The "Hydro-ABC model" (Vn 2.0): a simplified convective-scale model with moist dynamics

Jiangshan Zhu[1,*] and Ross Noel Bannister[2,*]

[1]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, P.R. China
[2]National Centre for Earth Observation, University of Reading, Reading, UK
[*]These authors contributed equally to this work.

Correspondence: Ross Bannister (r.n.bannister@reading.ac.uk)

Abstract. The prediction of convection (in terms of position, timing, and strength) is important to achieve in high-resolution weather forecasting. This problem not only requires good convective-scale models, but also data assimilation systems to give initial conditions which neither improperly hinders nor hastens convection in the ensuing forecasts. Solving this problem is difficult and expensive using operational-scale numerical weather prediction systems, and so a simplified model of convective-scale flow is under development. This paper extends the "ABC model" of dry convective-scale flow to include mixing ratios of vapour and condensate phases of water. This revised model is called "Hydro-ABC".

Hydro ABC includes transport of the vapour and condensate mixing ratios within a dynamical core, and transitions between these two phases via a micro-physics scheme. A saturated mixing ratio is derived from model quantities, which helps determine whether evaporation or condensation happens. Latent heat is exchanged with the buoyancy variable (ABC's potential temperature-like variable) in such a way to conserve total energy, where total energy is the sum of dry energy and latent heat. The model equations are designed to conserve the domain-total mass, water, and energy.

An example numerical model integration is performed and analysed, which shows the development of a realistic looking anvil cloud, and excitation of inertio-gravity and acoustic modes over a wide range of frequencies. This behaviour means that Hydro-ABC is a challenging model to allow experimentation with innovative data assimilation strategies in the next stage of work. Further, an ensemble of Hydro-ABC integrations is also performed in order to study the possible forecast error covariance statistics (necessary for data assimilation). These show patterns that are dependent on the presence of convective activity (at any model's vertical column), thus giving a taste of flow-dependent error statistics. Candidate indicators/harbingers of convection are also evaluated (namely relative humidity, hydrostatic imbalance, horizontal divergence, convective available potential energy, and convective inhibition), which appear to be reliable diagnostics concerning the presence of convection. These diagnostics will be useful in the selection of the relevant forecast error covariance statistics when data assimilation for Hydro-ABC is developed.

1 Introduction

Numerical models used for modern numerical weather prediction (NWP) services have evolved rapidly in sophistication and scale over recent decades. One of the most notable advances concerns the reduction in size of the models' grid boxes. Using

25   smaller grid sizes not only allows for finer flows to be resolved, but also permits new flow regimes, which could not be represented with coarser models (Clark et al., 2016; Yano et al., 2018). These flow regimes are more non-linear and inherently less predictable than larger-scale flows (Hohenegger and Schär, 2007; Leung et al., 2019; Lorenz, 1969) and thus there is an increased need for more observations to constantly correct for fast growing errors in order to produce useful forecasts (e.g. Banos et al. (2022)). This is not just a dynamical effect (for instance due to larger Rossby or Reynolds numbers), but is also

30   due to the role of phase transitions of water (namely evaporation and condensation, thus absorbing or releasing latent heat, (Errico et al., 2007)), which can vary over short length-scales, especially where moist processes are occurring (Montmerle and Berre, 2010).

  By resolving moist processes that lead to convective motion (partially achieved with km-scale grid lengths), some aspects of numerical models have become simpler, as much of the fine-scale transport of water and latent heat can be handled explicitly,

35   thus removing the need for a convection scheme. Conversely, the necessary techniques needed to realistically estimate the initial conditions (namely data assimilation, DA) have become more complicated owing to the breakdown in many of the assumptions that are made in traditional DA methods, like linearity of the models and observation operators, Gaussianity of the background and observation errors, and the homogeneity and quasi-static nature of the background error statistics (Bannister et al., 2020). These (and other) problems contribute to the difficulty of assimilating data (especially radar data) in such models

40   (Fabry and Meunier, 2020).

  High resolution models of the atmosphere and their DA counterparts remain very expensive to run, requiring costly super-computers. It is therefore desirable to study the convective-scale DA problem in simplified (or 'toy') models that are capable of being run at less cost on smaller computers. One of the first such simplified model was presented by Würsch and Craig (2014). Their model is based on the non-rotating 1D shallow water equations, comprising an equation each for fluid velocity,

45   u, fluid depth, h, and rain water, r. A further variable, the geopotential, ϕ, is tied to h in a non-trivial way. When the total height, $H + h$ (where H is the topography height) is less than a specified level, the system is considered sub-saturated and ϕ and h are related via $\phi = g(H + h)$ (where g is the acceleration due to gravity). When h is above the specified level, a cloud is assumed to form and an analogue of a 'low pressure' region is formed by redefining $\phi = gH + \phi_{\mathrm{c}}$ (where ϕ_{c} has a small value). This step encourages mass convergence and a form of convection to be initiated. Rain is then formed when h is above

50   a second specified level in regions of mass convergence. The model exhibits convection-like behaviour showing the sporadic appearance of clouds. Such a model, and its developments (Kent et al., 2017), provide useful first steps to provide a cheap and effective model on which to base developments in convective-scale data assimilation.

  The ABC model (Petrie et al., 2017) was separately developed as a low-cost model of convective-scale behaviour also to be used as a basis to investigate convective DA methods. The ABC model is based on the three-dimensional Euler equations, but is

55   reduced to two-dimensions (longitude and height). It is therefore more complex than the shallow water-based model of Würsch and Craig (2014), most notably in that it allows a height dependence. A variational DA capability was later developed for the ABC model (Bannister, 2020). This has since been used to show that enforcing geostrophic and hydrostatic balances as part of the background error covariance matrix (a key ingredient of the traditional variational DA problem) affects the DA's ability to analyse observations (Bannister, 2021). That study found that these balances, in the mid-latitude cases, are an advantage when

60 analysing the larger scales but are a disadvantage at small scales. The model has also been used in a tropical setting, where the DA method has been expanded to include a hybrid (ensemble-variational) capability (Lee et al., 2022). Until now however, the ABC model lacks a water variable, meaning that studies have been limited to dry dynamics only, where the behaviour of the model is shows little evidence of sporadic convective behaviour. In this paper we show how moist processes can be included in the ABC model and how such behaviour emerges, which is a more appropriate setting for DA schemes to be tested and

65 developed.

2 "Dry ABC" to "Hydro-ABC"

We seek modifications to the original ABC equations (Petrie et al., 2017) which are as simple as possible and requiring minimal change of the underlying numerical advection scheme. We refer to this moist model as "Hydro-ABC". The ABC model comprises five variables on a two-dimensional (longitude/height) grid: zonal wind u, meridional wind v, vertical wind

70 w, scaled density perturbation $\tilde{\rho}'$ (a pressure-like variable), and buoyancy perturbation b' (a potential temperature-like variable). Hydro-ABC contains two new variables, namely water vapour q, and condensate q_c mixing ratios, the latter of which may also be thought of a cloud variable. The model allows exchange between these two species according to the local thermodynamic conditions, with an associated latent heat exchange (see below). In order to keep the model as simple as possible, no rain is currently represented. The continuous dry equations support mass and total energy conservation, as shown in Petrie et al.

75 (2017). Hydro-ABC must maintain similar conservation properties, but with the addition of conservation of total water, $q + q_c$, and a total energy that includes a contribution from latent heat.

2.1 The Hydro-ABC model equations

The continuous Hydro-ABC equations are as follows:

$$\frac{\partial u}{\partial t} + B\mathbf{u} \cdot \nabla u + C\frac{\partial \tilde{\rho}'}{\partial x} - fv = 0, \tag{1a}$$

$$\frac{\partial v}{\partial t} + B\mathbf{u} \cdot \nabla v + fu = 0, \tag{1b}$$

$$\frac{\partial w}{\partial t} + B\mathbf{u} \cdot \nabla w + C\frac{\partial \tilde{\rho}'}{\partial z} - b' = 0, \tag{1c}$$

$$\frac{\partial \tilde{\rho}'}{\partial t} + B\nabla \cdot (\tilde{\rho}\mathbf{u}) = 0, \tag{1d}$$

$$\frac{\partial b'}{\partial t} + B\mathbf{u} \cdot \nabla b' + A^2 w = \mathcal{S}_{b'}, \tag{1e}$$

$$p' = C\rho_0 \tilde{\rho}', \tag{1f}$$

$$\frac{\partial \tilde{\rho}q}{\partial t} + B\nabla \cdot (\tilde{\rho}q\mathbf{u}) = \tilde{\rho}(Ev - Co), \tag{1g}$$

$$\frac{\partial \tilde{\rho}q_c}{\partial t} + B\nabla \cdot (\tilde{\rho}q_c\mathbf{u}) = -\tilde{\rho}(Ev - Co). \tag{1h}$$

Here A is the pure gravity wave frequency, B is the modulation of the advection and mass divergence terms, C is the inverse compressibility coefficient, t is time, x and z are positions in the longitude and height directions respectively, $\mathbf{u} = (u, v, w)$, f

is the Coriolis parameter, ρ_0 is a reference density, $\tilde{\rho}$ is the scaled density ($\tilde{\rho} = (\rho_0 + \rho')/\rho_0$), p' is the pressure perturbation

90 (diagnosed purely from $\tilde{\rho}'$), $\mathcal{S}_{b'}$ is the buoyancy source term (related to latent heat exchange, which is activated when there is a change in water phase), Ev is the evaporation rate, and Co is the condensation rate. The key differences between Hydro-ABC the dry ABC system are $\mathcal{S}_{b'}$ in (1e), and the additional equations (1g) and (1h). How $\mathcal{S}_{b'}$ and $Ev - Co$ are parametrised in Hydro-ABC is described below.

2.2 Conservation properties of the continuous equations

95 Petrie et al. (2017) (their Appendix B) uses Gauss' divergence theorem, the model's vertical boundary conditions (their Table 2), and periodic lateral boundary conditions, to show that the dry equations conserve total mass, $\int \int dx dz \, \rho$, and total dry energy, $\int \int dx dz \, E_{\mathrm{dry}}$, where the integration is done over the whole domain (E_{dry} is defined in Sect. 2.2.2). Here we show that the Hydro-ABC equations (1) also conserve total water, and total moist energy.

2.2.1 Conservation of total water

100 Total water at a particular location, q_{t}, is the sum of the vapour and condensate mixing ratios, $q_{\mathrm{t}} = q + q_{\mathrm{c}}$. Adding (1g) and (1h) gives the governing equation for q_{t}: $\partial(\tilde{\rho} q_{\mathrm{t}})/\partial t = B \nabla \cdot (\tilde{\rho} q_{\mathrm{t}} \mathbf{u}) = 0$. Multiplying this equation by the constant reference density ρ_0, recognising that $\rho = \rho_0 \tilde{\rho}$, and integrating over the domain gives the following equation for the total amount of water in the domain:

$$\frac{\partial}{\partial t} \int \int \rho q_{\mathrm{t}} dx dz + B \int \int \nabla \cdot (\rho q_{\mathrm{t}} \mathbf{u}) \, dx dz = 0. \tag{2}$$

105 By following the same procedure to that in Petrie et al. (2017) (their Appendix B), it is straightforward to show that the second term is zero, thus demonstrating that the total amount of water in the domain is constant in time.

2.2.2 Conservation of total moist energy

From Petrie et al. (2017) the kinetic energy is $E_{\mathrm{k}} = \tilde{\rho}(u^2 + v^2 + w^2)/2$, the buoyancy energy is $E_{\mathrm{b}} = \tilde{\rho} b'^2/(2A^2)$, and the elastic energy is $E_{\mathrm{e}} = C\tilde{\rho}'^2/(2B)$. These make up the total dry energy, $E_{\mathrm{dry}} = E_{\mathrm{k}} + E_{\mathrm{b}} + E_{\mathrm{e}}$. In Hydro-ABC there is an

110 additional latent heat energy which we define as

$$E_{\mathrm{l}} = \tilde{\rho} L_{\mathrm{v}} q, \tag{3}$$

where L_{v} is the specific latent heat of vapourisation of water. In Hydro-ABC, the total moist energy is defined as the sum of the dry energy and latent heat, $E_{\mathrm{moist}} = E_{\mathrm{dry}} + E_{\mathrm{l}} = E_{\mathrm{k}} + E_{\mathrm{b}} + E_{\mathrm{e}} + E_{\mathrm{l}}$.

In its continuous form, Hydro-ABC is designed to conserve $\int \int dx dz \, E_{\mathrm{moist}}$. In order to show this, we note that Hydro-ABC

115 solves equations (1) over a timestep Δt in two parts: a 'dynamical core' in which $\mathcal{S}_{b'}$, Ev, and Co are each zero, followed by a micro-physics step. The dynamical core is equivalent to integrating the dry equations (with extra advection equations, for q

and q_c). The evolution of the total energy is represented by the following sum

$$\frac{\partial E_{\mathrm{moist}}}{\partial t} = \underbrace{\left(\frac{\partial E_{\mathrm{dry}}}{\partial t}\right)^{\mathrm{dc}} + \left(\frac{\partial E_{\mathrm{l}}}{\partial t}\right)^{\mathrm{dc}}}_{(\partial E_{\mathrm{moist}}/\partial t)^{\mathrm{dc}}} + \underbrace{\left(\frac{\partial E_{\mathrm{l}}}{\partial t}\right)^{\mathrm{mp}} + \left(\frac{\partial E_{\mathrm{b}}}{\partial t}\right)^{\mathrm{mp}}}_{(\partial E_{\mathrm{moist}}/\partial t)^{\mathrm{mp}}}, \tag{4}$$

where the large brackets indicate that the changes are performed over the dynamical core (dc) or the micro-physics (mp).

120   The first term represents point-wise changes of the dry energy by the dynamical core; this is the part that is shown in Petrie et al. (2017) to integrate to zero over the model's domain. The second term represents point-wise changes of the latent heat energy by the dynamical core, which is driven by transport of q. Although q (and hence E_l) can change at a particular point, the dynamical core does not change the phase of water, so this term also integrates to zero over the model's domain. The third and fourth terms are each non-zero, but the micro-physics scheme is designed so that their point-wise sum is zero (see Sect. 125   2.3). Hence integrating $\partial E_{\mathrm{moist}}/\partial t$ over the entire domain is zero. In Sect. 3.2 we examine how well such energy conservation is approximated in the time/space discretised equations.

2.3   The parametrisations of $\mathcal{S}_{b'}$, Ev and Co

As explained above, Hydro-ABC solves equations (1) over a timestep Δt by first assuming $\mathcal{S}_{b'}$ and $Ev - Co$ are each zero (the dynamical core), and then their effects are accounted for in a separate step (micro-physics). The conservation of energy, 130 $\Delta^{\mathrm{mp}} E_{\mathrm{moist}} = 0$ (where $\Delta^{\mathrm{mp}} \bullet = (\partial \bullet /\partial t)^{\mathrm{mp}} \Delta t$) point-wise over the micro-physics step is the basis for our micro-physics parametrisation, which is now described.

In the micro-physics, $\Delta^{\mathrm{mp}} E_l$ is related to the change in q, $\Delta^{\mathrm{mp}} q = (Ev - Co)\Delta t$. The rules for determining $Ev - Co$ and the corresponding change in the other model quantities are based on the constraint that $\Delta^{\mathrm{mp}} E_{\mathrm{moist}} = 0$, or put in an alternative way, $\Delta^{\mathrm{mp}} E_{\mathrm{dry}} = -\Delta^{\mathrm{mp}} E_l$. To facilitate closure, we assume that buoyancy energy is the only form of the dry energy that is 135   exchanged with the latent heat. This is a reasonable assumption because in ABC, buoyancy is a potential temperature-like variable. This means that $\Delta^{\mathrm{mp}} E_{\mathrm{dry}} = \Delta^{\mathrm{mp}} E_{\mathrm{b}}$, where the form of E_{b} is given in the opening paragraph of Sect. 2.2.2. This is why E_{b} appears in the fourth term of (4). Substituting the formulae for these energy forms into the conservation of energy requirement for the micro-physics yields

$$\frac{\tilde{\rho}}{2A^2}\Delta^{\mathrm{mp}}\left(b'^2\right) = -\tilde{\rho}L_{\mathrm{v}}(Ev - Co)\Delta t,$$

140    $$\text{or } \Delta^{\mathrm{mp}}\left(b'^2\right) = -2A^2 L_{\mathrm{v}}(Ev - Co)\Delta t. \tag{5}$$

Equation (5) will now be used with further rules (below) to determine how our micro-physics scheme will modify q and q_c (via $\Delta^{\mathrm{mp}} q = -\Delta^{\mathrm{mp}} q_c = (Ev - Co)\Delta t$), and how b' is modified (via $\Delta^{\mathrm{mp}} b' = \mathcal{S}_{b'}\Delta t$).
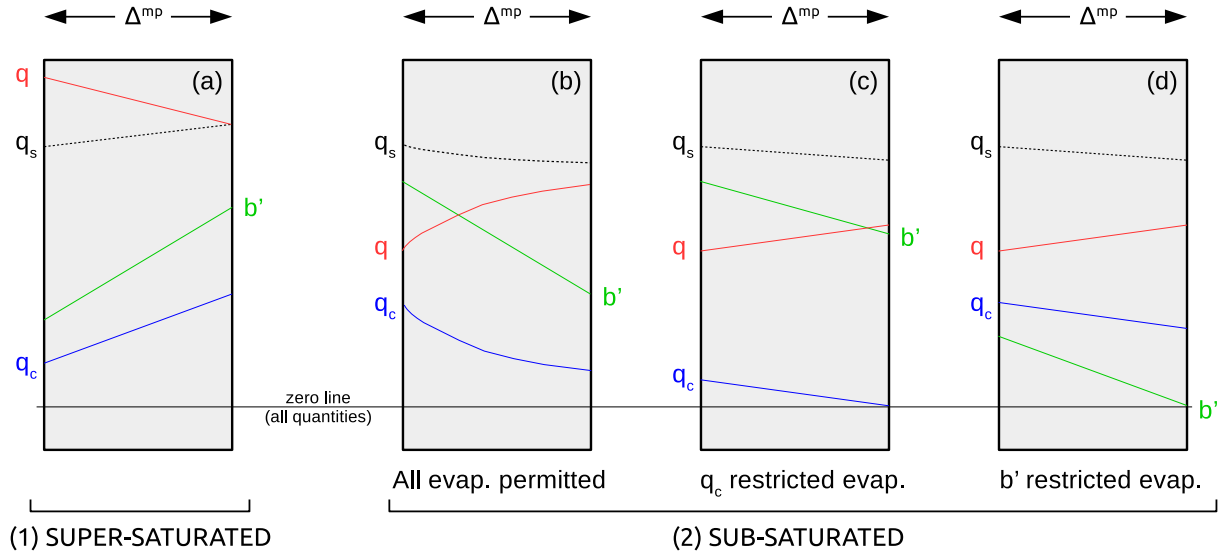
Figure 1. Graphical representation of Eq. (6), showing the change in q, q_{c}, and b' over a micro-physics step. Water vapour mixing ratio is shown in red, condensate is in blue, and buoyancy is in green. The water vapour mixing ratio according to saturation is the black dotted line. The zero line for all quantities spans all panels. Case (1) concerns condensation of water vapour due to super-saturation and case (2) concerns evaporation of condensate into vapour where there are three possibilities. Case (3) is not shown, as it represents no micro-physical changes to the fields.

Consider the following change in vapour over the micro-physics step, which depends on whether q is super- or sub-saturated after the dynamical core is run for one timestep:

$$
(Ev - Co)\,\Delta t =
\begin{cases}
-(q - q_{\mathrm{s}}) & \text{when } q \geq q_{\mathrm{s}} \text{ (super-saturated, applies for all } b') \\
& \text{and } \tilde{\rho} L_{\mathrm{v}} \left| q - q_{\mathrm{s}} \right| / E_{\mathrm{b}} > \Gamma \text{ (applies for } b' < 0 \text{ only)} \\
\min\left((q_{\mathrm{s}} - q)\left(1 - e^{-\Delta t/\tau}\right),\, q_{\mathrm{c}},\, {b'}^2/(2A^2 L_{\mathrm{v}}) \right) & \text{when } q < q_{\mathrm{s}} \text{ (sub-saturated) and } b' > 0 \\
0 & \text{otherwise,}
\end{cases}
$$

145

(6)

where q_{s} is the saturated mixing ratio (this is a function of the model state, see Sect. 2.4), Γ is a threshold for convection (see below), τ is the prescribed evaporation timescale, and the "min" function selects the smallest of its three arguments. This equation is represented in Fig. 1. The three cases of this equation are now discussed.

1. The first case concerns condensation of water vapour due to super-saturation (Fig. 1a). In this case q will instantaneously

150   decrease by the amount $q - q_{\mathrm{s}}$ (red line), and q_{c} will increase by the same amount (blue line). According to (5), associated with this step is an increase in ${b'}^2$ by the amount $\Delta^{\mathrm{mp}}\left({b'}^2\right) = 2A^2 L_{\mathrm{v}}(q - q_{\mathrm{s}})$. This is due to the release of latent heat, which is converted to buoyancy energy. Writing $\Delta^{\mathrm{mp}}\left({b'}^2\right)$ as ${b'}^2_{\mathrm{mp}} - {b'}^2$ (where b'_{mp} is the value of buoyancy after

micro-physics, and b' before), this equation translates to

$$b'_{\mathrm{mp}} = +\sqrt{b'^{2} + 2A^{2}L_{\mathrm{v}}(q - q_{\mathrm{s}})} \tag{7}$$

155   (green line). The positive square-root is assumed because b' is associated with potential temperature, which would rise after latent heat is released.

    (a) Note that q_{s} will be shown in Sect. 2.4 to be a function of $\tilde{\rho}'$ and b'. Changing b' as prescribed by (7) will therefore change q_{s}. Equation (7) therefore represents a non-linear equation. This is solved by finding the root of the function $f(q_{\mathrm{mp}}) = q_{\mathrm{mp}} - q_{\mathrm{s}}(\tilde{\rho}', b'_{\mathrm{mp}}(q_{\mathrm{mp}}))$ given that b'_{mp} is a function of q_{mp}, i.e. $b'_{\mathrm{mp}} = +\sqrt{b'^{2} + 2A^{2}L_{\mathrm{v}}(q - q_{\mathrm{mp}})}$,

160   and assuming that $\tilde{\rho}'$ is constant over the micro-physics step. When $b' \geq 0$ this is done using a Newton-Raphson procedure while assuming $\tilde{\rho}'$ is constant. The value of $(Ev - Co)\,\Delta t$ in (6) is then $-\left(q - q_{\mathrm{s}}(\tilde{\rho}', b'_{\mathrm{mp}}(q_{\mathrm{mp}}))\right)$.

    (b) This step, however, becomes potentially unphysical if b' is large and negative, and $q - q_{\mathrm{s}}$ is small, as this would flip the sign of b' to be positive, resulting is an unrealistically large increment in b'. For this reason, included in the first case is the additional condition that $\tilde{\rho}L_{\mathrm{v}}\,|q - q_{\mathrm{s}}|\,/E_{\mathrm{b}} > \Gamma$, where $\Gamma \gg 1$. This is the condition (which must

165   be met only when $b' < 0$) that, for condensation to happen, the release of latent heat must be much larger than the amount of buoyancy energy. The value of $\Gamma = 10$ is chosen for the results in this paper. Also, for simplicity, when $b' < 0$, the amount of latent heat released is based on the value of q_{s} at the pre-micro-physics value of b', and not on the value of q_{s} at the value of increased value of b' due to the latent heat release. This means that the non-linear equation in (a) above is not used in this case.

170   2. The second case concerns evaporation of condensate into vapour, where the q is assumed to relax exponentially upwards towards q_{s} with timescale τ (Fig. 1b, red line), and q_{c} relaxes downwards (blue line). According to (5), associated with this step is an change in b'^{2} by the amount $\Delta^{\mathrm{mp}}\left(b'^{2}\right) = -2A^{2}L_{\mathrm{v}}\left(Ev - Co\right)\Delta t$ (where $Ev - Co > 0$). This translates to a new value b'_{mp} according to the following:

$$b'_{\mathrm{mp}} = +\sqrt{b'^{2} - 2A^{2}L_{\mathrm{v}}\left(Ev - Co\right)\Delta t}, \tag{8}$$

175   where the positive square-root is taken, which corresponds to a reduction in b' due to evaporation (green line). This process is restricted to when $b' > 0$, because in the case of $b' < 0$, (8) dictates that b' will increase, even when the negative square-root is taken, which is unphysical.

  This process assumes that there is an adequate amount of condensate to evaporate, and an adequate amount of buoyancy energy to convert to latent heat during evaporation (when the first argument of the "min" function in (6) is the smallest).

180   The evaporation process is limited should either of these requirements not be met, which correspond to the second and third arguments of the "min" function. When the second argument is smallest, evaporation is limited by available q_{c} (panel c, where the micro-physics step evaporates all available q_{c}), and when the third argument is smallest, evaporation is limited by available buoyancy energy (panel d, where the micro-physics step converts all available E_{b}). As evaporation is much slower than condensation, $(Ev - Co)\,\Delta t$ (case 2 in (6)) is based on the value of q_{s} before evaporation lowers

185   the value of b' (and hence lowers q_{s}). This avoids the need to solve the non-linear equations mentioned previously.

3. The third case is invoked for the residual cases (i.e. for super-saturation but when the threshold for condensation is not met, or for sub-saturation but when $b' < 0$). In this case there are no micro-physics-related changes to the fields.

The change in buoyancy due to micro-physics, $b'_{\mathrm{mp}} - b'$, yields the value $\mathcal{S}_{b'}\Delta t$ in (1e), and the change in vapour, $q_{\mathrm{mp}} - q$, gives the value $(Ev - Co)\Delta t$ in (1g) and (1h).

2.4 The saturated vapour mixing ratio

The saturated mixing ratio, q_{s}, specifies the upper limit for water vapour. In standard atmospheric models it is a function of pressure and temperature. We use the following formula:

$$q_{\mathrm{s}} = \frac{380000}{p} \exp\left(\frac{17.3(T - 273.2)}{T - 35.9}\right), \tag{9}$$

which is based on Eq. (9.8) of Pielke (2002) (adapted so that pressure is in Pa, and mixing ratio is in gkg^{-1}). In (9), p and T (and hence q_{s}) have an implicit dependence on x, z, and t. Neither p nor T are ABC variables, and so we now take steps to produce physically reasonable formulae for p and T (needed for (9)) from ABC variables.

For pressure, a reference density profile, $\rho_0(z)$, is required. In the ABC model (Petrie et al., 2017), ρ_0 is a constant, but for the purposes of computing q_{s} we let this depend on z according to $\rho_0(z) = \rho_{00}\exp(-z/H)$, where ρ_{00} is the surface air density ($1.225\mathrm{kgm}^{-3}$) and H is the scale height (9km). For pressure, $p = p_0 + p'$, where p_0 (the reference pressure) is in hydrostatic balance with ρ_0: $dp_0(z)/dz = -\rho_0(z)g$, where g is the acceleration due to gravity. The solution to this equation, given $\rho_0(z)$ is $p_0(z) = p_{00} - H\rho_{00}g\left(1 - \exp(-z/H)\right)$, where p_{00} is the reference surface pressure. The relationship between surface reference pressure and surface reference density is $p_{00} = H\rho_{00}g$, which is derived from the previous equation with the condition $p_0(z \to \infty) \to 0$. ABC's equation of state (1f) relates the ABC variable $\tilde{\rho}'$ to p', and using the above form of $p_o(z)$ gives

$$p(x,z,t) = p_0(z) + p'(x,z,t) = p_{00} - H\rho_{00}g\left(1 - \exp(-z/H)\right) + C\rho_{00}\exp(-z/H)\tilde{\rho}'(x,z,t). \tag{10}$$

In Petrie et al. (2017), b' is related to a potential temperature increment θ' via $\theta' = (\theta_{\mathrm{R}}/g)b'$, where θ_{R} is a constant (value 273K)[1]. Hence we focus now on potential temperature. The potential temperature in Petrie et al. (2017) is defined as $\theta(x,z,t) = \theta_{\mathrm{R}} + \theta_0(z) + \theta'(x,z,t)$. To derive a reference profile, $\theta_0(z)$, we use the definition of the Brunt-Väisälä frequency $N^2 = (g/\theta_{\mathrm{R}})d\theta_0(z)/dz$ (see e.g. Holton and Hakim (2013)). The parameter A in (1e) takes the place of N, and so we solve the equation $A^2 = (g/\theta_{\mathrm{R}})d\theta_0(z)/dz$. The solution to this equation is $\theta_0(z) = \theta_{00} - \theta_{\mathrm{R}} + (A^2\theta_{\mathrm{R}}/g)z$, where θ_{00} is the surface potential temperature (300K). The potential temperature is therefore

$$\theta(x,z,t) = \theta_{\mathrm{R}} + \theta_0(z) + \theta'(x,z,t) = \theta_{00} + \frac{\theta_{\mathrm{R}}}{g}\left(A^2z + b'(x,z,t)\right). \tag{11}$$

---

[1]This relationship between b' and θ' does not imply that b' has the full properties of potential temperature in the ABC model (e.g. that b' is conserved during adiabatic processes). In Petrie et al. (2017) this relationship was used as a starting point in the derivation of the ABC equations from the Euler equations and it is re-used here as a pragmatic way of deriving temperature from the ABC variables.

The standard relationship for T in terms of θ and p is $T(x,z,t) = \theta(x,z,t)\,(p(x,z,t)/p_{00})^{\kappa}$ (Holton and Hakim, 2013), where κ is the constant R/c_{p} (where R is the gas constant for dry air and c_{p} is the specific heat capacity at constant pressure; value of κ is 0.286). Using (11) and (10), gives

$$T(x,z,t) = \left[\theta_{00} + \frac{\theta_{\mathrm{R}}}{g}\left(A^2 z + b'(x,z,t)\right)\right]\left[\frac{p_{00} - H\rho_{00}g\left(1 - \exp(-z/H)\right) + C\rho_{00}\exp(-z/H)\tilde{\rho}'(x,z,t)}{p_{00}}\right]^{\kappa}. \qquad (12)$$

Equations (10) and (12) can now be substituted into (9) to yield a physically reasonable function for the saturated vapour mixing ratio in terms of ABC variables b' and $\tilde{\rho}'$.

2.5 Numerical details

The ABC model stores variables on an Arakawa C grid in the horizontal and a Charney-Phillips grid in the vertical (see Petrie et al. (2017), Fig. 1). The two extra variables in Hydro-ABC, q and q_{c}, are each stored at the same location as $\tilde{\rho}$ and $\tilde{\rho}'$. This choice makes sense from a numerical perspective as each of q and q_{c} is multiplied by $\tilde{\rho}$ in (1g) and (1h), and most of the finite differences that approximate the derivatives in these equations evaluate at the correct locations on the grid.

In Hydro-ABC, (1g) and (1h) are solved as advection/source equations rather than as the continuity equations shown. By employing the product rule for differentiation, (1g) becomes $\tilde{\rho}\partial q/\partial t + q\partial\tilde{\rho}/\partial t + Bq\nabla\cdot(\tilde{\rho}\mathbf{u}) + B\tilde{\rho}\mathbf{u}\cdot\nabla q = \tilde{\rho}(Ev - Co)$. The second and third terms sum to zero due to the mass continuity equation (1d), leaving the advection/source equation for q:

$$\frac{\partial q}{\partial t} + B\mathbf{u}\cdot\nabla q = Ev - Co. \qquad (13)$$

A similar equation is also found for q_{c}. The advections of q and q_{c} are solved by means of the same upwind scheme used for other variables in ABC (see Petrie et al. (2017), Sect. 3.3), with $Ev - Co = 0$ for the dynamical core ($Ev - Co$ is subsequently found with the micro-physics parametrisation of Sect. 2.3).

3 Example run of the model

In this section we show example behaviour of Hydro-ABC. The initial conditions for the dry variables (u, v, w, $\tilde{\rho}'$, and b') are set according to the prescription in Petrie et al. (2017) (their Sect. 5.2). This involves extracting u and v from a longitude/height slice output from a 1.5km Unified Model (UM) run, which has the same grid staggering as the ABC model, and then adjusting u and v to have smooth period boundary conditions. Then $\tilde{\rho}'$ is set using geostrophic balance, b' is set using hydrostatic balance, and w is set by imposing a 3-D divergence-free condition. The initial water vapour is set as shown in the top panel of Fig. 2 (set so the maximum relative humidity is 0.98), and the initial $q_{\mathrm{c}} = 0$.

3.1 Reference run

The rest of Fig. 2 (rows 2, 3, and 4) shows the evolution of w (left column) and q, q_{c} (right column) over the first six hours of the integration using the reference parameters indicated in the caption. Before the first hour of the simulation has completed there are no obvious signs of convection (not shown). It is only at $t = 60$ min that localised upward motion starts to develop
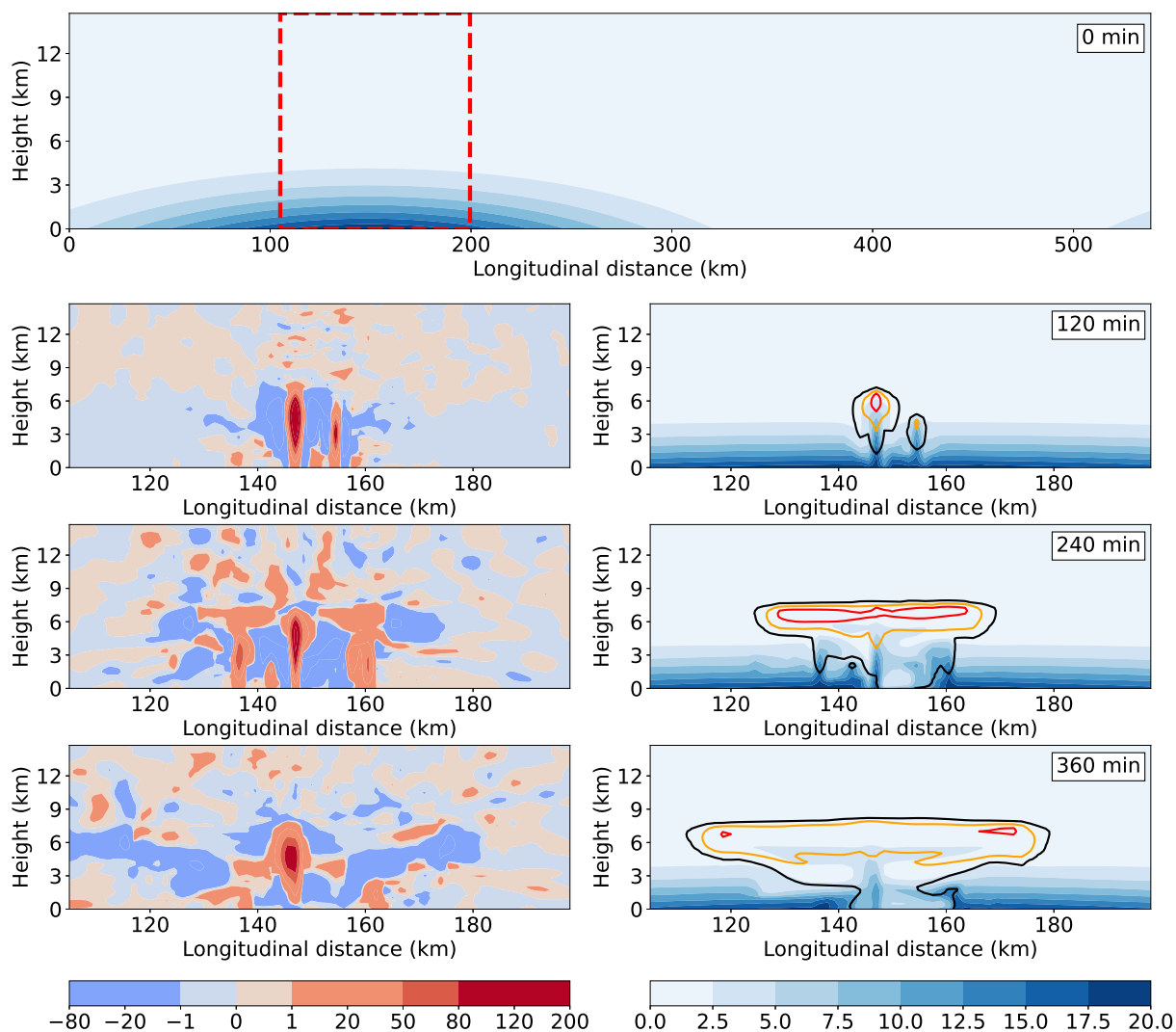
Figure 2. Initial conditions of the example reference run of Hydro-ABC ($A = 0.01\mathrm{s}^{-1}$, $B = 0.05$, $C = 2 \times 10^4 \mathrm{m}^2\mathrm{s}^{-2}$, $f = 10^{-4}\mathrm{s}^{-1}$, $\Delta t = 0.1\mathrm{s}$, $\tau = 1000\mathrm{s}$, $L_\mathrm{v} = 2500\mathrm{Jg}^{-1}$, and $\Gamma = 10$). The initial conditions of the dry variables are set as described in the text, the initial values of q are shown in the top panel (set so the maximum relative humidity is 98%), and the initial $q_\mathrm{c} = 0$. Snapshots of the subsequent integration of Hydro-ABC are shown in the remaining panels, zoomed into the region indicated by the red box in the top panel for: 120, 240, and 360 min. w is shown in the left column (ms^{-1}), and q (shading) and q_c (contours) are shown in the right column. The contours are the black ($1\mathrm{gkg}^{-1}$), orange ($6\mathrm{gkg}^{-1}$) and red ($12\mathrm{gkg}^{-1}$) lines.

at ground level most strongly at $x = 148.5$km, but also at 156km(also not shown). By 120 min, these have developed into striking convective plumes, which transport vapour upwards. This coincides with the position of the water vapour maximum in the initial conditions. In this region there is significant conversion of vapour to condensate, thus releasing latent heat, which is converted to buoyancy energy, thus inducing strong upward motion. The magnitude of the plume's vertical motion is very high (peaking at ~200ms^{-1} at the strongest convecting point), but in the ABC model this speed is moderated by the small value of B (0.05), so it is equivalent to a peak vertical mass transport of 10ms^{-1}, which is realistic. The vertical grid spacing is ~ 250m, making the vertical Courant number 0.08 for the unmodulated motion and 0.004 allowing for the B-modulation. These values are well within the Courant–Friedrichs–Lewy condition for the Courant number to be less than unity.

By 240 min, an anvil cloud of vapour and condensate has established itself between heights of ~ 5.5 to 8.5km, which spreads out laterally in both directions. The zonal winds at these levels peak at around ± 100ms^{-1} (not shown), but again once B moderation is taken into account this is equivalent to a horizontal mass transport of ± 5ms^{-1}. At this time, the second smaller plume has been extinguished by downdrafts from the main plume, but smaller convective plumes 11 to 13km have developed either side of the main plume. Between 240 and 360 min, the smaller plumes are themselves impeded by down welling air on each side of the main plume. By 360 min the condensate anvils have continued to develop laterally, and the central plume starts the process of weakening.

3.2 Numerical effects on the total energy evolution

Hydro-ABC in its continuous form, as described by (1) and the algorithm in Sect. 2.3, will conserve total moist energy, $\int \int dx dz\, E_{\mathrm{moist}}(t) = \mathrm{constant}$. Discretisation of the equations in space and time will though lead to loss of energy conservation. Five numerical experiments – based on the reference parameters, but of different Δt from 0.1 to 1.0s – are performed to analyse how in practice the total moist energy deviates from conservation. Figure 3 shows the evolution of relative total moist energy (relative to the total moist energy at $t = 0$), where the labels are the Δt values of the model runs. All runs show fluctuations in this quantity, which become visible on the plot's scale when convection starts just after $t = 60$ min. As expected $\Delta t = 0.1$s deviates from conservation the least, and $\Delta t = 1.0$s the most. Our assumption is that the $\Delta t = 0.1$s setting represents an acceptable deviation from conservation, especially given fast-moving nature of the features that are being simulated. This is the setting that has been used for the result in Fig. 2, and will be used for the remainder of this paper.

3.3 Excitation of waves by convection

The ABC model is capable of supporting Rossby-like, inertio-gravity and acoustic waves. The continuous equations can support all horizontal wavenumbers allowed by the periodic boundary conditions, and all frequencies. In practice though the maximum wavenumber and frequency are set by Δx (the grid box length) and Δt (the timestep) respectively. In runs of the dry model, high wavenumbers and frequencies have a very small weight, but their presence is still important in order to represent scales that would otherwise anomalously dissipate energy. Allowing moist convection, as facilitated by Hydro-ABC, potentially excites appreciable weight to modes of high wavenumber and frequency.
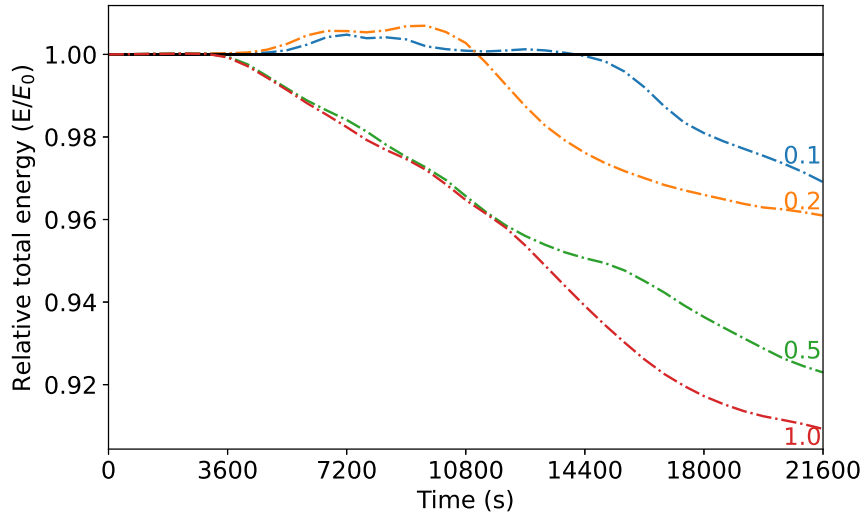
Figure 3. Evolution of the relative total energy of runs of hydro-ABC for different model integration timesteps as indicated (in s). The relative total energy is $\left[\int\int dxdz\, E_{\text{moist}}(t) \right] / \left[\int\int dxdz\, E_{\text{moist}}(0) \right]$, where $E_{\text{moist}} = E_{\text{k}} + E_{\text{b}} + E_{\text{e}} + E_{\text{l}}$ (Sect. 2.2.2).

Figure 4 shows w for level 30 (near the top of the anvil cloud) of two model runs: one for the dry model (left) and another for
275 Hydro-ABC (right). The top panels are Hovmöller plots of w for the longitude/time domain. Both versions of the model show westward and eastward propagating waves. The bottom panels of Fig. 4 show the spectral density of w at the same height. These plots are simply the magnitude bi-Fourier transform of the top panels. This reveals the dispersion relations of the dominant waves in the system. By analogy with waves in systems that represent more closely the real atmosphere, each feature on these plots represents a wave with a different vertical wavenumber (Salby (1996), his Fig. 14.9) and the lines separate into two
280 groups. The first group is the lower frequency inertio-gravity waves, whose frequencies increase with horizontal wavenumber and then level off at $A/(2\pi) = 1.59 \times 10^{-3}\text{s}^{-1}$. This is the wave frequency of the highest frequency inertio-gravity waves in the high wavenumber limit (the horizontal dashed lines) which is consistent with classical wave theory, e.g. Gill (1982); Salby (1996). The second group is the higher frequency acoustic waves, whose frequencies increase with horizontal wavenumber. The quantity $\sqrt{BC}$ is the pure sound wave speed (the gradient of the sloping dashed lines). These line should separate the inertio-
285 gravity waves and the acoustic waves and is occupied by Lamb waves in the real atmosphere (there is no visible evidence of Lamb waves in these (Hydro)-ABC runs, but we will still call this line the "Lamb line"). In the (Hydro)-ABC runs some of the low-wavenumber inertio-gravity waves have frequencies higher than the Lamb line, which may be due to numerical effects.

Comparing the dry ABC and Hydro-ABC spectra, the shapes of the dispersion curves look similar. This indicates that adding the adiabatic moisture scheme is not noticeably modifying the underlying dispersion relations of the dynamical core. Instead,
290 the striking effects that moist convection in Hydro-ABC causes are (i) the excitation of stronger inertio-gravity and acoustic modes across the spectrum, and (ii) broadening of their frequencies. Effect (i) is visible at all wavenumbers and frequencies,
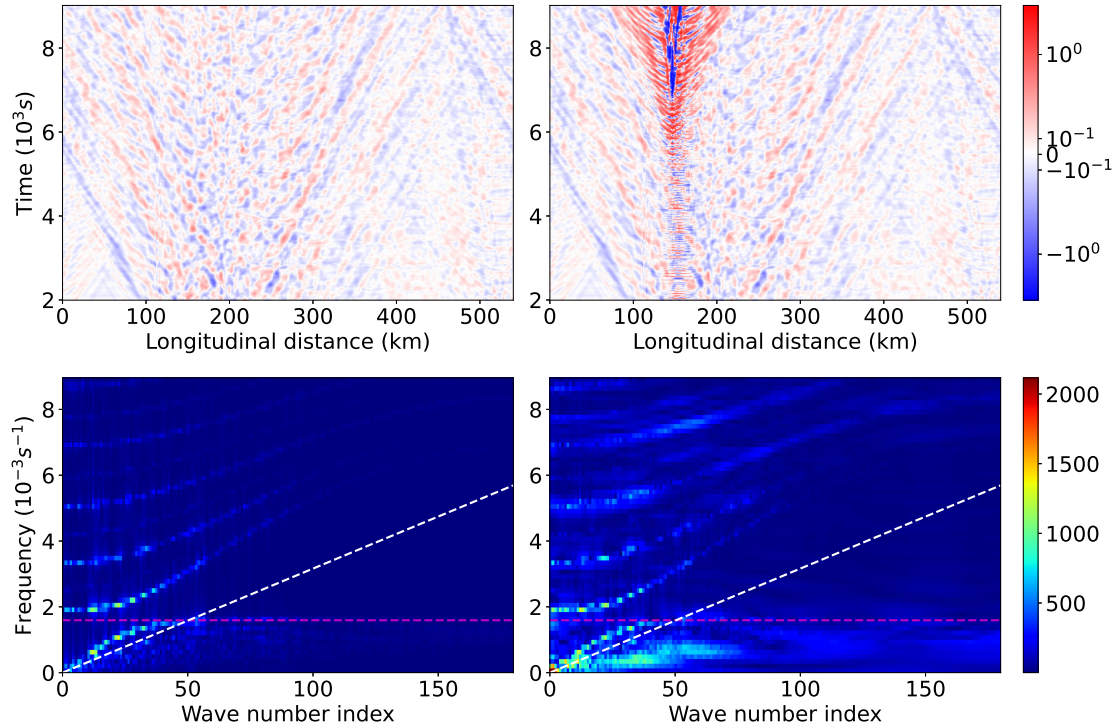
Figure 4. Vertical wind in the longitude/time (top panels) and wavenumber/frequency (bottom panels) domains for level 30 in the model (7.7km height). The left panels are for dry ABC and the right panels are for hydro-ABC. The initial conditions of u, v, w, $\tilde{\rho}'$, and b' are identical for ABC and Hydro-ABC, and the additional initial conditions for q and q_c are the same as those used in Fig. 2. In the bottom panels, the horizontal dashed lines are at the frequency $A/(2\pi)$ (the pure gravity wave frequency), and the sloped dashed lines have gradient of $df/dk = \sqrt{BC}$ (the pure sound wave speed; note that k is wavenumber and the slope with respect to the wavenumber index, n, is $df/dn = L^{-1}df/dk$, where L is the length of the domain, 540×10^3m).

especially for inertio-gravity waves for wavenumbers less than ~ 70. Effect (ii) is likely due to the gradual onset of convection, where the degree of broadening is inversely proportional to the timescale of the onset of convection. In the top-right panel of Fig. 4, there is a gradual onset of the strength of convection, which leads to a convolution of the spectra with this 'low frequency' onset.

295

4 Convection-dependent covariances and identification of areas prone to convection

It is well-known that latent heat release can lead to convective instabilities. Such rapidly changing conditions present challenges for DA, not least because the background error covariances can become highly flow-dependent, e.g. Lorenc (2007); Montmerle and Berre (2010); Sun et al. (2014); Bannister et al. (2020). In this section we show how vertical error statistics of

300   model variables depend on the background conditions for hydrodynamically stable and unstable regions. An important issue in convective-scale DA is the identification of convective regions, so that the (otherwise static) background error covariance matrix can be modified with the relevant vertical covariance statistics should the model be used with data assimilation.

4.1   How forecast statistics are affected by convectively-prone regions

In order to give a taste of moisture-affected covariances and correlations between different variables and positions, and in
305   different regimes in hydro-ABC, we generate an ensemble of hydro-ABC forecasts. Longitude/height slices from a 93-member ensemble of convective-scale UM forecasts (Bannister et al., 2017) are each processed in the same way as described in Sect. 3 to give a 93-member ensemble of hydro-ABC forecasts (run out to 360 min). Each ensemble member is prepared with the same q and q_c fields (as in Fig. 2), but the relative humidity (q/q_s, the field that matters for convection) will be different between members because q_s is a function of b' and $\tilde{\rho}'$ (see Sect. 2.4). The example hydro-ABC run in Sect. 3 is one of these members,
310   and convection happens around the same place in all but one member allowing the vertical statistics (covariances between different vertical levels) to be based on slightly different versions of the convective situation, which is centred at $x = 148.5\mathrm{km}$ as in Fig. 2.

    The covariances of and between different model variables will have different units and magnitudes. In order to simplify comparison of covariances, we first normalise each model variable by its (constant) maximum standard deviation, where the
315   maximum is calculated over all positions and times. This will result in covariances (referred to as normalised covariances) that are bounded between -1 and 1. These values are not correlations though as correlations are based on normalisation by the local standard deviations (local in space and time), although we do later also show correlations of different quantities.

    Figure 5 shows the normalised covariances for a selection of quantities (see Fig. caption) at different times in the convection zone ($x = 148.5\mathrm{km}$). Near the start of the integrations ($t = 40\mathrm{min}$), all covariances shown are too small to be detected on
320   the scale used. At 80 min, convection is just starting, which appears in a region of non-zero covariances within $4.5\mathrm{km}$ of the surface. There are positive covariances between w and itself at different levels, and with q (and more weakly with q_c), and split negative/positive covariances with u. These results show that larger w is associated with greater vertical advection of vapour, and the formation of a shear zone in zonal wind, with westward wind in the lower part and eastward wind in the upper part of this $4.5\mathrm{km}$ layer. By 120 min the non-zero covariances region has grown to within $9\text{-}10\mathrm{km}$ of the surface, including
325   growth of the wind shear zone. An increase in the vertical length-scales in q errors is also seen in the WRF-based study of Michel et al. (2011) and in the AROME-based study of Montmerle and Berre (2010), in each case when the underlying conditions are raining (which is assumed to be comparable to the convective situation here where the condensate would fall as rain if precipitation processes were included). Additionally, q_c has developed strong positive covariances with w in this region indicating that condensate has formed, with larger amounts with increasing w. As indicated in Fig. 2 (contours in right panels),
330   the condensate concentrations increase with height up to $\sim 9\mathrm{km}$. This suggests that the region of positive q_c-w covariances just mentioned is not caused by advection (the covariances would have negative sign if they were), but instead by the process of more condensate being produced in association with stronger w. At 160 min a shallow surface layer of negative q_c-w covariances has developed, which may be due to a circulation being induced by the convection, which is drawing in air with
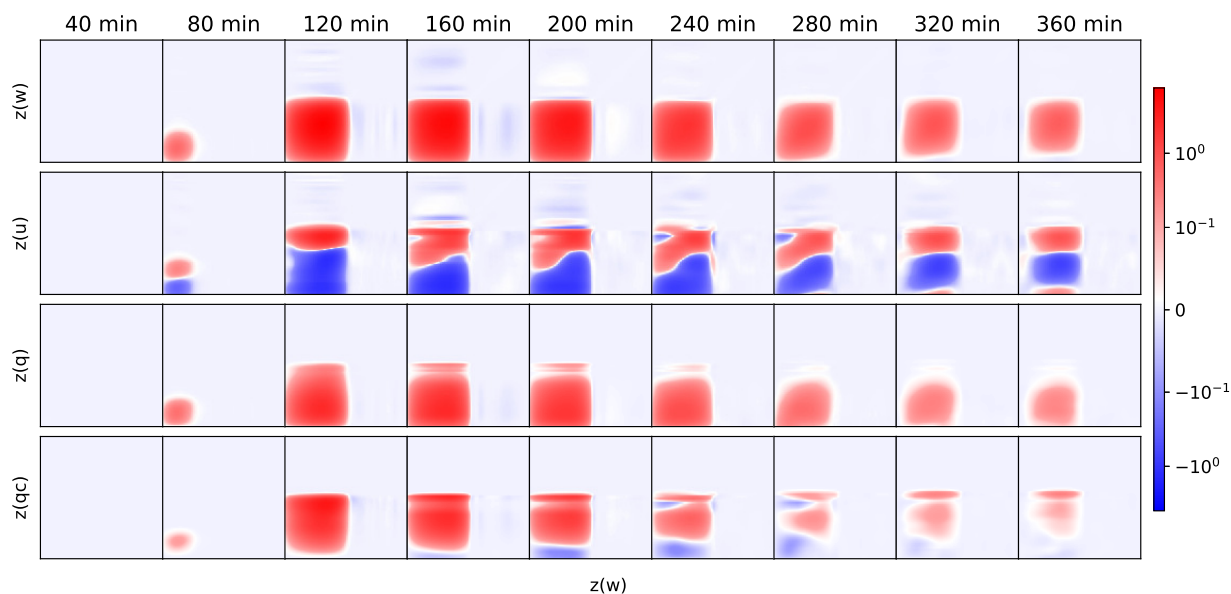
Figure 5. Evolution of a selection of normalised vertical covariances in the convection zone ($x = 148.5$km) at times between 40 and 360 minutes. The rows are for w-w, u-w, q-w, and q_c-w covariances. In each plot of p_1-p_2 covariances, the y-axis is for p_1 and the x-axis is for p_2, and the surface is at the bottom left of each square. Red (blue) indicates positive (negative) covariances according to the logarithmic scale shown.

lower q_c from the surroundings. There is an additional region of weak, but non-zero, w-w and u-w correlations above the top of the convection, the latter indicating a further (but weak and shallow) shear region at the higher levels, which continues to 200 min. From 240 min the covariances develop more complicated, but weaker, patterns, including the development of a further possible u shear layer near the surface. There is also an influence of the convection in vertical covariances at locations other than at $x = 148.5$km, including at locations within and beyond the moist 'bubble' set-up in the initial conditions (covariances not shown, but correlations are shown below).

340   Figure 6 shows vertical correlations between a different selection of quantities. The top three rows show auto-correlations of $\tilde{\rho}'$ for the convection, moist, and dry zones respectively (see Fig. caption for their lateral positions). At $t = 40$min the different zones show similar patterns. The model variable $\tilde{\rho}'$ appears to comprise two layers (one below 6km, and the other above), where in each $\tilde{\rho}'$ is strongly correlated vertically, but only weakly (and negative) across each layer. In the convection zone there are some small-scale vertical oscillations, which are not present in the other zones shown and so are likely to be due to early 345   signs of convection. At 80 min in the convection zone, the lower layer level itself splits into two negatively correlated sub-layers, which are associated with the establishment of upper and lower pressure changes formed in connection with the u-w covariances in Fig. 5 at this time (the wind shearing). The covariances in the moist and dry zones have not changed qualitatively very much at this time. At 120 min and 160 min in the convection zone the sub-layering has disappeared (probably because the shear zone has moved upwards), and the vertical lengthscales above 6km have dramatically reduced. This shortening of the
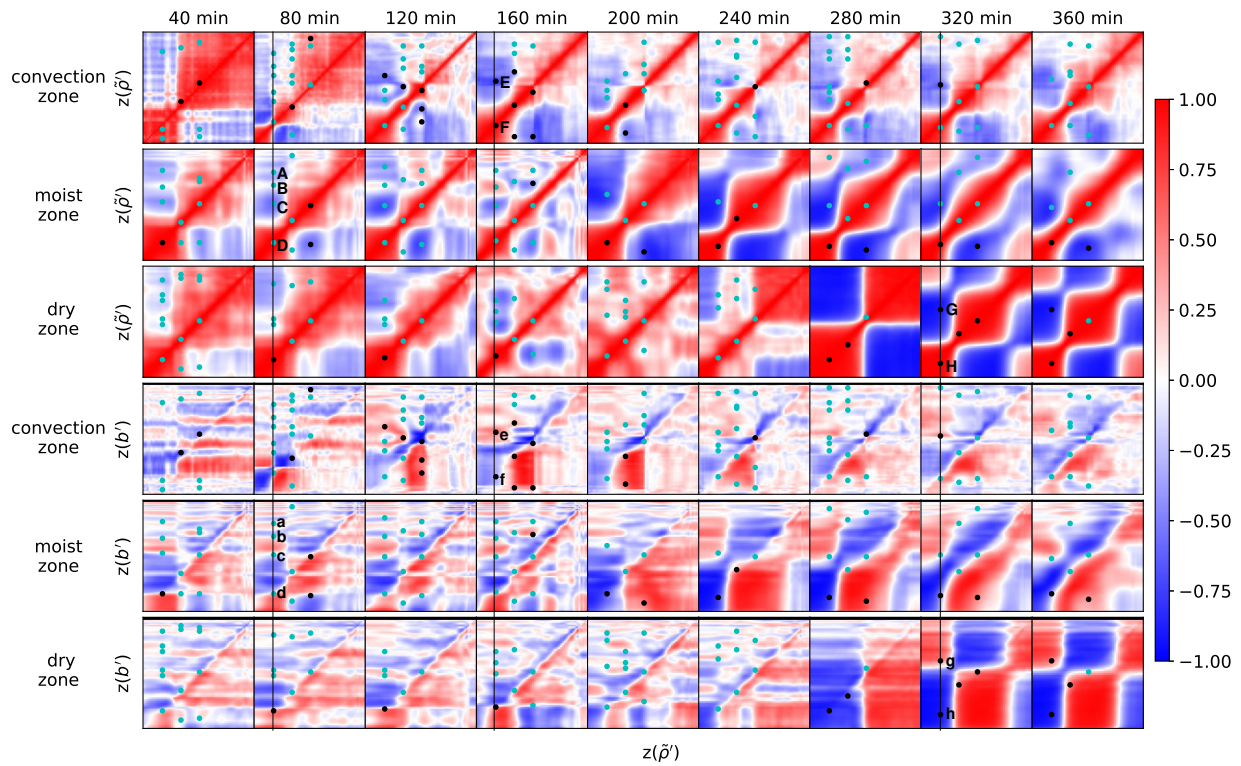
Figure 6. Evolution of a selection of vertical correlations at times between 40 and 360 minutes. The top set of three rows is for $\tilde{\rho}'$-$\tilde{\rho}'$ correlations, and the bottom set is for b'-$\tilde{\rho}'$ correlations. For each set of three rows, the rows are in the convection zone ($x = 148.5\mathrm{km}$ – the same position as used for Fig. 5), the moist zone (286.5km), and the dry zone (421.5km) respectively. In each plot of p_1-p_2 correlations, the y-axis is for p_1, the x-axis is for p_2, and the surface is at the bottom left of each square. Red (blue) indicates positive (negative) correlations. The dots represent points where consistency with hydrostatic balance is tested. These are placed at a selection of heights in the top three rows where there is a local maximum or minimum in $\tilde{\rho}'$-$\tilde{\rho}'$ correlation with respect to the y-axis. Each dot is also plotted in the corresponding bottom three rows to see if it associated with a 'small' b'-$\tilde{\rho}'$ correlation. Cyan dots are associated with small b'-$\tilde{\rho}'$ correlation (≤ 0.3, associated with hydrostatic balance) and black dots otherwise.

350 vertical lengthscales is also seen in the moist and dry zones, but with a delay of tens of minutes. From 200 min onwards the vertical covariances in the convection zone do not change very much in structure, but the other zones develop strong large-scale (vertically) correlation structures presumably due to outward propagation of disturbances from the convection zone.

The bottom three rows of Fig. 6 show vertical correlations between b' at difference heights (y-axis) and $\tilde{\rho}'$ at different heights (x-axis). These are included in combination with the $\tilde{\rho}'$ auto-correlations to look for signatures of hydrostatic balance (HB). The
355 HB equation is found from (1c) after the Lagrangian derivative terms are neglected, resulting in $C\partial\tilde{\rho}'/\partial z = b'$. This relates b' to vertical gradients in $\tilde{\rho}'$. This property is seen qualitatively in most of the correlation functions in Fig. 6. A sample of points is selected where a signature of HB is tested. In the top three rows dots are placed where there are stationary points in the vertical correlation $(\partial(\tilde{\rho}'\text{-}\tilde{\rho}'\text{ correlation})/\partial z(\tilde{\rho}') = 0$. The corresponding dots are placed at the same positions in the bottom three panels. Each dot is coloured cyan if it has a small b'-$\tilde{\rho}'$ correlation (≤ 0.3), which are associated with HB, or black otherwise.
360 At 80min for instance in the moist zone, there are four stationary points highlighted (A, B, C, and D). These correspond to near zero values of b'-$\tilde{\rho}'$ correlations (a, b, c, and d). There are only a few points in the moist and dry zones in the early stages of the evolution that do not satisfy this condition (black dots). At 160min in the convection zone, there are two stationary points highlighted (E and F). These do not correspond to near zero values of b'-$\tilde{\rho}'$ correlations (e and f), suggesting a clear breakdown of HB. In the first half of the forecast period, there are more non-hydrostatic signatures in the convection zone than in the other
365 zones. Later in the convection zone, the hydrostatic signatures return as the system stabilises. In the second half of the forecast period, the moist and then dry zones experience more non-hydrostatic signatures as a result of the outward propagation of disturbances from the convection zone. At 320min in the dry zone for instance, there are two stationary points highlighted (G and H). These do not correspond to near zero values of b'-$\tilde{\rho}'$ correlations (g and h), suggesting a propagated breakdown of HB. Bannister et al. (2011) and Vetra-Carvalho et al. (2012) similarly found evidence of non-hydrostatic patterns in the covariance
370 patterns in rainy regions (compared to dry regions) of forecast errors of the convective-scale UM. These statistics encourage the development of data assimilation systems that do not always constrain the background error covariances to be hydrostatic when convection is happening. Hydrostatic imbalance is further mentioned in Sect. 4.2.

4.2 Possible indicators and harbingers of convection

Since there are significant differences in the covariance and correlation structures between convecting and non-convecting
375 regions, convective-scale data assimilation systems should gain advantages by being aware of the convection 'status' of each geographical region, and use appropriate background error statistics for each. This is done automatically in ensemble-based systems (e.g. ensemble Kalman filters or ensemble-variational methods, (Bannister, 2017)), and partially in hybrid systems (where the static background error covariances of traditional variational methods are averaged with ensemble-derived covariances, ibid.). It is well known though that the use of ensemble-based covariances leads to rank deficiency and under-sampling
380 (Houtekamer and Mitchell, 1998; van Leeuwen, 1999; Hamill et al., 2001; Houtekamer and Mitchell, 2005; Ehrendorfer, 2007; Zhang and Zhang, 2012; Houtekamer and Zhang, 2016), and so any progress that allows the (otherwise) static background error covariance matrix to gain flow dependence is still valuable. This is the strategy behind a number of studies, including Montmerle and Berre (2010); Ménétrier and Montmerle (2011); Michel et al. (2011); Montmerle (2012); Yang et al. (2022), although

these studies used the strength of precipitation or the presence of fog (rather than convection) as a criterion for selection of the
385   most appropriate covariance statistics. Here, we explore ways to determine locations of convecting and non-convecting flows.
Our strategy is to determine these regions before convection has fully developed, so that a data assimilation system that uses
the right covariances with new observations does not lead to an analysis state that will impede the development of convection
in the ensuing forecast.

Three contenders (and associated quantities) are considered here to help in the discrimination: relative humidity (RH),
390   hydrostatic imbalance (HI), and convective available potential energy (CAPE). These quantities are described below, and
plotted in Fig. 7.

1. In this work the RH at any horizontal location is defined as the maximum value of q/q_s in the vertical column, where
q_s is the saturated specific humidity defined in Sect. 2.4. Large values of RH are expected during and just before moist
convection.

395   2. The HI at any horizontal location is defined as the largest absolute value of $(C\partial\tilde{\rho}'/\partial z - b')/(\mathrm{RMS}(C\partial\tilde{\rho}'/\partial z) + \mathrm{RMS}(b'))$
in the vertical column, where RMS is the root-mean-square calculated over the whole domain. A value of zero repre-
sents perfect hydrostatic balance, so convection is associated with non-zero HI. Associated with the instability of HI is
horizontal wind divergence, defined as the maximum value of $\partial u/\partial x$ in the column (recall there is no y-dependence in
ABC).

400   3. Convective Available Potential Energy (CAPE) is a vertically integrated quantity, which can be used to help determine if
an atmospheric column is 'primed' for convection. Appendix A shows how CAPE can be derived from model quantities,
as well as the converse quantity convective inhibition (CIN).

Figure 7 shows how these quantities vary with horizontal position in the domain at $t = 10\mathrm{min}$, which is before convection has
developed. The RH (green) in this case is a smooth function of position, peaking at unity around the convection zone. The
405   HI (magenta) peaks at the same point, but is sharper than the RH. The horizontal divergence (cyan) is a noisy field which has
generally higher values close to the convection (although the value drops at the specific convection point). The CAPE (red)
has increases in value towards the convection zone, but is not a smooth function of position (and drops slightly at the specific
convection point). The CAPE drops abruptly at $\sim 0\mathrm{km}$ and $\sim 285\mathrm{km}$, even though these points are within the region where
$q > 0$ in the initial conditions. Reassuringly, the CIN, also changes abruptly at these positions, but in the opposite way: the CIN
410   is small where the CAPE is large and vice-versa. Further work will be required to test the most reliable indicator of convection
for background error statistics selection purposes in data assimilation.

5   Conclusions

Described in this paper is how the simplified ABC model of convective-scale flow has been extended to include moist processes,
namely evaporation and condensation, yielding a revised model called Hydro-ABC. This is the next step in the long-term aim
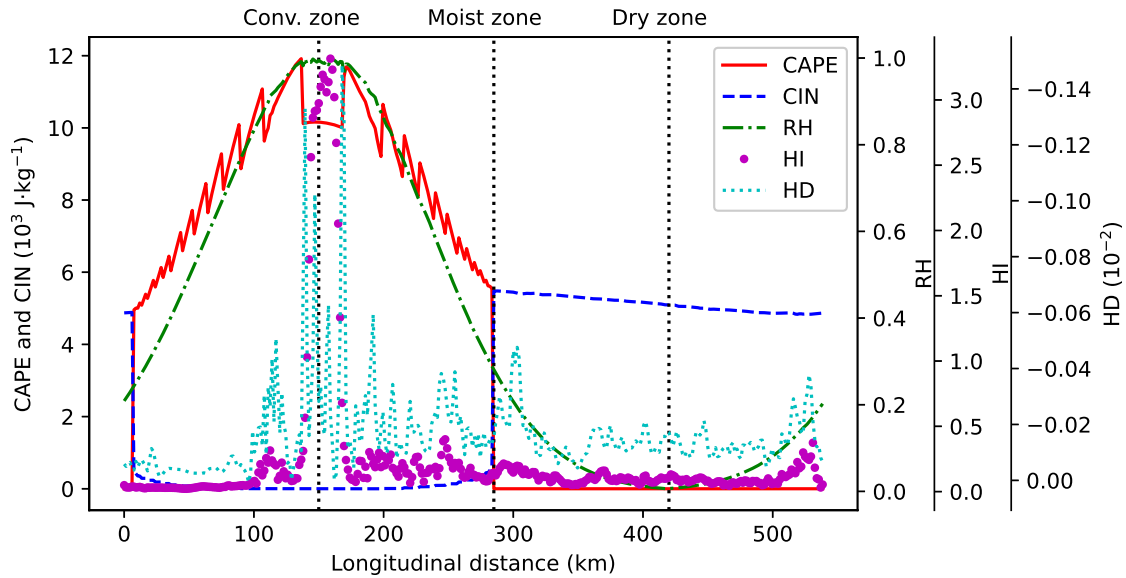
Figure 7. The geographical variation of various indicators/harbingers of convection at $t = 10\text{min}$. Plotted are: relative humidity (dot-dashed green line), hydrostatic balance (magenta dots), horizontal divergence (cyan dotted line), convective available potential energy (red continuous line), and convective inhibition (blue dashed line). The black vertical lines mark locations that represent the convection, moist, and dry zones.

415   to provide a low cost framework to study convective-scale data assimilation (DA) strategies. The following new variables, terms, and parameters have been introduced.

  – The model's vapour mixing ratio field is q (the gaseous phase of water), Eq. (1g).

  – The model's condensate mixing ratio field is q_c (the solid and/or liquid phases of water), Eq. (1h).

  – The net rate of source of vapour (or equivalently the net rate of sink of condensate) is $Ev - Co$ (the evaporation minus
420   the condensation rates) for each grid point. Only the difference needs to be known and not the values of Ev and Co separately.

  – Latent heat is released (absorbed) for net condensation (evaporation). The latent heating rate is $\mathcal{S}_{b'}$, and is a source term for buoyancy (latent heat is assumed to be exchanged directly only with buoyancy energy, Eq. (1e)). $\mathcal{S}_{b'}$ and $Ev - Co$ are not determined explicitly, but are the result of potentially solving a non-linear system of equations as set out in Sect.
425   2.3. The latent heat of vapourisation, L_v (set to 2500Jg^{-1}) is introduced which relates changes of vapour to changes in latent heat energy, Eq. 3.

  – Condensation happens when situations are super-saturated ($q > q_\text{s}$), in which case q drops immediately (practically over a model timestep) to a new q_s, and b' is increased from its previous value due to latent heat release. This is the reason

for the non-linear equations as mentioned above, see Eq. (7). When $b' < 0$ condensation is not allowed when the ratio of latent heat release to buoyancy energy is greater than Γ (set to 10). This is to avoid excessively large jumps to buoyancy.

- Evaporation happens when situations are sub-saturated, when there is a finite amount of condensate to evaporate, and when b' is positive. The last condition is imposed as otherwise b' will increase (i.e. become less negative) when evaporation happens, which is unphysical. Evaporation happens at a slower rate (timescale τ, set to 1000s) than condensation, see Eq. (6).

- The saturated mixing ratio, q_{s}, sets a maximum q that can be supported. In real systems q_{s} is a function of temperature and pressure. Since these quantities do not exist in the ABC model, analogous quantities are derived based on an assumed reference density profile, hydrostatic balance, and the correspondence between buoyancy and potential temperature. A formula is derived for q_{s} in terms of ABC variables b' and $\tilde{\rho}'$ as set out in Sect. 2.4. The conditions used to relate b' and $\tilde{\rho}'$ to q_{s} do not need to be exactly relevant to the ABC system, but are used only as a basis for a sensible relationship.

Hydro-ABC is constructed to satisfy certain conservation properties over the whole domain. These are for total water ($q + q_{\mathrm{c}}$, Sect. 2.2.1) and total moist energy (defined as dry energy plus latent heat energy, Sect. 2.2.2). There are no precipitation processes in Hydro-ABC; this is to keep the system as simple as possible while allowing highly non-linear processes.

An example run of Hydro-ABC shows a strong convective event driven by latent heat release. This forms a realistic anvil cloud (Sect. 3.1), and the excitation of inertio-gravity and acoustic modes, higher in frequency than those normally generated by the dry ABC model (Sect. 3.3).

A very important concern in DA applied to convective-scale flows involves how the background error covariance statistics are affected by the flow situation. This is an issue in real forecasting and DA systems where attempts have been made in the literature to use statistics that depend on the presence of background properties like rain and fog. This study gives an example of this by demonstrating how flow-dependent statistics (found from a 93-member ensemble) can depend on the presence of convection. It is found that there can be significant differences between the covariances in convecting and non-convecting regions, including the breakdown of hydrostatic balance. Many operational systems base their static background error covariance model, or the training data used to calibrate it, on hydrostatic balance (Berre, 2000; Bannister, 2008; Gustafsson et al., 2018; Bannister et al., 2020), and so this result significant in order to urge changes to this practice.

Complementary to knowledge of how the error covariances depend on the presence of convection, is a reliable indicator concerning whether convection is happening (or is impending) in any vertical column of the model. This can be usefully applied to the background state in a DA application, in order to decide the most relevant set of error covariances. Using climatologically-derived (static) error covariances, or covariances that support hydrostatic flow, may well impede the progress of convection prediction. A number of convection diagnostics have been proposed here as an indicator or harbinger of convection, namely relative humidity, hydrostatic imbalance, horizontal divergence, convective available potential energy, and convective inhibition (Sect. 4.2). It is found in the example shown that all of these quantities are sensitive to convection. Further work though is required to decide which quantity or combination of quantities is best, and which threshold values should be used. This will be part of the next stage of this work, namely to adapt the dry ABC-DA system to assimilate water-related information.

Appendix A: Computation of the convective available potential energy (CAPE)

In the real atmosphere CAPE is proportional to the vertical integral of the difference between a moist parcel's temperature and
465 the environmental temperature (e.g. Sect. 7.4.1 of Salby (1996)). It represents the amount of energy that can be converted to
kinetic energy by convective instability. A positive value indicates energy is available for convection, and is thus a measure of
instability. In Hydro-ABC there is no temperature variable, but we can derive a physically reasonable CAPE-like quantity by
analysing the energetics of an air parcel.

Consider an air parcel, whose properties are indicated with a 'hat'. The Lagrangian derivative of $\hat{w}$ is

470
$$\frac{\partial \hat{w}}{\partial t} + B\hat{\mathbf{u}} \cdot \nabla \hat{w} = \frac{D\hat{w}}{Dt} = \frac{D^2 \hat{z}}{Dt^2},\tag{A1}$$

where $\hat{z}$ is the vertical position of the parcel. Equation (A1) is the vertical acceleration of the parcel. We can substitute for
$\partial \hat{w}/\partial t + B\hat{\mathbf{u}} \cdot \nabla \hat{w}$ using the vertical momentum equation (1c). The parcel equation is then

$$\frac{D^2 \hat{z}}{Dt^2} = -C\frac{\partial \hat{\tilde{\rho}}}{\partial z} + \hat{b}'.\tag{A2}$$

The environmental profile (found by averaging the model fields over 11 points horizontally centred on the profile of interest)
475 is supposed to be in hydrostatic equilibrium:

$$-C\frac{\partial \tilde{\rho}'_{\text{env}}}{\partial z} + b'_{\text{env}} = 0.\tag{A3}$$

Taking the difference between (A2) and (A3) gives

$$\begin{aligned}
\frac{D^2 \hat{z}}{Dt^2} &= -C\frac{\partial \left(\hat{\tilde{\rho}} - \tilde{\rho}'_{\text{env}} \right)}{\partial z} + \left(\hat{b}' - b'_{\text{env}} \right) \\
&\approx \hat{b}' - b'_{\text{env}},
\end{aligned}\tag{A4}$$

480 where in the second line we have assumed mechanical equilibrium (that $\hat{\tilde{\rho}} = \tilde{\rho}'_{\text{env}}$). This gives a formula for the parcel's
acceleration (or specific force), which is used in the CAPE formula below.

The parcel's buoyancy profile, $\hat{b}'(z_i)$, is found by raising a hypothetical parcel from the surface. At the surface the parcel's
buoyancy and vapour mixing ratio take the model's values, $\hat{b}'(z_0) = b'(z_0)$ and $\hat{q}(z_0) = q(z_0)$ respectively. Start with level
$i = 1$. By raising the parcel to the next level, the total buoyancy is assumed to be conserved, $\hat{b}(z_i) = \hat{b}(z_{i-1})$. The total buoyancy
485 is the sum of the reference state and the perturbation, namely $\hat{b}(z_i) = \hat{b}_0(z_i) + \hat{b}'(z_i)$. By definition (Petrie et al., 2017), A^2 is
the rate of increase of reference buoyancy with height, namely $\hat{b}_0(z_{i+1}) - \hat{b}_0(z_i) = A^2(z_{i+1} - z_i)$ in discrete form. Putting the
last three equations together gives the parcel's perturbation value at z_i:

$$\hat{b}'(z_i) = \hat{b}'(z_{i-1}) - A^2(z_i - z_{i-1}).\tag{A5}$$

The parcel's mixing ratio is also assumed to be conserved, $\hat{q}(z_i) = \hat{q}(z_{i-1})$.
490 Should the parcel become supersaturated, the same non-linear equations as used by hydro ABC's micro-physics scheme to
adjust $\hat{b}'(z_i)$ and $\hat{q}(z_i)$ towards saturation are used here. This calculates revised values of $\hat{b}'(z_i)$ and $\hat{q}(z_i)$ (call these $\hat{b}'_{\text{mp}}(z_i)$

and $\hat{q}_{\mathrm{mp}}(z_i)$ respectively). The equations are given in point 1a of Sect. 2.3, where a 'hat' should be added to those equations to make them relevant here. Once $\hat{b}'_{\mathrm{mp}}(z_i)$ and $\hat{q}_{\mathrm{mp}}(z_i)$ have been found, $\hat{b}'(z_i)$ and $\hat{q}(z_i)$ are then over-written respectively by these adjusted values. This process is repeated by replacing $z_i \rightarrow z_{i+1}$ until the top of the model is reached.

495   Once this procedure is complete, two special levels are found. Starting from the bottom of the domain, the first level encountered that has the property $\hat{b}' \geq b'_{\mathrm{env}}$ is called the level of free convection, z_{LFC}, and the first level above that where $\hat{b}' \leq b'_{\mathrm{env}}$ is called the level of neutral buoyancy, z_{LNB}. The CAPE is then computed from the integral

$$\mathrm{CAPE} = \int_{z_{\mathrm{LFC}}}^{z_{\mathrm{LNB}}} \left(\hat{b}'(z) - b'_{\mathrm{env}}(z) \right) dz, \tag{A6}$$

which has units of Jkg^{-1}. The related quantity convective inhibition (CIN) can also be found from the parcel's profile:

500      $$\mathrm{CIN} = - \int_{0}^{z_{\mathrm{LFC}}} \left(\hat{b}'(z) - b'_{\mathrm{env}}(z) \right) dz. \tag{A7}$$

Should a value of z_{LFC} not be found (owing to $\hat{b}' < b'_{\mathrm{env}}$ for all levels), CAPE would be zero, and CIN would be found by integrating over the entire profile. Should a value of z_{LNB} not be found (owing to $\hat{b}' > b'_{\mathrm{env}}$ for all levels above z_{LFC}), then CAPE would be found by integrating to the top of the model.
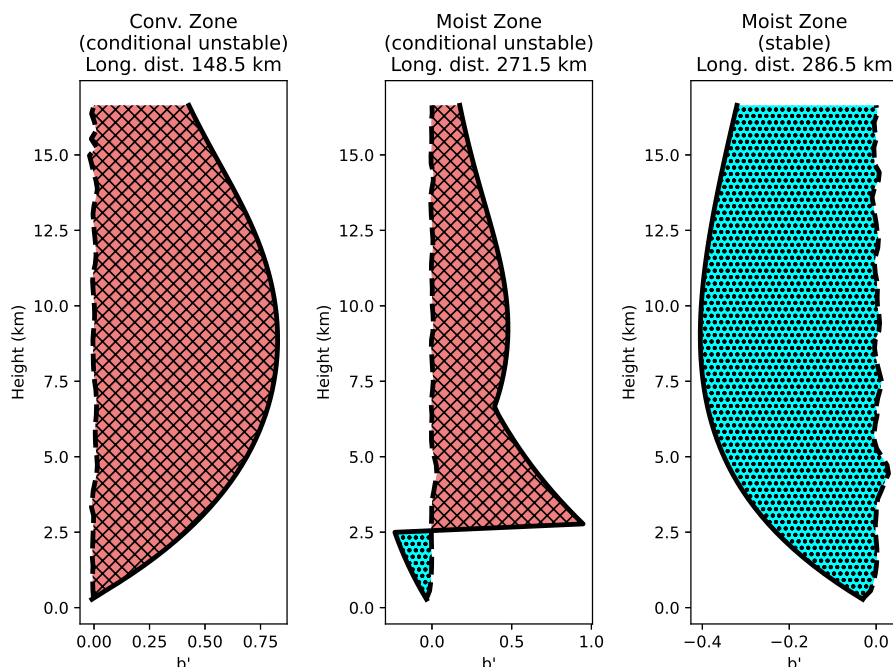
Figure A1. Example profiles of a parcel's computed profile ($\hat{b}'$, solid line) and the environmental profile (b'_{env}, dashed line) for three horizontal positions in the domain (calculated at $t = 10\mathrm{min}$). The horizontal positions are in the convection zone ($x = 148.5\mathrm{km}$), and two points within the moist zone ($271.5\mathrm{km}$, and $286.5\mathrm{km}$). The area of the red hatched region is the CAPE, and the area of the cyan hatched region is the CIN.

Example profiles of the parcel and environmental buoyancy are shown in Fig. A1 for three positions in the domain at
505 $t = 10\mathrm{min}$. In the convection zone at 148.5 km (left panel), $\hat{b}' > b'_{\mathrm{env}}$ for all levels, so the CAPE has a large value, consistent with the strong convection there. Two further profiles are shown for the moist zone. One, at 271.5 km (middle panel), is where the CAPE remains large but where a value of z_{LFC} is found. There is therefore a non-zero CIN associated with this position, but the moist instability still has influence. The other, at 286.5 km (right panel), is just past the position where the CAPE drops away (see Fig. 7). Here $\hat{b}' < b'_{\mathrm{env}}$ for all levels, so there is no CAPE, but a large value of CIN. This is the situation into the dry
510 zone too, where a very similar plot is found, for instance at 420 km (not shown).

Code availability. The Hydro-ABC V2.0 code is written in Fortran-90 and is freely available via a GitHub repository (Zhu and Bannister, 2022). The Hydro-ABC V2.0 user documentation is available in the same place.

Author contributions. JZ and RNB jointly designed the moisture framework. JZ modified the dry ABC code to include the moist processes described in this paper, ran the model and produced Figs. 2, 3, 4, 5, 6, 7, and A1. RNB wrote the paper and produced Fig. 1.

515 *Competing interests.* The authors declare that they have no competing interests.

Acknowledgements. The authors would like to thank the National Natural Science Foundation of China for funding JZ (grant ref 41875136), and the Natural Environment Research Council, which provides national capability funding for RNB via the National Centre for Earth Observation (contract number PR140015). The authors would also like to thank Mike Cullen for correspondence during the development of Hydro-ABC.

References

Bannister, R.: A review of operational methods of variational and ensemble-variational data assimilation, Quarterly Journal of the Royal Meteorological Society, 143, 607–633, https://doi.org/DOI:10.1002/qj.2982, 2017.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, Quarterly Journal of the Royal Meteorological Society, 134, 1971–1996, 2008.

Bannister, R. N.: The ABC-DA system (v1.4): a variational data assimilation system for convective scale assimilation research with a study of the impact of a balance constraint, Geoscientific Model Development, 13, 3789–3816, https://doi.org/10.5194/gmd-13-3789-2020, 2020.

Bannister, R. N.: Balance conditions in variational data assimilation for a high-resolution forecast model, Quarterly Journal of the Royal Meteorological Society, 147, 2917–2934, https://doi.org/DOI: 10.1002/qj.4106, 2021.

Bannister, R. N., Migliorini, S., and Dixon, M.: Ensemble prediction for nowcasting with a convection-permitting model–II: Forecast error statistics, Tellus A, 63, 497–512, 2011.

Bannister, R. N., Migliorini, S., Rudd, A. C., and Baker, L. H.: Methods of investigating forecast error sensitivity to ensemble size in a limited-area convection-permitting ensemble, Geoscientific Model Development Discussions, pp. 1–38, 2017.

Bannister, R. N., Chipilski, H., and Martinez-Alvarado, O.: Techniques and challenges in the assimilation of atmospheric water observations for numerical weather prediction towards convective scales, Quarterly Journal of the Royal Meteorological Society, 146, 1–48, https://doi.org/10.1002/qj.3652, 2020.

Banos, I. H., Mayfield, W. D., Ge, G., Sapucci, L. F., Carley, J. R., and Nance, L.: Assessment of the data assimilation framework for the Rapid Refresh Forecast System v0. 1 and impacts on forecasts of a convective storm case study, Geoscientific Model Development, 15, 6891–6917, 2022.

Berre, L.: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model, Monthly Weather Review, 128, 644–667, 2000.

Clark, P., Roberts, N., Lean, H., Ballard, S. P., and Charlton-Perez, C.: Convection-permitting models: a step-change in rainfall forecasting, Meteorological Applications, 23, 165–181, 2016.

Ehrendorfer, M.: A review of issues in ensemble-based Kalman filtering, Meteorologische Zeitschrift, 16, 795–818, 2007.

Errico, R. M., Bauer, P., and Mahfouf, J.-F.: Issues regarding the assimilation of cloud and precipitation data, Journal of the Atmospheric Sciences, 64, 3785–3798, 2007.

Fabry, F. and Meunier, V.: Why are radar data so difficult to assimilate skillfully?, Monthly Weather Review, 148, 2819–2836, 2020.

Gill, A. E.: Atmosphere-Ocean Dynamics, Academic Press, 1982.

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., et al.: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres, Quarterly Journal of the Royal Meteorological Society, 144, 1218–1256, 2018.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Monthly Weather Review, 129, 2776–2790, 2001.

Hohenegger, C. and Schär, C.: Atmospheric predictability at synoptic versus cloud-resolving scales, Bulletin of the American Meteorological Society, 88, 1783, 2007.

Holton, J. and Hakim, G.: An Introduction to Dynamic Meteorology, 5th edition, Academic Press, Waltham, MA, 2013.

Houtekamer, P. and Zhang, F.: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 144, 4489–4532, 2016.

Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Monthly Weather Review, 126, 796–811, 1998.

Houtekamer, P. L. and Mitchell, H. L.: Ensemble Kalman filtering, Quarterly Journal of the Royal Meteorological Society, 131, 3269–3289, 2005.

Kent, T., Bokhove, O., and Tobias, S.: Dynamics of an idealized fluid model for investigating convective-scale data assimilation, Tellus A: Dynamic Meteorology and Oceanography, 69, 1369 332, 2017.

Lee, J. C. K., Amezcua, J., and Bannister, R. N.: Hybrid ensemble-variational data assimilation in ABC-DA within a tropical framework, Geoscientific Model Development, 15, 6197–6219, https://doi.org/DOI:10.5194/gmd-15-6197-2022, 2022.

Leung, T. Y., Leutbecher, M., Reich, S., and Shepherd, T. G.: Atmospheric predictability: revisiting the inherent finite-time barrier, Journal of the Atmospheric Sciences, 2019.

Lorenc, A.: A study of ob monitoring statistics from radiosondes, composited for low-level cloud layers, Met Office NWP Forecasting Research Technical Report, 504, 1–32, 2007.

Lorenz, E. N.: The predictability of a flow which possesses many scales of motion, Tellus, 21, 289–307, 1969.

Ménétrier, B. and Montmerle, T.: Heterogeneous background-error covariances for the analysis and forecast of fog events, Quarterly Journal of the Royal Meteorological Society, 137, 2004–2013, 2011.

Michel, Y., Auligné, T., and Montmerle, T.: Heterogeneous convective-scale background error covariances with the inclusion of hydrometeor variables, Monthly Weather Review, 139, 2994–3015, 2011.

Montmerle, T.: Optimization of the assimilation of radar data at the convective scale using specific background error covariances in precipitation, Monthly Weather Review, 140, 3495–3506, 2012.

Montmerle, T. and Berre, L.: Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale, Quarterly Journal of the Royal Meteorological Society, 136, 1408–1420, 2010.

Petrie, R. E., Bannister, R. N., and Cullen, M. J. P.: The ABC model: a non-hydrostatic toy model for use in convective-scale data assimilation investigations, Geoscientific Model Development, 10, 4419, https://doi.org/10.5194/gmd-10-4419-2017, 2017.

Pielke, R.: Mesoscale Meteorological Modeling, Academic Press, San Diego, California, 2002.

Salby, M. L.: Fundamentals of atmospheric physics, vol. 61, Academic press, San Diego, California, 1996.

Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., Joe, P., Barker, D. M., Li, P.-W., Golding, B., et al.: Use of NWP for nowcasting convective precipitation: Recent progress and challenges, Bulletin of the American Meteorological Society, 95, 409–426, 2014.

van Leeuwen, P. J.: Comment on 'Data assimilation using an ensemble Kalman filter technique', Monthly Weather Review, 127, 1374–1377, 1999.

Vetra-Carvalho, S., Dixon, M., Migliorini, S., Nichols, N. K., and Ballard, S. P.: Breakdown of hydrostatic balance at convective scales in the forecast errors in the Met Office Unified Model, Quarterly Journal of the Royal Meteorological Society, 138, 1709–1720, 2012.

Würsch, M. and Craig, G. C.: A simple dynamical model of cumulus convection for data assimilation research, Meteorologische Zeitschrift, 23, 483–490, 2014.

Yang, Y., Gao, S., Wang, Y., and Shi, H.: Impact of Feature-Dependent Static Background Error Covariances for Satellite-Derived Humidity Assimilation on Analyses and Forecasts of Multiple Sea Fog Cases over the Yellow Sea, Remote Sensing, 14, 4537, 2022.

Yano, J.-I., Ziemiański, M. Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S. L., et al.:
595   Scientific challenges of convective-scale numerical weather prediction, Bulletin of the American Meteorological Society, 99, 699–710,
      2018.

Zhang, M. and Zhang, F.: E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area
      weather prediction model, Monthly Weather Review, 140, 587–600, 2012.

Zhu,   J.   and   Bannister,   R.:   Hydro-ABC   model   software,   https://github.com/rossbannister/Hydro-ABC_vn2.0,
600   https://doi.org/10.5281/zenodo.7418510, 2022.