steady state' column denotes whether or not the parameter values allow convergence to a steady state or, for the β-plane, steadily propagating state, at large times.

all of the two-dimensional simulations discussed in the paper are given in Table 1. See Sect. 4.3 for brief discussion of the non-zero values of α, λ and f used later in the paper.

3.2 Simulation results

We first consider the case $\epsilon = 0$, without nonlinear advection of moisture. The evolution of the moisture distribution with time is shown in Fig. 2. The qualitative behaviour is similar to CMWC. There is an initial adjustment phase on the time scale μ^{-1} as the small-scale noise grows. During this initial phase WTG applies only up to $L \sim \mu^{-1}c$, about 10^6m for the parameters

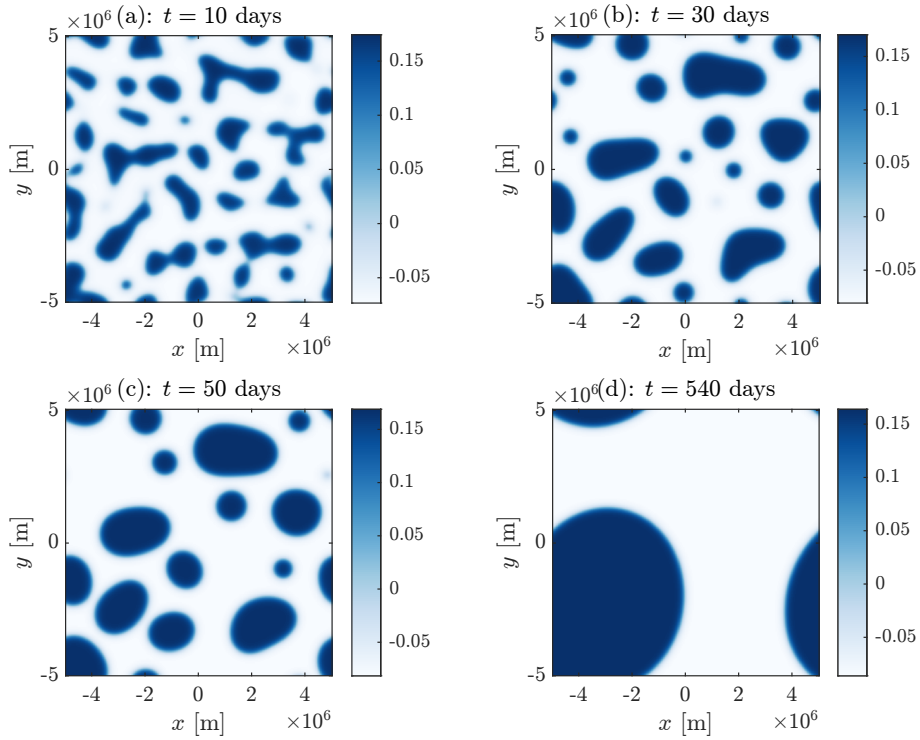

Figure 2. A series of snapshots of the perturbation moisture distribution q/Q from a numerical integration with no rotation or damping. Note that the final panel in this series is very close to, but has not reached, steady state, which would be a perfectly round moist region.

chosen. Hence distinct regions of enhanced and suppressed moisture on this scale, or less, form, with values corresponding to the effective stable fixed points $q_\pm$ such that $G_{hq} = 0$, $\partial G_{hq}/\partial q < 0$.

Once this has occurred, the coarsening process proceeds, with the scale of moist and dry regions slowly evolving, consistent with understanding of the reaction-diffusion system as discussed in the previous section. In particular the evolution of the boundaries occurs on a slow time scale determined by the reaction-diffusion velocity c_{RD} and the smaller curvature-associated velocity κ/R, both of which are smaller than the gravity wave speed c. Hence the WTG approximation continues to hold as the scale of moist and dry regions increases. In this regime the proportion of the domain filled by each of the moist and dry state is changing, so there is a slow evolution of the mean heating $\overline{F_h(q)}$. This causes a slow change in the locations of the stable moisture fixed points $q_\pm$, however this is on a longer timescale than those determined by G_{hq} so the q distribution quickly adjusts to the new stable values. These features of the long term evolution of the moisture distribution, the rapid adjustment to values of q close to $q_\pm$ and the subsequent slow evolution of the values of $q_\pm$, is shown in Fig. 3. This diffusive growth proceeds to the domain scale, when the areas of the regions are such that there is net zero heating and precipitation, and, consistent with theory (Rubinstein et al., 1989) the length of the boundary is minimised (forming either a circular or a band shaped structure, depending on geometry of the computational domain). At this point a steady state has been reached.

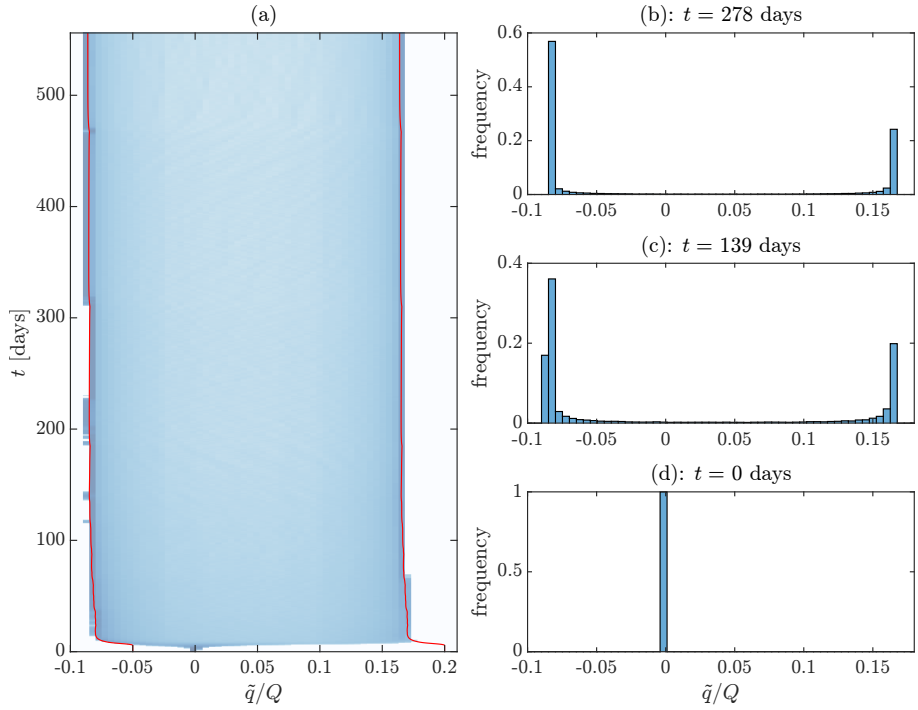

Figure 3. Panel (a): A histogram of the perturbation moisture distribution q/Q, plotted against time. The shading corresponds to the frequency distribution of the moisture. In panels (b)–(d) the histogram bars have been shown at selected times.

We will now consider briefly the system with advective nonlinearity in the moisture equation, $\epsilon > 0$. As discussed in noted in Sect. 2.2 above, we expect this to lead to reduction in the spatial scale of moist regions, increase in the spatial scale of dry regions and to a distinct advective mechanism for aggregation. The simulations show that the effect of the latter is that the system evolves towards a steady state, as was the case for $\epsilon = 0$. However advection changes the geometry of this final steady state, as may be seen from the steady state distributions for different values of ϵ shown in Fig. 4. Unlike with $\epsilon = 0$, where the steady state is governed only by reaction-diffusion, the final state is now governed by a balance between reaction-diffusion and advection. On the square periodic domain considered in paper this leads to the moisture forming either a band or a cross shape. At smaller values of ϵ ($\epsilon = 0.5$ is shown) a band forms and at larger ϵ ($\epsilon = 1$ is shown) a cross is preferred. However at intermediate values either shape, according to details of the initial conditions, may be reached and will persist. The changes in geometry associated with increasingly strong advection shown here may be relevant to realistic atmospheric flows even if the precise steady-state configurations are not.

4 Breakdown of WTG and implications for aggregation

When frictional and thermal damping and rotation are included in the system, then, as noted previously, there will be an upper limit L_{dyn}, depending on the dynamical parameters α, λ and f, on the scale to which WTG balance can apply. It is expected that

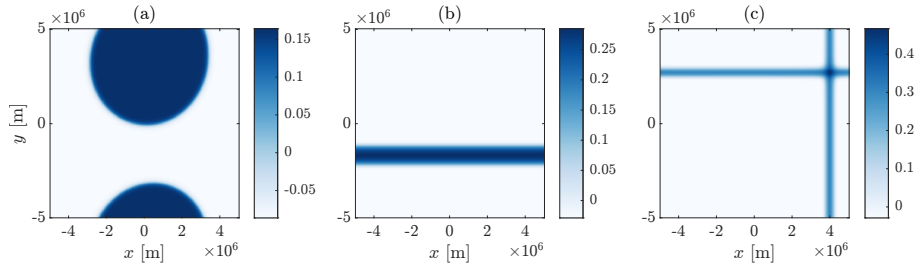

Figure 4. The steady state moisture distribution, showing plotting q/Q, in two dimensional simulations with no rotation or damping and (a) $\epsilon = 0$, (b) $\epsilon = 0.5$, and (c) $\epsilon = 1$

the coarsening to the domain scale exhibited in the previous section will be substantially modified, and perhaps halted, when the scale L_{dyn} is reached. Since the growth of disturbances to the RCE state is the pre-cursor to the coarsening phase we will begin this section by considering in Sect. 4.1 explicit solution of the linear instability problem when α, λ and f are non-zero. Then in Sect. 4.2 we will present semi-quantitative scaling arguments that are potentially relevant to the evolution beyond the linear instability phase. We then present results from numerical simulations in Sect. 4.3, both for $\epsilon = 0$ and for $\epsilon > 0$, including a regime diagram that summarises the overall pattern of behaviour as the dynamical parameters vary.

4.1 Key insights from the linear instability problem

Since the growth of disturbances to the RCE state is the pre-cursor to the coarsening phase we will begin this section by considering in Sect. 4.1 explicit solution of the linear instability problem when α, λ and f are non-zero.

There have been previous theoretical studies (e.g. Adames et al., 2019) of linear wave propagation and linear instability in systems equivalent to Eq. (1)–(3), but it is useful to establish some of the basic properties of the particular model system that we consider in this paper. For the case discussed in Sect. 2.2 and illustrated in Sect. 3, where $\alpha = \lambda = f = 0$ and the WTG approximation is valid, the linear stability properties of the RCE state are very straightforward and determined by the the sign of the derivative with respect to q of $G_{hq}(q, 0)$. We consider the linear stability problem in more detail for α, λ and f non-zero since this gives insight into the behaviour of the full nonlinear system as revealed by numerical simulation. Since the f-plane is isotropic, we can assume that perturbations vary only in the x-direction. Assuming small amplitude perturbations of the form $u = \Re\{\hat{u}\exp(\sigma t + ikx)\}$, with analogous notation for other variables, the linearised forms of Eq. (1), Eq. (2) and Eq. (3) are

$$\sigma\hat{u} - f\hat{v} = -ikg\hat{h} - \alpha\hat{u}, \tag{14}$$

$$\sigma\hat{v} + f\hat{u} = -\alpha\hat{v}, \tag{15}$$

$$\sigma\hat{h} - iHk\hat{u} = -\mu_2\hat{q} - \lambda\hat{h}, \tag{16}$$

$$\sigma\hat{q} - iQk\hat{u} = -\mu_1\hat{q} - \kappa k^2\hat{q}, \tag{17}$$

where $\mu_2 = -F_h'(0)$ and $\mu_1 = -F_q'(0)$, matching the notation used in Eq. (13) and Eq. (12). As is standard, these define an eigenvalue problem, the solution of which leads to a dispersion relation in the form of a quartic equation for σ, given explicitly in Appendix B as Eq. (B1).

An important simple case is the strict WTG limit with $\alpha = \lambda = f = 0$. This may be considered directly by neglecting the $\sigma\hat{h}$ and $-\lambda\hat{h}$ terms in Eq. (16) and then substituting for $\hat{u}$ in Eq. (17) to deduce, neglecting the κk^2 term,

$$\sigma = -\mu_1 + \frac{Q\mu_2}{H}. \tag{18}$$

This expression for σ motivates the previously noted definition of the normalised gross moist stability for the moist shallow water equations,

$$M = 1 - \frac{Q\mu_2}{H\mu_1}. \tag{19}$$

As noted previously, in this paper parameter values will be chosen such that $M < 0$, implying $\sigma > 0$ and moisture mode instability with inverse time scale $\mu = \mu_1|M|$. Note that the WTG approximation applies at small scales, $k \gg \mu/c$. At large scales, with $k \ll \mu/c$, the moisture adjusts on a timescale shorter than that of the dynamics and a steady state balance in the moisture equation, Eq. (17), neglecting the term $\sigma\hat{q}$ is appropriate. (The large-k and small-k limits in this problem correspond to the moisture-mode and gravity-mode limits identified by Adames et al. (2019)). Using the moisture equation to eliminate the moisture dependence in the height equation, Eq. (16), then gives the standard shallow water equations with the gravity wave speed adjusted from c^2 to

$$c^2(1 - \frac{\mu_2 Q}{H\mu_1}) = c^2 M. \tag{20}$$

This defines the moist gravity wave speed $c_m^2 = c^2 M$ and implies unstable growth rather than propagation if $M < 0$. $M < 0$ may be therefore identified as a criterion for instability whether or not the WTG approximation is valid. Re-including the effects of moisture diffusion will potentially inhibit instability on length scales comparable to or smaller than $\sqrt{\kappa/\mu_1}$.

Analysis of the full dispersion relation ,Eq. (B1), given in Appendix B shows that, if $M < 0$, there is a real positive root for σ only if

$$\kappa \leq (\sqrt{1-M} - 1)^2 \mu_1 \frac{c^2\alpha}{\lambda(\alpha^2 + f^2)} \tag{21}$$

(Note that the final factor represents the square of a length scale which we will identify in the following subsection as the dynamical length scale L_{dyn}.) However this does not give complete information on when instability is possible because there may be non-zero complex conjugate roots for σ with positive real part. Some analytical progress can be made in describing the dependence of roots on the different parameters, but the algebra is complicated. Further details are given in Appendix B.

The behaviour found is illustrated in Fig. 5 which maps out different regions of the (α, λ) plane for a specific choice of κ and for six different choices of f, including $f = 0$. Other parameters, μ_1, μ_2, c^2, Q and H take the same values as specified in Sect. 3. The (α, λ)-plane may be divided into four regions, each corresponding to a different regime of behaviour. Regime I is where $\Re(\sigma) > 0$ occurs for some k only when σ is real. Regimes IIa and IIb are where there are complex σ with $\Re(\sigma) > 0$ and

with non-zero imaginary parts and Regime III is where there is no instability, i.e. $\Re(\sigma) < 0$ for all k. The distinction between IIa and IIb is that in IIb σ corresponding to fastest growth over all k has non-zero imaginary part.

For $f = 0$ only Regimes I and III are present and the region of instability simply corresponds to Eq. (21). For non-zero values of f that are not too large the boundary between Regimes I and III is again described by Eq. (21), but, Regimes IIa and IIb exist in some part of the region $\alpha < f$. Note that Regimes IIa and IIb are therefore confined to smaller and smaller values of α as $f \to 0$. For the largest value of f shown, again only Regimes I and III are present, with Regime I corresponding to Eq. (21). It may be shown in general that if the inequality

$$f > c(\mu_1/\kappa)^{1/2}(\sqrt{1-M} - 1). \tag{22}$$

is satisfied there are no Regimes IIa and IIb and the transition between stability and instability is described completely by Eq. (21). For the parameters used to generate Fig. 5, the disappearance of regions IIa and IIb between $f = 6 \times 10^{-5}\mathrm{s}^{-1}$ and $f = 10^{-4}\mathrm{s}^{-1}$ is consistent with Eq. (22).

Whilst the parameter dependence that is found in the linear stability problem is complicated, two overall rules that seem to hold are first that (not surprisingly) increasing κ inhibits instability and second that, for given κ, increasing f also tends to inhibit instability, i.e. the region of the (α, λ) plane in which there is instability reduces as these parameters increase. In particular, for any specified non-zero values of α, λ and κ there is an f_{stab} such that $f > f_{\text{stab}}$ implies stability. The solid curves shown in Fig. 5 bounding Regime III are therefore, for the chosen value of κ, contours of the function $f_{\text{stab}}(\alpha, \lambda)$ in the (α, λ) plane. The existence of f_{stab} will be exploited in the description of the equatorial β-plane behaviour in the following section. (Note in particular that f_{stab} depends on κ as well as on α and λ, but we leave the dependence on κ unstated since in the simulations to be discussed κ is in practice kept fixed and only α and λ varied.)

4.2 Dynamical arguments

The description above is focused on the linear instability problem and cannot be assumed to extend to the evolution once the growing unstable disturbances have saturated, e.g. in an aggregation phase. More general insight in the evolution can be obtained by considering possible balances in the equations at horizontal scale L. Assume that the time scale of evolution is T_q. In the linear instability phase $T_q \sim \mu^{-1}$. However after the unstable growth has saturated T_q may be larger than this. In the aggregation of moist and dry regions described previously, for example, T_q is determined by the diffusivity κ and is large if κ is small.

If $\alpha \neq 0$ and $\lambda \neq 0$ then, assuming that $T_q \gg \alpha^{-1}, \lambda^{-1}$, a quasi-steady state balance is possible in the dynamical equations, i.e. $\boldsymbol{u}$ is instantaneously determined by the q field according to the quasi-steady balance

$$\alpha\delta - f\zeta = -g\nabla^2 h \tag{23}$$

$$\alpha\zeta + f\delta = 0 \tag{24}$$

$$H\delta + \lambda h = F_h(q), \tag{25}$$

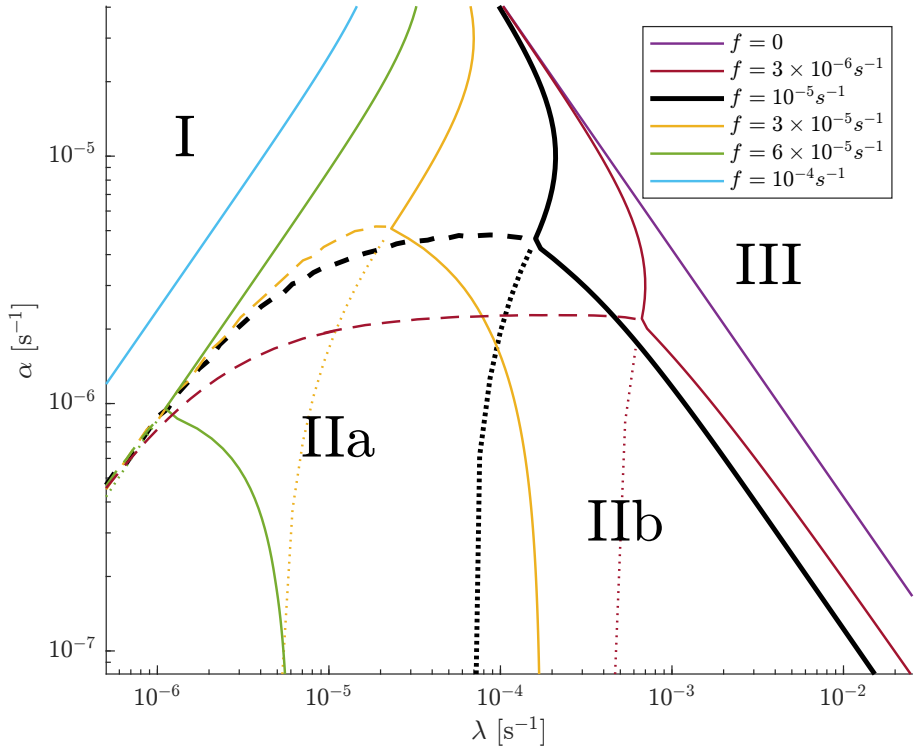

Figure 5. A regime diagram of the linear instability behaviour of the model on the f-plane. The black curves mark the boundaries between regimes for a value of $f = 10^{-5}\mathrm{s}^{-1}$, and the black regime labels correspond to this curve. The solid black curve marks the boundary between the unstable regimes and the globally stable regime (Regime III). The dashed black curve denotes the boundary separating Regime I on the left, where all unstable modes have zero frequency and Regime IIa on the right, where some unstable modes have a non-zero frequency, and so are not stationary. The dotted black curve separates Regime IIa, to the left, from Regime IIb, to the right, where the fastest growing linear mode is no longer stationary. The curves of four other colours show corresponding boundaries for different values of f, with no equivalent of Regime II appearing when $f = 0$ or when f is sufficiently large. A detailed description of the regime structure is given within the text.

where δ is divergence and ζ is vorticity. In this respect we can identify this case with Regime I in the previous section. Eliminating ζ gives that

$$\delta = -g\alpha(\alpha^2 + f^2)^{-1}\nabla^2 h. \tag{26}$$

Substituting into Eq. (25) implies that the local h, and hence the local $\boldsymbol{u}$, is determined by q in a surrounding region of scale

$$L_{\mathrm{dyn}} = c(\alpha/\lambda)^{1/2}(\alpha^2 + f^2)^{-1/2}. \tag{27}$$

This defines a dynamical length scale L_{dyn}. WTG balance, the local balance between divergence and heating applies only length scales smaller than L_{dyn}. On length scales larger than L_{dyn} the dominant balance in Eq. (25) is between λh and $F_h(q)$ and hence,

from Eq. (26) the divergence is proportional to $\nabla^2 F(q)$ rather than to $F(q)$. Another implication of the above balance, from Eq. (24), is that the flow will be dominated by the rotational component if $f \gg \alpha$ and by the irrotational component if $f \ll \alpha$.

The quasi-steady balance above cannot hold when $\alpha = 0$, when Eq. (26) would imply $\delta = 0$. However if q is to be maintained away from the $q = 0$ steady state, which is known to be unstable, Eq. (4), neglecting the term multiplied by ϵ, requires δ non-zero. This suggests that a distinct dynamical argument is required when α is small, analogous to the distinct nature of Regime II discussed in the previous section. If $f \neq 0$, a possible balance assuming small Rossby number, i.e. that the time scale of evolution of the q field is much larger than f^{-1}, is

$$-f\zeta = -g\nabla^2 h \tag{28}$$

$$\zeta_t + f\delta = 0 \tag{29}$$

$$h_t + H\delta + \lambda h = F_h(q), \tag{30}$$

implying

$$(h - (c^2/f^2)\nabla^2 h)_t + \lambda h = F_h(q) \tag{31}$$

i.e. a form of the quasi-geostrophic potential vorticity equation with the potential vorticity changing through the effects of heating $F_h(q)$ and thermal damping. So in this system there are two prognostic equations, one for potential vorticity and one for moisture. In this case it is the second term in the time derivative that corresponds to divergence, so that WTG applies if $L \ll (c/f)\min(1, (\lambda T_q)^{-1/2})$. Here the length scale appearing is the Rossby radius c/f and the flow may be expected to evolve on the time scale λ^{-1} at large times. Note that in this case the rotational part of the velocity field will be stronger than the divergent part.

In both the above cases it appears that the relation between moist heating and divergence becomes non-local at sufficiently large scale, L_{dyn} when α is large enough to bring the dynamical balance to a quasi-steady state and c/f when α is smaller. Therefore the aggregation behaviour seen previously is likely to be halted, or at least strongly modified, when these scales are reached. In Sect. 4.3 below the nature of this modification is examined by numerical simulation.

The scale L_{dyn} defined above decreases as f increases, suggesting that the scale of aggregated moist and dry regions will also decrease as f increases. Furthermore the underlying instability of the RCE state requires WTG dynamics to apply and L_{dyn} therefore also represents an upper limit on the scale of the instability. As f increases L_{dyn} will reduce to the diffusion scale $\sqrt{\kappa/\mu}$ and the instability of the RCE state will disappear. This provides an estimate for f_{stab},

$$\sqrt{\frac{\alpha}{\lambda}}\frac{c}{\sqrt{f_{\text{stab}}^2 + \alpha^2}} = \Gamma\sqrt{\frac{\kappa}{\mu}}, \tag{32}$$

where Γ is a non-dimensional parameter. The expression, Eq. (21), provided by the linear stability calculation is consistent with this reasoning and provides an explicit expression for Γ as equal to $\sqrt{-M}/(\sqrt{1-M} - 1)$. For the case when α is small, corresponding to Regime II, we have no expression for L_{dyn} or for the boundary of stability between Regimes II and III in Fig. 5 so a corresponding analytic result has not been determined. We do expect a qualitatively similar situation where the

system being stabilised by diffusion once the maximum length scale L_{dyn} becomes sufficiently small. However the fact that the behaviour as f changes of the boundary between Regime IIb and Regime III shown in Fig. 5 is geometrically complicated suggests that a simply dynamical estimate of the form of the boundary will be difficult to find.

4.3 Numerical simulations

The link between the maximum length scale L_{dyn}, determined by non-zero values of f, α and λ, over which WTG is expected to apply and the spatial scale of aggregation is now illustrated using numerical simulation. The same numerical scheme as in Sect. 3 is used. The parameters, μ_1, μ_2, c^2, Q and H take the same values as specified in Sect. 3, unless otherwise stated. f, α and λ are chosen so that there is linear instability, corresponding to Regimes I, IIa and IIb in Fig. 5. Recall that details of the parameter values chosen for these simulations are given in Table 1. The values of f are chosen for illustrative purposes. ($f = 10^{-5}\text{s}^{-1}$ corresponds to about 5°.) Representative values of α and λ are chosen to be similar to those in other papers on large-scale tropical dynamics, e.g. Sugiyama (2009b) takes $\alpha = 2 \times 10^{-6}\text{s}^{-1}$ and $\lambda = 5 \times 10^{-7}\text{s}^{-1}$, Adames and Wallace (2014) take $\alpha = 10^{-6}\text{s}^{-1}$ and $\lambda = 10^{-6}\text{s}^{-1}$, Adames and Kim (2016) take $\alpha = 3 \times 10^{-6}\text{s}^{-1}$ and $\lambda = 3 \times 10^{-6}\text{s}^{-1}$, but we deliberately vary away from these representative values in some simulations to investigate the effect of varying α and λ in different ways. Note that the justification of values for linear friction coefficients, or indeed the inclusion of linear friction at all, in models for tropical circulation remains a topic of active discussion (e.g. Romps, 2014).

4.3.1 $\epsilon = 0$ (advective nonlinearity excluded)

We begin with $\epsilon = 0$, i.e. excluding nonlinear advection of moisture. A selection of time series of the moisture distribution of the system for different choices of f, α and λ is shown in Fig. 6, with each simulation corresponding to a row. In all cases the moisture field q was initialised with small-scale random noise. The first case, in row A, corresponding to Regime I in the (α, λ) plane shown in Fig. 5 the system initially evolves similarly to the case without damping or rotation, with the formation of distinct moist regions, which evolve and enlarge through aggregation. However the aggregation does not proceed to the domain scale but halts at a smaller scale, with quasi-steady circular moist regions. This is as expected from the previous dynamical discussion. The inclusion of non-zero f, α and λ implies that WTG balance can hold only up the scale L_{dyn} and aggregation halts at this scale. Other simulations with f, α and λ corresponding to Regime I in Fig. 5 show similar evolution.

Rows B and C of Fig. 6 correspond to Regime II in the (α, λ) plane, B to IIa and C to IIb. For case B there is again an initial segregation and then aggregation process leading to distinct moist and dry regions at some finite scale. However, the long-time distribution is no longer stationary, but continues to evolve in time (without there being any further systematic increase in scale). In case C, whilst there is segregation, there is no clear aggregation stage and the moist and dry regions evolve in time in a manner that is more wave-like that that seen in case B.

To establish that the division of the (α, λ) plane, originally motivated by the linear instability properties, provides a useful guide to the behaviour of the ultimate nonlinear evolution, Fig. 7 repeats the depiction of the (λ, α) plane shown previously, for a single value of f, with superimposed symbols indicating whether the nonlinear evolution was aggregated and quasi-steady, as case A above, aggregated and unsteady, as in case B, or propagating, as in case C. Regime IIa, (case B) has been split into two

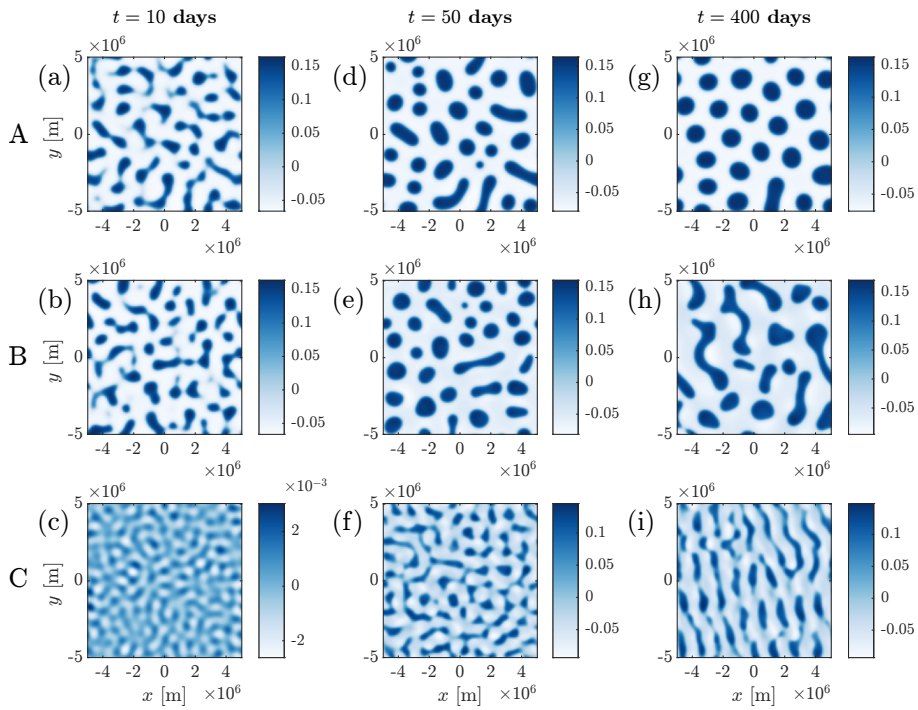

Figure 6. A series of snapshots of the perturbation moisture distribution q/Q of a two dimensional simulation, this time with rotation and damping. Row A corresponds to case I in Sect. 4.1. This has $f = 10^{-5}\mathrm{s}^{-1}$, $\alpha = \lambda = 4 \times 10^{-6}\mathrm{s}^{-1}$, giving $L_{\mathrm{dyn}} = 1.6 \times 10^{6}\mathrm{m}$. Rows B and C correspond to case II. Row B has $\alpha = 10^{-7}\mathrm{s}^{-1}$ and $\lambda = 10^{-6}\mathrm{s}^{-1}$, and C has $\alpha = 10^{-6}\mathrm{s}^{-1}$ and $\lambda = 10^{-4}\mathrm{s}^{-1}$.

sub-regimes, with a slow regime corresponding to transitional behaviour in which aggregated regions form but propagation is sufficiently slow that these remain round.

A possible interpretation of the apparent relation between the properties of the linear instability problem and the evolution observed in the numerical simulations is as follows. In Regime I, as illustrated by simulation A, the linear instability behaviour is essentially that described by the WTG approximation, with the relevant unstable mode having real σ. Therefore the system evolves through the the instability to the segregated state determined by the bistability. In Regime IIa, as illustrated by simulation B, the relevant unstable mode is similar to that in A, with σ real and the process of segregation is correspondingly similar. However the existence of slower growing propagating (complex-σ) unstable modes at larger scales is relevant to the nonlinear evolution post-segregation (even if the linear instability modes themselves do not provide a complete description of the behaviour). (A reduced mathematical model describing the evolution of the segregated state might make this relevance clearer). In Regime IIb, illustrated by simulation C, there is a pair of fastest growing modes with complex conjugate σ rather than a single mode with real σ and the mechanism for growth is therefore completely different to that described by the WTG approximation. In fact these modes are better considered as moisture-destabilized inertial waves, and are not moisture modes since they rely on the fact that the dynamics is not slaved to the moisture field. Consequentially the nonlinear evolution is not

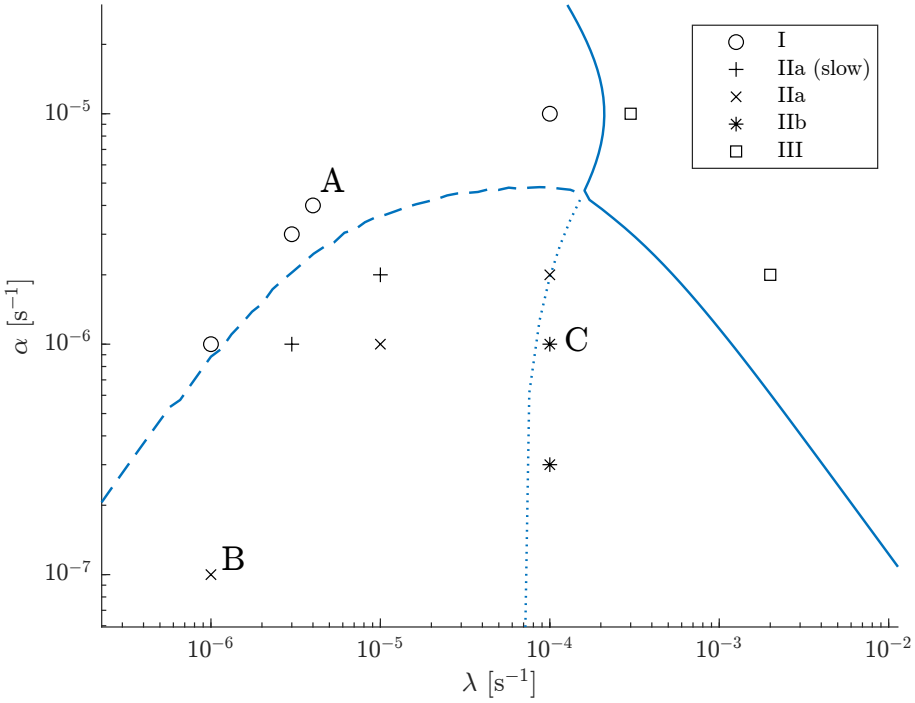

Figure 7. The regime diagram curve for $f = 10^{-5}\mathrm{s}^{-1}$ from Fig. 5, overlaid with observational data. Each point marked on the Figure correspond to the parameter values for a simulation which was then categorised into one of four regimes. The points corresponding to the moisture distributions shown in Fig. 6 are labelled A, B and C.

so clearly a segregation into the two states allowed by bistability and instead is better characterised as an evolving field of nonlinear moisture-inertial waves. Note that systematic propagation in this case C and the clear anisotropy of the instantaneous q distribution visible in Fig. 6(i) are an indication of spontaneous symmetry breaking rather than of any systematic anisotropy of the system as specified.

We now focus on the behaviour of the model with parameters chosen from Regime I, where there is aggregation to a quasi-steady state. The behaviour can be usefully summarised by using the spatial auto-correlation. The auto-correlation scale L_{auto} is defined as the minimum radius at which the spatial autocorrelation is a factor of $1/e$ less than its maximum value. The time evolution of the L_{auto} for simulations with various damping and rotation rates, and diffusivities, is shown in Fig. 8.

The two cases (a) and (b) have $\alpha = \lambda = f = 0$ so, as previously demonstrated in Sect. 2.2–3, aggregation is expected eventually to proceeds to the domain scale. The evolution of L_{auto} for both cases (a) and (b) is consistent with this expectation. For case (b) L_{auto} reaches a limiting value within the time period shown in the Figure. For case (a) a limiting value is not reached, but L_{auto} continues to increase throughout the period shown. The difference between (a) and (b) can explained by the fact that κ for (b) is $4\times$ larger that for (a) and therefore, recalling the established theory on reaction-diffusion systems noted in Sect. 2.2, that the rate at which aggregation proceeds is more rapid as κ increases.

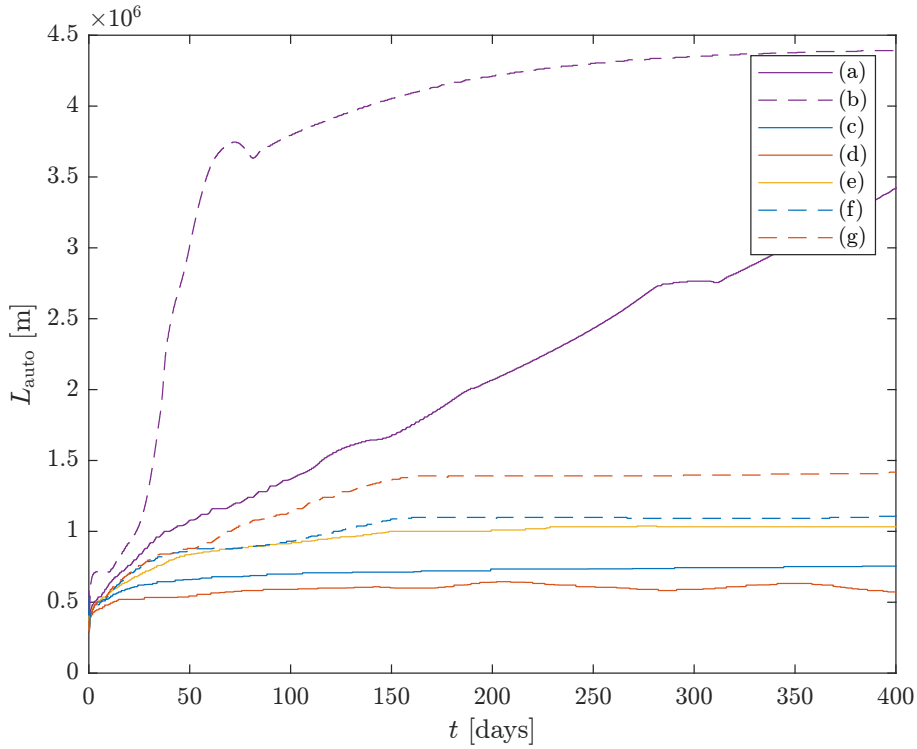

Figure 8. A measure of autocorrelation length scale plotted against time for various parameter values, within case I. Curves (a) and (b) have $\alpha = \lambda = f = 0$. Curve (a) has $\kappa = 10^5 \mathrm{m}^2\mathrm{s}^{-1}$ and (b) has $\kappa = 4 \times 10^5 \mathrm{m}^2\mathrm{s}^{-1}$. Curve (c) has $f = 10^{-5}\mathrm{s}^{-1}$ and $\alpha = \lambda = 4 \times 10^{-6}\mathrm{s}^{-1}$ and $\kappa = 10^5 \mathrm{m}^2\mathrm{s}^{-1}$. Curves (d) and (e) show the effect of reducing α and λ respectively by a factor of 4. Curves (f) and (g) have parameters as (c) and (d) respectively, but with $f = 0$. The length scale varies consistently with the value of L_{dyn}.

Other cases have non-zero values of α, λ and f. All these show approach to a finite limiting value indicating that aggregation ceases. A candidate value for the length scale at which this occurs is $L_{\mathrm{dyn}} = c(\alpha/\lambda)^{1/2}(\alpha^2+f^2)^{-1/2}$. The corresponding values deduced from Fig. 8 are consistent with this expression in the sense that the ordering as parameters are changed is consistent with the expression. Note in particular that for given α and λ the scale is smaller with $f > 0$ than it is with $f = 0$.

4.3.2 $\epsilon > 0$ (advective nonlinearity included)

The effects of advective nonlinearity in the presence of non-zero f, α and λ are illustrated in Fig. 9. Note that each of these cases have (λ, α) corresponding to Regime I. We have previously noted that the effect of advective nonlinearity is in the two-dimensional case to give moist regions that a more filamentary than quasi-circular. (For example recall the steady-state moisture distributions, where aggregation has proceeded to the domain scale, for $\alpha = \lambda = f$ shown in Fig. 4.) The effect of α, λ and f non-zero is both to limit any aggregation to a finite scale and to determine the flow pattern resulting from the moisture distribution. When $f \neq 0$ this flow has a substantial rotational component and the advective effect of this on the filamentary

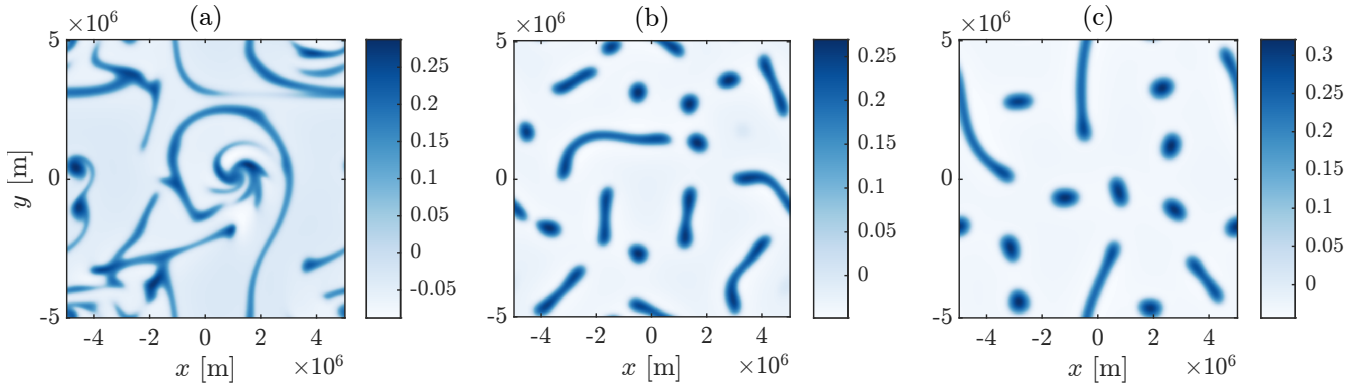

Figure 9. The perturbation moisture distribution, q/Q, after 400 days for a selection of two dimensional nonlinear simulations. (a) has $f = \alpha = \lambda = 10^{-6}\mathrm{s}^{-1}$, (b) has $f = \alpha = \lambda = 10^{-5}\mathrm{s}^{-1}$, and (c) has $\alpha = \lambda = 10^{-5}\mathrm{s}^{-1}$ but $f = 0$.

moisture structures is apparent in Fig. 9(a) and (b). The example with $f = 0$, shown in Fig. 9(c), where the advecting flow is irrotational, is distinctly different, with any curvature of the filamentary structures weaker and resulting from deformation by a spatially structured irrotational flow. As has been noted previously advective narrowing of moist regions means that the maximum magnitude q is affected by diffusion and not simply equal to the predicted value q_+.

5 Equatorial β-plane

In this section we will consider the model on the equatorial β-plane, i.e. with $f = \beta y$. It has been shown that on the f-plane aggregation tends to be inhibited by rotation, in two ways: (i) The upper limit of the scale for the underlying instability of the system is a decreasing function of f and the lower limit is a increasing function of κ, therefore when κ is non zero the instability disappears altogether for $f > f_{\text{stab}} = \beta y_{\text{stab}}$, with the latter equality defining y_{stab}. (ii) The upper limit on the aggregation scale is a decreasing function of f. This suggests the possibility on the β-plane of disturbances largely confined to some equatorial band with $|y| < y_{\text{stab}}$. Such disturbances do indeed form, and we will go on to describe their behaviour.

The regime diagram shown in Fig. 5 in Sect. 4.1 provides some insight into how the dynamics might vary with latitude. At the equator, with $f = 0$, either Regime I or Regime III must apply, with Regime III implying that the RCE state is stable. For large f Regime III applies. Whether or not the transition from Regime I to Regime III passes through Regime II will be determined by the values of α and λ. Generally speaking this will occur when α/λ is relatively small. If Regime II is visited then this is likely to be manifested as more complicated behaviour (recall Fig. 6) as y_{stab} is approached. However we will see later in this section that there are effects on the β-plane that are not captured by the f-plane behaviour as described in Sect. 4.

It was noted in the previous section that, on the f-plane, whilst there was sometimes evidence of a selection of a preferred direction (recall Fig. 6C) this selection is purely random. On the β-plane, however, there is a genuine east-west asymmetry (e.g. as manifested in the well-known Matsuno-Gill steady response to localised heating, which has been generalised to the

case where α and λ are not equal by Wu et al. (2001)) and it will be of particular interest to determine whether this leads to zonal propagation of moist and dry regions and how such propagation varies with model parameters.

We will begin this section discussing an adjustment to the previous, constant-f, dynamical arguments and its impact on the local distribution of aggregated regions. We will then discuss the effects of the larger scale equatorial circulation. We will then go on to discuss the behaviour of a series of numerical experiments, in both the $\epsilon = 0$ and $\epsilon > 0$ cases.

5.1 Implications of equatorial β-plane dynamics

Much of the scale analysis of the f-plane equations presented in Sect. 4.2 was based on a quasi-steady balance in the dynamical equations which led to the relation, Eq. (26), between δ and h and hence an estimate $L_{\text{dyn}} = c(\alpha/\lambda)^{1/2}(\alpha^2 + f^2)^{-1/2}$ for the scale on which WTG breaks down and hence as an effective upper limit on scale of aggregation. The same approach, of assuming a quasi-steady balance in the dynamical equations, will now be applied to the β-plane. The scale L_{dyn} as defined previously will still be useful, but will now vary with latitude. It is convenient to use the notation $L_{\text{dyn},f}$ to represent the value of L_{dyn} for a particular value of f. At this stage it is also useful to note that the system being considered has no imposed inhomogeneity in x, i.e. in longitude, and there is therefore no systematic change in character of the disturbances in x, i.e. there is representation of a 'warm pool' range of longitudes in which moisture has a stronger role in the dynamics than elsewhere. Correspondingly, any reference below to integration over the domain means the entire domain, not restricted to a particular range of x-values.

The balance in Eq. (26) must be modified on the β-plane because $\partial f/\partial y$ is non-zero and becomes

$$\delta = -\frac{g\alpha}{\alpha^2 + f^2}\nabla^2 h + \frac{g\beta}{f^2 + \alpha^2}\hat{\boldsymbol{n}} \cdot \nabla h, \tag{33}$$

where $\beta = \partial f/\partial y$ and $\hat{\boldsymbol{n}} = (\alpha^2 - f^2, 2f\alpha)/(f^2 + \alpha^2)$ is a unit vector. Substituting Eq. (33) into the thickness equation, Eq. (2), gives the steady response to a heating as

$$\lambda h + H\delta = \lambda h - \frac{gH\alpha}{\alpha^2 + f^2}\nabla^2 h + \frac{gH\beta}{f^2 + \alpha^2}\hat{\boldsymbol{n}} \cdot \nabla h = F_h(q). \tag{34}$$

One important difference from the corresponding equations for the f-plane is there is now a preferred direction in the relation between δ and h, allowing a systematic anisotropy. A second difference is that the coefficients in the equations are now functions of y. A local analysis, treating f as constant, may therefore not always be valid. The expressions above suggest a change in character of the system when the length scale is larger than α/β. Below this scale the first, isotropic, term in Eq. (33) is dominant, and, furthermore, the coefficient appearing in this term does not vary significantly on this scale, so a local description will be valid. For length scales larger than α/β, however, the second term is dominant, suggesting significant anisotropy in the dynamics. The coefficients and the vector $\hat{\boldsymbol{n}}$ will also vary significantly with f so significant latitudinal variation is expected and, more fundamentally, a local description may not be valid.

Along with the above it must be taken into account that aggregation is initiated by instability of the RCE state and that this instability is confined to $|y| < y_{\text{stab}}$. Additionally, whilst it is expected that moist anomalies are large confined within this region, note that Eq. (34) implies that dynamical effects extend outside, on a length scale of $L_{\text{dyn},f=\beta y_{\text{stab}}}$. When M is large and

negative, Eq. (32) suggests this scale is $\sqrt{\kappa/\mu}$: a diffusive response extending outside of the unstable equatorial region. When M is closer to zero, however, this length scale may be significantly larger.

The f-plane analysis predicts L_{dyn} as an upper limit on the scale for aggregation. This suggests $L_{\text{dyn},f=0} = c/\sqrt{\lambda\alpha}$ as the corresponding scale at the equator on the β-plane. Therefore, on the basis of the arguments above, aggregation at the equator may be isotropic if $L_{\text{dyn},f=0} < \alpha/\beta$ and $y_{\text{stab}} > L_{\text{dyn},f=0}$. If the second condition is not satisfied then the geometry will not allow isotropy. Furthermore, since $L_{\text{dyn},\beta y}$ is a decreasing function of y, isotropy will extend to $y = y_{\text{stab}}$ but the characteristic length scales will decrease as $|y|$ increases. In other words we expect the aggregation to be qualitatively similar to the f-plane case, except that it will be confined to a region slightly larger than $|y| < y_{\text{stab}}$ and that the characteristic length scale with vary with y. If $L_{\text{dyn},f=0} < \alpha/\beta$ and $y_{\text{stab}} < L_{\text{dyn},f=0}$, on the other hand, then the aggregated structures in moisture will be largely confined to $|y| < y_{\text{stab}}$ and therefore extended in x relative to their scale in y. However the dynamical signatures will extend to $|y| = L_{\text{dyn},f=0}$. This suggests two distinct regimes of behaviour in a parameter space defined by $y_{\text{stab}}\beta/\alpha$ and $L_{\text{dyn},f=0}\beta/\alpha = c\beta\alpha^{-3/2}\lambda^{-1/2}$. We will label these as Regime A ($L_{\text{dyn},f=0} < \alpha/\beta$, $L_{\text{dyn},f=0} < y_{\text{stab}}$) and Regime E ($L_{\text{dyn},f=0} < \alpha/\beta$, $y_{\text{stab}} < L_{\text{dyn},f=0}$). These two Regimes are shown in the $(L_{\text{dyn},f=0}, y_{\text{stab}})$ plane in Fig. 10. We will add further regimes to this diagram and discuss them in more detail below.

We now consider the case where $L_{\text{dyn},f=0} > \alpha/\beta$. The non-isotropic terms in Eq. (33) and Eq. (34) must be taken into account. Furthermore close to the equator a local analysis is no longer adequate. A natural y-scale close to $y = 0$, obtained by requiring a balance between the first and second terms in the middle expression in Eq. (34) is $L_{\text{eq}} = (c/\beta)^{1/2}(\alpha/\lambda)^{1/4}$ as obtained by Wu et al. (2001) in their generalisation of the Matsuno-Gill problem. The corresponding x-scale, however, agrees with the length scale from the local analysis at the equator, $L_{\text{dyn},f=0}$. Note that $L_{\text{eq}} = \sqrt{L_{\text{dyn},f=0}\alpha/\beta}$ and that L_{eq} therefore always lies between α/β and L_{dyn}, i.e. when $L_{\text{dyn},f=0} > \alpha/\beta$, $\alpha/\beta < L_{\text{eq}} < L_{\text{dyn},f=0}$. The structure of Eq. (34) implies that a moisture anomaly localised within region $|y| < L_{\text{eq}}$ will force a dynamical anomalies that extend across the region $|y| \sim L_{\text{eq}}$. This implies a distinct Regime D ($\alpha/\beta < L_{\text{dyn},f=0}$, $y_{\text{stab}} < L_{\text{eq}}$) similar to Regime E, with latitudinal confined moisture anomalies driving broader dynamical anomalies.

Now consider $(L_{\text{dyn},f=0} > \alpha/\beta,\ y_{\text{stab}} > L_{\text{eq}})$. In this case, since $y_{\text{stab}} > L_{\text{eq}}$, the unstable region is large enough to allow multiple aggregated structures in the y-direction. Since we have $L_{\text{dyn},0} > \alpha/\beta$ there is anisotropy at the equator, but since L_{dyn} is a decreasing function of f the anisotropic region is expected to extend only to y such that $L_{\text{dyn},\beta y} = \alpha/\beta$, hence

$$y = y_{\text{iso}} = L_{\text{dyn},f=0}\left(1 - \frac{\alpha^2}{\beta^2 L_{\text{dyn},f=0}^2}\right)^{\frac{1}{2}}, \tag{35}$$

which defines y_{iso}. For $|y| > y_{\text{iso}}$ we expect isotropic aggregation. This implies two further regimes, Regime B ($L_{\text{dyn},f=0} > \alpha/\beta$, $y_{\text{stab}} > \max\{L_{\text{eq}}, y_{\text{iso}}\}$) in which there is anisotropic aggregation in $|y| < \max\{L_{\text{eq}}, y_{\text{iso}}\}$ and isotropic aggregation in $\max\{L_{\text{eq}}, y_{\text{iso}}\} < |y| < y_{\text{stab}}$ and Regime C ($L_{\text{dyn},f=0} > \alpha/\beta$, $L_{\text{eq}} < y_{\text{stab}} < y_{\text{iso}}$) in which there is anistropic aggregation across the entire region $|y| < y_{\text{stab}}$.

Figure 10 shows all five of the Regimes A–E. It should additionally be noted that if y_{stab} is very small compared to dynamical length scales then instability may be inhibited. This might justify defining a further distinct regime, but since the priority has

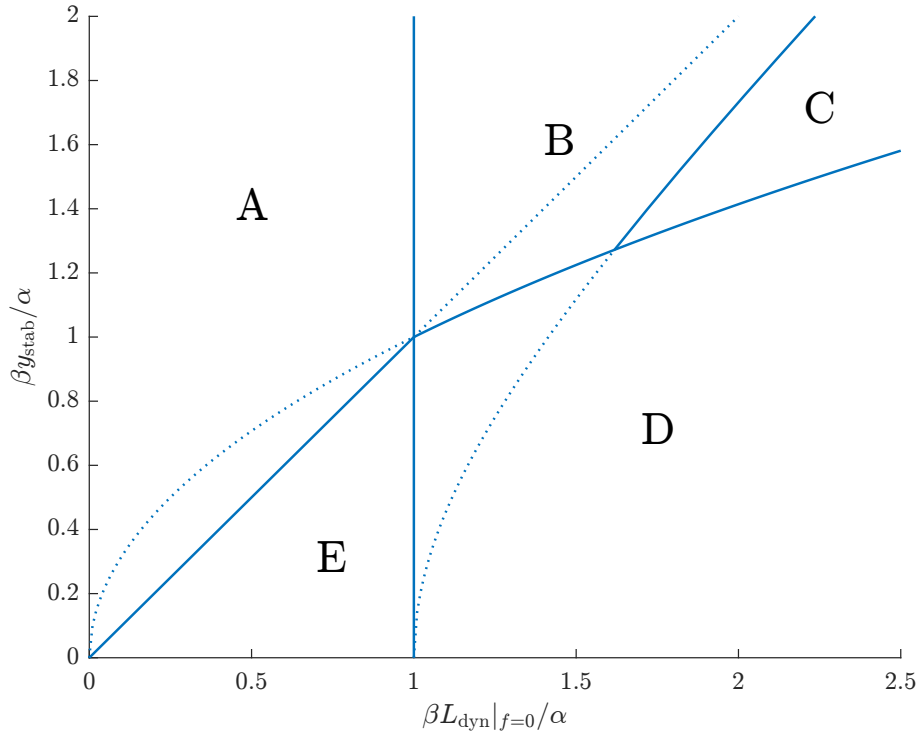

Figure 10. A schematic plot of the expected regimes of behaviour of our model on the equatorial β plane, shown against varying the latitudinal scale of the equatorial wave response $L_{\text{dyn}}|_{f=0} = c/\sqrt{\alpha\lambda}$ and the latitudinal limit of bistability y_{stab}, both in units of β/α. Solid lines denote regime boundaries, and the remainder of each curve is dotted. The straight line dividing A from E is given by $y_{\text{stab}} = L_{\text{dyn}}|_{f=0}$. The curve dividing B and C from D is given by $y_{\text{stab}} = L_{\text{eq}}$. The curve separating B from C is given by $y_{\text{stab}} = y_{\text{iso}}$. A brief summary of the characteristics of each Regime is as follows. Regime A: Local f plane analysis valid, isotropic aggregated regions extend to y_{stab}. Regime B: Aggregated regions on equator with y-scale L_{eq}, transition to isotropic aggregrated regions at larger $|y|$. Regime C: Aggregated regions on equator with y-scale L_{eq}, anisotropic aggregated regions at larger $|y|$. Regime D: Aggregated regions centred on equator with y-scale L_{eq}. Regime E: Aggregated regions centred on equator with y-scale y_{stab}. In Regimes A, B and C there are multiple aggregated regions in latitude. Regions at different latitudes propagate in the x-direction at different velocities. In Regimes D and E all aggregated regions are centred on the equator and there is coherent propagation in the x-direction.

been to interpret behaviour seen in cases where there is instability the criteria for such a regime has not been investigated in detail.

Having identified the different dynamical regimes that characterise the system on the β-plane we now consider the implications for spatial propagation of the aggregated moist and dry regions. It is useful to consider the moisture evolution equation. Using Eq. (25) to substitute for $\nabla.\boldsymbol{u}$, Eq. (4) becomes

$$q_t + \epsilon\nabla\cdot(\boldsymbol{u}[q]q) - \kappa\nabla^2 q = F_q(q) - \frac{Q}{H}(F_h(q) - \overline{F_h(q)}) + \frac{\lambda Q}{H}(h - \overline{h}). \tag{36}$$

This differs from the WTG form, Eq. (7), by the final term on the right-hand side, which is non-zero unless h is spatially uniform. Note that the contributions to q_t that arise under WTG are not expected to lead to systematic propagation since the relation between these terms and q is isotropic.

The extra term $(\lambda Q/H)(h-\overline{h})$ can potentially cause systematic propagation if its relation to q, as expressed by Eq. (34) is anisotropic. As has been noted previously this relation is isotropic on the f-plane, implying no propagation in that case. Under circumstances where a local analysis of Eq. (34) is appropriate, it is useful to exploit the analogy between Eq. (34) and a damped advection-diffusion equation, with a source term $F_h(q)$. The advecting velocity is $gH\beta/(f^2+\alpha^2)\hat{\boldsymbol{n}}$. For a positive q anomaly, given that $F_h(q)$ is negative, h will therefore be negative in the direction $\hat{\boldsymbol{n}}$ and hence, according to Eq. (36), q_t will be negative in the $\hat{\boldsymbol{n}}$ direction and positive in the $-\hat{\boldsymbol{n}}$ direction, implying propagation of the q anomaly in the $-\hat{\boldsymbol{n}}$ direction. This direction is westward if $f<\alpha$ and eastward if $f>\alpha$.

Local analysis of Eq. (34) applies for all y in Regimes A and E. These have been identified previously as isotropic at leading-order, but there is weak anisotropy and this will lead to westward propagation of aggregated moist regions close to the equator. It follows that a sufficient condition for westward propagation at the equator is $L_{\text{dyn},f=0}<\alpha/\beta$. In case A the instability and hence aggregation, may extend to y such that $|f|>\alpha$, i.e. $|y|>\alpha/\beta$, and there propagation will be eastward. The same analysis is relevant to regions such that $|y|>y_{\text{iso}}$ in Regime B. Since this corresponds to $|f|>\alpha$ the propagation in these regions will be eastward.

Since the equatorial region $|y|<L_{\text{eq}}$ in Regimes A, B and C is broader than α/β, and the direction of the vector $\hat{\boldsymbol{n}}$ will change direction when varied across this region, the implication of the local analysis is that there will be a relative westward shift of the moisture anomalies near the equator and a relative eastward shift elsewhere, generating a "<" shape. The overall propagation speed is likely to depend on the latitudinal extent of the equatorial moist region, i.e. on the ratio $L_{\text{eq}}\beta/\alpha$. A wider region, with a larger value of this ratio will have a larger proportion in which $-\hat{\boldsymbol{n}}$ is directed in the eastward region shift, and hence is likely to have greater more eastward propagation.

The dependence of propagation speed on latitude suggests that when there are multiple aggregated structures in latitude, i.e. in Regimes A, B and C, there will not be a single propagation speed. If there is a dominant speed for structures at the equator then that will be different from, and more westward, those that for structures at higher latitudes.

An alternative way to understand the behaviour near the equator in Regimes B, C and D is through decomposition into equatorial waves, following Wu et al. (2001). This offers a different approach to describing solutions of Eq. (34), for h given $F_h(q)$ that does not require regarding the coefficients as locally constant. For reference, the method of calculation is set out in Appendix C. In this case the correction $(\lambda Q/H)(h-\overline{h})$ appearing in Eq. (36) to the WTG convergence associated with a moisture-driven equatorial heating anomaly will be divergent responses to the east from Kelvin waves, acting to shift the moist anomaly to the west, and to the west from Rossby waves, acting to shift the moist anomaly to the east. There will also be a similar response, with convergence rather than divergence, associated with the negative moisture, i.e. dry, anomaly. It is the combination of these two wave responses that determines the net propagation of moist regions (and dry regions) in this system, with a stronger Rossby wave response implying eastward propagation. The fact that the Kelvin wave response decays away from the equator on a scale L_{eq} whereas the Rossby wave response can be excited at any latitudinal scale suggests

that the Kelvin wave response will be relatively weaker and therefore there will be eastward propagation if moist regions are significantly wider that L_{eq} (though it should be noted that it has previously been argued that L_{eq} is the natural latitudinal scale of aggregated moist anomalies).

The arguments presented above may be used to formulate a quantification of contributions of different processes to zonal propagation that may be applied to the simulations discussed in the following section. This quantification is useful where the disturbances may be considered to be coherently propagating at a speed U, which is possible when the latitudinal structure of the disturbances is dominated by a single moist anomaly centred on the equator, corresponding to Regimes D and E. In a reference frame moving at constant speed U in the zonal direction, the moisture distribution will then be steady and the moisture equation Eq. (4) may be re-written as

$$-Uq_x = -Q\nabla.\boldsymbol{u} - \epsilon\nabla.(q\boldsymbol{u}) + F_q + \kappa\nabla^2 q. \tag{37}$$

Multiplying by q_x and integrating over the domain implies that the propagation speed U is given by

$$U = -\iint \left(-Q\nabla.\boldsymbol{u} - \epsilon\nabla.(q\boldsymbol{u}) + F_q + \kappa\nabla^2 q\right) q_x \, dA \Big/ \iint q_x^2 \, dA. \tag{38}$$

This may be interpreted as an expression of the propagation speed as a sum of contributions from different individual terms in the moisture equation. We can exploit this, decomposing the divergence into a weak temperature gradient part D_{WTG} defined from Eq. (6) and its departure from WTG, D'.

$$U = -\iint \left(-QD' - \epsilon\nabla.(q\boldsymbol{u}) - Q/H(F_h - \bar{F}_h) + F_q + \kappa\nabla^2 q\right) q_x \, dA \Big/ \iint q_x^2 \, dA. \tag{39}$$

Since F_q and F_h are functions of q only and $\bar{F}_h$ is constant, the corresponding terms in the integral are total derivatives and so vanish when integrated over a periodic domain. We can also write the diffusive term as $q_x\nabla^2 q = (\frac{1}{2}[q_x^2 - q_y^2])_x + (q_x q_y)_y$. The diffusive term therefore also does not contribute to the integral. The coherent propagation of moist regions can therefore decomposed into parts due to the departure of the divergence from WTG and nonlinear advection:

$$U = -\iint \left(-QD' - \epsilon\nabla.(q\boldsymbol{u})\right) q_x \, dA \Big/ \iint q_x^2 \, dA. \tag{40}$$

Other terms in the moisture equation do not contribute. Note however that this does not mean that the propagation speed is independent of the value of κ, for example. Diffusivity still plays a role in setting the width of the unstable region $|y| < y_{\text{stab}}$ and the shape of individual moist regions. The term including $-QD'$ may be further decomposed into separate contributions $-QD'_{\text{Rossby}}$ and $-QD'_{\text{Kelvin}}$ from, respectively, Rossby and Kelvin wave parts of the dynamical response to the heating implied by the moisture field. The method for calculating the Rossby and Kelvin wave responses is set out in Appendix D.

5.2 Numerical Simulations

As with the f-plane case previously, much understanding of the behaviour on an equatorial β-plane can be gleaned from numerical simulations. Numerical details are again in Appendix A and details of parameter values for simulations for which results are displayed are included in Table 1.

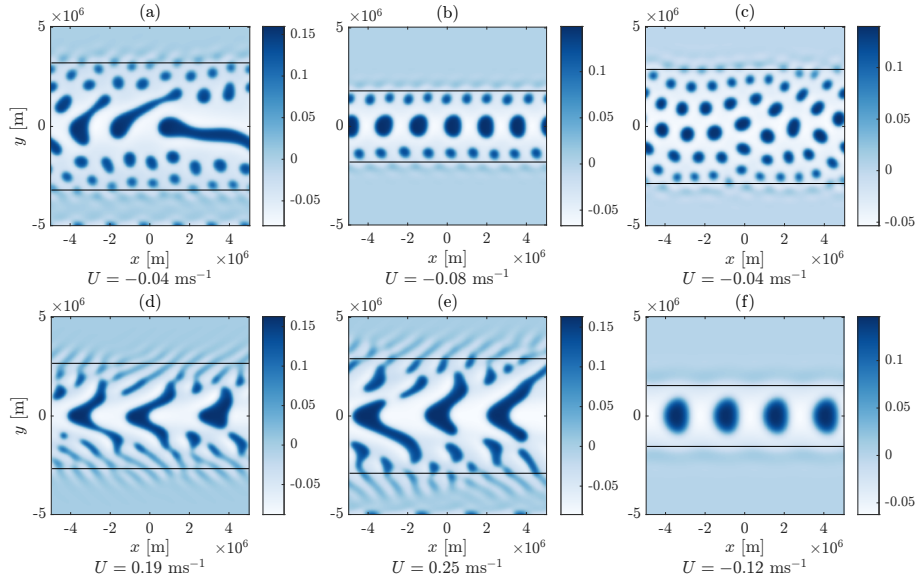

Figure 11. A selection of perturbation moisture distributions q/Q from various beta plane simulations, annotated with the propagation speed U at the equator estimated from the simulation (see text for details). All cases shown 400 days after initialization from a state with small-amplitude random q field. (a) has damping rates $\alpha = \lambda = 10^{-5}\mathrm{s}^{-1}$. (b) is as (a) but with λ increased by a factor of 3, this has $\alpha = 10^{-5}\mathrm{s}^{-1}$, $\lambda = 3 \times 10^{-5}\mathrm{s}^{-1}$.. (c) then has α also increased by a factor of 3, to $\alpha = \lambda = 3 \times 10^{-5}\mathrm{s}^{-1}$. (d) is as (a) but with α reduced by a factor of 10 for $\alpha = 10^{-6}\mathrm{s}^{-1}$, and $\lambda = 10^{-5}\mathrm{s}^{-1}$. (e) then has $\lambda = 3 \times 10^{-6}\mathrm{s}^{-1}$, and $\alpha = 10^{-6}\mathrm{s}^{-1}$. (f) is as (a) but with the diffusivity κ increased by a factor of 4 to $\kappa = 4 \times 10^5\mathrm{m}^2\mathrm{s}^{-1}$. The red horizontal lines denote $y = y_{\mathrm{stab}}$.

A particular focus in analysing the simulations will be on the zonal propagation of disturbances. For each simulation a zonal phase speed U is estimated by identifying the dominant zonal wavenumber in the moisture field at the equator and then tracking its time evolution. The estimated values U are noted in the figures showing results from the simulations. In some cases, where the latitudinal structure of the disturbances is dominated by a single moist anomaly centred on the equator, the disturbances may be considered to be uniformly propagating. When the latitudinal structure is more complicated, with multiple moist regions at different latitudes, the assumption of uniform propagation is not applicable, since different moist regions may propagate at different speeds. The estimated phase speed U then corresponds to the phase speed of moisture features along the equator and with features further from the equator tending to propagate eastward relative to those at the equator. The expression, Eq. (40), that potentially diagnoses the mechanisms for propagation is unlikely to be useful in cases with multiple phase speeds.

5.2.1 $\epsilon = 0$

We will again start with the case where nonlinear advection is excluded. The results of a set of simulations for various diffusivities and damping rates such that $L_{\mathrm{eq}} < y_{\mathrm{stab}}$ are shown in Fig. 11. In all panels the red horizontal lines denoting $y = y_{\mathrm{stab}}$, calculated on the basis of the f-plane linear instability analysis, mark the boundary between the low-latitude region of insta-

bility of the RCE state and the higher latitude region where RCE is stable. The early time evolution and up-scale growth is similar to the constant-f cases, controlled by the value of κ (recall Fig. 8). The aggregation ceases at a scale determined by the dynamics, as discussed previously for the f-plane in Sect. 4.2 and developed further for the β-plane in Sect. 5.1. At the equator, where $f = 0$, we expect the scale of aggregated regions to follow $c/\sqrt{\alpha\lambda}$, up to quantisation by the domain size. This is consistent with the structure seen in panels (a)–(c), where the predicted scale decreases by a factor of $\sqrt{3}$ between each panel and the zonal scale decreases accordingly. Also as expected, the scale of aggregated regions decreases away from the equator as f increases. This is especially clear in panel (c) and also visible in panels (a) and (b). As in the f-plane cases, the timescale for upscale growth becomes very slow at large scales. The plots here are all shown after 400 days of simulation even though there may at that time be continuing systematic growth of the spatial structure.

In panels (a), (b), (d), (e) and (f) the spatial scale of moist regions at the equator is sufficiently large, $L_{\text{dyn},f=0} > \alpha/\beta$, so that the anisotropic term becomes dominant in Eq. (33). According to the classification defined in Fig. 10 Panels (a) and (b) reside in Regime B. Consistent with this a region of non-isotropic aggregation forms near the equator and extends to over part, corresponding to $|y| < y_{\text{iso}}$, of the unstable region, though in panel (b) the anisotropic effect is weak. The anisotropic equatorial moist regions are locally shifted in the local direction of $-\hat{\boldsymbol{n}}$, in contrast to the quasi-circular structures seen at larger $|y|$ where isotropy applies. The increase in damping rates from panel (b) to panel (c), so that (c) lies in Regime A, decreases the scale of equatorial moist regions sufficiently that the aggregation is isotropic across the entire unstable region.

Panels (d), (e) and (f) reside in regime C and the extension of the anisotropic region up to $|y| = y_{\text{stab}}$ is consistent with this. Panel (f) is an example of the system with increased diffusivity and all other parameters the same as those in panel (a). The main differences in (f) relative to (a) are the decrease in y_{stab}, and the slight increase in scale due to the increased with of the diffusive boundaries between the moist and dry states. Panel (f) shows, in some sense, marginal behaviour between regimes C and D. Whilst the aggregated regions near the equator are confined to $|y| < y_{\text{stab}}$, they are sufficiently wide that no further disturbances form at larger y. The aggregation takes the form of a series of uniformly propagating regions, qualitatively similar to regime D.

Evidence of the f-plane Regimes may also be seen in the change in structure with latitude. Panels (a)–(c) and (f) all have a direct transition from Regime I to III at $y = y_{\text{stab}}$, and accordingly aggregated regions remain circular up to this boundary. Panels (d) and (e), however, have an intermediate range of y for which the local f-plane behaviour is in Regime II. The moist and dry regions near the boundaries in these cases are no longer circular and are far more transient, similar in character to the structure seen in Fig. 6C.

A case with $L_{\text{eq}} > y_{\text{stab}}$ is shown in Fig. 12. This has parameters chosen to lie in regime D. The value of Q has been decreased relative to previous cases shown, to 10.5 m, and this has reduced the magnitude of the negative gross moist stability at RCE, hence reducing y_{stab} to be less than L_{eq}. The unstable region is then sufficiently narrow that the equatorial wave response to the moist heating anomalies arising from the instability spread into $y > y_{\text{stab}}$. The associated convergence drives moisture anomalies and eventually a self-consistent balance between the dynamical and moisture fields is reached, with similar length scales for each, with the zonal scale given by $c/\sqrt{\alpha\lambda}$ and the latitudinal scale by L_{eq}.

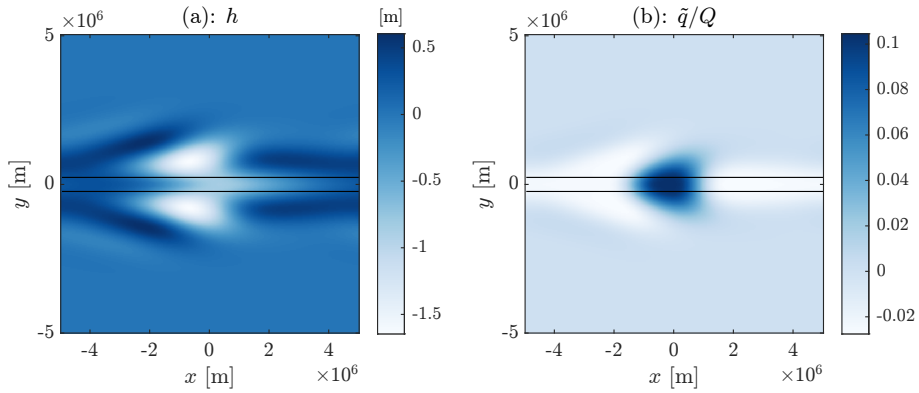

Figure 12. The thickness h (a) and perturbation moisture q/Q (b) fields for a simulation of the system on an equatorial beta-plane, 400 days after initialization from a state with small-amplitude random q field, with $\alpha = \lambda = 10^{-6}\mathrm{s}^{-1}$. The unstable region has been narrowed relative to the cases shown in Fig. 11 by decreasing the magnitude of the negative GMS at RCE. For this, we have set $Q = 10.5$ m. The horizontal lines show $y = y_{\text{stab}}$. This disturbance propagates to the east at a speed of $0.15\mathrm{ms}^{-1}$.

In contrast to the more complicated cases shown in Fig. 11(a-e), the structures shown in Fig. 11(f) and in Fig. 12 can be regarded as uniformly propagating and the decomposition expressed by Eq. (40) may be usefully applied. Given that $\epsilon = 0$ it is exclusively the $Q\nabla.\boldsymbol{u}$ term that is of interest and, as noted previously and using a method set out in Appendix D, this can be divided into the non-local quasi-steady Rossby and Kelvin wave responses to the heating $F_h(q)$. The separate Rossby and Kelvin wave contributions to the propagation speed of equatorial moist regions for these two cases, calculated on the basis of Eq. (40) with the convergence term split into the two contributions is shown in Fig. 13. Row (a)–(b) is the case originally shown in Fig. 11(f) and row (c)–(d) is the case in Fig. 12.

As expected the propagation direction does indeed depend on the relative strengths of the Kelvin wave response, which gives divergence to the east of the moisture anomaly and hence the heating anomaly, and the Rossby wave response which gives divergence to the west, with, in both cases, corresponding convergence associated with the dry anomaly. Note that the patterns shown in Fig. 13 are made up of both the divergence associated with the moist anomaly and the convergence associated with the dry anomaly. Row (a)–(b) has a stronger equatorial Kelvin wave response, and hence the moist and dry regions travel westward, whereas in row (c)–(d) the Rossby wave response is stronger and eastward propagation is observed. In both cases the contributions from each of the wave types is of a similar magnitude, and there is significant cancellation. This has two implications: the propagation speeds will tend to be smaller than expected from a dynamical scaling argument, and the propagation speed and direction is sensitive to the latitudinal structure of the moisture, since the details of that structure determine the relative strength of the Kelvin and Rossby wave responses. Note that the mechanism described above that associates the Rossby-wave response with eastward propagation is very different to those that are relevant to a similar association in other recent papers. The papers by Yano and Tribbia (2017); Rostami and Zeitlin (2019) describe a strongly nonlinear Rossby wave which propagates eastwards as a vortex pair, relying on nonlinearity in the vorticity and hence the momentum equations, which is absent in our model equations. The paper by Hayashi and Itoh (2017) describes a kind of diabatic Rossby wave, in which

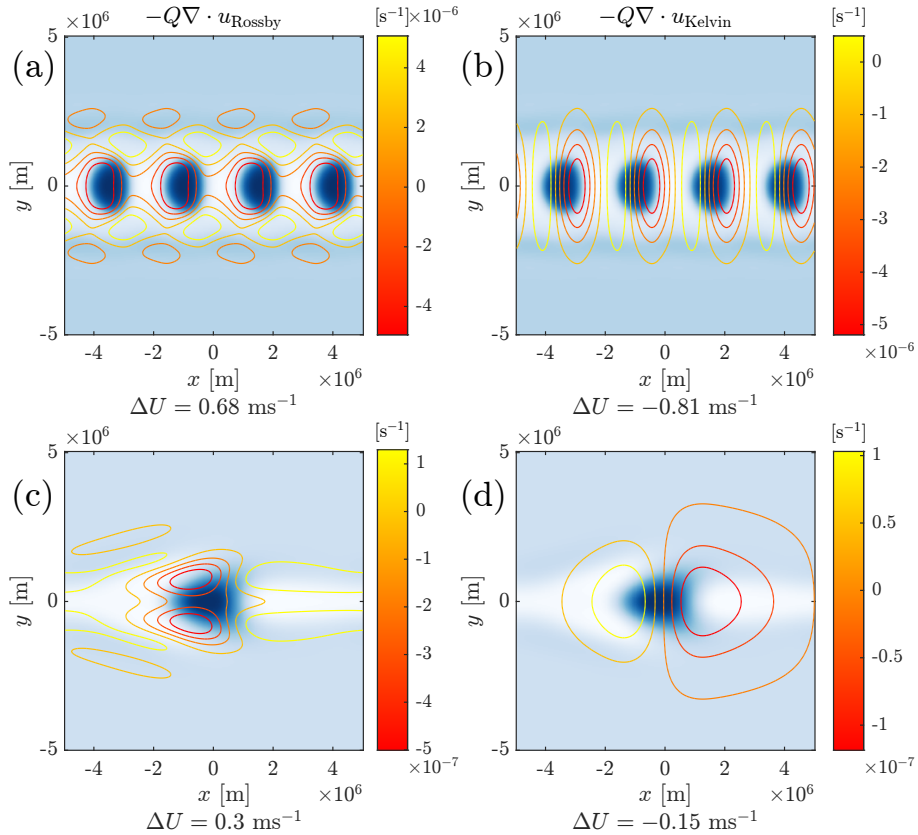

Figure 13. The spatial structure (contours) of the Rossby (left column) and Kelvin (right column) wave response to the moisture distribution (shading) of a uniformly propagating structure in two beta-plane simulations. The contribution of each wave, estimated according to the method given in Appendix D to the propagation speed is labelled. Row (a)–(b) corresponds to the state of the simulation in Fig. 11f, and row (c)–(d) corresponds to that shown in Fig. 12. Note that the colourbar on each figure has a different scale.

the vorticity field organises convective heating such that the resulting vorticity forcing implies propagation to the east. This mechanism is allowed in principle by our equations, but is not consistent with the fact that the evolution as simulated can be explained by a quasi-steady dynamical balance.

The arguments in Sect. 5.1 suggest that the parameter $c\beta/\alpha^{3/2}\lambda^{1/2}$ is important, with eastward propagation preferred when this is large. This is broadly consistent with the numerical simulations shown in Fig. 11 and Fig. 12, though this parameter certainly does not explain the propagation speed. For example, the system also shows a weak, complicated, dependence of propagation speed and direction on diffusivity, with increasing κ both widening the boundaries between moist and dry regions and reducing y_{stab}. Regions which propagate to the east, seen in Fig. 11(d) and (e) tend to have a '<' shaped spatial structure. This structure, also noted in Sugiyama (2009b), may be explained by the same asymmetry of the Kelvin and Rossby responses that determine the direction of propagation. The Kelvin wave divergence to the east is necessarily localised near the equator, however the Rossby wave response can be at any latitudinal scale. Hence, to the east of the moist region there is reduced

moisture convergence near the equator compared to further from the equator, however to the west of the moist region there is reduced moisture convergence at all latitudes. There is therefore a tendency for off-equatorial ($y > L_{eq}$) regions of moisture to shift to the east. For a large moist region at the equator extending into this region, an eastward tilt will be generated away from the equator.

5.2.2 $\epsilon > 0$

A set of examples with $\epsilon = 1$ is shown in Fig. 14. The change in spatial structure of the moisture field from $\epsilon = 0$ to $\epsilon = 1$ is largely similar to that seen previously in the constant-f case (Fig. 6 versus Fig. 9). The associated convergence causes moist regions to narrow, and rotational flows at larger f lead to spirals in the moisture distribution.

Of the cases shown, (a), in regime C, and (d), in regime D, may be considered as leading to a coherent steadily propagating disturbance (although the off-equatorial round regions in (a) are still unsteady). Panel (b) corresponds to regime B and is closer to local quasi-isotropic aggregation, with features at different latitudes propagating at different speeds. This is highly unsteady, though the narrow propagation of the nonlinear advective convergence close to the equator is preferentially in the zonal direction. Panel (c) resides near the boundary of regime B and C. The evolution is still very unsteady, with more evidence of rotational advection, particularly away from the equator where f is non-zero.

For the steadily propagating cases the same technique as previously, based on Eq. (40), may be used to decompose the different contributions to propagation as shown in Fig. 15. The nonlinear moisture advection term $\epsilon\nabla.(q\boldsymbol{u})$ term is now present. It is helpful to separate out the contribution from any zonal mean zonal flow. This is present because the segregation/aggregation process, with the form chosen for $F_h(q)$ and $F_q(q)$ tends to lead to a systematic zonal mean latitudinal structure in q, with greater q at low latitudes. This leads to a corresponding structure in h and hence in u. This zonal mean structure is present when $\epsilon = 0$, but then does not have any advective effect on q. The zonal mean u usually took the form of two off-equatorial jets, centred where the latitudinal gradient in h was largest. This zonal flow can potentially have a strong effect on propagation, but in the simulations shown here the generation of a strong zonal flow is avoided by choosing λ sufficiently large that variations in h are small. In practice this means $\lambda = \mathcal{O}(10^{-5}\mathrm{s}^{-1})$, well within the range discussed in the uniform rotation case.

For reference, the zonal mean u for the simulation in Fig. 15 is shown in Fig. 16. In this simulation the global mean u is $0.030\mathrm{ms}^{-1}$. The damping still allows a non-zero zonal mean u, however this is significantly reduced. The domain mean u is constrained to be near zero so there is no advection by a uniform background flow.

Returning to the decomposition into different contributions to propagation shown in Fig. 15, it may be seen that the linear divergence term (which has not been separated into Rossby and Kelvin contributions in this case, but the fact that the corresponding ΔU is positive implies that the Rossby contribution dominates), the term associated with advection by the zonal mean flow and the remaining nonlinear advection term are all comparable in magnitude.

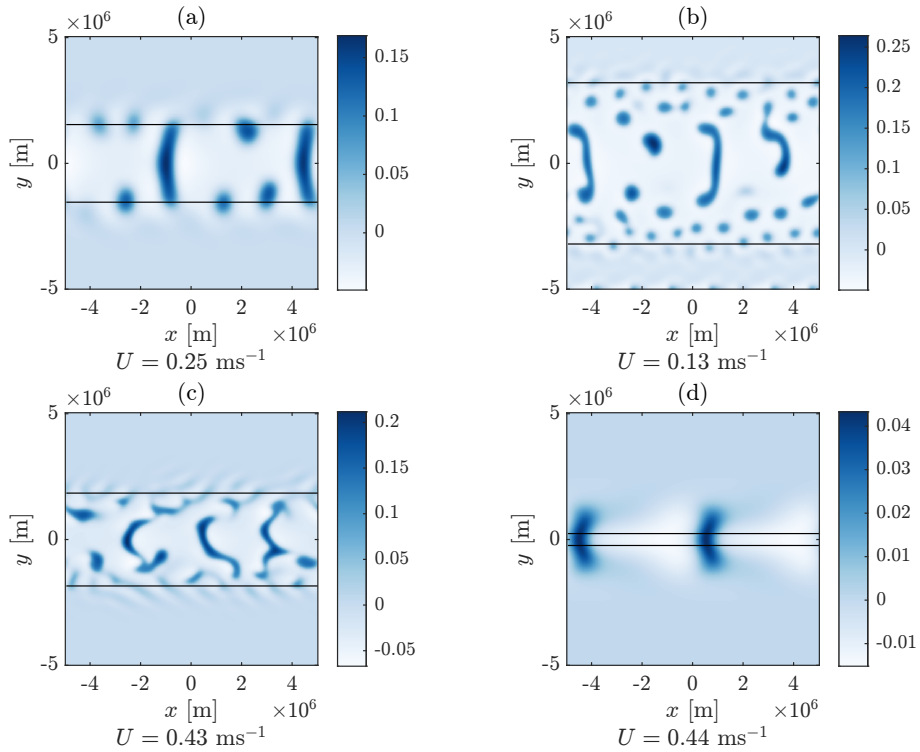

Figure 14. Snapshots of the normalised perturbation moisture distribution q/Q after 400 days in simulations with nonlinear dynamics ($\epsilon = 1$) on an equatorial beta-plane with (a) $\kappa = 4 \times 10^5 \mathrm{m}^2\mathrm{s}^{-1}$, $\alpha = \lambda = 10^{-5}\mathrm{s}^{-1}$, the same as Fig. 11f. Panel (b) has $\kappa = 10^5 \mathrm{m}^2\mathrm{s}^{-1}$ and $\alpha = \lambda = 10^{-5}\mathrm{s}^{-1}$, the same as Fig. 11a and panel (c) has $\kappa = 10^5 \mathrm{m}^2\mathrm{s}^{-1}$, $\alpha = 3 \times 10^{-6}\mathrm{s}^{-1}$ and $\lambda = 3 \times 10^{-5}\mathrm{s}^{-1}$, and does not correspond directly to a previous figure. Panel (d) illustrates the case with $L_{\mathrm{eq}} > y_{\mathrm{stab}}$, and has the same parameters as Fig. 12. These reside in regimes C, B, near the border of B and C, and D respectively.

6 Conclusions

In this study we have presented a single layer model for convective aggregation and its connection to large-scale dynamics. The linearised shallow water equations, governing the dynamics, are augmented with a moisture equation. The moisture field affects the dynamics via a heating term, i.e. a forcing term in the thickness equation, and the dynamics affects the moisture through convergence alone ($\epsilon = 0$) or with the additional effect of horizontal advection ($\epsilon > 0$). The form of the moisture-dependent precipitation term and the moisture-dependent heating term are such that under the WTG approximation the spatially homogeneous state in which both precipitation and heating are zero (the 'RCE state') is unstable and the system is bistable with moist and dry stable states (interpreted respectively as convecting and non-convecting). In this regime and with $\epsilon = 0$ the behaviour is described by a nonlinear reaction-diffusion equation for q, very similar to that presented by Craig and Mack (2013) and Windmiller and Craig (2019) (CMWC). As discussed by CMWC, the system exhibits a well-known spatial coarsening that may be interpreted as a representation of convective aggregation.

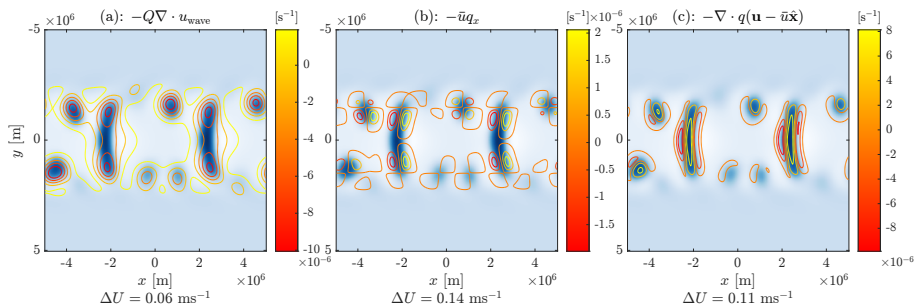

Figure 15. The spatial structure (contours) of the (a) linear wave, (b) advection by the zonal mean flow $\bar{u}$ and (c) nonlinear advection by the wave component contributions of the dynamical response to a moisture distribution (shading) of a propagating structure in a simulation on the beta-plane with nonlinear advection. This is the same case as shown in Fig. 14a. The contribution of each term to the propagation speed is labelled, and has been calculated using Eq. (38).

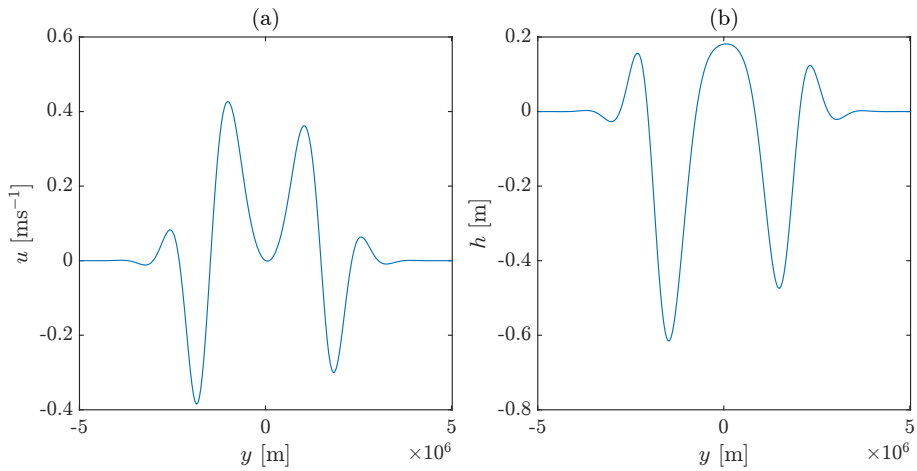

Figure 16. The zonal mean velocity (left) and height (right) fields for the state of the nonlinear beta-plane simulation shown in Fig. 15

.

The difference between our model system and that in the CMWC case is that in the latter the form of the reaction term is determined solely through the dependence of precipitation on q. In our case it is determined in addition by the q-dependent heating, which couples to the moisture equation through the dynamics. This allows us to include more general dynamical effects beyond the WTG approximation, including the impact of thermal and frictional damping, nonlinear advection, and rotation, including the extension to the equatorial β-plane. Under WTG dynamics with $\epsilon = 0$ the system, as that studied by CMWC, coarsens to the largest available scale. This remains the case when nonlinear advection ($\epsilon > 0$) is included, however the coarsening process is modified and the end state is different. Thermal and frictional damping and f-plane rotation in combination set a dynamical scale L_{dyn} which is an upper limit on the validity of WTG. The result is that coarsening proceeds only to this scale and then ceases. Depending on the relative values of thermal and mechanical damping and rotation the final

state may be essentially steady, or it may be unsteady, with, in some cases, a symmetry breaking leading to loss of spatial isotropy and to propagating structures. The nature of the linearly unstable modes provides some guidance to the type of the final state that is observed. If f is large enough, with the critical value f_{stab} depending on moisture diffusivity κ as well as other parameters defining the system, then the RCE state is stable.

Many aspects of the behaviour of a β-plane may be interpreted in terms of the previously discussed f-plane behaviour. A latitude y_{stab} may be defined by $f_{\text{stab}} = \beta y_{\text{stab}}$. Disturbances resulting from the instability of the RCE state are largely confined within the latitudinal band $(-y_{\text{stab}}, y_{\text{stab}})$, with some penetration outside of this band as a result of the non-locality of the dynamics, typically with a scale of L_{eq} or the local value of L_{dyn}, whichever is the larger. When thermal and mechanical damping are strong enough, specifically when $c\beta\alpha^{-3/2}\lambda^{-1/2} < 1$, the evolution of the system is similar to that observed on the f-plane, with spatial modulation corresponding to the spatial variation of the value of f. The weak anistropy introduced by the β-effect leads to zonal propagation. The propagation is incoherent, with structures at different latitudes propagating at different speeds, however those on the equator propagate to the west. When $c\beta\alpha^{-3/2}\lambda^{-1/2} > 1$ the structures that form in the q-field are more strongly anistropic and coherently propagating and it is helpful to formulate a description of the dynamics in terms of equatorial Kelvin and Rossby waves. If the former dominate the dynamical response to the q-anomalies there is propagation to the west, if the latter there is propagation to the east. Nonlinear moisture advection enhances eastward propagation but the enhancement is not large. Whilst the model equations allow time-dependent dynamics, the propagating disturbances that are seen in simulations may essentially be explained by considering the quasi-steady velocity fields forced by moisture-determined heating and then the effect of those velocity fields, through convergence/divergence and nonlinear advection on the moisture field and in that sense are 'moisture modes'.

The β-plane results that we present, particularly for the $c\beta\alpha^{-3/2}\lambda^{-1/2} < 1$ regime, have significant common ground with those presented by Sugiyama (2009b). The behaviour observed in both studies is similar, with small wavenumber, slowly propagating, moist regions forming at the equator. Indeed, as noted previously, one might argue that we are re-examining aspects of the models developed by Sugiyama (2009a, b). These models include WISHE, which we do not, and have a physically derived formulation of the forcing terms F_q and F_h. We on the other hand emphasise the general implications of the bistability of the system, as implied by the form of $F_q(q) - (Q/H)F_h(q)$ and then choose very simple ad hoc forms of $F_q(q)$ and $F_h(q)$ to provide such bistability. We also deliberately trace the behaviour of the system through a sequence starting with WTG dynamics, and hence the reaction-diffusion behaviour discussed by CMWC, and finishing with the β-plane with thermal and mechanical damping and nonlinear moisture advection that was the focus of Sugiyama (2009b), thereby provide a new perspective on the latter.

One difference between the conclusions of this study and those of Sugiyama (2009b) is that the latter suggests that diffusion is an important contributor to the eastward propagation of moist regions in the nonlinear advection case. We, however, conclude that the propagation is due to a combination of nonlinear advection and the displaced convergence associated with the Rossby wave response. This difference in interpretation arises in part from the diagnostic approaches used to measure the impact of different terms of the moisture tendency. We assume a uniformly propagating disturbance and then calculate a contribution from each moisture tendency term to the speed of propagation using Eq. (40), taking into account the entire spatial distribution.

As noted previously, according to that approach the net contribution of the diffusive term to propagation is zero. In contrast, whereas Sugiyama (2009b) compares the tendencies at $y = 0$. Only comparing at $y = 0$ will overestimate the contribution of the diffusive tendency to the overall propagation. In a "<" shaped aggregated region, the negative curvature at the equator to the east increases the diffusive speed of the boundary whereas the opposite is true away from the equator where the sign of the curvature changes. The fact that according to Eq. (15) the net contribution of diffusion to propagation is identically zero does not, of course, rule out the possibility that the form of the terms that do contribute to net propagation are affected by diffusivity. Certainly there is evidence of diffusivity dependence of propagation speed in the results presented, for example cases (a) and (f) in Fig. 11, where increasing diffusivity increases westward propagation speed and cases (a) and (b) in Fig. 14, where increasing diffusivity increases eastward propagation speed. A further difference from Sugiyama (2009b) is that we have not assumed h-dependence in F_h and F_q. As has been noted previously, this follows other moisture-mode models including Sobel and Maloney (2012, 2013); Adames and Kim (2016), but a preliminary assessment of the effect of including h-dependence is given in Appendix E and concludes that the formation, evolution and propagation of moist and dry regions, as described in Sect. 5 above remains broadly unchanged when this is included.

Recent review articles on convective aggregation, for example Wing et al. (2017) and Muller et al. (2022), include the CMWC reaction-diffusion model in their discussion of different mechanisms and models. The model we have presented above generalises CMWC by linking the moisture and large-scale dynamical equations. In the categorisation given by Wing et al. (2017) our model fits the category of a long-wave radiation feedback with an additional advective process feedback included if $\epsilon > 0$. The CMWC model on the other hand, which relies on the form of the moisture dependence of precipitation, is categorised as a moisture feedback. Unlike the CMWC model, our model, by including dynamics, provides an upper limit on the scale of aggregation scale, which is finite and determined by thermal and momentum damping rates when $f = 0$ and reduces as f increases. How this relates to evidence from GCM and CRM simulations (Muller et al., 2022) boundary layer processes are important in determining the upper limit on aggregation scale remains to be determined.

Whilst most high-resolution three-dimensional modelling of convective aggregation has focused on the non-rotating case, there has been some recent investigation of the f-plane and β-plane cases, Carstens and Wing (2022, 2023). Some aspects of the behaviour reported in those papers is seen in our much simpler model. On the f-plane circular moist, i.e. convecting, regions form and the scale of these regions decreases as f increases, scaling as $1/f$ in both models at large rotation rates. However the physics of this behaviour is likely to be very different between the two models. In the CRM studies of Carstens and Wing (2022) the structure for larger values of f is dominated by tropical cyclones. These have no clear analogue in our model which neglects advective nonlinearity in the momentum equation. Carstens and Wing (2022) identify an intermediate range of f within which convective aggregation simply does not occur. One possibility is that this corresponds to our $f > f_{\text{stab}}$, with formation of tropical cyclones being a distinct process occurs that occurs at larger f value in the CRM, but which is simply absent in our model. On the β-plane (Carstens and Wing, 2023) identify the dominant structures that arise out of convective aggregation at low latitudes as convectively coupled Kelvin waves, whereas in our model the structures are clearly moisture-mode in character, with quasi-steady dynamical fields. In future work it would be interesting to investigate further whether

there are parameter regimes in which our model also shows moisture-modified Kelvin waves as the dominant low-latitude structures.

One goal for formulating and studying the model presented in this paper was as a basis for understanding the MJO. But in several respects, with the model in its current form, this goal has not been met. One aspect of this is the phenomenon of aggregation, which might provide an explanation for the large longitudinal scale of the MJO, without requiring a scale-selection mechanism for an underlying instability, such as radiative-convective instability, that is consistent with such a large scale. However the rate of increase in scale due to aggregation is in our model determined by diffusivity of moisture and, with the value assumed, $\kappa = 10^{-5}\mathrm{m}^2\mathrm{s}^{-1}$ the time taken to reach scales of 1000s of kilometers is (at least) many tens of days. This does not seem consistent with the observed MJO evolution in which convection appears to develop over a large, perhaps 10^4km region of the Indian Ocean on a time scale of a few days. A larger value of diffusivity would reduce the time required but might be difficult to justify – e.g. see the arguments in Biagioli and Tompkins (2023). It might be that aggregation acts alongside large-scale-selective processes such as the radiative destabilization suggested by Adames and Kim (2016).

A second aspect is the horizontal structure and propagation characteristics of disturbances predicted by the model on an equatorial β-plane. Our model in this form is similar to that considered previously by Sugiyama (2009b) and our conclusions are also similar. There is organisation into distinct moist and dry regions on the equator and these propagate in the zonal direction. As Sugiyama (2009b), we emphasise that these propagating disturbances are fundamentally nonlinear and their structure and propagation characteristics are not captured by a linear stability analysis. However, the direction of propagation can be westward or eastward and the speed of propagation is consistently much less than 1 ms^{-1} when the observed MJO phase speed is around 5 ms^{-1}. Furthermore, even when the propagation is eastward propagating disturbances, the spatial structure of eastward propagating moist regions also necessarily forms a '<' shape, compared to the '>' structure of the observed MJO (Adames and Wallace, 2014).

It seems likely that, for a model of this nonlinear moisture-mode form to produce more realistic MJO-like behaviour, further physics needs to be included (though it should not be forgotten that MJO-like behaviour can also be produced by models with a different form). Recent simple one-dimensional moisture-mode models for the MJO (Sobel and Maloney, 2013; Adames and Kim, 2016), which are based on prescribed latitudinal structure of the flow variables, have included extra effects such as synoptic eddy drying, boundary-layer convergence of moisture (Adames and Wallace, 2014) and latitudinal gradients in background moisture (Adames and Kim, 2016) and have shown that these effects lead to enhanced eastward propagation. However the results we have presented in Sect. 5 of this paper demonstrate that propagation speed and direction are sensitive to latitudinal structure, so there are disadvantages to prescribing this *a priori*. We are currently including of some of these effects in the two-dimensional model described in this paper with the aim of capturing the behaviour of nonlinear MJO-like disturbances in a model where the two-dimensional structure emerges rather than being prescribed. Such a model could also potentially be used to study a broader class of moist tropical variability (e.g. Wang and Sobel, 2022).

Code availability. TEXT

Data availability. TEXT

Code and data availability. TEXT

Sample availability. TEXT

Video supplement. Videos associated with the figures shown in this paper are available at https://drive.google.com/drive/folders/1AG7Mjn0Nk4Gp_1wv0la9NutXAHdUFQuz?usp=sharing

Appendix A: Numerical Details

To solve Eq. (1)–(3) numerically, we will use standard methods for solving the shallow water equations. The equations will be discretised on an Arakawa C-grid with the grid spacing chosen to be sufficiently small as to avoid numerical instability. In general we take a grid spacing of $d = 4 \times 10^4$ m and a time step of 112.5 s. The system is then numerically stepped forward in time using a third order Adams-Bashforth scheme.

On the f-plane, we will take (doubly) periodic boundaries, whereas on the equatorial β-plane rigid north and south boundaries will be used along with periodic east and west boundaries. At rigid boundaries, a sponge layer will be included to avoid edge effects. This takes the form of a linear damping term included in the velocity and thickness fields. On the periodic channel we use on the β-plane this takes the form

$$\alpha_{\text{sponge}} = 1 \times 10^{-5} \left[\exp\left(-\frac{70(L_y - 2y)}{L_y} \right) + \exp\left(\frac{70(L_y + 2y)}{L_y} \right) \right] \text{s}^{-1}. \tag{A1}$$

The magnitude decays over a scale of $1/70^{\text{th}}$ of the domain. Any numerical results do not depend on the magnitude or decay scale of this damping term.

Appendix B: Details of the f-plane linear stability problem

This Appendix gives further details of the linear stability problem considered in Sect. 4.1. Considering small amplitude perturbations of the form $u = \Re\{\hat{u} \exp(\sigma t + ikx)\}$, with analogous notation for other variables, the linearised forms of Eq. (1), Eq. (2) and Eq. (3) lead to Eq. (14), Eq. (15), Eq. (16) and Eq. (17), with $\mu_2 = -F_h'(0)$ and $\mu_1 = -F_q'(0)$, matching the notation used in Eq. (13) and Eq. (12). These define an eigenvalue problem for the growth rate σ, which is a root of the quartic equation

$$\sigma^4 + [\lambda + 2\alpha + \mu_1 + \kappa k^2]\sigma^3$$
$$+ [(\mu_1 + \kappa k^2)(\lambda + 2\alpha) + f^2 + c^2k^2 + \alpha^2 + 2\alpha\lambda]\sigma^2$$
$$+ [(\mu_1 + \kappa k^2)(f^2 + c^2k^2 + \alpha^2 + 2\alpha\lambda) + c^2k^2\alpha + (f^2 + \alpha^2)\lambda - g\mu_2 Q k^2]\sigma$$
$$+ (\mu_1 + \kappa k^2)c^2k^2\alpha - g\mu_2 Q k^2\alpha + (f^2 + \alpha^2)\lambda(\mu_1 + \kappa k^2) = 0, \quad \text{(B1)}$$

with coefficients that are either linear or quadratic in k^2. Simple limits of the above in the case where $f = \alpha = \lambda = 0$ have been noted previously, in particular in the WTG limit valid at small scales such that $k \gg \mu/c$ and the corresponding large-scale results for $k \ll \mu/c$.

To analyse further the behaviour of the full system as described by Eq. (B1), first consider the case where α and λ are non-zero, but there is no rotation, $f = 0$. It is helpful to note the solutions for σ in the large-scale (k small) and small-scale (k large) limits. Neglecting diffusion, it is straightforward to show that at small k the roots of Eq. (B1) are $\sigma \simeq -\lambda, -\alpha, -\alpha, -\mu_1$ (i.e. the $\sigma \simeq -\alpha$ root is repeated in the limit as $k \to 0$), implying stability at small k. Correspondingly at large k there are roots $\sigma \simeq \pm ick - \lambda - \mu_2 Q/H, -\alpha, \mu_2(Q/H) - \mu_1 = -\mu_1 M$ representing, respectively, two gravity waves damped by moisture effects and thermal damping, a vorticity disturbance damped by friction and a moisture mode (stable or unstable according to whether M is positive or negative). There is therefore stability at small k and, consistent with the WTG analysis, instability at large k if $M > 0$ with the latter limit unaffected by α and λ being non-zero. Introducing diffusion by taking $\kappa > 0$ will lead to all roots have negative real part as $k \to \infty$, i.e. the system is stabilised by diffusion of moisture.

Given that all roots for σ have negative real part for $k \to 0$ it is possible to deduce whether or not there is instability by seeking conditions, in particular a value of k, under which one of the roots has zero real part. One possibility is $\sigma = is$ with s real and non-zero. Substituting this form for σ into Eq. (B1), with $f = 0$, shows that such a root is not possible. (A key simplification in this case $f = 0$ that can be exploited here is that 4th-order polynomial appearing in Eq. (B1) has a factor $(\sigma + \alpha)$.) The other possibility is that $\sigma = 0$, implying the condition

$$(\mu_1 + \kappa k^2)c^2k^2\alpha - g\mu_2 Q k^2\alpha + \alpha^2\lambda(\mu_1 + \kappa k^2) = 0. \quad \text{(B2)}$$

This is a quadratic for k^2 which has real roots only if

$$\kappa \leq \frac{c^2\mu_1}{\lambda\alpha}(\sqrt{1 - M} - 1)^2 \quad \text{(B3)}$$

where as before $M = 1 - \mu_2 Q/\mu_1 H$. If this condition is satisfied then there is instability. If it is not satisfied then there is no instability. Furthermore it may also be deduced that roots with positive real part must be real, and that for any value of k there can be at most one such root. Given that roots crossing the real axis must do so at $\sigma = 0$, complex roots with positive real part would be possible only if there were two such crossings, from below to above the real axis, as k increased, and those real roots then combined to give a complex conjugate pair. But this is not possible since the above shows that there are at most two crossings for any k and, given that all roots are below the real axis for small k and (with $\kappa > 0$) for large k, then one crossing is from below the real axis to above and the other is from above to below.

When f is non-zero then the conditions for instability are more complicated. The large-k limits of the roots of Eq. (B1) are unaffected from the expressions given above (which were for $\kappa = 0$, the same holds for $\kappa > 0$. However the small-k limits

are now $\sigma \simeq -\lambda, \pm if - \alpha, -\mu_1$, implying stability at small k, but note that the previous repeated root $-\alpha$ now becomes the complex conjugate pair $\pm if - \alpha$ (i.e. frictionally damped inertial oscillations). The transition between small-k and large-k now depends on the values of α, λ, f, κ and other parameters. The expression, Eq. (B3), for real roots of Eq. (B1) to cross the real axis generalises to (repeating Eq. (21)

$$\kappa \leq \frac{c^2 \mu_1 \alpha}{\lambda(\alpha^2 + f^2)} (\sqrt{1 - M} - 1)^2, \tag{B4}$$

however this does not give complete information on when instability is possible because with f non-zero complex conjugate roots may also cross the real axis. Some analytical progress can be made in describing the dependence of roots on the different parameters, but the algebra is complicated.

As described in Sect. 4.1, numerical investigation shows that, for given κ, the (α, λ)-plane may be divided into four regions, each corresponding to a different regime of behaviour. Regime I is where $\Re(\sigma) > 0$ occurs for some k only when σ is real. Regimes IIa and IIb are where there are complex σ with $\Re(\sigma) > 0$ and with non-zero imaginary parts and Regime III is where there is no instability, i.e. $\Re(\sigma) < 0$ for all k. The distinction between IIa and IIb is that in IIb σ corresponding to fastest growth over all k has non-zero imaginary part. (An example is given in Fig. 5.) For $f = 0$ only Regimes I and III are present and Regime I, the region of instability, simply corresponds to Eq. (B3).

The boundary between Regime I and Regime IIa, indicated by the dashed curves in Fig. 5, is predicted by the condition that Eq. (B1) has a double root at $\sigma = 0$. This is because close to the $\lambda = 0$ axis a real root crosses the real axis. As λ increases a bifurcation point emerges on the solution branch joining to that root and then moves toward the real axis. A transition between crossing of the real axis by a real root and crossing by a complex conjugate pair occurs when the bifurcation point reaches the real axis, i.e. at the point in there is a double root at $\sigma = 0$. For all f such that Regimes IIa and IIb exist the boundary between Regimes IIa and Regime I asymptotes to $\alpha \simeq \lambda$ as $\lambda \to 0$. The condition for a double root at $\sigma = 0$ must be compatible with the condition given by Eq. (21). Given that the gradient of the dashed curves is observed to reduce as λ increases, there will not be compatibility if the gradient of the curve defined by equality in Eq. (21) is larger than 1 as $\lambda \to 0$, i.e. if

$$f > c(\mu_1/\kappa)^{1/2}(\sqrt{1 - M} - 1). \tag{B5}$$

When this inequality is satisfied there are no Regimes IIa and IIb and the transition between stability and instability is described completely by Eq. (B4) or, equivalently Eq. (21). The double root at $\sigma = 0$ occurs when both the constant term and the coefficient of the linear term in Eq. (B1) vanish simultaneously. Eliminating k^2 gives a constraint on λ and α (for given values of other parameters such as f and κ) which defines the boundary between regions I and IIa in Fig. 5. For illustration, the resulting expression when $\kappa = 0$ is

$$\lambda = \lambda_*(\alpha) = \frac{(\alpha^2 + f^2)(Q\mu_2/H - \mu_1)\mu_1 \alpha}{(Q\mu_2/H - \mu_1)(f^2\mu_1 - f^2\alpha - \alpha^2\mu_1 - \alpha^3) - \alpha\mu_1(f^2 + \alpha^2)}. \tag{B6}$$

In the small-α limit it corresponds to $\lambda \simeq \alpha$. Since this expression for $\lambda_*(\alpha)$ ignores diffusivity it will provide an accurate description of the boundary between regions I and IIa only when that is approximately independent of diffusivity, which, from inspection of Figure 5 appears to be the case when λ and α are small.

The boundary between regions IIb and III corresponds to a double root of Eq. (B1) of the form $\sigma = is$, $s \neq 0$. Some information about the small-α part of this boundary may be obtained by considering the special case $\alpha = 0$. Then there is a root that is $\mathcal{O}(k^2)$ for small k, with real part that has the same sign as $\lambda\mu_1(\mu_2(Q/H) - \mu_1) - f^2(\lambda + \mu_2(Q/H))$. It follows that for $f^2 < (\mu_2(Q/H) - \mu_1)\mu_1$ there is instability at small k for large λ, therefore stabilization requires non-zero α. On the other hand for $f^2 > (\mu_2(Q/H) - \mu_1)\mu_1$ then there is stability at large λ independent of the value of α. This explains the different forms of the boundary between regions IIb and III seen in Figure 5 for $f = 0$ and $f = 10^{-5}\mathrm{s}^{-1}$ on the one hand and $f = 3 \times 10^{-5}\mathrm{s}^{-1}$ and $f = 6 \times 10^{-5}\mathrm{s}^{-1}$ on the other.

Finally we consider the regime where α and λ are both small. This is potentially described by setting $\alpha = \lambda = 0$ in Eq. (B1), leading to

$$\sigma^3 + \left(\kappa k^2 + \mu_1\right)\sigma^2 + \left(f^2 + c^2k^2\right)\sigma + f^2\mu_1 + c^2\kappa k^4 + f^2\kappa k^2 + c_m^2 k^2 \mu_1 = 0, \tag{B7}$$

with the moist gravity wave speed $c_m^2 = c^2 M < 0$ for the potentially-unstable case. Denote the roots of this polynomial as σ_1, σ_2 and σ_3.

First assume that there is a purely imaginary root $\sigma = is$, with s real. Substituting into Eq. (B7) and separating the real and imaginary parts gives

$$s^3 = (f^2 + c^2k^2)s, \tag{B8}$$

$$(\mu_1 + \kappa k^2)s^2 = \mu_1 f^2 + \mu_1 c_m^2 k^2 + f^2\kappa k^2 + c^2\kappa k^4. \tag{B9}$$

Eliminating s gives that

$$(\mu_1 + \kappa k^2)(f^2 + c^2k^2) = \mu_1 f^2 + \mu_1 c_m^2 k^2 + f^2\kappa k^2 + c^2\kappa k^4. \tag{B10}$$

For $k > 0$, this is true only in the dry case $c_m^2 = c^2$. Since we are interested in $c_m^2 < 0 < c^2$, purely imaginary roots are not possible. Therefore any change in stability happens at $\sigma = 0$.

For small k, Eq. (B7) has roots $\sigma = -\mu_1$ and $\sigma = \pm if - k^2(2g\mu_2\mu_1 f^2 Q \pm 2if(f^2c^2 + c_m^2\mu_1^2))/4(f^4 + f^2\mu_1^2)$, both of which have negative real part. For instability, a change of stability to occur at some $k > 0$. This requires the constant coefficient of Eq. (B7) to change sign, i.e.

$$\mu_1 f^2 + \mu_1 c_m^2 k^2 + f^2\kappa k^2 + c^2\kappa k^4 = 0. \tag{B11}$$

If the system is to be stable, this must have no real roots $k^2 > 0$. Nondimensionalise with $\hat{f} = f\sqrt{\kappa/\mu_1}/c$ and $\hat{k} = k/\sqrt{\kappa/\mu_1}$, for

$$\hat{f}^2 + M\hat{k}^2 + \hat{f}^2\hat{k}^2 + \hat{k}^4 = 0. \tag{B12}$$

The transition from two roots to none happens when the discriminant of this expression (as a quadratic in k) vanishes, or

$$(M + \hat{f})^2 - 4\hat{f}^2 = 0 \tag{B13}$$

This has four roots,

$$\hat{f} = \pm(1 - \sqrt{1 - M}) \tag{B14}$$

corresponding to the change in global stability, and

$$\hat{f} = \pm(1 + \sqrt{1 - M}), \tag{B15}$$

which are a spurious solution arising from roots with $k^2 < 0$. (When $\hat{f}^2 > 1 - M$, the left-hand side of Eq. (B12) is strictly positive for k^2>0).

With zero damping, and assuming positive we therefore expect instability if and only if

$$f < f_{\text{stab}} = \frac{c}{\sqrt{\kappa/\mu_1}}(\sqrt{1 - M} - 1), \tag{B16}$$

where f_{stab} is defined by this equality. Since only one mode may be unstable in this case the unstable mode must have a purely real growth rate and therefore cannot propagate.

When f_{stab} is defined by Eq. (21), we can consider the limit as $\alpha \to 0$ and $\lambda \to 0$. This limit is not well defined, it depends on the value of α/λ.

The condition $f = f_{\text{stab}}(\lambda, \alpha)$ defines a function on the (λ, α) plane, as the maximum f for given α and λ (with dependence on other parameters such as κ suppressed) for which the moist shallow water equations, Eq. (1), Eq. (2) and Eq. (3), permit unstable modes. Contours of this function are shown in Fig. 5, as the bounding curves for Regime III. This surface is expected to be continuous at all points but the origin, $\alpha = \lambda = 0$.

There is a value f_I such that the region of (λ, α) space defined by $f_{\text{stab}}(\lambda, \alpha) > f_I$ contains no parameters corresponding to Regime II, i.e. the unstable roots of Eq. (B1) are real for all k. Within this region, the stability boundary is given by Eq. (21), rewritten here as a curve in the (λ, α) plane,

$$\lambda = \frac{c^2 \mu_1 \alpha}{\kappa(\alpha^2 + f^2)}(\sqrt{\mu_2 Q/\mu_1 H} - 1)^2. \tag{B17}$$

(Note that as f is made arbitrarily large this curve moves closer to the α-axis, but does not meet it, so there is always some range of α such that there is instability for $\lambda = 0$.) Within this region, the transition with varying k of a single real mode from unstable to stable corresponds to the change in sign of the constant term of Eq. (B1). When $\alpha = \lambda = 0$, this constant term is identically zero. We therefore have a degenerate case where the previously unstable mode is zero for all k. The limit as $\alpha \to 0$ and $\lambda \to 0$ of f_{stab} in this region is therefore not expected to be well defined.

In the other section of the (α, λ) plane, where $f_{\text{stab}} < f_I$, instability both near the origin and near $f = f_{\text{stab}}$ relies on a combination of two complex roots of Eq. (B1). If $\alpha \to 0$ and $\lambda \to 0$ only one of these roots can become identically zero so we expect the other to become the unstable mode, and hence the limit of f_{stab} should be well defined.

Since f_{stab} is continuous, apart from at the origin, we also expect the limit along the boundary between each of the cases discussed above, $f_{\text{stab}} = f_I$ to give the correct value of $f_{\text{stab}}(0, 0)$. This is a curve of constant f_{stab} and therefore

$$f_I = f_{\text{stab}}(0, 0) = \left(\frac{c^2(\sqrt{\mu_2 Q/\mu_1 H} - 1)^2}{\kappa/\mu_1} \right)^{1/2}. \tag{B18}$$

The curve $f_{\text{stab}}(\lambda,\alpha) = f_{\text{stab}}(0,0)$ has the expected form, $\alpha = \lambda$, when $\alpha \ll f$.

Appendix C: The Equatorial Wave Response to a Moist Heating

The full derivation of linear equatorial wave response to a heating with general thermal damping λ and friction α is derived by Wu et al. (2001). The setup and relevant results are repeated here for reference.

We look for equatorially trapped steady state solutions to the dynamical equations Eq. (1) and $Eq.$(3), under the long wave approximation. These satisfy

$$\alpha u - \beta y v = -g h_x, \tag{C1}$$

$$\beta y u = -g h_y, \tag{C2}$$

$$\lambda h + H \nabla \cdot \boldsymbol{u} = F, \tag{C3}$$

with u, v and $h \to 0$ as $y \to \pm\infty$. For simplicity we will take a domain of effectively infinite zonal length, i.e. with length much greater than the extent of the equatorial wave response. We may also assume that $F(x,y)$ has mean zero, as a non-zero spatial mean part will be balanced by a change in the mean h and not contribute to the divergence or equatorial wave response.

We must first decompose the heating in terms of parabolic cylinder functions $\tilde{D}_n$ as

$$F = \sum_{n\geq 0} F_n \tilde{D}_n\left(\frac{y}{L_{eq}}\right), \tag{C4}$$

with $L_{eq} = \sqrt{c/\beta}\,\sqrt[4]{\alpha/\lambda}$.

The relevant parts of the solutions are then the equatorial Kelvin wave, which has

$$h_0 = -\frac{1}{2}\int_{-\infty}^{x} \sqrt{\frac{\alpha}{\lambda}} c^{-1} \exp\left(-\frac{\sqrt{\alpha\lambda}(x-u)}{c}\right) F_0(u)\, du\, \tilde{D}_0\left(\frac{y}{L_{eq}}\right). \tag{C5}$$

and the equatorial Rossby waves, for $n = 1, 2, 3, \ldots$,

$$h_n = -\frac{1}{2}\int_{x}^{\infty} \sqrt{\frac{\alpha}{\lambda}} c^{-1} \exp\left(\frac{(2n+1)\sqrt{\alpha\lambda}(x-u)}{c}\right) (F_{n-1}(u) + 2nF_{n+1}(u))\, du \left(\tilde{D}_{n+1}\left(\frac{y}{L_{eq}}\right) + 2(n+1)\tilde{D}_{n-1}\left(\frac{y}{L_{eq}}\right)\right). \tag{C6}$$

The divergence is then calculated using Eq. (C3). This gives

$$D[F] = F/H - \lambda \sum_{n\geq 0} h_n/H \tag{C7}$$

For our purposes it is worth noting that for a heating, $F < 0$, the equatorial wave response has $h < 0$, and hence the associated divergence is positive. This is the opposite sign to the WTG divergence associated with a heating.

Appendix D: Decomposition of the Divergence into Equatorial Wave Components

For diagnostics of numerical simulations we will want to calculate the expressions in Appendix C numerically, including separating the divergence into weak temperature gradient and equatorial Kelvin and Rossby components. The expression for the divergence, Eq. (C7), can simply be split into parts

$$D[F] = D_{\text{WTG}}[F] + D_{\text{Kelvin}}[F] + D_{\text{Rossby}}[F], \tag{D1}$$

where

$$D_{\text{WTG}} = \frac{1}{H}\left(F\right), \tag{D2}$$

$$D_{\text{Kelvin}} = -\lambda h_0[F]/H, \tag{D3}$$

and

$$D_{\text{Rossby}} = -\lambda \sum_{n\geq 1} h_n[F]/H. \tag{D4}$$

The first two of these are simple to calculate numerically, using the expressions in Appendix C with $F = F_h - \bar{F}_h$. The Rossby wave component will then be calculated as the residual

$$D_{\text{Rossby}} = D - D_{\text{WTG}} - D_{\text{Kelvin}}. \tag{D5}$$

This will introduce some error as the numerical fields are not exactly quasi-steady, however it also sidesteps a few issues with calculating the Rossby wave response. The first of these is that simple theoretical expressions for the Rossby wave response to a given heating assume the long-wave approximation, whereas this assumption is not needed for the Kelvin wave and may not be valid for certain simulations. The second is that the Rossby wave response requires forcing due to all terms in Eq. (C4). For a finite discretised domain these will be both poorly resolved and cut off by boundaries. This leads to errors in the projection, since the modes are no longer orthogonal.

Appendix E: The dependence of precipitation parameterisation on temperature

The impact of including a temperature dependence in the precipitation, as done by Sugiyama (2009a, b), will be briefly discussed in this section. The coupling terms between the moisture and thickness equations in these papers are assumed to be functions of the precipitation P, which is assumed to depend only on the moist static energy, $c_pT + L_vq$ in standard units, and the nonlinear combination of P with a surface flux term. The terms thus take the form $F_h(q - h/\gamma, h)$ and $F_q(q - h/\gamma, q)$, where γ is proportional to L_v/c_p. This form is motivated in Sugiyama (2009a, b) as a Betts-Miller type representation, but

the alternative interpretation as a representation of the dependence of precipitation on free-tropospheric humidity, which was emerging at the time of those papers, is noted and that interpretation would not require the h-dependence.

The forcing in this paper has been defined differently, we have taken a simple linear form for all forcing terms except for a nonlinear $F_h(q)$, which provides an effective limit on the magnitude of the moisture variable. Two initial questions which arise from the temperature dependence of F_h and F_q are: how does the behaviour change if the nonlinear term limits $q - h/\gamma$ rather than q, and, what is the effect of a linear h term in the moisture equation?

In first case we can get some insight by redefining the moisture variable. In this case the moisture terms become $F_h(q-h/\gamma)$ and $F_q(q-h/\gamma)$. Subtracting $1/\gamma$ times Eq. (2) from Eq. (3) gives

$$(q-h/\gamma)_t + (Q-H/\gamma)\nabla\cdot\boldsymbol{u} = F_q(q-h/\gamma) - \gamma^{-1}F_q(q-h/\gamma) + \lambda h/\gamma + \kappa\nabla^2 q. \tag{E1}$$

Defining a new variable $q' = q - h/\gamma$, along with corresponding $Q' = Q - H/\gamma$ and $F_{q'} = F_q - F_h/\gamma$ we can write this as

$$q'_t + Q'\nabla\cdot\boldsymbol{u} = F_{q'}(q') + \kappa\nabla^2 q' + (\lambda + \nabla^2)h/\gamma. \tag{E2}$$

We have returned to the original form of the equations to be studied with a different interpretation of the moisture variable and some terms due to h on the right hand side of the moisture equation. Physically, it is unclear whether the diffusive forcing should act on moisture q or on moist static energy q', so the $\nabla^2 h$ term may be unnecessary. If this term is neglected we end up in the situation described by the second question above. The distinction between the two setups can therefore also be interpreted as whether the diffusivity is believed to act upon the moisture or moist static energy.

The situation described by Sugiyama (2009b) is one step further, with $F_h = F_h(q,h)$ and $F_q = F_q(q,h)$. We will now investigate this case, but for simplicity we will continue to take $\epsilon = 0$. In the strict WTG limit, $h = h(t)$, using Eq. (2) to eliminate the divergence from Eq. (3) gives

$$q_t = F_q(q,h) - \frac{Q}{H}(F_h(q,h) - \overline{F_h(q,h)}) + \kappa\nabla^2 q. \tag{E3}$$

This is still a reaction-diffusion equation for q, however now the stable states depend on h, $q_\pm = q_\pm(h)$. These vary slowly as h varies slowly with time.

When L_{dyn} is larger than the domain size, the arguments in Sect. 2 hold and aggregation proceeds as expected. Otherwise, WTG cannot be valid across the entire domain. Moist regions, with their associated heating, will have decreased h and the opposite will apply to dry regions. On a β-plane the region of reduced h associated with a heating at the equator will extend zonally with the equatorial Rossby and Kelvin wave responses. It is therefore expected that there will be local variation of the stable values of q. Of particular interest will be whether the adjustment to the fixed points due to the equatorial wave response to heating will facilitate moistening to the east or west of an equatorial moist region.

We know that the fixed points will depend on h, but what do we expect this variation to look like? We will discuss two informative special cases. For the first, we assume that the moisture and thickness coupling terms depend only on the moist static energy, $F_q(q-h/\gamma)$ and $F_h(q-h/\gamma)$. Note that since the spatial mean of F_h must be subtracted when applying the WTG approximation, an additional linear relaxation term in the thickness equation will not affect the WTG moisture equation.

In this case the stable fixed points are related to the stable fixed points $q_{\pm 0}$ when $h = 0$ by

$$q_{\pm}(h) = q_{\pm 0} + h/\gamma. \tag{E4}$$

Thus, the stable values of the moisture increase with h.

For the second case we will assume that the effect of h in the moisture equation is weak and therefore may be represented by a linear term. Taking inspiration from te previous case but Taylor expanding the moisture about RCE, the moisture equation, Eq. (3), is

$$q_t + Q\nabla \cdot \boldsymbol{u} = F_q(q) - F_q'(Q)h/\gamma + \kappa\nabla^2 q. \tag{E5}$$

Note that $F_q' < 0$. Now, rather than the condition that at fixed points the reaction function $G_{hq} = 0$, we need that $G_{hq} = F_q'(0)h/\gamma$. This corresponds to a translation of the curve shown in Fig. 1 upwards with increasing h. Therefore as h increases the stable moisture values will correspondingly increase.

These two situations agree on the expected effect of h dependent forcing in the height equation on the moisture distribution, however the impact on the propagation speed of aggregated regions is unclear. Since the deviation from isotropy in the beta-plane case is due to the equatorial wave response to the heating we expect significant cancellation between the equatorial Rossby and Kelvin wave components, as seen in the divergence response discussed in Sect. 5. We will investigate this numerically.

Examples of numerical simulations from both cases is shown in Fig. E1. The first case, with both coupling terms functions of only moist static energy leads to only very small changes from the corresponding case in Fig. 11—the spatial distribution and wavenumber are similar, and the speed has increased by $0.1\mathrm{ms}^{-1}$. The 'specific heat' parameter $\gamma = 22.2$ has been chosen to match the combination of geopotential and temperature in Sugiyama (2009b). The fact that the effect is small is consistent with the large value of γ. The height dependence in the second case has had a larger effect, with the wavenumber reduced from 4 to 3, and the speed reduced by a factor of 3.

Author contributions. Both authors contributed equally to this work.

Competing interests. The authors have no competing interests to declare.

Disclaimer. TEXT

Acknowledgements. MD is grateful for funding from the UK EPSRC via a PhD studentship. Both authors thank the reviewers for valuable comments that helped us to improve the paper in revision.

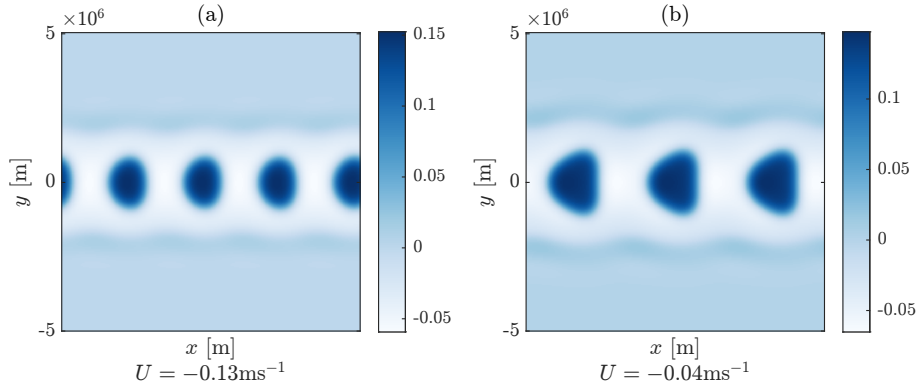

Figure E1. The late time behaviour of a beta-plane simulation with parameters as in panel (f) of Fig. 11, but with h dependent heating. Panel (a) has h dependence in both $F_h(q - h/\gamma)$ and $F_q(q - h/\gamma)$. Panel (b) has $F_h(q)$ and $F_q(q - h/\gamma)$. The 'specific heat' parameter $\gamma = 22.2$ has been chosen to match the combination of geopotential and temperature in Sugiyama (2009b).

References

Adames, Á. F. and Kim, D.: The MJO as a dispersive, convectively coupled moisture wave: Theory and observations, Journal of the Atmospheric Sciences, 73, 913–941, https://doi.org/10.1175/JAS-D-15-0170.1, 2016.

Adames, Á. F. and Wallace, J. M.: Three-dimensional structure and evolution of the vertical velocity and divergence fields in the MJO, Journal of the Atmospheric Sciences, 71, 4661–4681, https://doi.org/10.1175/JAS-D-14-0091.1, 2014.

Adames, A. F., Kim, D., Clark, S. K., Ming, Y., and Inoue, K.: Scale analysis of moist thermodynamics in a simple model and the relationship between moisture modes and gravity waves, Journal of the Atmospheric Sciences, 76, 3863–3881, https://doi.org/10.1175/JAS-D-19-0121.1, 2019.

Arnold, N. P. and Randall, D. A.: Global-scale convective aggregation: Implications for the Madden-Julian Oscillation, Journal of Advances in Modeling Earth Systems, 7, 1499–1518, https://doi.org/https://doi.org/10.1002/2015MS000498, 2015.

Beucler, T., Cronin, T., and Emanuel, K.: A Linear Response Framework for Radiative-Convective Instability, Journal of Advances in Modeling Earth Systems, 10, 1924–1951, https://doi.org/10.1029/2018MS001280, 2018.

Biagioli, G. and Tompkins, A. M.: A Dimensionless Parameter for Predicting Convective Self-Aggregation Onset in a Stochastic Reaction-Diffusion Model of Tropical Radiative-Convective Equilibrium, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 231, https://doi.org/https://doi.org/10.1029/2022MS003231, e2022MS003231 2022MS003231, 2023.

Bray, A. J., Sluckin, T. J., McLeish, T. C., Blumenfeld, R., Hinch, E. J., Magerle, R., and Ball, R. C.: Coarsening dynamics of phase-separating systems, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 361, 781–792, https://doi.org/10.1098/rsta.2002.1164, 2003.

Bretherton, C. S., Blossey, P. N., and Khairoutdinov, M. F.: An energy-balance analysis of deep convective self-aggregation above uniform SST, Journal of the Atmospheric Sciences, 62, 4273–4292, https://doi.org/10.1175/JAS3614.1, 2005.

Carstens, J. D. and Wing, A. A.: A Spectrum of Convective Self-Aggregation Based on Background Rotation, Journal of Advances in Modeling Earth Systems, 14, 1–14, https://doi.org/10.1029/2021MS002860, 2022.

Carstens, J. D. and Wing, A. A.: Regimes of Convective Self-Aggregation in Convection-Permitting Beta-Plane Simulations, Journal of the Atmospheric Sciences, https://doi.org/https://doi.org/10.1175/JAS-D-22-0222.1, 2023.

Craig, G. C. and Mack, J. M.: A coarsening model for self-organization of tropical convection, Journal of Geophysical Research Atmospheres, 118, 8761–8769, https://doi.org/10.1002/jgrd.50674, 2013.

Emanuel, K., Wing, A. A., and Vincent, E. M.: Radiative-convective instability, Journal of Advances in Modeling Earth Systems, 6, 75–90, https://doi.org/https://doi.org/10.1002/2013MS000270, 2014.

Hayashi, M. and Itoh, H.: A New Mechanism of the Slow Eastward Propagation of Unstable Disturbances with Convection in the Tropics: Implications for the MJO, Journal of the Atmospheric Sciences, 74, 3749 – 3769, https://doi.org/10.1175/JAS-D-16-0300.1, 2017.

Hirt, M., Craig, G. C., Schäfer, S. A., Savre, J., and Heinze, R.: Cold-pool-driven convective initiation: using causal graph analysis to determine what convection-permitting models are missing, Quarterly Journal of the Royal Meteorological Society, 146, 2205–2227, https://doi.org/10.1002/qj.3788, 2020.

Holloway, C. E. and Neelin, J. D.: Moisture Vertical Structure, Column Water Vapor, and Tropical Deep Convection, Journal of the Atmospheric Sciences, 66, 1665 – 1683, https://doi.org/https://doi.org/10.1175/2008JAS2806.1, 2009.

Jiang, X., Adames, Á. F., Kim, D., Maloney, E. D., Lin, H., Kim, H., Zhang, C., DeMott, C. A., and Klingaman, N. P.: Fifty Years of Research on the Madden-Julian Oscillation: Recent Progress, Challenges, and Perspectives, Journal of Geophysical Research: Atmospheres, 125, 1–64, https://doi.org/10.1029/2019JD030911, 2020.

Khairoutdinov, M. F. and Emanuel, K.: Intraseasonal variability in a cloud-permitting near-global equatorial aquaplanet model, Journal of the Atmospheric Sciences, 75, 4337–4355, https://doi.org/10.1175/JAS-D-18-0152.1, 2018.

Muller, C., Yang, D., Craig, G., Cronin, T., Fildier, B., Haerter, J. O., Hohenegger, C., Mapes, B., Randall, D., Shamekh, S., and Sherwood, S. C.: Spontaneous Aggregation of Convective Storms, Annual Review of Fluid Mechanics, 54, 133–157, https://doi.org/10.1146/annurev-fluid-022421-011319, 2022.

Muller, C. J. and Bony, S.: What favors convective aggregation and why?, Geophysical Research Letters, 42, 5626–5634, https://doi.org/10.1002/2015GL064260, 2015.

Neelin, J. D. and Zeng, N.: A Quasi-Equilibrium Tropical Circulation Model—Formulation, Journal of the Atmospheric Sciences, 57, 1741–1766, https://doi.org/https://doi.org/10.1175/1520-0469(2000)057<1741:AQETCM>2.0.CO;2, 2000.

Raymond, D. J.: The Hadley circulation as a radiative-convective instability, Journal of the Atmospheric Sciences, 57, 1286–1297, https://doi.org/10.1175/1520-0469(2000)057<1286:THCAAR>2.0.CO;2, 2000.

Raymond, D. J., Sessions, S., Sobel, A. H., and Fuchs, Ž.: The Mechanics of Gross Moist Stability, Journal of Advances in Modeling Earth Systems, 1, n/a–n/a, https://doi.org/10.3894/james.2009.1.9, 2009.

Romps, D. M.: Rayleigh Damping in the Free Troposphere, Journal of the Atmospheric Sciences, 71, 553 – 565, https://doi.org/10.1175/JAS-D-13-062.1, 2014.

Rostami, M. and Zeitlin, V.: Eastward-moving convection-enhanced modons in shallow water in the equatorial tangent plane, Physics of Fluids, 31, 021 701, https://doi.org/10.1063/1.5080415, 2019.

Rubinstein, J., Sternberg, P., and Keller, J.: Fast Reaction, Slow Diffusion, and Curve Shortening, SIAM Journal on Applied Mathematics, 49, 116–133, https://doi.org/10.1137/0149007, 1989.

Sobel, A. and Maloney, E.: Moisture modes and the eastward propagation of the MJO, Journal of the Atmospheric Sciences, 70, 187–192, https://doi.org/10.1175/JAS-D-12-0189.1, 2013.

Sobel, A. H. and Maloney, E. D.: An idealized semi-empirical framework for modeling the Madden-Julian oscillation, Journal of the Atmospheric Sciences, 69, 1691–1705, https://doi.org/10.1175/JAS-D-11-0118.1, 2012.

Sobel, A. H., Nilsson, J., and Polvani, L. M.: The weak temperature gradient approximation and balanced tropical moisture waves, Journal of the Atmospheric Sciences, 58, 3650–3665, https://doi.org/10.1175/1520-0469(2001)058<3650:TWTGAA>2.0.CO;2, 2001.

Sobel, A. H., Bellon, G., and Bacmeister, J.: Multiple equilibria in a single-column model of the tropical atmosphere, Geophysical Research Letters, 34, 1–5, https://doi.org/10.1029/2007GL031320, 2007.

Sugiyama, M.: The moisture mode in the quasi-equilibrium tropical circulation model. Part I: Analysis based on the weak temperature gradient approximation, Journal of the Atmospheric Sciences, 66, 1507–1523, https://doi.org/10.1175/2008JAS2690.1, 2009a.

Sugiyama, M.: The moisture mode in the quasi-equilibrium tropical circulation model. Part II: Nonlinear behavior on an equatorial β plane, Journal of the Atmospheric Sciences, 66, 1525–1542, https://doi.org/10.1175/2008JAS2691.1, 2009b.

Wang, S. and Sobel, A. H.: A Unified Moisture Mode Theory for the Madden–Julian Oscillation and the Boreal Summer Intraseasonal Oscillation, Journal of Climate, 35, 1267–1291, https://doi.org/https://doi.org/10.1175/JCLI-D-21-0361.1, 2022.

Windmiller, J. M. and Craig, G. C.: Universality in the spatial evolution of self-aggregation of tropical convection, Journal of the Atmospheric Sciences, 76, 1677–1696, https://doi.org/10.1175/JAS-D-18-0129.1, 2019.

Wing, A. A., Emanuel, K., Holloway, C. E., and Muller, C.: Convective Self-Aggregation in Numerical Simulations: A Review, Surveys in Geophysics, 38, 1173–1197, https://doi.org/10.1007/s10712-017-9408-4, 2017.

Wu, Z., Sarachik, E. S., and Battisti, D. S.: Thermally driven tropical circulations under Rayleigh friction and Newtonian cooling: Analytic solutions, Journal of the Atmospheric Sciences, 58, 724–741, https://doi.org/10.1175/1520-0469(2001)058<0724:TDTCUR>2.0.CO;2, 2001.

Yang, D.: Boundary Layer Diabatic Processes, the Virtual Effect, and Convective Self-Aggregation, Journal of Advances in Modeling Earth Systems, 10, 2163–2176, https://doi.org/10.1029/2017MS001261, 2018.

Yang, D.: A shallow-water model for convective self-aggregation, Journal of the Atmospheric Sciences, 78, 571–582, https://doi.org/10.1175/JAS-D-20-0031.1, 2021.

Yano, J.-I. and Tribbia, J. J.: Tropical Atmospheric Madden–Julian Oscillation: A Strongly Nonlinear Free Solitary Rossby Wave?, Journal of the Atmospheric Sciences, 74, 3473 – 3489, https://doi.org/10.1175/JAS-D-16-0319.1, 2017.

Zeng, N., Neelin, J. D., and Chou, C.: A Quasi-Equilibrium Tropical Circulation Model—Implementation and Simulation, Journal of the Atmospheric Sciences, 57, 1767 – 1796, https://doi.org/10.1175/1520-0469(2000)057<1767:AQETCM>2.0.CO;2, 2000.

Zhang, C., Adames, Á. F., Khouider, B., Wang, B., and Yang, D.: Four Theories of the Madden-Julian Oscillation, Reviews of Geophysics, 58, https://doi.org/10.1029/2019RG000685, 2020.