The hectometric modelling challenge

Gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models

Lean, Humphrey W.; Theeuwes, Natalie E.; Baldauf, Michael; Barkmeijer, Jan; Bessardon, Geoffrey; Blunn, Lewis; Bojarova, Jelena; Boutle, Ian A.; Siebesma, Pier; More Authors

DOI
10.1002/qj.4858

Publication date
2024

Document Version
Final published version

Published in
Quarterly Journal of the Royal Meteorological Society

Received: 18 March 2024 | Revised: 3 August 2024 | Accepted: 5 August 2024

DOI: 10.1002/qj.4858

REVIEW ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

The hectometric modelling challenge: Gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models

Humphrey W. Lean[1] | **Natalie E. Theeuwes**[2] | **Michael Baldauf**[3] | **Jan Barkmeijer**[2] | **Geoffrey Bessardon**[4] | **Lewis Blunn**[1] | **Jelena Bojarova**[5] | **Ian A. Boutle**[1] | **Peter A. Clark**[6] | **Matthias Demuzere**[7,8] | **Peter Dueben**[9] | **Inger-Lise Frogner**[10] | **Siebren de Haan**[2] | **Dawn Harrison**[1] | **Chiel van Heerwaarden**[11] | **Rachel Honnert**[12] | **Adrian Lock**[1] | **Chiara Marsigli**[3] | **Valéry Masson**[12] | **Anne Mccabe**[1] | **Maarten van Reeuwijk**[13] | **Nigel Roberts**[1] | **Pier Siebesma**[2,14] | **Petra Smolíková**[15] | **Xiaohua Yang**[16]

Correspondence
Natalie E. Theeuwes, Royal Netherlands Meteorological Institute (KNMI), de Bilt, the Netherlands
Email: natalie.theeuwes@knmi.nl

Funding information
HORIZON EUROPE Climate, Energy and Mobility, Grant/Award Number: 101137851; Deutsche Forschungsgemeinschaft, Grant/Award Number: 437467569

Abstract

For a number of years research has been carried out in several centres which has demonstrated the potential benefits of 100-m scale models for a range of meteorological phenomena. More recently, some meteorological services have started to consider seriously the operational implementation of practical hectometric models. Many, but by no means all, of the applications are likely to relate to urban areas, where the enhanced resolution has obvious benefits. This article is concerned with the issues that need to be addressed to bridge the gap between research at 100-m scales and practical models. We highlight a number of key issues that need to be addressed, with suggestions of important avenues for future development. An overarching issue is the high computational cost of these models. Although some ideas to reduce this are presented, it will always be a serious constraint. This means that the benefits of these models over lower resolution ones, or other techniques for generating high-resolution forecasts, will need to be clearly understood, as will the trade-offs with resolution. We discuss issues with model dynamical cores and physics–dynamics coupling. There are a number of challenges around model parameterisations, where some of the traditional problems (e.g., convection) become easier but a number of new

Humphrey W. Lean and Natalie E. Theeuwes contributed equally to this study.

For affiliations refer to page 29

challenges (e.g., around surface parameterisations) appear. Observational data at these scales present a challenge and novel types of observations will need to be considered. Data assimilation will be needed for short-range forecasts, but there is currently little knowledge of this, although some of the likely issues are clear. An ensemble approach will be essential in many cases (e.g., convection), but research is needed into ensembles at these scales and significant work on post-processing systems is required to make the best use of models at these grid lengths.

KEYWORDS

hectometric modelling, numerical weather prediction, parameterisations, urban

1 | INTRODUCTION

With the continuing increases in available computer power, there is growing interest within meteorological services in practical applications of hectometric models (HMs). A number of centres have started projects to develop their hectometric modelling capability and, in some cases, to implement routinely running models. There are two general classes of scientific benefit from running models at very high resolution. Most obviously, we expect benefits where improved atmospheric resolution improves the representation of important processes, for example, (deep) convection, which is often poorly resolved in km-scale models (e.g., Bryan *et al.*, 2003), effectively reducing the need to parameterise certain processes. Secondly, we expect benefits from use of higher resolution observations and surface data, especially associated with complex surface forcing such as orography and urban areas. Here, we will discuss the challenges associated with the practical use of HMs and the research still needed to realise the full benefits. For the purposes of this article, we are considering models in the turbulence-permitting regime, that is, with grid lengths ~500–50 m.

Convection-permitting, order ~km grid length models have been established in many meteorological services for over 10 years and are used extensively for weather forecasting applications (e.g., Baldauf *et al.*, 2011; Seity *et al.*, 2011; Tang *et al.*, 2013). Although developed somewhat later, convection-permitting regional climate models have also become important to downscale global climate models (e.g., Belušić *et al.*, 2020; Kendon *et al.*, 2012; Kendon *et al.*, 2021; Prein *et al.*, 2015; Schär *et al.*, 2020). Given the great success of these models, it is interesting to consider that there have, until now, been very few examples of further increases in resolution in operational models. In general, centres have not changed the resolution of their operational km-scale models in more than 10 years (e.g., Bengtsson *et al.*, 2017; Tang *et al.*, 2013). The reason for this is due to two closely linked factors. Firstly, there was a very strong driver to move regional forecast models from order 10 km grid length to closer to 1 km. The move from a parameterised to an explicit, albeit poorly resolved, representation of deep convection was often found to improve model performance greatly in ways important for applications (Clark *et al.*, 2016). In contrast, decreasing the grid length below 1 km has only shown incremental improvements to convection and the ability to resolve heterogeneous terrain, urban areas, and orography. However, it is unknown whether this is due only to incremental improvements of the representation of the physics, or whether the available observations and verification methods are insufficient to show the added benefit. Secondly, costs increase very rapidly with model resolution and additional computer resources have instead been used for other enhancements with clearer benefits (larger domains, longer runs, and running ensembles).

Over the last 10 years a number of centres have experimented with sub-km versions of their numerical weather prediction (NWP) models (Figure 1 shows an example). There have been a number of motivations for this work, including the desire to explore potential benefits such as better resolution of convection and the gains coming from higher resolution surface data. This research has demonstrated the benefits of 100-m scale models for a number of meteorological phenomena. Examples include cold pooling in small valleys (e.g Valkonen *et al.*, 2020; Vosper *et al.*, 2013), convection (e.g., Hanley *et al.*, 2015), fog (e.g., Boutle *et al.*, 2016), stratocumulus (e.g., Boutle *et al.*, 2014a), urban overheating and human thermal comfort (e.g., Ronda *et al.*, 2017), tornadoes (e.g., Hanley *et al.*, 2016), and urban pollution dispersion (Blunn *et al.*, 2023). Some centres have implemented routinely running $\mathcal{O}(100)$ m scale models (Boutle *et al.*, 2016; Joe *et al.*, 2018). These developments represent NWP models being run at resolutions previously reserved for relatively coarse large eddy simulations (LES), for

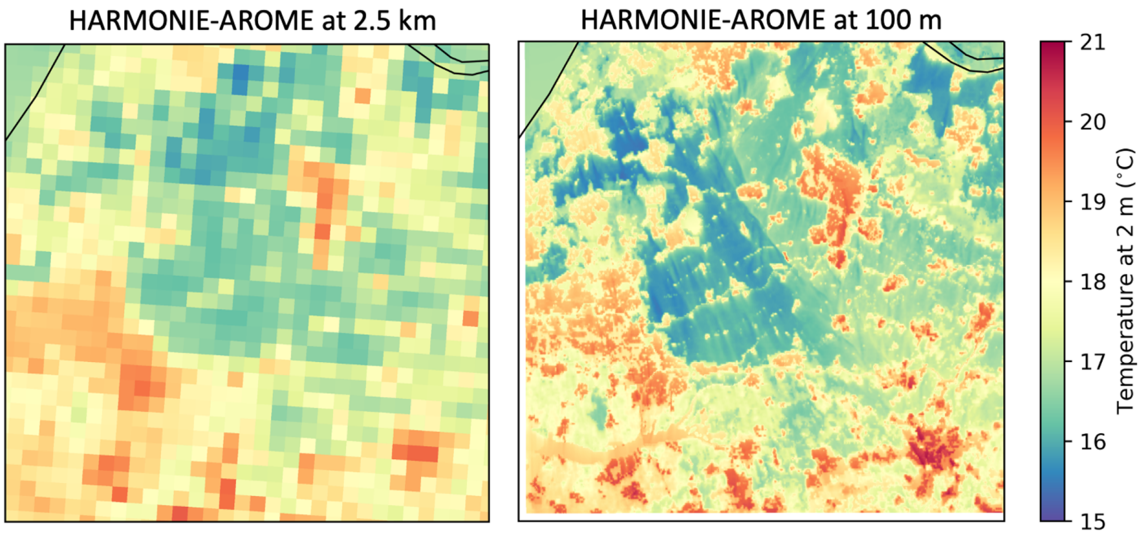

FIGURE 1 Example of 2-m temperature for two simulations over the Netherlands on September 14, 2020, 2100 UTC, showing a more detailed representation of the higher temperatures in urban areas. The figure shows the HARMONIE-AROME model (Bengtsson *et al.*, 2017) at operational grid spacings of (left) $2.5 \times 2.5\,\text{km}^2$ and (right) $100 \times 100\,\text{m}^2$. [Colour figure can be viewed at wileyonlinelibrary.com]

example, models for (deep) convection, often referred to as cloud-resolving models (CRMs), though the more recent term convection-permitting models (CPMs) is more accurate. At the same time, LES modelling has progressed to more realistic simulations on larger domains (e.g., Gehrke *et al.*, 2021; Schalkwijk *et al.*, 2015). These two communities now have the opportunity to work together to mutual benefit.

Despite this body of work demonstrating potential benefits of these models, there are still significant practical and scientific obstacles to their widespread adoption for routine weather and climate forecasting. The most important of these is their high cost, which is the context through which all other issues have to be seen. Without radical improvements to model numerics, increasing the horizontal resolution from, for example, 2 km to 200 m would require a factor of 1000 increase in computational costs (a factor of 10 for each horizontal dimension and time), plus any additional time from changes to the vertical resolution. Any practical system will need to justify these very high costs, which means that it is important to understand the benefits of different grid lengths (e.g., 100 m compared with 300 m) for a given application. One obvious cost mitigation, which has been used by several centres so far, is to run the model over a very small domain covering an area where hazards might have a large economic impact (e.g., city, airport, etc.).

All NWP and climate models can benefit from various forms of post-processing. This is particularly true near the surface, where a fully functional dynamical model may not be needed to infer the impact of known small-scale variation in surface forcing. An example is near-surface temperature, where many horizontal heterogeneity features come directly from the land surface, for example, parks in cities, rivers. These benefits could likely be realised more cheaply by a downscaling system using the detailed surface data with an offline land-surface model rather than the full atmospheric model. This emphasises a general point about applications: to be useful, the HMs need to be not only better when compared with lower resolution models but also better than cheaper downscaling systems, which do not involve running the full atmospheric model. The true benefits of more detailed surface information are likely to come in cases where the surface interacts in a complicated way with the atmosphere. An example of this might be orographic rain, where the higher resolution orography has to interact with the dynamical and microphysical processes in the atmosphere. Furthermore, if the downscaling model is a machine-learning algorithm, simulations of a conventional model may still be needed to provide training data.

An important benefit of a large-domain hectometric model is that it can capture interactions between different scales—for example, small-scale urban canopy effects, effects of parks and rivers with relatively larger scale effects such as sea breezes, frontal passages, etc. Examples include understanding the effects of greening certain areas of the city on the temperatures of the whole city, where the effects of different local surfaces interact via the effects on the larger scale flow, and the effects of the urban surface on the initiation of convection over a city.

It is important to note that each application of these HMs may bring about different challenges requiring various kinds of research. Some of the potential applications include the following.

- Orographic rain/wind: with higher resolution orography, the steepest slopes will be represented better, leading to small-scale peaks in rainfall or wind (gusts) being forecast better.
- Urban temperatures: detailed land-surface data will allow representation of temperatures across cities on neighbourhood scales.
- Urban winds/turbulence: better representation of the urban surface and boundary-layer turbulence will lead to improved forecasts of winds and turbulence across cities.
- Air quality/dispersion: better representation of the atmospheric boundary layer, horizontal heterogeneity, and mixing of scales can lead to better air-quality forecasts. However, since NWP models do not resolve buildings, very local dispersion will not be represented explicitly.
- Deep convective initiation: better forecasts of the initiation of convection from improved representation of surface heterogeneity and convergence lines, etc.
- Deep convection: better forecasts of convective clouds and rainfall and upscaling due to improved representation of updrafts and entrainment/detrainment and other storm-scale structures.
- Fog: better representation of the surface topography combined with the ability to represent higher resolution spatial variations in the boundary layer will lead to better forecasts.

It is expected that many of the applications of HMs will stem from being able to resolve urban areas better. Cities strongly modify the exchanges of momentum, heat, and moisture towards the atmosphere compared with rural areas, and create their own microclimate. In particular, after sunset and at night under calm conditions, cities often stay warmer than the countryside by up to 10–12 K for large cities (Masson *et al.*, 2020b; Oke *et al.*, 2017), but smaller scale heteorogenities also appear, for example, so-called "cool islands" within cities, which can only be modelled by HMs.

The rapid rise of machine learning (ML) will play a large and uncertain role in the evolution of HMs over the coming years. ML can be used to generate fast emulators of conventional parameterisations (e.g., Gettelman *et al.*, 2021, Meyer *et al.*, 2022a, 2022b, Rasp *et al.*, 2018) to form hybrid ML—conventional HMs that retain physical integrity, but with reduced computational cost. However, developments of pure ML models may also replace entire forecast models, as seen for global models trained on ERA5 reanalysis that are competitive with conventional models for deterministic and ensemble forecast scores (e.g., Bi *et al.*, 2023, Chen *et al.*, 2023, Keisler, 2022, Lam *et al.*, 2022, Pathak *et al.*, 2022). Such ML models can provide physically plausible behaviour (Hakim & Masanam, 2023), high accuracy (de Burgh-Day & Leeuwenburg, 2023), and large ensembles at small computational cost (Hu *et al.*, 2023; Price *et al.*, 2023; Weyn *et al.*, 2021). In addition, limited-area ML models have been developed that use limited-area NWP as training data, and either one-way nest within a coarser global model (Oskarsson *et al.*, 2023) or an adaptive mesh that is finer over the limited area (Nipen *et al.*, 2024). Similarly, ML models that are trained against in situ and remote sensing observations may also replace conventional HMs in the future. If the information from observations is sufficient to train entire ML models, ML models may not need conventional weather models (Espeholt *et al.*, 2022). Also, ML post-processing techniques have been shown to downscale NWP accurately to hectometre scale for some variables such as near-surface air temperature (Blunn *et al.*, 2024b; Wu *et al.*, 2021) and precipitation (Harris *et al.*, 2022). Given the immense computational expense of conventional HMs, their benefits over ML post-processing techniques would need to be demonstrated across applications. On the other hand, hectometric model output could also become an essential reference data source for the training of ML models if observations alone turn out to be insufficient.

This article is the result of discussions between a number of experts from different fields in NWP and LES and seeks to understand and present the gaps in the current NWP state of the art that represent barriers to the use of HMs. It is hoped that this article will stimulate further research and collaboration. It is important to understand that the solutions to some of these problems will be very dependent on the application considered. The following sections detail some of the gaps in science/understanding for specific areas in NWP and discuss the issues in more detail, along with suggestions for possible ways forward.

2 | DYNAMICAL CORE AND STABILITY

One might expect that the pure dynamical core is rather insensitive to the resolution. This is indeed the case if the *unapproximated* equation set is used, that is, *the non-hydrostatic, compressible Euler equations*(potentially together with less sensitive modifications such as the shallow atmosphere and the so-called traditional approximation, or the usually made spherical geopotential approximation). Today, probably most of the forecast models used operationally, in fact, use these unapproximated Euler equations, since they want to reduce the overhead of maintaining multiple model systems and favour a "unified model" approach. However, it is worth mentioning which further equation sets are adequate to hectometric-scale modelling and which are not. The hydrostatic approximation (only applicable for horizontal scales larger than a few km) and the Boussinesq approximation (only usable for very shallow flows in the boundary layer) are definitely not usable. The LES community, of course, has always operated at hectometric or even finer scales, and here use of anelastic approximations is common. Their basic idea is to neglect the time derivative in the continuity equation to filter out sound waves.

The equation set of Ogura and Phillips (1962) still makes a strong restriction to isentropic reference states, which was weakened in the set of Wilhelmson and Ogura (1972). Probably most often used is the anelastic equation set of Lipps and Hemler (1982), which additionally is energy-conserving and better suited for the moist atmosphere (also see the survey of Nance and Durran (1994)). Other equation sets that filter out sound waves and are usable for hectometric-scale modelling are the pseudo-incompressible equations (Durran, 1989), the unified anelastic, quasi-hydrostatic set (Arakawa & Konor, 2009), and its implementation by Voitus *et al.* (2019). There were several attempts to create unified systems including various equation sets through an introduction of control parameters (Benacchio & Klein, 2019; Klein & Benacchio, 2016). Then hydrostatic primitive equations, soundproof equations, or fully compressible equations may be evoked from the same framework. Such blended systems may be used for initialisation when undesirable unbalanced modes are filtered out at the beginning of the integration and the full set of equations is only used further into the integration; see (Chew *et al.*, 2022). In Smolíková and Vivoda (2023) a blended system is found, which allows one to slow down the acoustic modes while keeping the gravity modes unchanged, resulting in a solution preserving all the nonhydrostatic features of the flow.

The anelastic sets are advantageous for low Mach numbers; however, in a hectometric-scale model, jets and even gravity-wave breaking that produce very high velocities $>100\,\mathrm{m\cdot s^{-1}}$ can occur. This implies that the Mach numbers are not very low any more. Beyond this, nowadays iterative implicit Helmholtz solvers for the compressible equations can be designed to be as efficient (i.e., similar time-step size and wall-clock time usage) as iterative Poisson solvers for the anelastic equations (Smolarkiewicz *et al.* (2014); Kurowski *et al.* (2014)). Additionally, the principal issues of lacking mass conservation and of surface-pressure reconstruction of approximated equation sets are increasingly less tolerated. There are hints that local conservation properties become even more important for smaller scale modelling. On the other hand, some numerical stabilisation mechanisms, like the divergence damping needed in split-explicit compressible solvers, also can modify important properties; Baldauf *et al.* (2013) demonstrates comparable deviations in the dispersion relation of small-scale gravity waves for the compressible Euler equations with artificial divergence damping and the anelastic approximation.

The fundamental choices of discretisation in space and time and the treatment of the transport (the advection scheme) influence the affordable time step. Two approaches are possible: Eulerian advection allows only for short time steps and may be combined with horizontally explicit–vertically implicit (HEVI) discretisation techniques to treat fast vertically propagating sound waves implicitly, while semi-Lagrangian advection allows for relatively long time steps and is usually combined with semi-implicit time-stepping (ACCORD, UM) including an iterative process aimed at implicit treatment of the nonlinear residual (Bénard, 2003; Walters *et al.*, 2019).

Together with fine horizontal meshes comes the necessity for high vertical resolution, especially near the surface and in connection to the representation of the terrain. The slopes of orography in such conditions become relatively steep, imposing several challenges on numerical schemes solving transport and time evolution of model variables. It may be difficult to satisfy the Courant (CFL) criterion and thus to ensure the stability of numerical schemes used. With semi-implicit time-stepping, the orographically dominated terms may be necessarily included in the linear part to diminish the nonlinear residual, which would otherwise impose restrictions on the time step used or the number of iterations needed to reach stable integration. There are hints that higher spatial approximation order accompanied with higher temporal approximation order (as with IMEX-RK method) may improve stability properties (Baldauf, 2021). It must be explored whether this holds only for idealised test cases or for the full fledged model, too. Alternative vertical discretisation may potentially raise the accuracy order and improve the simulation quality, as shown for the finite elements using cubic spline basis functions, for example in Vivoda *et al.* (2018), or for the staggered nodal finite-element method in Guerra and Ullrich (2016).

After defining the horizontal grid, the associated vertical coordinate generally uses one of two approaches: the terrain-following coordinate or the cut-cell approach. In the latter, regular flat cells are modified when their edges cross the terrain. Even if the cut-cell method may create some very small cells, several techniques have been found that preserve cell length in the direction of flow and do not impose additional constraints on the time-step length (Shaw & Weller, 2016). Thus, several authors advocate cut-cell methods as being suitable for very fine mesh sizes (Steppeler *et al.*, 2002). In contrast, the distortion of cut-cell meshes necessary in the vicinity of steep terrain may cause numerical errors that do not reduce with resolution, and this could be the reason why most applications use terrain-following vertical coordinates. Here, to alleviate the errors connected to the calculation of the pressure-gradient term, either smoothing of coordinate surfaces with height (Klemp, 2011; Laprise, 1992) or interpolation onto flat layers (Zängl, 2012) is proposed. This may be combined with the distinction of small and large scales in orography, while short-scale imprints are mitigated in the top layers, avoiding possible large errors in advection and amplification of vertical motion above jet level (Husain *et al.*, 2020; Schär *et al.*, 2002). According to Shaw and Weller (2016), when combined with an appropriate time-stepping and suitable advection scheme, no significant problems arose in idealised tests with terrain-following grids.

Another advantage of terrain-following coordinates is the possibility of keeping high vertical resolution everywhere close to the ground without the need to keep it high over a wide vertical extent. On the other hand, with the cut-cell approach, to keep high vertical resolution from the lowest point to the highest mountain in the domain may be computationally demanding. The height of the bottom level is highly relevant for coupling with the parameterisation of turbulence and the surface scheme.

A hectometric-scale model is (at least for the foreseeable future) always a limited-area model (LAM) that must be nested in a larger scale model. If this nesting is done in a (supposedly smooth) two-way approach, no fundamental problems are expected, since waves or other signals can travel in and out of the domain almost freely without significant reflection. However, there is often a need for one-way nesting from practical requirements. Here, the task is to set boundary conditions (BCs) that both drive the LAM and avoid wave reflection from the LAM at this artificial boundary. The correct setting of BCs underlies well-posedness conditions (e.g., Davies, 2014; Oliger & Sundström, 1978), which requires a proper distinction of inflow and outflow directions in a compressible fluid. This well-posedness problem is to a certain extent reduced by modifying the equation system with relaxation terms that blend the solutions of the driving model and the LAM and damp waves travelling towards the boundary (Davies, 1976). Sound waves can be partially removed by the use of wave-permitting BCs (Durran, 1998), but this also reduces the driving ability of the outer model. Additionally, the use of upwind discretisations relaxes the well-posedness problem.

Although these approaches work rather well in current operational environments, hectometric-scale models face a distinct problem: their spatial extent might only be 100–300 km in both horizontal directions. However, the width L_r of relaxation zones is determined by the grid spacing of the driving model and, more seriously, by the scale of the majority of waves occurring in the LAM. Consequently, L_r does not necessarily shrink linearly with decreasing horizontal grid spacing and therefore covers a non-negligible part of the LAM, reducing the usable simulation area. Perhaps the most important remedy to avoid this is to mimic two-way nesting: increase the grid spacing in the LAM perpendicular to the boundary, so that the relaxation zone only needs relatively few grid points (Davies, 2014). Model systems that cannot immediately achieve such a horizontal stretching might go back to a multi-nest approach, in which the outer nests only have relaxation layers.

In detail, one can modify the application of the relaxation zone in the dependence of an (advective) inflow or outflow condition and the prognostic fields that are used: for example, one should not relax rain or snow at an outflow boundary if the driving model uses a deep convection scheme and does not predict rain or snow at all. The update frequency is of course an issue too; it is determined mainly by technical, operational requirements and in any case should be clearly smaller than one hour.

Another problem of hectometric-scale models is the fact that they are in principle LES models. This means that they have to be fed with a realistic large eddy turbulence field already at the boundary (a sufficiently fast self-development of such structures in the LAM itself is not possible for near-neutral or even stable stratifications due to the small domain size). For this purpose, turbulence pattern generators that generate resolved turbulent eddies that depend on local wind shear and stratification also in the relaxation zone should be developed (e.g., Munoz-Esparza *et al.*, 2014).

At the upper boundary, damping upward-travelling gravity waves is necessary to avoid artificial reflections. Again, a relaxation layer is mostly used, either for all prognostic variables or only for the vertical velocity w on a short time step (Klemp *et al.*, 2008). In contrast, gravity-wave radiating BCs (Klemp & Durran, 1983) may work less well for horizontally varying driving fields or for compressible solvers. Horizontal-diffusion layers may become unstable due to the high diffusion coefficients needed. A rule of thumb is to use about one third (in z) of the whole model domain for this relaxation layer. Due to strong vertical stretching, only a few grid cells are used for this artificial layer. Since even a hectometric-scale

Quarterly Journal of the Royal Meteorological Society RMetS

model needs to simulate at least the whole troposphere, the model top height H_{top} should not be reduced below about 25–30 km. Data assimilation requirements may even need higher H_{top}.

A hectometric-scale model might be initialised by interpolating a larger scale analysis (Short & Petch, 2022). In this case, and even if a data assimilation (DA) system is applied for the hectometric-scale model, then the system may suffer from too long spin-up times of the forecast model. The standard 2D or 3D divergence filters are often not damping enough to get rid of noisy gravity and sound waves excited by the analysis step. Therefore, efficient filters must be developed for this purpose. The alleviating techniques are discussed further in Section 7.

3 | PHYSICS–DYNAMICS COUPLING

The large computational cost of HMs represents an overarching concern about the viability of these models for operational use. Scientific aspects of physics–dynamics coupling start to become simpler at hectometric scales—the complicated time-stepping strategies and slow/fast process splitting often employed in NWP models with long time steps (Dubal *et al.*, 2005) become less important as the dynamical time-step reduces to LES scales, LES models usually employ explicit strategies for passing physics increments to the dynamics. Therefore physics–dynamics coupling represents one area where we can look for efficiency gains to reduce the computational cost.

The typical timescale for "fast" physics processes in NWP models (turbulence, condensation/evaporation) is of the order of 10 s. Therefore, as the dynamical time step reduces beyond this, consideration should be given to whether these processes can now be considered "slow" (alongside radiation) and coupled less frequently or less tightly to the dynamical core, with obvious cost savings. Similarly, as time steps become shorter, the requirement for an implicit coupling between the surface and the atmosphere becomes less important, potentially allowing easier explicit coupling to distributed canopy models (e.g., Bengtsson *et al.*, 2017).

Lack of parallelism in the time dimension provides a fundamental limit on the scalability of HMs. To achieve a 24-hour forecast requires an order of magnitude more time steps than it did at a kilometre scale, hence, even for a model that is perfectly scalable in the horizontal domain, time to completion will inevitably be longer. Additional scalability can be incorporated by concurrent calculation of physics and dynamics processes on their own dedicated processors (e.g., Heidari *et al.*, 2021), significantly reducing the time to completion. In the era of exascale computing, techniques like this will allow for not only efficient modelling strategies, but also significant improvements in physical parameterisations, for example, 3D radiative transfer running on a GPU coupled to a traditional dynamical core running on a CPU. Finally, in the long-term, research into parallel-in-time methods for the dynamical core (Christlieb *et al.*, 2010) may also provide useful extra scalability.

Traditionally, NWP models have coupled dynamics and physics using the same spatial grid, but recently there has been greater interest in breaking this paradigm and allowing coupling of different processes at different scales (e.g., Brown *et al.*, 2024). The cost savings that could be achieved by this are clear, for example, if we want to incorporate detailed aerosol or chemistry schemes into HMs, which have no need to be run at the grid scale of the dynamics. However, there are also significant opportunities for improvements in the coupling to be made through this increased flexibility. The resolved or filter scale of models is not the same as the grid length, and improvements to the grey-zone turbulence problem could be achievable by utilising this information (e.g., Germano, 1992) and allowing parameterisations to operate at the physical scale they are designed for, rather than the grid scale of the dynamics. Improved spatial coupling methods will also allow the exploration of higher order finite-element methods in dynamical cores, effectively allowing for better "resolution" of the dynamics for a given grid scale and allowing this additional information to be utilised by physical parameterisations.

4 | ATMOSPHERIC PARAMETERISATIONS

A key issue for hectometric modelling is to understand the advantages and challenges they pose in terms of the parameterisations required in contrast to km-scale models. In the vast majority of cases, km-scale models do not resolve even the largest eddies, so the turbulence is fully parameterised. The additional step of assuming forcing is only slowly varying in space leads to traditional one-dimensional (1D) column "planetary boundary layer" (PBL) parameterisations. However, as resolution increases, models can enter the "terra incognita" (Wyngaard, 2004) or the "turbulence grey zone" (TGZ, Honnert *et al.*, 2011, 2020), where larger turbulent eddies begin to be partially resolved, or "permitted." This is, in principle, the realm of LES, but the objective of true LES is to resolve the largest energy-containing eddies and so also avoid (where possible) the TGZ.

Experience with LES provides us with clear guidance as to what being in the TGZ means in practice. Under typical daytime convective boundary layer (CBL) conditions, for example, an LES will be within the inertial sub-range for most of the CBL, and so the coherent resolved structures accurately produce most of the turbulent fluxes. In this case, the remaining three-dimensional subfilter turbulence can be reasonably well parameterised as an eddy diffusivity in a scale-aware manner in all three dimensions. This is usually done via either a diagnostic (Smagorinsky) or a prognostic turbulent kinetic energy (TKE) equation and by assuming that the spectral transfer of turbulent kinetic energy is equal to the energy dissipation at the Kolmogorov scale. In principle, a grid length less than about $z_i/60$ is needed for a reasonably well-converged LES solution (Sullivan & Patton, 2011), where we have taken the "Smagorinsky constant," C_s, to be about 0.2. Even for a typical daytime peak boundary-layer depth, z_i, of around 1 km, then, an LES requires a 15–20 m grid or better. In practice, the CBL is quite "amenable" to simulation at coarser resolution (Mason, 1994), because such a large fraction of the vertical turbulent fluxes are carried by the largest eddies, and at least some situations can be simulated tolerably well with 100-m horizontal grid spacing. Thus, HMs can be considered as approaching LES.

However, the overlap is small at best and, in many circumstances (such as near the surface or top of convective boundary layers, or earlier in the diurnal cycle of the CBL), most of the turbulent transport is still subfilter, a source of uncertainty, and so HMs still lie in the TGZ. At the other end of the scale, stable boundary layers (SBLs) generally have much smaller energy-containing eddies and LES requires resolutions of only a few metres (Beare, 2006), so turbulence remains "fully parameterised" in HMs. Figure 2 is a schematic picture of how a model represents various boundary-layer depths and cloud sizes.

Thus, while some applications may avoid the turbulent grey zone, many operational NWP centres will soon be running routinely with hectometric grids that are very much in the TGZ, with phenomena where turbulent transports are critical but remain unresolved (from SBLs to the entrainment zone at the top of mixed layers, especially for stratocumulus clouds) to regimes of largely resolved turbulent variability. Ideally a turbulence scheme would move seamlessly from LES-type at the well-resolved end to a good, fully parameterised, 1D PBL scheme at the other. Because km-scale models start to resolve deep convective overturning, they mostly operate with only a shallow cumulus parameterisation, or no cumulus parameterisation at all, while deep convection is left to the resolved dynamics. Turbulent boundary-layer processes are still fully parameterised, typically using 1D turbulence schemes that represent vertical transport by the full spectrum of turbulent eddies. How to deal with horizontal turbulent fluxes in deep convective clouds is not well understood (e.g., Hanley *et al.*, 2015). Cloud condensation schemes still diagnose subfilter cloud fraction and subfilter condensed water, usually based on subfilter variability of temperature and moisture. All cloud microphysics schemes in these models (should) use prognostic hydrometeors (essentially because the timescale for lateral advection is typically short compared with the timescale for fallout), but differ in whether they prognose mass (one-moment) or also number density (two-moment), and in the choice of categories of ice. Radiative transfer

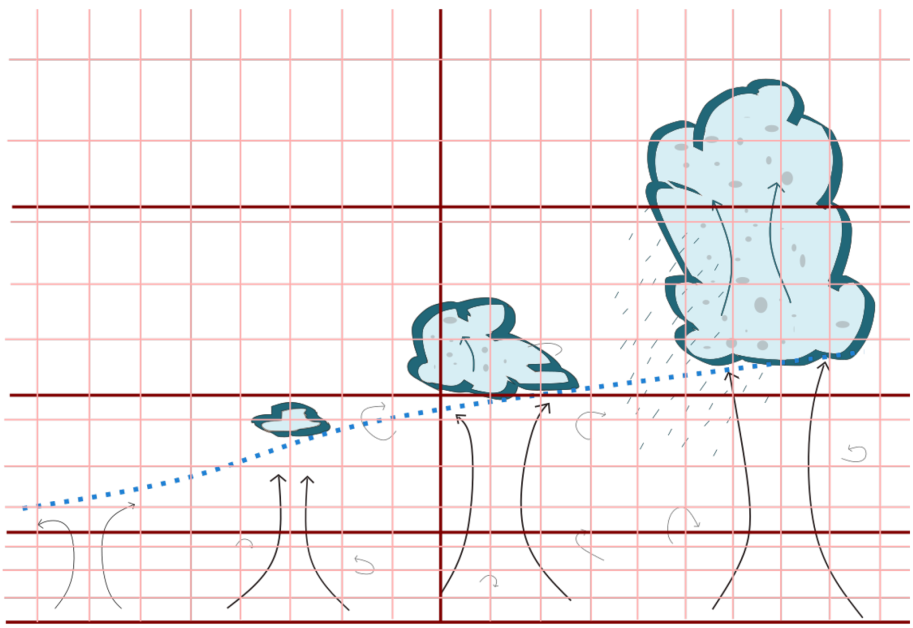

FIGURE 2 Schematic illustration of how a model represents various boundary-layer depths and different cloud sizes between km-scale grid lengths (dark red) and hectometric grid lengths (light red). As the boundary layer deepens, hectometric models are able to represent larger eddies. Similarly they are able to represent growing convective clouds once they become large enough. [Colour figure can be viewed at wileyonlinelibrary.com]

is calculated exclusively in the vertical direction, without taking into account explicit 3D effects.

HMs treat even more of cloud systems through the resolved dynamics. At the same time as treating 3D turbulence, unresolved cloud condensation processes need a consistent parameterisation. Because of their fine resolutions, LES models usually do not take into account subfilter variability of clouds, so grid boxes are simply assumed to be cloudy in the case of supersaturation and cloud-free otherwise. However, subfilter variability of condensation will need to be taken into account in the TGZ, although it is not yet clear how complex a scheme is required. The presence of clouds may thus exacerbate the problem of moving seamlessly from LES to fully parameterised processes.

The main advantage for the cloud microphysics at these resolutions is that the impacts can be picked up directly by the resolved dynamics. A prime example of this is how evaporative cooling of precipitation enhances the resolved downdrafts, which is the main driving mechanism of cold-pool structures. The cloud microphysical processes themselves, however, remain completely subfilter at these fine resolutions. Therefore the whole complexity of the microphysics still needs to be fully parameterised.

Three-dimensional effects of radiation become increasingly important at higher horizontal resolution. However, although three-dimensional codes for radiation are available now, the high computational cost of these parameterisations has prevented the interactive use of such codes. In future, machine learning emulators may be able to offer a route to doing this at much lower cost.

In short, the main difference between hectometric and km-scale model resolutions is that in the high-resolution case a shallow convection scheme is less likely to be required; a relatively simple subfilter cloud condensation scheme may suffice, but the one-dimensional turbulence scheme is replaced by a three-dimensional scheme. For the latter, the length scale used under convective conditions should asymptotically approach one, proportional to the grid scale.

4.1 | Turbulence representation

A key challenge for the representation of turbulence and for HMs more generally will be to handle the TGZ, that is, achieve an accurate transition from unresolved to resolved, both temporally (as for the morning transition from stable to convective PBL) and spatially (e.g., advecting from a cold rural surface to a relatively warm urban one). An overview is given by (Dudhia, 2022). Pragmatically, the starting point will likely be to run with existing parameterisations—either LES or km-scale NWP, but either is likely to result in systematic errors. For example, neither scheme is designed to give an accurate handover from parameterised subfilter turbulence to resolved, which may result in errors in the vertical structure of transitional PBLs. Similarly, neither is designed or tested specifically to represent turbulence outside the boundary layer at hectometric scales, which is likely to lead to errors in the growth rate of shallow cumulus clouds into precipitating convection, via the parameterisation of entrainment. This might also be manifest as a systematic dependence of statistics such as cloud-size distributions on the grid size.

A fundamental difficulty is that neither 1D PBL schemes nor eddy-diffusivity schemes in LES are designed to deal with the 3D nature of the largest overturning circulations. Their net vertical transport is often treated via "countergradient" terms in PBL schemes, while they are expected to be fully resolved in LES. It is likely that improved schemes will need to include non-downgradient transports in 3D. Some progress in this direction has been made over the last decade or so, through implementation of non-downgradient terms that have been derived from a variety of routes ("Leonard," "tilting" or "Hgrad" terms; (Hanley *et al.*, 2019; Moeng *et al.*, 2010; Verrelle *et al.*, 2017).

Related to this is the fact that the 1D PBL scheme is based upon the concept of quasi-equilibrium; turbulent kinetic energy is generated at "large" scales (on the scale of the PBL) and cascades down to small. The overall turbulent mixing is determined by the balance of this production and dissipation; the overall shape of the horizontally averaged turbulence spectrum and hence the subfilter component for any choice of spatial filter is then determined by the production length scale at any given height. This justifies the idea of blending PBL and LES schemes somehow (Boutle *et al.*, 2014a). However, one of the main drivers of the need for HMs is to treat rapid transitions in space and time, where such an approach cannot be justified; more local (3D) determination of turbulent dissipation (and hence length scale) is likely to be needed. This is especially true for cumulus clouds, especially deep clouds, where individual clouds may be far from any identifiable larger-scale equilibrium, and such an equilibrium may not even exist (Done *et al.*, 2006). Here, the so-called dynamic method (Germano *et al.*, 1991) may offer a way forward.

Finer horizontal grids raise the question of also refining the vertical resolution (see also Section 2). For many surface types, however, this is likely to require significant revision to how we couple the surface fluxes, as the surface elements become resolved in the vertical, discussed in Section 5. More generally, traditional surface exchange is based on averaged statistics of the turbulence and their applicability in a turbulence-permitting model is yet to be demonstrated. Finally, there is a significant technical challenge for many NWP models, where parameterisations are currently coded to operate only in the vertical (treating grid columns independently). A turbulence scheme in a model that resolves aspects of turbulent flows will need to know the full three-dimensional structure of the local flow and equally to transmit turbulent fluxes horizontally as well as vertically (Honnert & Masson, 2014).

It is helpful to consider developments from both ends of the spectrum (i.e., 1D PBL schemes and 3D eddy diffusivity). We start with the latter.

1. As the filter scale gets larger than the strict inertial sub-range (where mixing length is universally proportional to filter scale), how do we determine a mixing length in the eddy viscosity/diffusivity? This is particularly an issue for deep convective clouds.
    - Can we parameterise a priori in terms of the local resolved flow? If so, the (moist) stability of the resolved flow is likely to be a key parameter.
    - Will a local, online determination using the Germano dynamic approach provide better results (Efstathiou *et al.*, 2018; Germano *et al.*, 1991)?
    - How far from local equilibrium are the sub-filter TKE and scalar variance equations, and hence do we benefit from a prognostic approach?
    - Can we pragmatically benefit from and theoretically justify different horizontal (one- or two-component) and vertical diffusivities? If so, is this driven by anisotropy of turbulence (via the anisotropy of production) or is further anisotropy needed in the length scale?
    - If anisotropy of turbulence is dominant, can we benefit from a separate prognostic for the vertical component of TKE?
2. Do we need stochastic backscatter? Does the Germano dynamic approach provide sufficient inherent variability? Do we need or benefit from stochastic perturbations at lateral boundaries?
3. We know that turbulent fluxes are not always downgradient. How much do non-local scalar variance and tilting/Leonard/Hgrad terms (Hanley *et al.*, 2019; Moeng *et al.*, 2010; Verrelle *et al.*, 2017) contribute to these countergradient fluxes?
4. Assuming that the tilting/Leonard/Hgrad terms are important (which we believe has already been demonstrated), can we use the same length scale as in the eddy viscosity/diffusivity, or can we adapt the approach to provide an additional length scale?
5. Given that we anticipate the greatest need for improved schemes to be in cases with rapid transition in space and/or time, how do we evaluate enhancements to schemes?

The above can be phrased as, "How do we build a better scheme than Smagorinsky as we reduce resolution into the TGZ?" Structurally, the PBL scheme typically contains downgradient and non-downgradient terms. The former may be determined locally (in much the same way as Smagorinsky) or non-locally (recognising the overall PBL structure). The latter is generally determined non-locally, and often expressed as a flux profile or, more recently, using the mass-flux approach. The Leonard/tilting/Hgrad terms cannot appear explicitly (because, by construction horizontal gradients are assumed zero), though their vertical counterparts do appear in some formulations.

We now consider the opposite viewpoint of the PBL scheme.

1. How do we ensure that, as we increase resolution, we decrease parameterised fluxes in such a way that the fluid becomes unstable *and* develops the largest overturning structures with the right horizontal scale and amplitude? (This already applies to CPMs outside the PBL—we know that we are currently failing to develop deep clouds with the right scales and intensity.)
2. What role do the Leonard/tilting/Hgrad terms play in controlling these scales? Can we improve their formulation to improve this aspect?
3. How does the non-countergradient flux become divided between resolved flow and the components of the 3D scheme? Does this necessitate use of more complex formulations than just local downgradient flux (possibly including anisotropic diffusivity)?

Some of these questions can be answered by reference to extremely well-resolved LES; however, much has already been done to at least provide the tools to establish whether turbulence schemes will reproduce "standard" phenomena (convective PBLs, stratocumulus, cumulus) well. It is not at all clear that high-resolution reference simulations will be sufficiently reliable to help test and improve HMs for the complex situations we anticipate applying them to. Thus, there is a need for observations, including field campaigns, carefully designed to answer these questions. This is particularly true for turbulent interactions with complex, heterogeneous surfaces such as urban areas and complex terrain or, indeed, both, where we need further understanding of how surface-driven turbulence interacts with the atmosphere (e.g., through the urban canopy or high-rise buildings).

4.2 | Representation of clouds

One of the main drivers for HMs is the removal of the need for any kind of deep convection parameterisation. For the most part, shallow convection will be treated explicitly, but it will be desirable if the 3D turbulence scheme tends to a reasonable shallow cumulus and stratocumulus scheme, as their scales become so small they are unresolved. Thus, the treatment of "a cloud" will largely be through the parameterisation of subfilter variability. Although LES models often have no such scheme, it is arguable that this is an approximation that, at least, makes demonstration of grid convergence of solutions more difficult. However, as Smagorinsky becomes a good subfilter model, when the filter scale becomes (much) smaller than the cloud scale and clouds are reasonably well represented, a simple Gaussian scheme based on it (e.g., Deardorff, 1980) is likely to be sufficient. If the TGZ is entered and the filter scale includes both "cloud" and "environment," a more complex scheme might be needed, such as a Gaussian mixture scheme approaching the 1D schemes proposed for fully parameterised clouds (e.g., Larson *et al.*, 2001) or a bi-model approach that is relatively simple, whilst still adressing the clouds and entrainment issues (Van Weverberg *et al.*, 2021).

4.3 | Microphysics

If cloud updraughts and downdraughts are treated explicitly in HMs, the cloud microphysics parameterisation has the opportunity to respond to (hopefully) much more accurate forcing. Thus, the microphysics scheme becomes relatively more important in the prediction of clouds and precipitation. Double-moment schemes introduce two prognostic variables (typically mass and number) per hydrometeor class. It is therefore important to determine classes in which it is essential to have double moment, to prevent increasing memory usage and the computational costs of weather models unnecessarily. Previous studies (e.g., Bryan & Morrison, 2012; Field *et al.*, 2023) show that double-moment schemes can be superior to single-moment schemes, especially at higher spatial resolution in the simulation of deep convective precipitating systems. As always, more degrees of freedom does not, in itself, guarantee accuracy. It is also not always clear whether a single-moment scheme could be "tuned" to give equally good results, but this procedure is less justified with less parameterised forcing. However, some processes (such as auto-conversion in the initiation of warm rain) are difficult to represent well with a single moment. As previous sections have discussed, HMs are still not resolving all of the variability that is present in turbulent cloudy environments, and therefore we must still consider that subgrid variability could have a significant effect on microphysical process rates (e.g., Boutle *et al.*, 2014b).

While different approaches to turbulence will be developed, perhaps with strengths in different applications, there should be agreement on the desired "truth" to be compared with. This is likely not so for cloud microphysics, where many uncertainties remain. HMs should provide a much better environment for microphysics, and, combined with good field data, may help constrain microphysics schemes, but we may need to consider the need for multi-parameterisation approaches to microphysics. Some key processes are very hard to observe, especially in clouds (e.g., buoyant production of TKE). What diagnostics can we rely on to tell us our models are right or wrong? How do we assess these statistically? The development of "piggyback" methods for separating parameterisation from feedback might give insights (Grabowski *et al.*, 2019). Likewise, we already have the tools to implement interactive chemistry/aerosols in HMs, and they will provide a more realistic environment. The question is whether we gain predictive skill in doing so, and a great deal of work will be needed to demonstrate whether we do.

4.4 | Radiation representation

HMs will resolve much of the cloud-size distribution and can provide a far more detailed description of complex (urban) surfaces than current weather models. This poses challenges to the treatment of radiation, which is generally performed using vertically operating two-stream methods that cannot produce correctly positioned cloud shadows and shading induced by complex surfaces. The cloud size distribution is also potentially important for solar energy forecasting.

Increasing the amount of streams (Jakub & Mayer, 2015) or using ray tracing (Veerman *et al.*, 2022; Villefranque *et al.*, 2019) can solve these problems, but any of those solutions is for now too costly for forecasting purposes, as radiative transfer is already one of the most costly model components. Development of artificial intelligence (AI)-based emulators is likely to be a way forward and is already being pursued by a number of centres, but producing sufficient training data to cover all relevant cases is a significant constraint. The ultimate challenge in the field of radiation is hence to include only those aspects of the 3D nature of radiation that are expected to increase the quality of the model significantly. Based on Jakub and Mayer (2017) and Veerman *et al.* (2022), we suggest that providing the model with correct surface solar irradiance fields is the key challenge, while modifications to in-cloud heating and cooling rates are an interesting second step, but less crucial.

The higher level of detail in clouds in HMs also provides the opportunity for better forecasts of cloud-induced variability in radiative transfer, and will benefit from a detailed and consistent exchange of information on effective radii of cloud droplets and ice particles between microphysics and radiation.

5 | SURFACE REPRESENTATION

Hectometric resolution models are expected to have their largest benefits in the ability to resolve meteorological processes and in resolving processes resulting from surface heterogeneity. In terms of the latter, this is particularly relevant for complex areas such as cities and mountainous areas. A flip side of this benefit is that, unlike the case of parameterisations in the atmosphere, where HMs simplify the representation of some processes by resolving them better, generally surface parameterisations get more complicated because many different and complex processes at these small scales become important.

A typical scale of variability in cities is the neighbourhood scale, which is of the order of a couple of hundred metres; at this scale, one can suppose that the urban features are relatively homogeneous (e.g., all high-rise in Central Business Districts in American or Asian cities, residential houses in some suburban areas). This is why the concept of the "local climate zone" (LCZ) emerged (Stewart & Oke, 2012) to describe them and their impact on the local meteorology. LCZs can be very different from one another. However, this raises challenges, not only for parameterisations, which will be described here, but also for fine-scale urban parameters (see Section 6).

Mountains and places with complex orography will also benefit from HMs. The orography will be better represented, with steeper slopes, leading to a better mountain meteorology representation. In addition, benefits are expected in situations where even small relief impacts local meteorology, such as fog in small river valleys and hilly terrain.

Because of their application objectives, the numerical setup of HMs will also provide a constraint that will affect surface-process parameterisations. For many applications, there is a need for a first atmospheric layer that is very thin (1 or 2 m above the surface): for example, the representation of fog for airport security.

5.1 | Representation of urban areas

In order to reap the benefits of HMs, one needs urban canopy surface models. These urban canopy models are still not general in most forecast models, even at kilometric scale. State-of-the-art urban canopy models (Lipson *et al.*, 2023) represent the 3D shape of the city (e.g., using an urban canyon, urban cubes, statistical relationships of distance between walls, etc.), because the 3D structure is crucial to reproduce the most important processes: (1) radiative trapping due to multiple scattering of short- and long-wave radiation inside the canopy; (2) large surface areas with high thermal inertia that exchange heat and moisture with the atmosphere; and (3) friction. Such models exist and will be essential in HMs.

Anthropogenic sources of heat, due to domestic heating or air conditioning, also contribute to urban climate. While air-conditioning impact is generally limited compared with the role of the Sun, it can have a large influence on the surface energy balance at night. To improve the representation of anthropogenic processes, as well as potential energy consumption impact indicators, inclusion of a building energy module is necessary. It can be improved by considering human behaviour, that is, people's interaction with heating, ventilation, and air conditioning. This requires interdisciplinary approaches.

Wintertime urban heat waste impacts are less explored by the community. They are still of major importance for high-latitude cities, for which the urban heat island is not solar-driven in winter. There is often less snow in cities, and it is often removed mechanically, potentially leading to albedo contrasts with the snowy countryside.

It is important to consider urban vegetation, either grass or high vegetation. Its subgrid representation should be within the urban tile and interactions with other urban elements is recommended. This improves the geometric representation of the urban canopy, since the fraction of urban tile does not have to be decreased to allow for vegetated tile fraction in urban areas, thus parameters such as canyon height-to-width ratio remain accurate (and streets are not too narrow). There is a strong social driver for increasing our understanding of the effect of vegetation and park layout on the urban microclimate.

The best way to represent the 3D structure of buildings in surface models is still a research question. At 100-m grid length the building grey zone is entered, where the 3D structure of large buildings and the flow around them become partially resolved. This opens the question of whether seamless urban canopy parameterisations need to be developed, where subgrid turbulence and drag interactions are parameterised with conventional multi-layer urban canopy models, but larger buildings and the large-scale flow around them are resolved explicitly on the grid. The 3D structure of buildings also has an influence on the radiation budget. It should be noted that the choice of the simplified geometry in the urban canopy model induces relationships between form-derived parameters (e.g., wall surface and sky view factor), which makes them no longer independent. There is then the need to choose one of these parameters to define the building geometry as input to the model in order to ensure energy conservation. A long-term suggestion is to compute the energy exchanges with the full 3D shape of the city, including individual buildings, at sub-hectometric scale, even if other aspects (e.g., turbulence and drag) are still estimated at hectometric scale.

5.2 | Multi-layer coupling between buildings and the atmosphere

A great advantage of representing cities at the urban scale is that the need for representing multiple tiles in a single grid cell is reduced. However, there are scenarios when one would still require multiple tiles in a cell: for example, the representation of rivers and the edges of a park. However, the increase in horizontal resolution also comes at a cost. The inherent heterogeneity of the urban surface implies that horizontal exchanges between adjacent tiles—even within the canopy—become important. Most surface energy balance schemes have traditionally assumed that the horizontal scales are sufficiently large, such that the different land-use classes within the tiles have had the chance to "blend," implying that only vertical transport needs to be taken into account.

In addition, the lowest atmospheric level of NWP models is always assumed to be above the urban and vegetation canopy, allowing for a clean separation between the surface exchange processes, which are calculated in the surface model, and the atmospheric flow aloft. The need for including lateral transport implies that the lowest atmospheric level needs to be at ground level, such that the surface exchange scheme can be fed information about the flow in the lowest levels of the atmosphere. This has been implemented in Weather Research and Forecasting (WRF: (Martilli *et al.*, 2002)) and MesoNH (Schoetter *et al.*, 2020), and allows simulation particularly of high-rise cities (e.g., Hong Kong in the case of the latter reference).

Once the surface model and the atmospheric models overlap, the drag, sensible, and latent heat fluxes need to be distributed over height (Figure 3). This is particularly pertinent for high-rise areas, which can occupy a significant percentage of the boundary-layer depth. Drag forces tend to be concentrated near the top of buildings, particularly for buildings that are not shielded by other buildings (Sützl *et al.*, 2021b). This has a substantial effect on the surface stresses and 10-m wind in built-up areas (Sützl *et al.*, 2021a). The sensible and latent heat fluxes will also need to be distributed over the canopy depth, in a manner that is consistent with drag. Indeed, if only the drag is distributed but the other fluxes are imposed at ground level, atmospheric temperatures at ground level will be overestimated substantially (Schoetter *et al.*, 2020).

Explicit representation of the flow within the urban canopy also requires parameterisation of the vertical mixing inside the urban canopy. Within the canopy, two distinct transport processes can be discerned: turbulent fluxes and dispersive fluxes. The former occur due to turbulence, whereas dispersive fluxes arise from an inhomogeneity in the mean velocity due to the presence of obstacles. These vary depending on the neighbourhood (or LCZ) type (Nagel *et al.*, 2023). Both will need to be represented in surface exchange models, as buildings are not represented explicitly. These are typically based on mixing-length formulations (Blunn *et al.*, 2022; Martilli *et al.*, 2002; Schoetter *et al.*, 2020). In addition, the stability of the atmosphere will play a role—convective boundary layers have much higher levels of turbulence, which is transported down into the urban canopy, leading to more vigorous mixing (Grylls *et al.*, 2020). Furthermore, there is a difference between the mixing of momentum and scalars (Blunn *et al.*, 2022).

Some models use a fully implicit coupling between the surface model and the atmosphere (Joint UK Land Environment Simulator (JULES) + Unified Model (UM)), whereas others use an explicit coupling (MESO-NH). The latter is clearly much more straightforward to develop and maintain than the former. The main argument for using a fully implicit coupling is that there is no time-step restriction, which is advantageous for use in climate modelling predictions. A fully explicit approach for the town energy balance (TEB) urban canopy scheme is implemented within the Aire Limitée Adaptation dynamique Développement InterNational (ALADIN) regional climate model, which has been used successfully for grid sizes up to 10 km and time steps of 7.5 min (Daniel *et al.*, 2019).

There is a clear need for robust and accurate parameterisations to represent the effects of urban areas in these high-resolution and vertically distributed surface exchange schemes. Large eddy simulation can play an important role in developing these parameterisations (Blunn *et al.*, 2022; Nagel *et al.*, 2023; Sützl *et al.*, 2021a) using both idealised and realistic geometries. A particular challenge here is the representation of 3D radiation effects and how these need to be represented at hectometre scale (Schoetter *et al.*, 2023).

Similar to urban areas, areas with high vegetation also may need to be coupled within the atmospheric layers if the lowest model levels are closer to the surface and the blending-height assumption is no longer valid. For LES, these multi-layer coupling approaches have already been tested (e.g., Patton *et al.*, 2016). In addition, Bonan *et al.* (2021) summarised that, besides an improved representation of the physical processes, using a multi-layer plant canopy model also gives demonstrated improvements in radiation, wind speed, and temperature profiles. However, more research needs to be done about the best way to include multi-layer vegetation canopy models in hectometric NWP.

5.3 | Complex orography

Hectometric scales will represent complex terrains better, producing much steeper slopes and narrower valleys. Therefore, HMs will reproduce mountain meteorology and associated local phenomena such as the valley breeze system of katabatic/anabatic winds (e.g., Goger & Dipankar, 2024), cold pools, and eventually fog (Smith *et al.*, 2021) much better. Forcing of local convection above ridges, for example, by anabatic winds, and the resulting clouds will be much better represented.

While anabatic flows can be represented adequately with what we can now consider as a typical vertical grid resolution near the surface (5–10 m), katabatic flows require much finer vertical resolution. Brun *et al.* (2017) use a vertical resolution of 1 m (first mid layer at 50 cm) in order to represent a katabatic flow 10 m deep that was observed on a 35° slope. Katabatic wind thicknesses can be even smaller above snow, with a thickness of just a couple of metres. Therefore, a very high resolution is required; typically it is recommended to have the first layer at 1 m height to simulate mountain flows in all conditions.

However, as seen in the numerics (Section 2), steeper slopes combined with a very thin first atmospheric layer are challenging for dynamical cores in terms of stability and horizontal pressure gradient. Numerical diffusion along terrain-following layers instead of horizontally can also degrade the representation of the stability along the vertical (e.g., strong inversions above cold pools that can sometimes reach 10K in reality). Similarly, true horizontal diffusion close to the surface might be problematic where there is a strong inversion there. These issues need to be looked at in numerical setups.

Shadows due to topography also impact the differential temperatures between sunlit and shaded sides of a valley strongly. In winter, this leads to, and governs, very variable snow cover (altitude, snow depth, snow mantle state, etc.). The differential temperature between valley sides also evidently affects valley flows (as shown, e.g., near the Mont Blanc massif in Sabatier *et al.* (2020a, b)). Such influence has a seasonal response.

Therefore, these radiative effects should be taken into account. Shadow calculations and solar irradiance generally consider only local slope (in the grid mesh), but this should be extended to the shadows cast by nearby mountains. This therefore requires simulation of an effect propagating from one horizontal grid mesh to at least another one in the domain. These mountain 3D shadows can be simulated, correcting the incoming direct solar radiation provided by a classical 1D radiation scheme. In MesoNH, AROME, and UM, this is done by checking whether ray-tracing towards the Sun encounters a triangulated orography computed from the grid, or by using precalculated ridge lines above the horizon, respectively. A global correction on sunlit sloping points at the scale of the domain is then necessary to ensure energy conservation. This approach introduces an inconsistency of considering the shadows of clouds above, but mountains in the direction of the Sun. Therefore, for future research, it is highly recommended to study how to implement full 3D radiation schemes, either ordinate methods or Monte Carlo ones. In addition to much better physical coherence, it also allows consideration of both short-wave (solar) and long-wave (terrestrial) 3D radiative exchanges, within the complex surface at least, and also with 3D clouds at best. This is a strong challenge in terms of computation cost, coupling, and parallelisation methods.

The high spatial variability of snow that can be represented at hectometric scale requires consideration of the use of snowpack data assimilation. Satellite images provide snow cover at high resolution, and so will allow such implementation in the relatively foreseeable future. A complication, however, is that snow drift by the wind

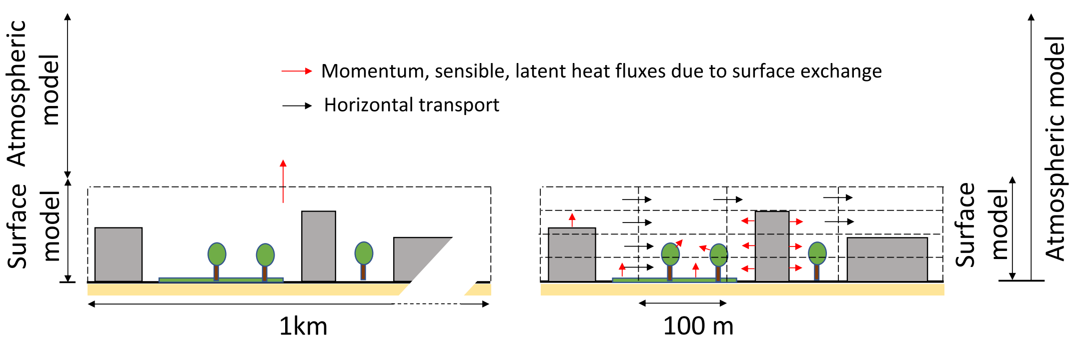

FIGURE 3 Surface exchange schemes at km scale versus hectometric scales. [Colour figure can be viewed at wileyonlinelibrary.com]

also modifies the snow mantle at hectometric scale. Snow already on the ground can be transported from peaks and mountain passes towards less exposed areas, that is, from one grid point to another. To implement resuspension of the snow in the atmosphere and drifting of snow at the surface, as in Vionnet *et al.* (2021), would then be pertinent. The question of subsurface water transfer due to local hydrological flows at small scale may also be an issue.

Of course, there are cities located by mountains and in valleys. In such cities, the mountain flow interacts with urban heat island effects. Generally, valley cold pools will lead to colder conditions in city centres located on the valley floor, while less dense suburban areas 100 or 200 m above along the slope may be above the inversion and much warmer (Masson *et al.*, 2020b). Polar cities located on small (50-m high) hills can, in contrast, encounter very strong urban heat islands compared with the tundra below (Konstantinov *et al.*, 2018), due to both topographic and anthropogenic heat effects. HMs will be of huge interest to represent and forecast the weather better in complex-terrain settlements.

5.4 | Other challenges

While land cover is generally static at kilometre scale, small temporary changes like flooded areas, seasonal lakes, or even tidal zones are "seen" at finer scales. Soil moisture variability also shows larger spatial heteorogeneity. The need to represent all these effects must be explored in the light of potential applications, and may raise issues in land-cover retrieval (see Section 6).

Wildfires provide a heat flux of several tens of thousands of W·m^{-2}, together with significant emissions, which influence the atmospheric flow and aerosols. They are very transient and occur on lines. Such wildfire can lead to pyro-convection and deep thunderstorms. Wildfire models able to be coupled to hectometric meteorological models already exist (Costes *et al.*, 2021).

At kilometric scale, there are several existing parameterisations describing the increase in drag and increase in TKE caused by a wind farm within the grid cell, which impact several layers (e.g., Fitch *et al.*, 2012; van Stratum *et al.*, 2022). When going towards a hectometric scale, we may need to consider individual wind turbines. In the near future, offshore windmills will be 200 m or even 300 m wide, therefore impacting several grid points. This can be done using actuator disc (or, better, rotating actuator disc) parameterisations, which can be validated against LES at 10-m resolution using actuator line parameterisations, which represent each blade individually (Joulin *et al.*, 2020). Moreover, wind turbines will orientate on wind direction, influencing different grid points depending on the meteorological conditions. This will require a parameterisation of the piloting of the wind-turbine direction and blade rotation speed.

6 | SURFACE DESCRIPTION

Accurate surface description is fundamental in NWP for calculating the exchanges of water, energy, gases, and other compounds between the surface and the lower boundary of the atmosphere. The main benefit of moving to a hectometric grid-length surface description is the potential for higher-detail information on surface heterogeneity. Heterogeneous environments include urban areas, mountains, boundaries between land and water, and tree canopies. We give particular attention here to urban areas, as they are the environment with simultaneously the most hectometric NWP applications (e.g., heat stress, climate resilience, air quality) and surface description challenges. We discuss state-of-the-art datasets, hectometric data requirements that are not currently met, and necessary future developments for land use and land cover (LUC) (Section 6.1), urban form, fabric, and function (Section 6.2), time-varying surface parameters (Section 6.3), and orography and soil properties (Section 6.4).

6.1 | Land use and land cover

In NWP and climate models, surface–atmosphere exchanges are represented by Land Surface Models (LSMs), which typically have a categorical description of the LUC. Each class is associated with an array of parameters describing its biogeophysical and chemical characteristics. LUC maps are typically produced by gathering a set of features of the surface into classes or labels, which are identifiable and which the map producer wants to distinguish. As such, there is no universal way of producing a LUC classification. This means that National Meteorological and Hydrological Services (NHMSs) use different LUC datasets in their models.

As described in Walsh *et al.* (2021), the Global Land-Cover Characteristics database (GLCC; Loveland *et al.*, 2000) is used in Integrated Forecasting System (IFS) cycle 47r1 (ECMWF, 2020) of the European Centre for Medium-Range Weather Forecasts (ECMWF). The UK Met Office Unified Model can be run with GLCC (also known as IGBP) or the European Space Agency (ESA) Climate Change Initiative (ESA-CCI: (version 2.0.7; Defourny *et al.*, 2017)) land-cover classification datasets for global and regional configurations, and the operational UK limited-area configuration uses the Institute of Terrestrial Ecology (ITE) land-cover classification dataset (Bunce *et al.*, 1990) for Great Britain and ESA-CCI for the outer domain. The HIRLAM consortia runs the HARMONIE-AROME canonical model configuration (CMC) of the shared ALADIN-HIRLAM NWP system for short-range operational weather forecasting, which uses the ECOCLIMAP global land-cover database developed by Météo-France in partnership with the scientific community (CNRM, 2018; Faroux *et al.*, 2013; Masson *et al.*, 2003). The Consortium for Small-Scale Modelling (COSMO: (Doms *et al.*, 2011)) and COSMO-CLM (Rockel *et al.*, 2008) models can use a variety of global land-use datasets (e.g., Global Land-Cover 2000 (Bartholome & Belward, 2005), ECOCLIMAP, and ESA-CCI), which are made available by the External Parameter for Numerical Weather Prediction and Climate Application (EXTPAR) tool (Asensio *et al.*, 2020).

The operational global datasets are typically coarser resolution than 100 m and are based on relatively "old" datasets. For example, the second generation of ECOCLIMAP (ECOCLIMAP-SG) introduced in 2018 has 300-m resolution and is based on ESA-CCI v1.6.1, which was developed in 2012. High-resolution remote sensing imagery and new techniques have led to the recent emergence of very high-resolution land-cover classification datasets such as Globland30 (Jun *et al.*, 2014), ESRI2020 (Karra *et al.*, 2021), and ESA WorldCover (Zanaga *et al.*, 2022). These sub-hectometric datasets provide the opportunity for a step change in land-cover representation in hectometric NWP. Land-cover class pixels within a hectometric NWP grid cell can be aggregated to calculate land-cover fractions, thus providing information on the sub-hectometric scale land-cover heterogeneity.

However, these very high-resolution datasets are not a land-cover panacea. Even at 10-m resolution, satellite remote-sensing based land-cover maps often do not detect in-canopy vegetation, since individual trees lining streets or small residential gardens are often not resolved (Figure 4c,d compared with Figure 4a,b). There are promising satellite remote-sensing approaches to tackling this based on calculating fractional rather than class-based land cover at the pixel level, using the relative amounts of vegetative and impervious spectral signature (Haase *et al.*, 2019; Shahtahmassebi *et al.*, 2021). Also, very high-resolution datasets have a smaller number of coarse classes (Globland30 (10), ESRI2020 (10), ESA WorldCover (11)) compared with the ones used operationally (ECOCLIMAP-SG (33), GLCC (17), CORINE (43)). Complementary information about trees, crop types and urban form, fabric, and function (Section 6.2) is required. Thus, there are still challenges that must be overcome to benefit fully from the recent availability of global sub-hectometric resolution land-cover classification datasets.

6.2 | Urban form, fabric, and function

Urban surface exchange schemes require information on the urban form (e.g., plan area density, wall area fraction, building orientation, building dimensions), fabric (e.g., albedo, emissivity, thermal capacity, window fraction), and function (e.g., land use, population density) (Masson *et al.*, 2020a; Oke *et al.*, 2017). Parameterisations assume that subgrid surface heterogeneity is the same throughout each grid cell (or tile in the case of tile schemes) and that the flow is in equilibrium with the grid cell (or tile) surface (Claussen, 1990; Coceal & Belcher, 2004; Essery *et al.*, 2003). The former condition is satisfied better at hectometric grid lengths where individual neighbourhoods are resolved, since, despite the urban surface being highly heterogeneous, each neighbourhood has similar heterogeneity characteristics. Thus, accounting for more urban form properties (e.g., building-to-building height variability and predominant building orientation) in parameterisations becomes appropriate. Also, the use of vertically distributed canopy schemes is prevalent in hectometric NWP. These need parameters as a function of height. This means hectometric NWP urban surface exchange schemes demand additional and higher-detail parameters describing urban form, fabric, and function.

Urban surface exchange schemes need at least plan area density and average building height to describe the urban form, as from these many other parameters can be estimated. Two notable global, hectometre-scale, open-access building description datasets have recently been developed. Esch *et al.* (2022) created the World Settlement Footprint 3D (WSF3D) dataset, which describes the fraction, total area, average height, and total volume of buildings on a global grid with 90-m cell size. The Global Human Settlement Layer (GHSL) dataset also contains such information (Pesaresi & Politis, 2023a, 2023b). These datasets have only just become available, so there is the opportunity for a step change in not only land-cover description but also building form description at hectometre scale.

A common approach to estimating average building height globally is to subtract remote-sensed digital elevation models (DEMs) from digital terrain models (DTMs), as demonstrated by Kent *et al.* (2019) for London using several DEM products. A common problem is estimation of the DTM where there are steep slopes or complex terrain coinciding with dense urban areas, due to too few ground-truth points. The WSF3D building-height estimation approach (Esch *et al.*, 2022) aims to alleviate this problem by analysing building-to-building height variation at vertical building edges. The GHSL dataset is calculated using a different methodology. A multiple linear regression model is trained on high-fidelity 1-m resolution

Quarterly Journal of the Royal Meteorological Society RMetS

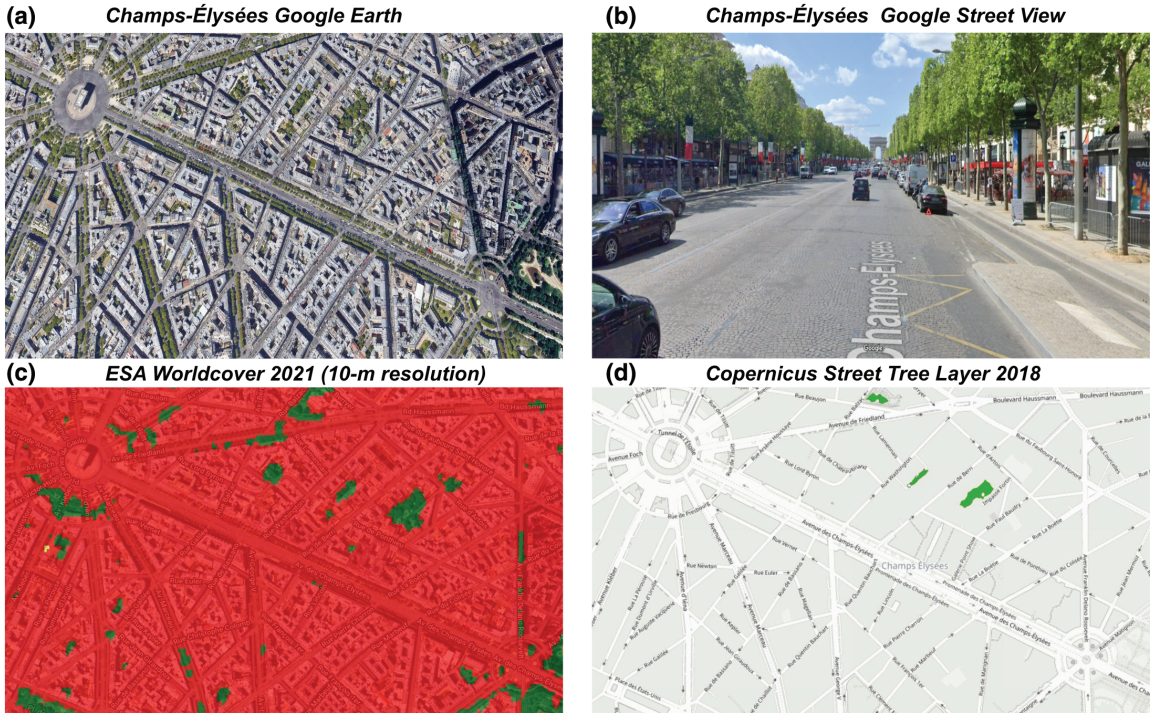

FIGURE 4 Illustration of poor sub-hectometric land-cover description exemplified for the Champs-Élysées, France. The in-canopy vegetation represented in (a) Google Earth imagery and (b) Google street view is under-represented in both (c) ESA Worldcover v200 (Zanaga *et al.*, 2022) and (d) Copernicus Street Tree Layer 2018 (Copernicus, 2018). [Colour figure can be viewed at wileyonlinelibrary.com]

cartography data available for several cities, and satellite remote sensed data including DEMs (Pesaresi *et al.*, 2021). The model takes the remote-sensed data as input to generate average building-height maps with global coverage.

The full set of parameters required to describe the 3D urban morphology in vertically distributed urban canopy schemes is usually derived from Geographic Information System (GIS) (Yu *et al.*, 2022) and lidar datasets (Bonczak & Kontokosta, 2019). Such datasets are increasingly becoming available for cities around the globe (Biljecki *et al.*, 2021), supported by emerging data sources and techniques developed to provide more accurate and detailed 3D information on a city's form (Middel *et al.*, 2022). Ching *et al.* (2019) developed a digital synthetic city tool for the generation of fit-for-purpose, at-scale urban canopy parameters. Geoscape has developed a 3D building and tree dataset (~1 m and ~2 m vertical resolution, respectively) covering the entirety of Australia using satellite and aerial-based remote sensing (Geoscape, 2022; Lipson *et al.*, 2022). Bocher *et al.* (2021) developed the open-source GeoClimate software that extracts data from an OpenStreetMap (OSM) crowd-sourced dataset (Haklay & Weber, 2008), to compute spatial indicators (e.g., building height, area, and fraction of wall share with other buildings) at three different urban scales (building, block, topographical spatial unit). Similarly, Lu *et al.* (2022) developed the open-source Python-based OSM2LES tool that extracts and rasterizes information from OSM to calculate key geometric parameters such as density, street connectivity, and entropy of facets.

Demuzere *et al.* (2022b) developed a global map of LCZs (Stewart & Oke, 2012), in which each LCZ class is associated with generic numerical descriptions of key urban canopy parameters that relate atmospheric processes to urbanisation. It is arguably better to use building-resolving than class-based datasets to derive urban form, fabric, and function parameters, as building-resolving datasets capture their distinct heterogeneous characteristics better (Lipson *et al.*, 2022). However, parameters related to urban fabric (e.g., thermal and radiative properties of buildings) are not currently available globally even at city scale. Typically, best-guess values for different urban fabric parameters are used based on classes such as region of the world, building density, and building height (Jackson *et al.*, 2010; Masson *et al.*, 2020a; Mills *et al.*, 2021). In such circumstances, where hectometric data are not available, a hybrid of continuous and class-based approaches is appropriate. For example, WUDAPT-to-WRF (W2W) (Demuzere *et al.*, 2022a) aggregates morphological information to give continuously varying gridded morphology, but, for unknown parameters like those relating to urban fabric, buildings are classified based on morphological values and class-based best-guess thermal and radiative parameter values are applied.

In the context of hectometric NWP, urban function is important for estimating anthropogenic heat emissions. Land use and the activity of humans (e.g., where they are located and their modes of transport) is important for estimating the amount of anthropogenic energy emitted spatially within a city. Detailed urban land use information is usually only available for individual cities from local authorities, but certain land use types can be inferred: for example, industry through remote sensing of night-time infrared radiation (Group, 2013). Some parameters that are available at kilometre scale, such as the day/night-time population density used in energy consumption models (Dobson *et al.*, 2000; Varquez *et al.*, 2021), might reasonably be downscaled to the hectometric scale using morphological parameters (e.g., building volume: (Blunn *et al.*, 2024a)).

The urban applications for which HMs will be used are global. The long-term ambition is to run HMs for any region of the world, with high-quality hectometric urban data, but it will likely take some time before such globally consistent and complete datasets become publicly available. A pragmatic approach to generating the best possible global open-access hectometric urban surface description database is to have global base datasets for each parameter that are derived in a mutually consistent manner (e.g., building further upon the above-mentioned tools and datasets), combined with a system that allows the local integration of higher fidelity (e.g., GIS and lidar) datasets. Since it is difficult to source high-fidelity datasets for different regions of the world and time-consuming to process them, there is scope for NHMSs to share resources, or work on a community-generated database, in line with the World Urban Database and Access Portal Tools (WUDAPT) community approach and philosophy to generate globally consistent urban data (Ching *et al.*, 2018).

6.3 | Time-varying surface description

In hectometric NWP, temporal changes associated with tides, lake extent, and agricultural LUC become better resolved spatially. It is therefore more important to have temporally varying information on LUC and surface parameters describing, for example, vegetation (leaf-area index (LAI), height) and water bodies (tides, lake depth). Surface characteristics can have significant variations on a daily (e.g., LAI, albedo, tides, and soil moisture), seasonal (e.g., lake depth), and yearly scale (e.g., tree height).

The newest remote-sensing techniques allow the creation of near-real-time LUC maps to reclassify crops into bare land after ploughing and represent changes in the extent of water bodies (seasonal lakes, estuaries), as demonstrated by Dynamic World, a 10-m resolution near-real-time land-cover dataset (Brown *et al.*, 2022). Due to cloud cover, these near-real-time global LUC map changes (e.g., Brown *et al.*, 2022) are not usable operationally. However, the recent release of the ESA Worldcereal (Tricht *et al.*, 2023) product, providing globally seasonally updated crop information at 10-m resolution, shows that seasonal or yearly LUC map updates are possible.

Daily and seasonally varying surface parameters should follow LUC changes (e.g., LAI drop after ploughing and ensuring lake depths are updated in temporally varying models such as fresh water lake model (FLake): (Kirillin *et al.*, 2011)) for a coherent representation of the surface evolution. These parameters also describe the temporal evolution of each LUC class (e.g., LAI increases when vegetation blooms and soil moisture increases following precipitation). Set values, either climatological or time-constant, are typically used to represent these parameters operationally. For example, LAI and albedo values used in complement with the ECOCLIMAP-SG LUC classification are rolling 10-day climatologies from the Copernicus global land service.

In agricultural areas, the assimilation of remotely sensed LAI and soil moisture data has potential for monitoring vegetation and predicting surface fluxes (carbon and water: (Tóth & Szintai, 2021)). Meanwhile, soil-moisture assimilation in urban areas is problematic, due to human activity contamination of the microwave signal. Since conventional satellite remote-sensed techniques (Wagner *et al.*, 2013) are not possible, new observational techniques need to be developed, or in situ observations made better to understand urban soil-moisture heterogeneity and temporal variability.

Automatic estimation of lake depth using remote-sensing data is complex due to varying water optical characteristics and sediment properties. The lake depth used to complement ECOCLIMAP-SG LUC classification is time-constant and taken from the global lake database (Kourzeneva *et al.*, 2012). While a method using lake water surface temperature exists, it has limitations (Balsamo *et al.*, 2009), and new observation techniques need to be developed to ensure modifications to lake extent result in realistic lake depth updates (e.g., Hou *et al.*, 2022).

6.4 | Other (time-constant) surface parameters

This subsection discusses parameters that vary on a geological scale, and are thus not affected by LUC updates (e.g., orography, soil texture). Time-constant values are used in NWP and climate models to represent them.

Current operational NWP configurations typically use global orography (or DEM) datasets, such as 250-m resolution Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010: (Danielson & Gesch, 2011), e.g., HARMONIE-AROME) and the 1-km resolution GLOBE30 (Hastings & Dunbar, 1998). For hectometric NWP, orography should be provided at a hectometric scale to achieve maximum improvements in resolved flow, but sub-hectometric resolution DEM datasets could give additional information to subgrid orographic drag schemes. Sub-hectometric orography datasets exist and can be global, such as the TanDEM-X ~12-m resolution DEM dataset (Wessel, 2016; Zink *et al.*, 2014), which is freely available at ~90-m resolution (DLR, 2018). Other sub-hectometric datasets have complete zonal coverage, but incomplete meridional coverage, such as the Shuttle Radar Topography Mission (SRTM: ~30-m resolution, 60°N–56°S coverage; (NASA, 2013)), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER: ~30-m resolution, 83°N–83°S; (NASA, 2001)), and *Advanced Land Observing Satellite* (*ALOS*) World 3D (AW3D30: ~30-m resolution, 60°N–60°S coverage; (ALOS, 2022)). Even higher resolution DEM datasets exist at national scale, such as the France ~1-m resolution DEM, “altimetry component of the référentiel à grande échelle” (RGE ALTI: (Institut national de l’information géographique et forestière, 2021)).

For soil texture, hectometric Met Office UM configurations use the 1-km resolution Harmonised World Soil Database (Nachtergaele *et al.*, 2010), while HARMONIE-AROME cycle 43 uses the 250-m resolution Soilgrids (Hengl *et al.*, 2017). Hall *et al.* (2024) demonstrated that there are potential benefits (e.g., for land-surface temperature) from moving to higher resolution soil texture datasets. However, to the best of the authors’ knowledge there are no global soil-texture datasets at sub-250-m resolution.

6.5 | Summary of research needs

To maximise the potential of hectometric NWP, it is necessary to provide an (urban) surface description that has detail commensurate with the complexity of hectometric NWP surface exchange models. This requires information on the 3D building morphology, thermal and radiative properties of building materials, in-canopy vegetation, soil moisture, and building energy consumption. At the time of writing, the first global datasets covering the most basic urban surface parameters (e.g., plan area density and average building height) are becoming available at hectometric scale, and rapid advances are being made through the use of remote sensing techniques, crowd-sourcing, and novel algorithms. We therefore propose that surface description systems should be developed that have globally consistent base datasets, and that can be updated locally with higher fidelity hectometric datasets as they become available. Where possible, to obtain maximum accuracy and reflect the continuously varying nature of the urban surface, its description should be generated from bottom-up approaches based on building/vegetation-resolving datasets.

7 | DATA ASSIMILATION

So far most HMs have been run without data assimilation—that is, simply using starting data and boundary conditions from lower resolution (usually km-scale) models. In this case, 100-m scale detail in the model either spins up due to physical processes or is forced by high-resolution surface data. However for short-range applications, for example, nowcasting of thunderstorms, it is likely that it will be necessary to develop data assimilation techniques for these models. Very few studies, so far, include any attempt to include data assimilation in HMs. One recent exception is an experimental study by (Koopmans *et al.*, 2023) of assimilation in a 100-m model for Amsterdam. In this study, observations were assimilated from World Meteorological Organization (WMO) sites around the city, radar data from a C-band Doppler radar, and also crowd-sourced personal weather stations for urban temperatures.

An important consideration is that data assimilation even at km scales is still not a mature field. For a comprehensive overview see (Gustafsson *et al.*, 2018). Many issues regarding hectometric-scale data assimilation will be similar to those at km scale, but probably more extreme. A key issue is likely to be the breakdown of linearity and Gaussian assumptions, which are made for global and current convective-scale data assimilation. Currently the majority of National Meteorological Services (NMSs) are in the process of making the transition to use flow-dependent data assimilation algorithms. The UK Met Office is already employing 4D-Var Milan *et al.* (2020), Deutscher Wetterdienst (DWD)/COSMO (Schraff *et al.*, 2016) is using Local Ensemble Kalman Filtering (LETKF), and the NMSs that are part of ACCORD are experimenting with 3D-EnVar (Montmerle *et al.*, 2018) or 4D-Var (Barkmeijer *et al.*, 2021; Gustafsson *et al.*, 2018). Given that the number of observations with high spatial (100 m) and temporal resolution (minutes) is increasing, it is now time to consider the opportunities and difficulties of performing data assimilation at the hectometric scale (50–500 m). In this section, we explore some of the topics that have to be addressed.

Models at current kilometre-scale resolutions already experience spin-up during the first hour of forecasts. It leads to unrealistic oscillations in variables and provokes processes such as spurious precipitation. The spin-up originates from imbalances in the analysis, for example, when adding an increment to a balanced model field (background field). Given the increased instability at the hectometre scale, it is likely that this model behaviour will worsen. Especially when moving to rapid-refresh sub-hourly analysis cycles, it is crucial that spin-up remains small and is only short-lived. Methods like incremental analysis updates can alleviate the shock caused by adding the increment to some extent; however, more research in this area is needed. The spatial balance constraints and the filtering of high-frequency oscillations derived for synoptic-scale processes to alleviate spin-up issues are not valid any more on sub-km scales (although the balance still needs to be satisfied on larger scales). Fast inertia–gravity waves and fast acoustic waves become important for many processes, including moist convection. The fast mode solutions such as inertia–gravity waves are the model devices that realise the process of adjustment of high-frequency atmospheric flow towards a slower and more balanced attractor, usually called the minimally imbalanced manifold. A deeper understanding of adjustment processes is needed in order to obtain the diagnostic relationships between model state components that could characterize the action of adjustment processes and could be used to project the analysis solution to the minimally imbalanced manifold. Solutions based on machine-learning methodologies such as generative adversarial network (GAN) or variational autoencoder (VAE) might be one way to go; however, research in this direction is in its infancy.

Spin-up in 4D-Var at the kilometre scale is not a concern yet. In the incremental approach of 4D-Var, the increment computation takes the model evolution during the observation window into account explicitly, albeit at a coarser grid resolution. Provided the linear assumption is still valid, this will help to soften the impact of the increment. Also, the main forecast starts from the analysis trajectory a few hours into the forecast, close to the centre of the observation window, when possible spin-up has usually disappeared. With increasing grid resolution in the nonlinear components of 4D-Var, higher resolution in the minimisation may also be required to reduce the discrepancy with the nonlinear model run used during minimisation. In doing so, performance difficulties of the 4D-Var algorithm itself may arise, in addition to the linear assumption, as convergence properties may be hampered by steeper gradients or by the consequences of correlated observation errors, which will become more apparent.

The length of the data assimilation window of the 4D-Var data assimilation scheme is determined by the length of two processes: the length of adjustment and the predictability limits. Ideally, the length of the data assimilation window should be large enough to incorporate the adjustment process and to allow disturbance introduced into one model's state components to propagate to the other model's state components. At the same time, the assimilation window should be short enough to stay within a nearly linear regime of model development. This constraint might be hard to meet when dealing with forecasting at 100-m horizontal resolution. At such resolutions, processes become highly nonlinear. An innovative multi-scale data assimilation scheme might be required, which would handle a varying length of data assimilation window dependent on the observation types to be assimilated. Measurements observing dynamical variables might require a relatively long assimilation window, while measurements affected by development of clouds might require a much shorter assimilation window acting on a nearly balanced model state. Quasi-continuous data assimilation might be an alternative, or novel re-linearisation techniques as in (Stappers & Barkmeijer, 2011) to extend the validity of the linearity assumption.

Ensemble-based data assimilation techniques like 3d(4D)-EnVar and LETKF are able to handle weakly nonlinear systems; however, they will have to deal with localisation. It is essential in filtering out unwanted sampling errors due to relatively small ensembles. The noisy background-error covariances that would otherwise result from the ensembles would lead to a sub-optimal analysis (Destouches *et al.*, 2021). The radius of localisation has to be large enough not to disturb balances at large scales. On the other hand, it should also be sufficiently small to capture phenomena for which hectometric data assimilation will just make the difference. The numerical cost of running ensembles at the hectometre scale will limit the number of ensemble members, thus complicating the localisation procedure. Space-scale dependent localisation schemes allow localisation to be performed in a more flexible way. The ensemble is decomposed into several overlapping scales and a different radius of localisation is applied on each sub-ensemble depending on the scales it contains.

There is a need for a closer connection between land surface and upper air assimilation when moving to the hectometric scale. There is no doubt of the role of the surface model in triggering convection. Treating the surface and upper air assimilation as separate model components is therefore not optimal and may lead to an unwanted divergence. For that reason, screen-level data like temperature and humidity at 2 m are already ingested in the upper air assimilation at various NMSs.

Historically, upper air data assimilation has received more attention and is on a more advanced level of algorithmic development than surface data assimilation does.

There are several reasons for this. The soil and surface processes are highly nonlinear and surface conditions are inhomogeneous. There is an obvious difficulty in describing surface and soil processes in a physically consistent way on a several kilometres large grid. In addition, the number of conventional surface and soil observations is limited and these have low representativity away from the point of measurement. Now, when the interest of the NWP community has shifted to higher resolutions, below hundreds of metres, and new satellite instruments sensitive to soil conditions and able to measure at high temporal and spatial resolution are becoming available, the design and development of surface data assimilation schemes receives more attention. Consistent design of upper air and land-surface data assimilation and, in particular, design of a genuine coupled land-surface–atmosphere data assimilation impose new challenges and bring new opportunities.

The use of high spatial and temporal observations in hectometre-scale analysis demands a more advanced treatment of observation-error covariances. At present, observation-error covariances are neglected. Therefore, methods need to be developed to be able to include these observation-error covariances in the present analysis systems in order to reach hectometre-scale initial conditions.

The signature of error correlation may be flow and weather-regime dependent. Observation errors have complex structure and are caused by many different sources, such as impact from unresolved scales, model simplifications, instrument errors, and processing procedures, and manifest themselves on a wide range of scales. One possibility is to split the observation-error statistics into large-scale (biases) and short-scale variation components, where systematic behaviour is attributed to large-scale biases and is modelled explicitly and short-scale variations are assumed to be static. Observations related to convective initiation, for example from (satellite) cloud information or measurements of the 3D wind vector, are crucial for hectometre analysis of convection. The surface fluxes play an important role in convective initialisation, and therefore coupling of surface and upper air analysis is important to obtain a balanced hectometre initial state.

For constraining scales that are resolved by hectometre modelling but not measured by the observation system, a non-homogeneous representation of forecast-error structures is essential. The homogeneous forecast-error covariance in physical space imposes uncorrelated errors in spectral space. Ensemble analysis allows us to capture non-homogeneity. Introduction of space-scale dependent localisation by defining overlapping spectral bands allows us to introduce non-homogeneities in the localisation process as well.

Data assimilation allows the implementation of a sequential feedback mechanism in the model by comparing it with observed quantities. In order for the feedback mechanism to be efficient, it is essential that the data assimilation captures model error structures in an adequate way. 3D/4D variational data assimilation schemes, as well as Kalman-filter-based analysis techniques, impose implicitly a close to linear error growth. Going to hectometric scales, processes enter quickly into a nonlinear regime of behaviour. In that case, a weakly nonlinear analysis might provide clearly sub-optimal results. The case of strongly nonlinear error growth, in particular when on–off processes are involved, requires optimisation of non-convex problems. It is a challenge to design a flexible and computationally efficient global minimisation algorithm that would avoid getting stuck in local minima.

8 | OBSERVATIONS AND VERIFICATION

8.1 | Current state and research gaps

Observations are required for three reasons: for process studies to develop models, for verification of model output (both routinely and for research), and for data assimilation. From the developers' perspective, it is important to find verification metrics that examine model simulations at the targeted spatial and temporal scales. A severe challenge that modellers face in hectometric-scale simulation is that, while some of the high-impact, extreme phenomena may manifest on rather small spatial and temporal scales, available observations typically lack representation for such scales that dominate the phenomena. In addition, verification methodology may need adaptation in order to assess the model skills adequately in prediction of weather phenomena with limited scales and predictability.

There are few observation sources that have the spatial resolution required for hectometric modelling. Traditional synoptic weather observations operated by NMSs have a generally sparse network density that is typically no better than around 20 km, often poorer in areas of challenging terrain and across large cities. For both of these areas, it tends to be difficult to find observation sites that fit the selection guidelines applied to synoptic observations, as specified by the WMO. However, many novel, non-conventional observations that offer high-density observations do exist, which may become useful for data assimilation, post-processing, and verification in hectometric modelling. Among these, ground-based scanning remote sensing systems such as Doppler weather radar and Doppler wind lidar (Filioglou *et al.*, 2022) can achieve sub-km horizontal resolution. Vertically pointing systems such as Raman and DIfferential Absorption Lidar (DIAL: (Flamant *et al.*, 2021; Gaffard *et al.*, 2021)), laser ceilometers, and slant delay from Global Navigation Satellite Systems (GNSS: Bender *et al.* (2009)) provide fine vertical resolution. In areas close to major airports, data from commercial aircraft, either in the form of Aircraft Meteorological Data Relay (AMDAR) reports or derived from Mode-S broadcasts (de Haan, 2011), provide wind and temperature data with good vertical resolution along flight paths. More profile data are to be expected from low-budget *CubeSat* and the forthcoming *Meteosat Third Generation*(*MTG*) satellite, the latter also providing data from its flexible combined imager (FCI) at 2.5-min frequency and with a sub-nadir resolution of 1 km. At hectometric scale, research campaign datasets such as eddy covariance measurements of surface fluxes, radiation, and soil measurements may become relevant for weather forecasting. High-density crowd-sourced data from private weather stations (PWS: (Coney *et al.*, 2022; Sgoff *et al.*, 2022)), smart phones (Hintz *et al.*, 2019), vehicles Bell (2022), and mobile communication networks (Doumounia *et al.*, 2014; Overeem *et al.*, 2013) may become significant data sources from which relevant weather parameters can be retrieved, hopefully to fill in the gap, especially regarding the smallest scale features. Obviously, much work can be expected in the coming years in research and development regarding use of this "non-conventional" information in hectometric modelling. On the other hand, for this emerging high-density observation information, more work is needed on quality assurance.

For developers, verification metrics should be able to examine model performance at the dominant spatial and temporal scales of the phenomena themselves. In the field of operational weather forecasting, synoptic observations have been widely used to verify weather forecasts near the surface at specific sites. For hectometric-scale modelling, it is important to realize that, compared with the targeted weather features with highly limited scales, the resolution of the synoptic network may be far too coarse to represent the essential scales fully. A typical synoptic rain-gauge network may very likely miss precipitation peaks and distribution patterns. Likewise, wind measurements at regular automatic weather stations (AWS) may miss wind extremes and are unable to detect horizontal variability. In both of these examples, the scale characteristics of the weather phenomena may be smaller than those resolvable by the observation network. Thus, verification using observations from coarse networks may at best serve the purpose of a gross sanity check, but is clearly insufficient to assess the added value of high-resolution modelling. In general, it is anticipated that, toward hectometric-scale modelling, the tendency with double penalty to discredit high-resolution modelling will become more pronounced as a consequence of insufficient observation-network density or misplaced features, affecting more forecast parameters. This is why neighbourhood methods are important for spatial verification of hectometric models.

8.2 | Observations in urban areas

A specific challenge in urban high-resolution modelling is the extensive lack of meteorological observations. NMS weather stations are typically located either at airports or (mostly) outside of cities. This is for two main reasons: first the need to provide meteorological observations at the synoptic or mesoscale, and secondly aviation support activities. Exceptionally, there may be stations installed within cities, but they are then typically placed in open spaces (e.g., parks). This is seen as a compromise with the WMO directives to install stations on grass areas and, if possible, unperturbed. None are placed in dense urban neighbourhoods. However, there are hopeful signs that things may improve in coming years, in two ways. First, city authorities are becoming increasingly concerned with urban heat issues and thermal comfort and adaptation to climate change, and hence may be motivated to install urban weather networks (e.g., Bassett *et al.* (2016)). This may be taken forward in collaboration with NMSs, research institutes, or private companies. Second, cities generally have good internet connectivity. There are many opportunities to collect observations of meteorological interest, either by citizen science or crowd-sourcing Muller *et al.* (2015), for example, using private weather stations on the internet (Meier *et al.*, 2017) or private connected cars (Marquès *et al.*, 2022). In the case of observations from private weather stations, there has already been a significant amount of research into the quality characteristics of data from these networks, and development of quality-control packages (e.g., automatic data quality control (TITAN): Båserud *et al.* (2020)) to address common issues. These data are already used by several NMSs in nowcasting and post-processing Nipen *et al.* (2020) and there has been experimentation with assimilation in km-scale NWP models Sgoff *et al.* (2022).

In addition to these issues with near-surface measurements, satellite imagery also has issues in providinf pertinent information on meteorological conditions (e.g., surface temperature) above cities. It is difficult to interpret surface temperatures observed at hectometric or kilometric scale by satellite, because of the extreme variability of the small-scale surfaces (roofs, roads, gardens, walls if observation is not at nadir, etc.) and shadowed/sunlit areas, all of them having potentially very variable surface temperatures. Lagouarde and Irvine (2008) and Lagouarde *et al.* (2012) observed from plane measurements a variation of more than 4 °C due to observation/solar angle configurations, even for a relatively homogeneous city centre neighbourhood. The same issue due to small-scale heterogeneity may also be present for types of cover other than cities, for example, mountainous areas.

8.3 | Short-term outlook on data and methodology

For high-resolution modelling, the capability gap needs to be filled in order to evaluate model performance adequately for high-impact small-scale features. Efforts will be made to improve collection and utilisation of measurement data at high resolution and high density. Further, observation data used for verification should be extended to a much broader range. In terms of verification objects, they may also be extended beyond the core meteorological variables (temperature, pressure, relative humidity, wind, and rainfall), examples of such being maximum and minimum temperature, visibility, wind gust speed, and cloud-layer characteristics. Meanwhile, verification techniques need to be adapted to uncover the potential of high-resolution simulations at targeted spatial and temporal scales. For urban modelling, a priority is the collection of temperature observations at high spatial resolution throughout the urban environment. This enables assessment of how HMs represent the temperatures and heat-stress hazards experienced by the urban population. There is also the need to verify how models represent 3D temperature variation with height within the urban canopy. Another useful observation is high spatial resolution 3D wind throughout the urban environment, ideally with sampled characteristics similar to the model, or at least at the resolution of features the model can be expected to resolve (i.e., 100–500 m for horizontal resolution). Although there are currently observing systems that meet or get close to these requirements, for example, Doppler radar and Doppler wind lidar, these do not typically give complete coverage over model domains or in all weather conditions. In order to understand surface energy balance, systems consisting of a sonic anemometer and infrared gas analyses have been developed for this purpose, allowing derivation of sensible and latent heat fluxes derived by eddy covariance techniques. Another option is using optical and microwave scintillometers. However, establishing such networks for routine verification on a larger scale is challenging in the urban environment, as well as costly.

In many NMSs, networks of laser ceilometers are operated to provide estimates of the cloud-base height and depth of cloud layers. In Europe, most weather radar networks operated by NMSs are C-band and have dual polarisation and Doppler capability (Huuskonen & Holleman, 2014). The precipitation-rate products, which typically have 1-km resolution and 5-minute observing cycle at the national scale, are used extensively for validation of precipitation rate and accumulations during high-impact events. The hydrometeor classification products (Al-Sakka *et al.*, 2013) are also used to validate the transition between rain and snow and the occurrence of hail and graupel. The sampling characteristics of weather radar, which allow the spatial variation of precipitation to be captured, are a key strength. However, radar measurement is an indirect estimate of precipitation rate requiring assumptions, for example, about the drop-size distribution, unlike a tipping bucket or weighing rain-gauge, which is a direct, albeit point, measurement. Obviously, a combination of weather radar and surface in situ observations offers the potential to facilitate verification for precipitation intensity, accumulation, and type (including very light rain and drizzle). Ideally, weather radar products at or close to 100-m resolution are the target, although that implies a denser radar network.

Doppler wind lidar Bonin *et al.* (2017) and sonic anemometers are both instruments that could potentially meet requirements for monitoring turbulence at low altitude, if operated in an adequate density network. These can be used to examine model representation and prediction of turbulence conditions close to the ground.

Cloud microphysical parameters such as cloud liquid water, cloud ice, cloud drop effective radius, and vertical velocity can be derived using dual-polarisation scanning cloud radars, operating in X, Ka, and W bands. However, these instruments are not common and typically are used only for research purposes.

Traditional verification compares model forecasts of key weather parameters with observations, which are typically in situ measurements such as surface SYNOP data and upper air radiosonde data. These comparisons have the advantage of validating model forecasts directly against in situ data. In high-resolution modelling research and applications, observational network density and sampling time interval are limiting factors whenever these become insufficient to represent the dominant scale of phenomena. For example, a rain-gauge network may easily miss precipitation peak values happening kilometres away from any of the gauges. In Greenland, gusty winds in coastal hilly areas are often missed from wind measurements installed in residential areas. On the other hand, considerable amounts of in situ measurement data remain to be utilised for point verification, especially the data sampled over moving platforms. Recently, work has been started in some of the European weather services to explore validation of observation data by ship and airborne measurements by making use of the screening procedure used in data assimilation (MetEirean, MeteoFrance). At Koninklijk Nederlands Meteorologisch Instituut (KNMI), verification infrastructure is being established to validate radiosonde data, taking into account horizontal drift during the ascents.

As reviewed in the previous section, collection of higher resolution, non-conventional data such as crowd-sourced data may help to reduce the gap between current and required horizontal resolution substantially. For example, over the past decades, the density of private weather stations in Europe has seen rapid growth. At least in some European countries, the density of PWS looks to be able to provide kilometre to sub-kilometre scale station density. However, much work remains in quality assurance to ensure that these data can be utilised to their full potential.

For point verification, much remains to be done to unlock the potential fully. In the ACCORD programme, work has been ongoing to make use of an assimilation-screening utility to extend verification to moving platforms such as data from ship and airborne observations. For upper air verification, horizontal drifting can now be taken care of. Observation operators used for data assimilation can be used to compare models with remote sensing observables for non-model parameters. HMs will be capable of simulating updrafts explicitly. The observation data suitable for this are weather radar and Doppler wind lidar. Radar observations will be a cornerstone of data assimilation for NWP nowcasting with short lead times.

For high-resolution modelling, reduced scales and predicability make the model output more vulnerable to phase errors, hence with higher risk of double penalty. The verification approaches that focus more on representation of relevant scales and with more tolerance to phase errors, such as spatial and object-based verification, become more relevant and effective. In spatial verification, model output is compared with 2D or 3D observations such as those derived from radar (accumulated precipitation composite), satellite remote sensing, and other composites (e.g., Integrated Nowcasting through Comprehensive Analysis (INCA) precipitation analysis). These data may potentially offer a spatial resolution more compatible with the model resolution, providing more directly information about skills on small scales. For spatial observations, quality assurance is crucial.

Hectometric-scale model evaluation will also take into account temporal aspects, as infrequent sampling from either model or observation sides risks missing extreme situations, especially when weather phenomena are becoming more local. For such situations, characterisation of observation uncertainties become more important when the sampling error tends to be larger. Use of more than one dataset may mitigate the risk. From the model output side, more frequent output become necessary to reflect the temporal variability.

Compared with regular forecasts, emphasis in hectometric-scale modelling may be focused on a narrower range of forecast parameters. Sometimes the potential of modelling products may need to be released via post-processing (e.g., upscaling, impact forecast, heterogeneous neighbourhood forecast). Accordingly, verification needs to be adapted to reflect such change in the model products.

9 | PREDICTABILITY

Despite the appearance of improved accuracy given by the fine-scaled details of HMs, it is unrealistic to expect all aspects of an observed feature—such as the location, intensity, and duration of a convective storm—to be captured by a single forecast. Uncertainties arise from both the inherently chaotic nature of the atmosphere (Lorenz, 1969) and the limitations of our modelling systems (Lorenz, 1996). This problem is exacerbated at hectometric scales, because the scales being forecast are small compared with the scales of predictbility of, say, convective storms.

To account for this uncertainty, many weather centres run an ensemble of km-scale forecasts where each individual member has slightly different initial conditions, boundary conditions, and model physics. The aim is to have a skilful ensemble that has, among other qualities, both resolution and reliability. To achieve this, it is important to have a good representation of all the relevant sources of uncertainty and an understanding of how these uncertainties interact and upscale within the modelling system. This can be challenging in practice and remains an open research question at both synoptic and convective scales. Additionally, the number of affordable ensemble members is usually far fewer than would be necessary to describe the full probability distribution, but instead offers a sample of solutions that approximates the pdf (post-processing techniques can be used to enhance the probability density function: see Section 10). At the hectometric scale, all the challenges of designing a reliable ensemble remain, but with the addition of much higher computational costs and a faster loss of predictability at the scales of interest.

The scales at which an ensemble is reliable and usable are determined by the limit of predictability. Predictability as a function of scale is investigated in many studies (e.g., Boer, 2003; Frogner *et al.*, 2019; Hohenegger & Schar, 2007;

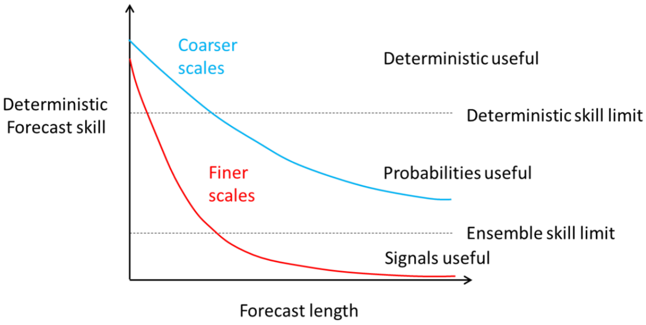

FIGURE 5 A conceptual view of predictability and scales and how they differ between a deterministic and ensemble system. The blue line represents the coarser scales and the red line the finer scales forecast. [Colour figure can be viewed at wileyonlinelibrary.com]

Lorenz, 1969; Surcel *et al.*, 2015; Zhang *et al.*, 2003). For high wavenumbers (small scales), the correlation between a forecast and the "truth" (measured by observations or an analysis) goes rapidly to zero. Skill is lost at lower wavenumbers (larger scales) with increased forecast range, with errors upscaling more rapidly at smaller spatial scales. Therefore, the forecast lead time at which the predictability limit is reached depends on the scales of interest, with shorter lead times for smaller scales. This predictability limit may be enhanced by resolved forcings in the forecast, such as orography, coastlines, and buildings, or synoptic-scale systems. Hohenegger and Schar (2007) found that error growth rates are about ten times larger for convective scales than for synoptic scales, and because of this the limit of predictability is about ten times shorter. Although the authors are not aware of any studies that quantify the limit of predictability at hectometric scales, theory suggests it will be much shorter again. This rapid loss of skill highlights the need to run an ensemble at the hectometric scale, even for very short lead times—the limit of predictability is much shorter for a deterministic forecast than the probabilistic output (Figure 5). Even with an ensemble, we can expect a rapid loss of skill at the grid scale, and this will need to be taken into account when considering the application of hectometric forecasts and communicating with users.

The speed at which the errors propagate to larger scales also depends on both the regime and the domain size. The domain size needs to be sufficiently large to allow for upscale growth and multi-scale evolution of features within the domain, which depends on the flow and how quickly features are leaving the domain. For example, Vié *et al.* (2011) found that, when the synoptic-scale circulation is weaker and the convective system is driven mainly by local and mesoscale processes, convective-scale initial perturbations have more impact on the simulated precipitating systems. In addition, the larger scales entering from the lateral boundaries will soon dominate if the domain is too small. It was shown by McCabe (personal communication, 2023) and (McCabe, personal communication 2023, and Frogner *et al.*, 2022) that for convective-scale ensembles the lateral boundary perturbations dominate after about 30 h for domains where each side is approximately 2000 km; this will be much shorter for the small domains envisaged to be necessary for hectometric forecasts to be affordable in an operational setting.

The computational cost associated with running an ensemble is of course an even larger issue when considering ensemble forecasts at hectometric scales. Only a few centres have started running hectometric-scale ensemble forecasts in research mode. At the UK Met Office, a relocatable 300-m ensemble was run in near-real time for a winter testbed (Bain *et al.*, 2022) to explore the practicalities of running on-demand sub-km ensemble forecasts for nine specific events of interest. While providing more detailed features, particularly over orography, the experiments showed limited impact on ensemble spread between the kilometre and sub-km scales, attributed in part to the limited size of the domain. More recently, Hanley and Lean (2024) have run a 300-m ensemble for a continuous three-month period. The domain covered a 540 km^2 area with an inner area over London at 300-m horizontal resolution, stretching out to 1.5 km at the boundaries using variable resolution. By using variable resolution, Hanley and Lean (2024) were able to increase the size of the domain compared with previous sub-km forecasts (London model Boutle *et al.*, 2016) at a quarter of the cost of running with 300-m resolution over the whole area. With this larger domain, smaller scale features were able to spin up and several cases show a number of members in the sub-km ensemble developing bands of intense rainfall, seen in the radar but missed by the km-scale driving model. The benefit of the higher model resolution was to improve the representation of convection, while the use of an ensemble was necessary to capture the observed features (otherwise missed by a single hectometric-scale forecast).

Although the initial experiments with hectometric-scale ensembles by Hanley and Lean (2024) are promising, there are many aspects of ensemble design at these scales that are still unknown. As mentioned above, although the theory is understood, the relationship between the spatial and temporal scales of the predictability at hectometric scales needs to be quantified, along with the scales at which the hectometric ensemble diverges from the driving ensemble. Additionally, while we understand that we need to represent uncertainties in the initial conditions, boundaries, and the model itself, we need to establish how the nature of these uncertainties might change at smaller spatial scales and how best to represent them. For applications at small spatial scales and short lead times, we would require frequently updated initial conditions. This is particularly important in seamless prediction systems, where nowcasting and forecasting are combined in a probabilistic framework and newer observations are constantly ingested in the system. Ensemble-based data assimilation (DA) systems may also be appropriate (see Section 7 for more information on DA techniques at the hectometric scale). Regardless of the choice of DA system, some thought will need to be given to how the ensemble perturbations interact with the DA, and vice versa.

The role of boundary conditions and their diversity is crucial in ensemble forecasting with limited-area models. For the small domains necessary at the hectometric scale, the effect of the boundary conditions will be felt throughout the domain in a relatively short period of time. This means that the diversity in the boundaries—representative of the uncertainty at the scale of the driving model—needs to be sufficient in the very short range. If only a small number of ensemble members are affordable, the question arises of how to optimise the perturbations to sample as wide a range of the pdf as possible. In this case, methods such as cluster analysis and machine learning may support selection of the most appropriate perturbations.

As we move to increased horizontal resolution, some processes that are currently parameterised will be resolved explicitly and are sufficiently variable between members, removing the need for some of the model perturbations. At the same time, new higher resolution processes may require new or adapted parameterisation schemes. In this way, some of the "known" uncertainty of the current model setup will be removed and replaced with new but "unknown" model uncertainties. The perturbation of specific processes, as currently attempted for the km scale (Clark *et al.*, 2021; Fleury *et al.*, 2022; Hirt *et al.*, 2019; Kober & Craig, 2016), could also be an effective strategy for the hectometric scale. Research should also focus on the feasibility and usability of intrinsic stochastic parameterisations (Craig & Cohen, 2006; Plant & Craig, 2008; Sakradzija *et al.*, 2015, 2016; Sakradzija & Klocke, 2018). While such schemes do not currently add much to ensemble spread, this may be different for hectometric scales when other processes are parameterised. A multi-physics ensemble may also be of use where there are several different schemes available, each with their own strengths and weaknesses. There may also be a need to consider the perturbations to the model dynamics, as more processes are resolved and the forecast relies less on parameterisations.

Addressing the research gaps in ensemble design at hectometric scales will need to be done in conjunction with the development of other areas outlined in this article. In the short term, simple experiments could be performed to help us start to quantify the issues of running an ensemble at these scales. The first and most important of these is to compare the benefits of a hectometric ensemble over its convective-scale driving model—specifically, do we gain sufficient improvements from the finer scales to justify the increased cost? To make such an evaluation, we need to include other (cheaper) methods for enhancing convective-scale forecasts, such as post-processing techniques (see Section 10). As part of such a comparison, consideration should be given to the balance of the computational cost of ensemble member numbers, domain size, and horizontal resolution, and how each of these impacts the limit of predictability at different scales. While the uncertainty in the model physics at the hectometric scale is currently unknown, a simple initial experiment would be to use the same stochastic physics scheme used in the driving model and maintain the same spatial and temporal patterns over the hectometric domain. This would give some indication of the likely impact of a perturbed parameter scheme, for example, and how that impact may be expected to vary with lead time.

Initially, these experiments could be performed with a simple downscaled ensemble. When a hectometric DA system is available and the necessary modifications have been made to the model, these experiments will naturally need to be revisited. The benefit of the early experiments will be to develop a robust framework for testing and evaluating ensemble systems, along with a useful dataset to inform other developments. A challenge for all aspects of development at the hectometric scale will be the selection of appropriate diagnostics, along with verification and evaluation methods (see Section 8). For ensembles, specific methods will be needed to evaluate extremes at these scales, as well as methods to interpret low probabilities at short lead times. The data produced by these early experiments can also be used to start addressing some of the practical considerations of running hectometric ensembles operationally. Given that the ensembles will be run out to relatively short lead times compared with the driving model, they will need to be delivered to operational meteorologists and automated products in a timely fashion. Another consideration is the amount of data that will be generated and how meteorological services will handle this. As with all ensemble applications, tools and techniques may also need to be developed to communicate the probabilistic information to a range of different users covering the full range of required applications.

In the longer term, there will be a need to refine the perturbation methods currently used at the convective scale to make them applicable to the hectometric scale based on our new understanding of model and land-surface uncertainty. In addition, the link with DA will need to be revisited as new and refined DA methods, along with better observations, become available. HMs can help inform machine-learning approaches, which in return could reduce the computational cost, which would be important to be able to run a sufficiently large domain and number of ensemble members. In the field of ensemble forecasting, which is particularly computationally demanding, beside pre- and post-processing, use of machine learning has many possibilities, such as parameterisation emulation and improvement of the construction of ensemble perturbations designed to represent model error. Uncertainty quantification can also be done by so "end to end" machine-learning models.

10 | POST-PROCESSING

HMs will provide forecasts with unprecedented detail, but that comes with substantial computational cost. If these models are going to provide cost-effective benefit beyond our current capabilities, there must be an emphasis on applying post-processing (PP) to extract the key signals that end users will want to incorporate into their decision-making. The choice of PP approaches depends on the implications of running at such fine resolution, as well as user requirements. One benefit of very fine grid spacing is that the grid cells are sufficiently small that representativeness errors associated with grid size largely disappear. Thus, the output can be directly related to point observations for most purposes, and PP that deals with forecast biases associated with representativeness error (Bouallegue *et al.*, 2020) becomes much less necessary. More of a concern for HMs is the rapid loss of predictability at very fine scales as described in the predictability section. This means that local detail may not be believable even if represented very accurately. For example, the evolution and positioning of a particular shower will almost certainly be wrong in detail. Whilst this is also true of kilometre-scale forecasts (Clark *et al.*, 2016), the situation is made worse by the more rapid growth of errors and perception of accuracy that comes from ultra-fine detail. It means that a probabilistic approach is essential, but the cost of sub-kilometre grid spacing makes it impossible to run a large enough ensemble to capture all the fine-scale uncertainty. Therefore, the neighbourhood processing (NP) approach that is already applied to kilometre-scale models and ensembles (Schwartz & Sobash, 2017) is even more essential for HMs to deal with that undersampling. NP is a methodology that uses surrounding points as additional ensemble members, thus effectively increasing the ensemble size and creating more smoothly varying probabilities that account for undersampled spatial differences. The two main types of NP are (1) the mean in neighbourhood, which provides a probability of occurrence at each grid cell, and (2) the maximum value in the neighbourhood, which provides a probability of occurrence within the vicinity of each grid cell. The latter is useful for highlighting areas at significant risk even if grid-scale probabilities are small. The problem with conventional neighbourhood methods is that they tend to use a static neighbourhood size and assume equal likelihood of an occurrence around each point. As a result, they may spuriously smear out more predictable local signals that are precisely what HMs are aiming to capture, such as fog in valleys or cooler temperatures in parks. This issue brings a greater need to develop further neighbourhood methods that allow the neighbourhood size to adapt to the spatial ensemble spread (Blake *et al.*, 2018; Dey *et al.*, 2016; Flack *et al.*, 2021) as well as the underlying terrain and surface type (Roberts *et al.*, 2023). An alternative or complimentary approach may be to use machine-learning emulation to create additional members in a more inherently heterogeneous way.

If we want to extract maximum benefit from HMs, we will require development of physically based or ML algorithms to detect fine-scale weather features of concern or interest. These might include indicators of severe weather such as storm modes, tornadic circulations, or flood-producing rainfall totals, as well as more ordinary conditions such as sea breezes (Cafaro *et al.*, 2019) or boundary-layer structure, fog patches or temperature variations associated with the land surface. The features should be identifiable as objects that can be incorporated into probabilistic PP, with visualisation designed to alert users to important signals that may not be detectable without the use of HMs. This sort of object detection can provide useful domain-wide information even if predictability is very low, and if appropriately applied to coarser resolutions as well, it allows a means of evaluating the value added by HMs.

Since HMs are likely to be run with small domains operationally, especially if run as ensembles, there will be geographical regions benefiting from both HM and kilometre-scale model outputs with larger areas only covered by kilometre-scale outputs. The question is how to deal with that in the best way for downstream users. It may be that for many applications it is fine to use both independently, on the understanding that there are differences, much as happens now with regional kilometre-scale forecasts. However, if a seamless product is required for consistency to avoid a patchwork of outputs, to reduce downstream data volumes, or for other reasons, there will need to be blending between the different resolutions. Blending should weight more towards the HM in the interior of the HM domain(s) to take advantage of the finer resolution. There may also be a need for blending over a time window at the end of the HM forecast(s) to transition into the kilometre-scale forecasts. Any seamless product should be provided on a common grid, which makes it necessary to either average the HM output or downscale the kilometre-scale forecasts, or both. This presents problems. Averaging the HM output may mean removing the fine-scale information it is intended to provide and hence the point of running it. Downscaling the kilometre-scale forecasts could generate huge data volumes and would potentially involve complex processing to generate fine-scale information the HM is already producing. One way of averaging without loss of detail is to convert to probabilities. A probability distribution can be constructed from the grid cells inside a neighbourhood using a suitable set of physical thresholds. The probabilities from two different resolutions can then be blended threshold by threshold, thus retaining any information about outliers or extremes contained in the HM. In the end, the need for blending or not, and the approaches taken, will need to be defined by each NMS based on its own priorities.

Another consequence of being restricted to small domains or few members is a need to try to capture the plausible range of larger-scale driving conditions that could lead to different fine-scale outcomes in the HM. For example, any scenarios that might import or lead to severe local storms should not be missed because the wrong driving members were chosen. Methodologies will be needed to identify local or mesoscale weather regimes or weather of interest from the km-scale forecasts. This may require new algorithms based on conventional meteorological understanding or use of ML or a combination. Clustering methods are needed either to select appropriate members that try to represent the driving ensemble distribution best or to focus on situations that would benefit most from HM forecasts (Sharma *et al.*, 2023). This is really a form of pre-processing, with the requirement becoming more important as models become increasingly costly compared with computational resources. The same methods could also be employed to make automated real-time decisions on whether to run HMs or not, or where best to place a HM domain for greatest benefit on the day.

Post-processing is most closely associated with the use of statistical calibration methods to correct NWP forecasts. The current state-of-the-art is described in Vannitsem *et al.* (2020), with one of the challenges being moving to finer resolution NWP. The challenge of applying statistical calibration methods to HMs is huge, because of the fine scales involved and enormous dimensionality. There also may be less need because of the greater representativeness, but HMs will still have biases and these could still be large, especially when considering fine scales. Very long training datasets would be needed to capture statistical relationships across all regimes and that would be hugely costly, especially in an ensemble context. Reforecasts of HMs are probably out of the question in an operational context. In addition, very densely distributed and accurate observations in space and time would be needed at the scales the HM models are trying to represent, and such observations are not easily available. As discussed in the Observations section, the use of citizen observations may help, but quality control is a challenge. Different forms of observations may be inconsistent with each other at such small scales in rapidly evolving situations. Some kind of upscaling based on groupings of similar grid points would likely be necessary to reduce dimensionality and utilise sparse observations. The game-changer now for post processing, and NWP in general, is the use of ML and it is expected that its use will increasingly become standard practice. However, the same limitations that affect more conventional statistical methods also make ML more difficult for HMs, since both require a large sample of accurate training information. It is conceivable, though, that emerging ML methods will be effective even with observational limitations as HM datasets become more extensive. This may be especially true for the correction of observed point locations, although it may be possible to achieve better results using longer training datasets from less noisy, coarser models. One area in which ML and HMs are expected to combine is in the use of HM outputs to provide fine-scale gridded "truth" from which coarser-resolution models can be calibrated or for emulating HM scales and potentially creating additional "synthetic" HM members from coarser driving conditions.

The nature of the post-processing applied will ultimately depend on the purpose of the HM. If an HM is used for rapidly updated nowcasts, an ensemble might not be affordable at rapid update frequency and some time lagging might be useful. Upscaling would be limited, because very local forecasts in space and time are the requirement. The HM nowcasts might benefit from blending with observation extrapolation nowcasts even if only for a few tens of minutes. If the HM is operating for longer forecast periods and more as a downscaler from kilometre-scale forecasts, the upscaling methods become more useful, as space–time accuracy becomes less easy to achieve. As discussed above, ML is likely to become an increasingly important part of many aspects of post-processing HMs. We could also see effective use of multi-resolution ensembles incorporating HMs, as demonstrated at much coarser resolution by Leutbecher and Ben Bouallègue (2020).

11 | CONCLUSIONS

In principle, HMs give many benefits in terms of resolving convection, turbulence, and surface heterogeneities better and opening up the possibility of finer scale forecasts within predictability constraints. However, as we have seen in preceding sections, a number of major challenges remain to realise this potential in practical models.

The foremost challenge is the high computational cost of these models. Although we have discussed above some ways to ameliorate this to some degree, for example, by running the physics at lower temporal and spatial resolution, this will remain a dominant problem. This has the consequence that HMs will have to demonstrate very significant benefits over other approaches to achieve similar benefits such as using downscaling/post-processing/machine learning for practical applications. This also means that it will be essential to understand the grid-length dependence of the benefits for particular applications to enable models to run with as coarse a mesh as is feasible.

In terms of developing the models themselves, the parameterisation problem gets easier, in that it will not be so important to parameterise processes like deep convection, cloud schemes, and the largest convective boundary-layer structures. One might hope that the microphysics parameterisation will become more relevant and impactful as clouds become better represented. However, different parameterisations will need further development. For example, it becomes very important to have a scale-aware turbulence parameterisation, both to aid the representation of explicit convection and for the boundary layer. A scale-aware shallow convection scheme is likely to be needed to represent the initiation of convection in particular. A second general issue with parameterisation is the need to consider 3D processes in order to represent the high-resolution horizontal heterogeneity.

Land-surface parameterisation is likely to become more complicated at hectometric scales with new, fine-scale, and lateral processes becoming more important. Examples include mountain-related processes, soil hydrology, and time-varying characteristics such as coastal wetting and drying, wind farms, and wildfires.

Given many of the applications of HMs are expected to be for urban areas, parameterising the urban surface is particularly important. HMs may provide some simplification, in that a tiled scheme is less essential, but also a complication that each grid box is more likely to need to take account of adjacent ones. Parameterisations are being developed for urban aspects such as anthropogenic heat, the urban canopy, and urban vegetation. An important practical question for meteorological services is which applications can be satisfied by non-building-resolving HMs and which require building-resolving LES.

Surface parameterisations such as those described above will need good sources of data to describe the surface as a function of location on hectometric scales. This is also a significant problem, which requires the development of new datasets.

A similar issue is the requirement for meteorological observations for models at hectometric scales, for development of the models, verification and, eventually, data assimilation. There are a number of potential novel observation types which may help here, along with great interest in crowd-sourced observations, which usually offer greater observation density at some cost of accuracy. Verification techniques need to be adapted to take account of the enhanced resolution.

Data assimilation is likely to be required in HMs if forecasts are required at very short lead times. We have almost no experience of this at present, although some ideas as to what the key issues are likely to be are set out above.

Predictablity becomes an even bigger issue at hectometric scales than at lower resolutions, partly because the scales being forecast are smaller relative to the scales of predictability for a given lead time. This means that, maybe other than for strongly surface forced phenomena, HMs will need to be used in an ensemble context. Similarly to km-scale models, work is required to develop hectometric model ensembles with the correct amount of spread at the scales of interest. As with km-scale models, post-processing of HMs and ensembles will be essential to get the most out of them for applications, and new techniques will need to be developed to take account of their characteristics.

The challenges set out above and in previous sections are formidable, but it is important to remember that some meteorological services are already on the path to implementing operational HMs (albeit at the coarser end of the scale) for practical applications. It is clear that these challenges will be addressed, at least to some degree, and it is hoped that having set them out in this article will help the process.

AFFILIATIONS

[1]UK Met Office, Exeter, UK

[2]Royal Netherlands Meteorological Insititute (KNMI), de Bilt, the Netherlands

[3]Deutscher Wetterdienst, Offenbach, Germany

[4]Met Éireann, Dublin, Ireland

[5]Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

[6]University of Reading, Reading, UK

[7]Urban Climatology Group, Department of Geography, Ruhr-University Bochum, Bochum, Germany

[8]B-Kode, Ghent, Belgium

[9]European Centre for Medium-Range Weather Forecasting, Reading, UK

[10]Norwegian Meteorological Institute, Oslo, Norway

[11]Meteorology and Air Quality Group, Wageningen University, Wageningen, the Netherlands

[12]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

[13]Department of Civil and Environmental Engineering, Imperial College London, UK

[14]Faculty of Civil Engineering and Geosciences, Technical University Delft, Delft, the Netherlands

[15]Czech Hydrometeorological Institute, Prague, Czech Republic

[16]Danish Meteorological Institute, Copenhagen, Denmark

ACKNOWLEDGEMENTS

M. Demuzere is supported by the ENLIGHT project, funded by the German Research Foundation (DFG) under grant no. 437467569, and by the European Union's HORIZON Research and Innovation Actions under grant agreement No 101137851, project CARMINE (Climate-Resilient Development Pathways in Metropolitan Regions of Europe, https://www.carmine-project.eu/).

CONFLICT OF INTEREST STATEMENT

Authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

No data was used for this review article.

ORCID

Humphrey W. Lean https://orcid.org/0000-0002-1274-4619
Natalie E. Theeuwes https://orcid.org/0000-0002-9277-8551
Michael Baldauf https://orcid.org/0000-0001-7353-1118
Geoffrey Bessardon https://orcid.org/0000-0002-9067-7167
Lewis Blunn https://orcid.org/0000-0002-3207-5002
Peter A. Clark https://orcid.org/0000-0003-1001-9226
Inger-Lise Frogner https://orcid.org/0000-0002-3446-9873
Adrian Lock https://orcid.org/0000-0001-6289-030X
Chiara Marsigli https://orcid.org/0000-0003-3376-1896

REFERENCES

ALOS. (2022) ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)". https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30&uscore;e.htm

Al-Sakka, H., Boumahmoud, A.-A., Fradon, B., Frasier, S.J. & Tabary, P. (2013) A new fuzzy logic hydrometeor classification scheme applied to the French X-, C-, and S-Band Polarimetric Radars. *Journal of Applied Meteorology and Climatology*, 52, 2328–2344. Available from: https://doi.org/10.1175/JAMC-D-12-0236.1

Arakawa, A. & Konor, C.S. (2009) Unification of the Anelastic and quasi-hydrostatic systems of equations. *Monthly Weather Review*, 137, 710–726.

Asensio, H., Messmer, M., Luthi, D., Osterried, K. & Jucker, J. (2020) External Parameters for Numerical Weather Prediction and Climate Application EXTPAR v5_4 User and Implementation Guide. Tech. rep. http://www.cosmo-model.org/content/support/software/ethz/EXTPAR&uscore;user&uscore;and&uscore;implementation&uscore;manual.pdf

Bain, C., Bush, M., Buxmann, J., Davies, P., Finnenkoetter, A., Flack, D. et al. (2022) The met office winter testbed 2020/2021: Experimenting with an on-demand 300-m ensemble in a real-time environment. *Meteorological Applications*, 29, e2096.

Baldauf, M. (2021) A horizontally explicit, vertically implicit (HEVI) discontinuous Galerkin scheme for the 2-dimensional Euler and Navier-Stokes equations using terrain-following coordinates. *Journal of Computational Physics*, 446, 110635.

Baldauf, M., Fuhrer, O., Kurowski, M., de Morsier, G., Müllner, M., Piotrowski, Z. et al. (2013) The COSMO priority project 'conservative dynamical core' - final report. Technical report, MeteoSwiss.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. & Reinhardt, T. (2011) Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Monthly Weather Review*, 139, 3887–3905.

Balsamo, G., Dutra, E., Stepanenko, V., Viterbo, P., Miranda, P. & Mironov, D. (2009) Deriving an effective lake depth from satellite lake surface temperature data: A feasibility study with modis data. https://www.ecmwf.int/node/7923

Barkmeijer, J., Lindskog, M., Gustafsson, N., Bojarova, J., Azad, R., Monteiro, I. et al. (2021) HARMONIE-AROME 4D-Var. Tech. rep. ALADIN Consortium.

Bartholome, E. & Belward, A.S. (2005) Glc2000: a new approach to global land cover mapping from earth observation data. *International Journal of Remote Sensing*, 26, 1959–1977.

Båserud, L., Lussana, C., Nipen, T., Seierstad, I., Oram, L. & Aspelien, T. (2020) Titan automatic spatial quality control of meteorological in-situ observations. *Advances in Science and Research*, 17, 153–163.

Bassett, R., Cai, X., Chapman, L., Heaviside, C., Thornes, J.E., Muller, C.L. et al. (2016) Observations of urban heat island advection from a high-density monitoring network. *Quarterly Journal of the Royal Meteorological Society*, 142, 2434–2441. Available from: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2836

Beare, R.J. & M.M.-H.A.e.a. (2006) An intercomparison of large-eddy simulations of the stable boundary layer. *Boundary-Layer Meteorolpgy*, 118, 247–272.

Bell, Z., D.S.L. & W.J.A. (2022) Exploring the characteristics of a vehicle-based temperature dataset for kilometre-scale data assimilation. *Meteorological Applications*, 29(3), e2058. Available from: https://doi.org/10.1002/met.2058

Belušić, D., de Vries, H., Dobler, A., Landgren, O., Lind, P., Lindstedt, D. et al. (2020) HCLIM38: a flexible regional climate model applicable for different climate zones from coarse to convection-permitting scales. *Geoscientific Model Development*, 13, 1311–1333.

Benacchio, T. & Klein, R. (2019) A semi-implicit compressible model for atmospheric flows with seamless access to soundproof and hydrostatic dynamics. *Monthly Weather Review*, 147, 4221–4240.

Bénard, P. (2003) Stability of semi-implicit and iterative centered-implicit time discretizations for various equation systems used in nwp. *Monthly Weather Review*, 131, 2479–2491.

Bender, M., Dick, G., Wickert, J., Ramatschi, M., Ge, M., de Gendt, G. et al. (2009) Estimates of the information provided by GPS slant data observed in Germany regarding tomographic applications. *Journal of Geophysical Research*, 114. Available from: https://api.semanticscholar.org/CorpusID:14907403

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W. et al. (2017) the HARMONIE–AROME model configuration in the Aladin–Hirlam NWP system. *Monthly Weather Review*, 145, 1919–1935.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X. & Tian, Q. (2023) Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619, 533–538.

Biljecki, F., Chew, L.Z.X., Milojevic-Dupont, N. & Creutzig, F. (2021) Open government geospatial data on buildings for planning sustainable and resilient cities. arXiv preprint, arXiv:2107.04023.

Blake, B.T., Carley, J.R., Alcott, T.I., Jankov, I., Pyle, M.E., Perfater, S.E. et al. (2018) An adaptive approach for the calculation of ensemble gridpoint probabilities. *Weather and Forecasting*, 33, 1063–1080.

Blunn, L., Xie, X., Grimmond, S., Luo, Z., Sun, T., Perera, N. et al. (2024a) Spatial and temporal3 variation of anthropogenic heat emissions in Colombo, Sri Lanka. *Urban Climate*, 54, 101828.

Blunn, L.P., Ames, F., Croad, H.L., Gainford, A., Higgs, I., Lipson, M. et al. (2024b) Machine learning bias correction and downscaling of urban heatwave temperature predictions from kilometre to hectometre scale. *Meteorological Applications*, 31, e2200.

Blunn, L.P., Coceal, O., Nazarian, N., Barlow, J.F., Plant, R.S., Bohnenstengel, S.I. et al. (2022) Turbulence characteristics across a range of idealized urban canopy geometries. *Boundary-Layer Meteorology*, 182, 275–307.

Blunn, L.P., Plant, R.S., Coceal, O., Bohnenstengel, S.I., Lean, H.W. & Barlow, J.F. (2023) The influence of resolved convective motions on scalar dispersion in hectometric scale numerical weather prediction models. *Quarterly Journal of the Royal Meteorological Society*, 150(759), 976–994.

Bocher, E., Bernard, J., Wiederhold, E.L.S., Leconte, F., Petit, G., Palominos, S. et al. (2021) Geoclimate: A geospatial processing toolbox for environmental and climate studies. *Journal of Open Source Software*, 6, 3541.

Boer, G. (2003) Predictability as a function of scale. *Atmosphere-Ocean*, 41, 203–215.

Bonan, G.B., Patton, E.G., Finnigan, J.J., Baldocchi, D.D. & Harman, I.N. (2021) Moving beyond the incorrect but useful paradigm: reevaluating big-leaf and multilayer plant canopies to model biosphere-atmosphere fluxes–A review. *Agricultural and Forest Meteorology*, 306, 108435.

Bonczak, B. & Kontokosta, C.E. (2019) Large-scale parameterization of 3D building morphology in complex urban landscapes using aerial LiDAR and city administrative data. *Computers, Environment and Urban Systems*, 73, 126–142.

Bonin, T.A., Choukulkar, A., Brewer, W.A., Sandberg, S.P., Weickmann, A.M., Pichugina, Y.L. et al. (2017) Evaluation of turbulence measurement techniques from a single Doppler Lidar. *Atmospheric Measurement Techniques*, 10, 3021–3039. Available from: https://amt.copernicus.org/articles/10/3021/2017/

Bouallegue, Z.B., Haiden, T., Weber, N.J., Hamill, T.M. & Richardson, D.S. (2020) Accounting for representativeness in the verification of ensemble precipitation forecasts. *Monthly Weather Review*, 148, 2049–2062.

Boutle, I., Eyre, J. & Lock, A. (2014a) Seamless stratocumulus simulation across the turbulent gray zone. *Monthly Weather Review*, 142, 1655–1668.

Boutle, I., Finnenkoetter, A., Lock, A. & Wells, H. (2016) The London model: Forecasting fog at 333 m resolution. *Quarterly Journal of the Royal Meteorological Society*, 142, 360–371.

Boutle, I.A., Abel, S.J., Hill, P.G. & Morcrette, C.J. (2014b) Spatial variability of liquid cloud and rain: observations and microphysical effects. *Quarterly Journal of the Royal Meteorological Society*, 140, 583–594.

Brown, A., Bendall, T.M., Boutle, I., Melvin, T. & Shipway, B. (2024) Physics-dynamics-chemistry coupling across different meshes in LFRic-Atmosphere: Formulation and idealised tests. *Quarterly Journal of the Royal Meteorological Society*, 1–21. https://arxiv.org/abs/2310.01255

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J. et al. (2022) Dynamic world, near real-time global 10 m land use land cover mapping. *Scientific Data*, 9, 1–17.

Brun, C., Blein, S. & Chollet, J.-P. (2017) Large-eddy simulation of a katabatic jet along a convexly curved slope. Part I: Statistical results. *Journal of the Atmospheric Sciences*, 74, 4047–4073.

Bryan, G.H. & Morrison, H. (2012) Sensitivity of a simulated squall line to horizontal resolution and parameterization of microphysics. *Monthly Weather Review*, 140, 201–225.

Bryan, G.H., Wyngaard, J.C. & Fritsch, J.M. (2003) Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review*, 131, 2394–2416.

Bunce, R., Barr, C., Clarke, R., Howard, D. & Lane, A. (1990) ITE land classification of Great Britain 1990. https://doi.org/10.5285/ab320e08-faf5-48e1-9ec9-77a213d2907f

Cafaro, C., Frame, T.H., Methven, J., Roberts, N. & Bröcker, J. (2019) The added value of convection-permitting ensemble forecasts of sea breeze compared to a Bayesian forecast driven by the global ensemble. *Quarterly Journal of the Royal Meteorological Society*, 145, 1780–1798.

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J. et al. (2023) Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. arXiv preprint, arXiv:2304.02948.

Chew, R., Benacchio, T., Hastermann, G. & Klein, R. (2022) A one-step blended soundproof-compressible model with balanced data assimilation: Theory and idealized tests. *Monthly Weather Review*, 150, 2231–2254.

Ching, J., Aliaga, D., Mills, G., Masson, V., See, L., Neophytou, M. et al. (2019) Pathway using WUDAPT's digital synthetic city tool towards generating urban canopy parameters for multi-scale urban atmospheric modeling. *Urban Climate*, 28, 100459.

Ching, J., Mills, G., Bechtel, B., See, L., Feddema, J., Wang, X. et al. (2018) WUDAPT: An urban weather, climate, and environmental modeling infrastructure for the anthropocene. *Bulletin of the American Meteorological Society*, 99, 1907–1924.

Christlieb, A.J., Macdonald, C.B. & Ong, B.W. (2010) Parallel high-order integrators. *SIAM Journal on Scientific Computing*, 32, 818–835.

Clark, P., Halliwell, C. & Flack, D. (2021) A physically based stochastic boundary layer perturbation scheme. Part I: Formulation and evaluation in a convection-permitting model. *Journal of the Atmospheric Sciences*, 78, 727–746.

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

Claussen, M. (1990) Area-averaging of surface fluxes in a neutrally stratified, horizontally inhomogeneous atmospheric boundary layer. *Atmospheric Environment. Part A. General Topics*, 24, 1349–1360.

CNRM. (2018) Wiki - ECOCLIMAP-SG - CNRM Open Source Site. https://opensource.umr-cnrm.fr/projects/ecoclimap-sg/wiki

Coceal, O. & Belcher, S.E. (2004) A canopy model of mean winds through urban areas. *Quarterly Journal of the Royal Meteorological Society*, 130, 1349–1372.

Coney, J., Pickering, B., Dufton, D., Lukach, M., Brooks, B. & Neely, R.R. (2022) How useful are crowdsourced air temperature observations? An assessment of Netatmo stations and quality control schemes over the United Kingdom. *Meteorological Applications*, 29(3), e2075. Available from: https://doi.org/10.1002/met.2075

Copernicus. (2018) Urban Atlas Street Tree Layer 2018. https://doi.org/10.2909/205691b3-7ae9-41dd-abf1-1fbf60d72c8c

Costes, A., Rochoux, M.C., Lac, C. & Masson, V. (2021) Subgrid-scale fire front reconstruction for ensemble coupled atmosphere-fire simulations of the FireFlux I experiment. *Fire Safety Journal*, 126, 103475. Available from: https://doi.org/10.1016/j.firesaf.2021.103475

Craig, G.C. & Cohen, B.G. (2006) Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *Journal of the Atmospheric Sciences*, 63, 1996–2004.

Daniel, M., Lemonsu, A., Déqué, M., Somot, S., Alias, A. & Masson, V. (2019) Benefits of explicit urban parameterization in regional climate modeling to study climate and city interactions. *Climate Dynamics*, 52, 2745–2764.

Danielson, J. & Gesch, D. (2011) Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011-1073, 26. http://pubs.usgs.gov/of/2011/1073/pdf/of2011-1073.pdf

Davies, H. (1976) A lateral boundary formulation for multi-level prediction models. *Quarterly Journal of the Royal Meteorological Society*, 102, 405–418.

Davies, T. (2014) Lateral boundary conditions for limited area models. *Quarterly Journal of the Royal Meteorological Society*, 140, 185–196.

de Burgh-Day, C.O. & Leeuwenburg, T. (2023) Machine learning for numerical weather and climate modelling: A review. *Geoscientific Model Development*, 16, 6433–6477.

de Haan, S. (2011) High-resolution wind and temperature observations from aircraft tracked by Mode-S air traffic control radar. *Journal of Geophysics Research*, 116, D10111. Available from: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015264

Deardorff, J. (1980) Stratocumulus-capped mixed layers derived from a three-dimensional model, 18, 495–527.

Defourny, P., Lamarche, C., Bontemps, S., De Maet, T., Van Bogaert, E., Moreau, I. et al. (2017) *Land Cover Climate Change Initiative-Product User Guide v2. Issue 2.0*, Tech. Rep. Belgium: ESA-UCLouvain-Geomatics, Louvain-la-Neuve. Available from: http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2.2.0.pdf

Demuzere, M., Argüeso, D., Zonato, A. & Kittner, J. (2022a) W2W: A Python package that injects WUDAPT's Local Climate Zone information in WRF. *Journal of Open Source Software*, 7, 4432.

Demuzere, M., Kittner, J., Martilli, A., Mills, G., Moede, C., Stewart, I.D. et al. (2022b) A global map of local climate zones to support earth system modelling and urban-scale environmental science. *Earth System Science Data*, 14, 3835–3873.

Destouches, M., Montmerle, T., Michel, Y. & Ménétrier, B. (2021) Estimating optimal localization for sampled background error covariances of hydrometeor variables. *Quarterly Journal of the Royal Meteorological Society*, 147, 74–93.

Dey, S.R., Roberts, N.M., Plant, R.S. & Migliorini, S. (2016) A new method for the characterization and verification of local spatial predictability for convective-scale ensembles. *Quarterly Journal of the Royal Meteorological Society*, 142, 1982–1996.

DLR. (2018) The TanDEM-X 90m digital elevation model. https://geoservice.dlr.de/web/dataguide/tdm90/

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C. & Worley, B.A. (2000) LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, 66, 849–857.

Doms, G., Förstner, J., Heise, E., Herzog, H., Mironov, D., Raschendorfer, M. et al. (2011) A description of the nonhydrostatic regional COSMO model. Part II: Physical parameterization. Deutscher Wetterdienst, Germany.

Done, J.M., Craig, G.C., Gray, S.L., Clark, P.A. & Gray, M.E.B. (2006) Mesoscale simulations of organized convection: Importance of convective equilibrium. *Quarterly Journal of the Royal Meteorological Society*, 132, 737–756. Available from: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.04.84

Doumounia, A., Gosset, M., Cazenave, F., Kacou, M. & Zougmore, F. (2014) Rainfall monitoring based on microwave links from cellular telecommunication networks: First results from a West African test bed. *Geophysical Research Letters*, 41, 6016–6022.

Dubal, M., Wood, N. & Staniforth, A. (2005) Mixed parallel–sequential-split schemes for time-stepping multiple physical parameterizations. *Monthly Weather Review*, 133, 989–1002.

Dudhia, J. (2022) Challenges in sub-kilometer grid modeling of the convective planetary boundary layer. *Meteorology*, 1, 313–402. Available from: 10.3390/meteorology1040026

Durran, D.R. (1989) Improving the Anelastic approximation. *Journal of the Atmospheric Sciences*, 46, 1453–1461.

Durran, D.R. (1998) *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*, Vol. 32. Springer Science & Business Media, New York.

ECMWF. (2020) Physical Processes, no. 4 in IFS Documentation and Part, IV. https://www.ecmwf.int/node/19748

Efstathiou, G., Plant, R. & Bopape, M.-J. (2018) Simulation of an evolving convective boundary layer using a scale-dependent dynamic smagorinsky model at near-gray-zone resolutions. *Journal of Applied Meteorology and Climatology*, 57, 2197–2214.

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A. et al. (2022) World settlement footprint 3D-A first three-dimensional survey of the global building stock. *Remote Sensing of Environment*, 270, 112877.

Espeholt, L., Agrawal, S., Sønderby, C., Kumar, M., Heek, J., Bromberg, C. et al. (2022) Deep learning for twelve hour precipitation forecasts. *Nature Communications*, 13, 1–10.

Essery, R., Best, M., Betts, R., Cox, P.M. & Taylor, C.M. (2003) Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *Journal of Hydrometeorology*, 4, 530–543.

Faroux, S., Kaptué Tchuenté, A., Roujean, J.-L., Masson, V., Martin, E. & Le Moigne, P. (2013) ECOCLIMAP-II/Europe: A twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models. *Geoscientific Model Development*, 6, 563–582.

Field, P., Hill, A., Shipway, B., Furtado, K., Wilkinson, J., Miltenberger, A. et al. (2023) Implementation of a double moment cloud microphysics scheme in the UK met office regional numerical weather prediction model. *Quarterly Journal of the Royal Meteorological Society*, 149(752), 703–739.

Filioglou, M., Preissler, J., Troiville, A., Thobois, L., Vakkari, V., Auvinen, M. et al. (2022) Evaluating modelled winds over an urban area using ground-based doppler Lidar observations. *Meteorological Applications*, 29(2), e2052. Available from: https://doi.org/10.1002/met.2052

Fitch, A.C., Olson, J.B., Lundquist, J.K., Dudhia, J., Gupta, A.K., Michalakes, J. et al. (2012) Local and mesoscale impacts of wind farms as parameterized in a mesoscale NWP model. *Monthly Weather Review*, 140, 3017–3038.

Flack, D., Clark, P., Halliwell, C., Roberts, N., Gray, S., Plant, R. et al. (2021) A physically based stochastic boundary layer perturbation scheme. Part II: Perturbation growth within a superensemble framework. *Journal of the Atmospheric Sciences*, 78, 747–761.

Flamant, C., Chazette, P., Caumont, O., Di Girolamo, P., Behrendt, A., Sicard, M. et al. (2021) A network of water vapor Raman Lidars for improving heavy precipitation forecasting in southern France: introducing the WaLiNeAs initiative. *Bulletin of Atmospheric Science and Technology*, 2, 1–21.

Fleury, A., Bouttier, F. & Couvreux, F. (2022) Process-oriented stochastic perturbations applied to the parametrization of turbulence and shallow convection for ensemble prediction. *Quarterly Journal of the Royal Meteorological Society*, 148, 981–1000.

Frogner, I.-L., Andrae, U., Ollinaho, P., Hally, A., Hämäläinen, K., Kauhanen, J. et al. (2022) Model uncertainty representation in a convection-permitting ensemble–SPP and SPPT in HarmonEPS. *Monthly Weather Review*, 150, 775–795.

Frogner, I.-L., Singleton, A.T., Køltzow, M.Ø. & Andrae, U. (2019) Convection-permitting ensembles: Challenges related to their design and use. *Quarterly Journal of the Royal Meteorological Society*, 145, 90–106.

Gaffard, C., Li, Z., Harrison, D.L., Lehtinen, R. & Roininen, R. (2021) Evaluation of a prototype broadband water-vapour profiling differential absorption Lidar at Cardington, UK. *Atmosphere*, 12(11), 1521. Available from: https://doi.org/10.3390/atmos12111521

Gehrke, K.F., Sühring, M. & Maronga, B. (2021) Modeling of land–surface interactions in the PALM model system 6.0: land surface model description, first evaluation, and sensitivity to model parameters. *Geoscientific Model Development*, 14, 5307–5329.

Geoscape. (2022) Geoscape Australia. https://www.google.com/?q=https://geoscape.com.au/

Germano, M. (1992) Turbulence: the filtering approach. *Journal of Fluid Mechanics*, 238, 325–336.

Germano, M., Piomelli, U., Moin, P. & Cabot, W.H. (1991) A dynamic subgrid-scale eddy viscosity model. *Physics of Fluids A: Fluid Dynamics*, 3, 1760–1765.

Gettelman, A., Gagne, D.J., Chen, C.-C., Christensen, M., Lebo, Z., Morrison, H. et al. (2021) Machine learning the warm rain process. *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002268.

Goger, B. & Dipankar, A. (2024) The impact of mesh size, turbulence parameterization, and land-surface-exchange scheme on simulations of the mountain boundary layer in the hectometric range. *Quarterly Journal of the Royal Meteorological Society*, 150(763), 3853–3873.

Grabowski, W.W., Morrison, H., Shima, S.-I., Abade, G.C., Dziekan, P. & Pawlowska, H. (2019) Modeling of cloud micro1563 physics: Can we do better? *Bulletin of the American Meteorological Society*, 100, 655–672.

Group, E.O. (2013) VIIRS nightfire nighttime detection and characterization of combustion sources. https://eogdata.mines.edu/download&uscore;viirs&uscore;fire.html

Grylls, T., Suter, I. & van Reeuwijk, M. (2020) Steady-state large-eddy simulations of convective and stable urban boundary layers. *Boundary-Layer Meteorology*, 175, 309–341.

Guerra, J.E. & Ullrich, P.A. (2016) A high-order staggered finite-element vertical discretization for non-hydrostatic atmospheric models. *Geoscientific Model Development*, 9, 2007–2029.

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H. et al. (2018) Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, 144, 1218–1256.

Haase, D., Jänicke, C. & Wellmann, T. (2019) Front and back yard green analysis with subpixel vegetation fractions from earth observation data in a city. *Landscape and Urban Planning*, 182, 44–54.

Hakim, G.J. & Masanam, S. (2023) Dynamical tests of a deep-learning weather prediction model. arXiv preprint, arXiv:2309.10867.

Haklay, M. & Weber, P. (2008) Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 7, 12–18.

Hall, T.W., Blunn, L., Grimmond, S., McCarroll, N., Merchant, C.J., Morrison, W. et al. (2024) Utility of thermal remote sensing for evaluation of a high-resolution weather model in a city. *Quarterly Journal of the Royal Meteorological Society*, 150, 1771–1790.

Hanley, K. & Lean, H. (2024) The performance of a variable resolution 300m ensemble for forecasting convection over London. *Quarterly Journal of the Royal Meteorological Society*, 150(763), 3737–3756.

Hanley, K., Whitall, M., Stirling, A. & Clark, P. (2019) Modifications to the representation of subgrid mixing in kilometre-scale versions of the unified model. *Quarterly Journal of the Royal Meteorological Society*, 145, 3361–3375.

Hanley, K.E., Barrett, A.I. & Lean, H.W. (2016) Simulating the 20 May 2013 Moore, Oklahoma tornado with a 100-metre grid-length NWP model. *Atmospheric Science Letters*, 17, 453–461.

Hanley, K.E., Plant, R.S., Stein, T.H., Hogan, R.J., Nicol, J.C., Lean, H.W. et al. (2015) Mixing-length controls on high-resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society*, 141, 272–284.

Harris, L., McRae, A.T., Chantry, M., Dueben, P.D. & Palmer, T.N. (2022) A generative deep learning approach to stochastic downscaling of precipitation forecasts. *Journal of Advances in Modeling Earth Systems*, 14, e2022MS003120.

Hastings, D.A. & Dunbar, P. (1998) Development & assessment of the global land one-km base elevation digital elevation model (GLOBE). *ISPRS*, 4, 218–221.

Heidari, M.R., Song, Z., Degregori, E., Behrens, J. & Bockelmann, H. (2021) Concurrent calculation of radiative transfer in the atmospheric simulation in ECHAM-6.3. 05p2. *Geoscientific Model Development*, 14, 7439–7457.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A. et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, 12, 1–40. Available from: https://doi.org/10.1371/journal.pone.0169748

Hintz, K.S., Vedel, H. & Kaas, E. (2019) Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation. *Meteorological Applications*, 26, 733–746.

Hirt, M., Rasp, S., Blahak, U. & Craig, G.C. (2019) Stochastic parameterization of processes leading to convective initiation in kilometer-scale models. *Monthly Weather Review*, 147, 3917–3934.

Hohenegger, C. & Schar, C. (2007) Atmospheric predictability at synoptic versus cloud-resolving scales. *Bulletin of the American Meteorological Society*, 88, 1783–1794.

Honnert, R., Efstathiou, G.A., Beare, R.J., Ito, J., Lock, A., Neggers, R. et al. (2020) The atmospheric boundary layer and the "gray zone" of turbulence: A critical review. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD030317.

Honnert, R. & Masson, V. (2014) What is the smallest physically acceptable scale for 1D turbulence schemes? *Frontiers in Earth Science*, 2, 27.

Honnert, R., Masson, V. & Couvreux, F. (2011) A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometric scale. *Journal of the Atmospheric Sciences*, 68, 3112–3131.

Hou, J., Van Dijk, A.I.J.M., Renzullo, L.J. & Larraondo, P.R. (2022) GloLakes: a database of global lake water storage dynamics from 1984 to present derived using laser and radar altimetry and optical remote sensing. *Earth System Science Data Discussions*, 2022, 1–20. Available from: https://essd.copernicus.org/preprints/essd-2022-266/

Hu, Y., Chen, L., Wang, Z. & Li, H. (2023) SwinVRNN: A Data-Driven Ensemble Forecasting Model via Learned Distribution Perturbation. *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003211.

Husain, S.Z., Girard, C., Separovic, L., Plante, A. & Corvec, S. (2020) On the progressive attenuation of finescale orography contributions to the vertical coordinate surfaces within a terrain-following coordinate system. *Monthly Weather Review*, 148, 4143–4158.

Huuskonen, A., E.S. & Holleman, I. (2014) The operational weather radar network in Europe. *Bulletin of the American Meteorological Society*, 95, 897–907. Available from: https://doi.org/10.1175/BAMS-D-12-00216.1

Institut national de l'information géographique et forestière. (2021) RGE ALTI 1m.

Jackson, T.L., Feddema, J.J., Oleson, K.W., Bonan, G.B. & Bauer, J.T. (2010) Parameterization of urban characteristics for global climate modeling. *Ann Assoc Am Geogr*, 100, 848–865. https://doi.org/10.1080/00045608.2010.497328.

Jakub, F. & Mayer, B. (2015) A three-dimensional parallel radiative transfer model for atmospheric heating rates for use in cloud resolving models–the TenStream solver. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 163, 63–71.

Jakub, F. & Mayer, B. (2017) The role of 1-d and 3-d radiative heating in the organization of shallow cumulus convection and the formation of cloud streets. *Atmospheric Chemistry and Physics*, 17, 13317–13327.

Joe, P., Belair, S., Bernier, N., Bouchet, V., Brook, J., Brunet, D. et al. (2018) The environment Canada pan and parapan American science showcase project. *Bulletin of the American Meteorological Society*, 99, 921–953.

Joulin, P.-A., Mayol, M.L., Masson, V., Blondel, F., Rodier, Q., Cathelain, M. et al. (2020) The actuator line method in the meteorological LES model meso-NH to analyze the horns rev 1 wind farm photo case. *Frontiers in Earth Science*, 7, 350.

Jun, C., Ban, Y. & Li, S. (2014) Open access to earth land-cover map. *Nature*, 514, 434.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M. & Brumby, S.P. (2021) Global land use/land cover with Sentinel 2 and deep learning. In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. Manhattan, NY, USA: IEEE, pp. 4704–4707.

Keisler, R. (2022) Forecasting global weather with graph neural networks. arXiv preprint, arXiv:2202.07575.

Kendon, E., Prein, A., Senior, C. & Stirling, A. (2021) Challenges and outlook for convection-permitting climate modelling. *Philosophical Transactions of the Royal Society A*, 379, 20190547.

Kendon, E.J., Roberts, N.M., Senior, C.A. & Roberts, M.J. (2012) Realism of rainfall in a very high-resolution regional climate model. *Journal of Climate*, 25, 5791–5806.

Kent, C.W., Grimmond, S., Gatey, D. & Hirano, K. (2019) Urban morphology parameters from global digital elevation models: Implications for aerodynamic roughness and for wind-speed estimation. *Remote Sensing of Environment*, 221, 316–339.

Kirillin, G., Hochschild, J., Mironov, D., Terzhevik, A., Golosov, S. & Nützmann, G. (2011) Flake-global: Online lake model with worldwide coverage. *Environmental Modelling & Software*, 26, 683–684.

Klein, R. & Benacchio, T. (2016) A doubly blended model for multiscale atmospheric dynamics. *Journal of the Atmospheric Sciences*, 73, 1179–1186.

Klemp, J., Dudhia, J. & Hassiotis, A. (2008) An upper gravity-wave absorbing layer for NWP applications. *Monthly Weather Review*, 136, 3987–4004.

Klemp, J.B. (2011) A terrain-following coordinate with smoothed coordinate surfaces. *Monthly Weather Review*, 139, 2163–2169.

Klemp, J.B. & Durran, D.R. (1983) An upper boundary condition permitting internal gravity wave radiation in numerical mesoscale models. *Monthly Weather Review*, 111, 430–444.

Kober, K. & Craig, G.C. (2016) Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of the Atmospheric Sciences*, 73, 2893–2911.

Konstantinov, P., Varentsov, M. & Esau, I. (2018) A high density urban temperature network deployed in several cities of Eurasian arctic. *Environmental Research Letters*, 13, 075007.

Koopmans, S., van Haren, R., Theeuwes, N., Ronda, R., Uijlenhoet, R., Holtslag, A. et al. (2023) The set-up and evaluation of fine-scale data assimilation for the urban climate of Amsterdam.

Quarterly Journal of the Royal Meteorological Society, 149(750), 171–191.

Kourzeneva, E., Asensio, H., Martin, E. & Faroux, S. (2012) Global gridded dataset of lake coverage and lake depth for use in numerical weather prediction and climate modelling. *Tellus A: Dynamic Meteorology and Oceanography*, 64, 15640.

Kurowski, M.J., Grabowski, W.W. & Smolarkiewicz, P.K. (2014) Anelastic and compressible simulation of moist deep convection. *Journal of the Atmospheric Sciences*, 71, 3767–3787.

Lagouarde, J.-P., Hénon, A., Irvine, M., Voogt, J., Pigeon, G., Moreau, P. et al. (2012) Experimental characterization and modelling of the nighttime directional anisotropy of thermal infrared measurements over an urban area: Case study of Toulouse (France). *Remote Sensing of Environment*, 117, 19–33. Available from: https://www.sciencedirect.com/science/article/pii/S0034425711002793

Lagouarde, J.-P. & Irvine, M. (2008) Directional anisotropy in thermal infrared measurements over Toulouse city centre during the CAPITOUL measurement campaigns: first results. *Meteorology and Atmospheric Physics*, 102, 173–185.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F. et al. (2022) GraphCast: Learning skillful medium-range global weather forecasting. arXiv preprint, arXiv:2212.12794.

Laprise, R. (1992) The euler equations of motion with hydrostatic pressure as an independent variable. *Monthly Weather Review*, 120, 197–207.

Larson, V.E., Wood, R., Field, P.R., Golaz, J.-C., Haar, T.H.V. & Cotton, W.R. (2001) Small-scale and mesoscale variability of scalars in cloudy boundary layers: One-dimensional probability density functions, 58, 1978–1994. Available from: https://doi.org/10.1175/1520-0469(2001)058<1978:SSAMVO>2.0.CO;2

Leutbecher, M. & Ben Bouallègue, Z. (2020) On the probabilistic skill of dual-resolution ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 146, 707–723.

Lipps, F.B. & Hemler, R.S. (1982) A scale analysis of deep moist convection and some related numerical calculations. *Journal of the Atmospheric Sciences*, 39, 2192–2210.

Lipson, M., Grimmond, S., Best, M., Abramowitz, G., Coutts, A., Tapper, N. et al. (2023) Evaluation of 30 urban land surface models in the urban-plumber project: Phase 1 results. *Quarterly Journal of the Royal Meteorological Society*, 150(758), 126–169.

Lipson, M.J., Nazarian, N., Hart, M.A., Nice, K.A. & Conroy, B. (2022) A transformation in city-descriptive input data for urban climate models. *Frontiers in Environmental Science*, 10, 866398.

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

Lorenz, E.N. (1996) Predictability – a problem partly solved. In: *Seminar on Predictability*, Vol. 1. Reading, United Kingdom: ECMWF, pp. 1–18.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L. et al. (2000) Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, 21, 1303–1330.

Lu, J., Nazarian, N. & Hart, M.A. (2022) OSM2LES - A python-based tool to prepare realistic urban geometry for LES simulation from OpenStreetMap (0.1.0). https://doi.org/10.5281/zenodo.6566202

Marquès, E., Masson, V., Naveau, P., Mestre, O., Dubreuil, V. & Richard, Y. (2022) Urban heat island estimation from crowdsensing thermometers embedded in personal cars. *Bulletin of the American Meteorological Society*, 103, E1098–E1113.

Martilli, A., Clappier, A. & Rotach, M.W. (2002) An urban surface exchange parameterisation for mesoscale models. *Boundary-Layer Meteorology*, 104, 261–304.

Mason, P. (1994) Large-eddy simulation: A critical review of the technique. *Quarterly Journal of the Royal Meteorological Society*, 120, 1–26.

Masson, V., Champeaux, J.-L., Chauvin, F., Meriguet, C. & Lacaze, R. (2003) A global database of land surface parameters at 1-km resolution in meteorological and climate models. *Journal of Climate*, 16, 1261–1282.

Masson, V., Heldens, W., Bocher, E., Bonhomme, M., Bucher, B., Burmeister, C. et al. (2020a) City-descriptive input data for urban climate models: Model requirements, data sources and challenges. *Urban Climate*, 31, 100536.

Masson, V., Lemonsu, A., Hidalgo, J. & Voogt, J. (2020b) Urban climates and climate change. *Annual Review of Environment and Resources*, 45, 411–444.

Meier, F., Fenner, D., Grassmann, T., Otto, M. & Scherer, D. (2017) Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, 19, 170–191.

Meyer, D., Grimmond, S., Dueben, P., Hogan, R. & van Reeuwijk, M. (2022a) Machine learning emulation of urban land surface processes. *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002744.

Meyer, D., Hogan, R.J., Dueben, P.D. & Mason, S.L. (2022b) Machine learning emulation of 3D cloud radiative effects. *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002550.

Middel, A., Nazarian, N., Demuzere, M. & Bechtel, B. (2022) Urban climate informatics: An emerging research field. *Frontiers in Environmental Science*, 10, 867434.

Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M. et al. (2020) Hoy 4D-Var in the Met Office UKV operational forecast model. *Quarterly Journal of the Royal Meteorological Society*, 146, 1281–1301.

Mills, G., Ching, J. & Bechtel, B. (2021) Characterising urban morphology for urban climate modelling. In: *Urban Climate Science for Planning Healthy Cities*. Springer, Cham, pp. 339–354.

Moeng, C.-H., Sullivan, P., Khairoutdinov, M. & Randall, D. (2010) A mixed scheme for subgrid-scale fluxes in cloud-resolving models. *Journal of the Atmospheric Sciences*, 67, 3692–3705.

Montmerle, T., Michel, Y., Arbogast, E., Ménétrier, B. & Brousseau, P. (2018) A 3d ensemble variational data assimilation scheme for the limited-area arome model: Formulation and preliminary results. *Quarterly Journal of the Royal Meteorological Society*, 144, 2196–2215.

Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G. et al. (2015) Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, 35, 3185–3203.

Munoz-Esparza, D., Kosovic, B., Mirocha, J. & van Beeck, J. (2014) Bridging the transition from mesoscale to microscale turbulence in numerical weather prediction models. *Boundary-Layer Meteorology*, 153, 409–440.

Nachtergaele, F., van Velthuizen, H., Verelst, L., Batjes, N., Dijkshoorn, K., van Engelen, V. et al. (2010) The harmonized world soil database. In: *Proceedings of the 19th World Congress of Soil*

Science, Soil Solutions for a Changing World, Brisbane, Australia, 1-6 August 2010, pp. 34–37.

Nagel, T., Schoetter, R., Bourgin, V., Masson, V. & Onofri, E. (2023) Drag coefficient and turbulence mixing length of local climate zone-based urban morphologies derived using obstacle-resolving modelling. *Boundary-Layer Meteorology*, 186, 737–769.

Nance, L.B. & Durran, D.R. (1994) A comparison of the accuracy of three Anelastic systems and the pseudo-incompressible system. *Journal of Atmospheric Sciences*, 51, 3549–3565.

NASA. (2001) ASTER DEM Product. https://lpdaac.usgs.gov/products/ast14demv003/

NASA. (2013) Nasa shuttle radar topography mission global 1 arc second. https://lpdaac.usgs.gov/products/srtmgl1v003/

Nipen, T., Chantry, M. et al. (2024) Data-driven regional modelling. https://www.ecmwf.int/en/about/media1751centre/aifs-blog/2024/data-driven-regional-modelling

Nipen, T., Seierstad, I., Lussana, C., Kristiansen, J. & Hov, Ø. (2020) Citizen observations for weather forecasts. *Bulletin of the American Meteorological Society*, 101, 137–142.

Ogura, Y. & Phillips, N.A. (1962) Scale analysis of deep and shallow convection in the atmosphere. *Journal of the Atmospheric Sciences*, 19, 173–179.

Oke, T.R., Mills, G., Christen, A. & Voogt, J.A. (2017) *Urban Climates*. Cambridge, United Kingdom: Cambridge University Press.

Oliger, J. & Sundström, A. (1978) Theoretical and practical aspects of some initial boundary value problems in fluid dynamics. *SIAM Journal on Applied Mathematics*, 35, 419–446.

Oskarsson, J., Landelius, T. & Lindsten, F. (2023) Graph-based neural weather prediction for limited area modeling. arXiv preprint, arXiv:2309.17370.

Overeem, A., Leijnse, H. & Uijlenhoet, R. (2013) Country-wide rainfall maps from cellular communication networks. *Proceedings of the National Academy of Sciences*, 110, 2741–2745.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M. et al. (2022) Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. arXiv preprint, arXiv:2202.11214.

Patton, E.G., Sullivan, P.P., Shaw, R.H., Finnigan, J.J. & Weil, J.C. (2016) Atmospheric stability influences on coupled boundary layer and canopy turbulence. *Journal of the Atmospheric Sciences*, 73, 1621–1647.

Pesaresi, M., Corbane, C., Ren, C. & Edward, N. (2021) Generalized Vertical Components of built-up areas from global Digital Elevation Models by multi-scale linear regression modelling. *PLoS One*, 16, e0244478.

Pesaresi, M. & Politis, P. (2023a) GHS-BUILT-H R2023A - GHS building height, derived from AW3D30, SRTM30, and Sentinel2 composite (2018). European Commission, Joint Research Centre (JRC).

Pesaresi, M. & Politis, P. (2023b) GHS-BUILT-V R2023A - GHS built-up volume grids derived from joint assessment of Sentinel2, Landsat, and global DEM data, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC).

Plant, R. & Craig, G.C. (2008) A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, 65, 87–105.

Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K. et al. (2015) A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53, 323–361.

Price, I., Sanchez-Gonzalez, A., Alet, F., Ewalds, T., El-Kadi, A., Stott, J. et al. (2023) Gencast: Diffusion-based ensemble forecasting for medium-range weather. arXiv preprint, arXiv:2312.15796.

Rasp, S., Pritchard, M.S. & Gentine, P. (2018) Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115, 9684–9689.

Roberts, N., Ayliffe, B., Evans, G., Moseley, S., Rust, F., Sandford, C. et al. (2023) IMPROVER: the new probabilistic postprocessing system at the Met Office. *Bulletin of the American Meteorological Society*, 104, E680–E697.

Rockel, B., Will, A. & Hense, A. (2008) The regional climate model COSMO-CLM (CCLM). *Meteorologische Zeitschrift*, 17, 347–348. Available from: https://doi.org/10.1127/0941-2948/2008/0309

Ronda, R., Steeneveld, G., Heusinkveld, B., Attema, J. & Holtslag, A. (2017) Urban finescale forecasting reveals weather conditions with unprecedented detail. *Bulletin of the American Meteorological Society*, 98, 2675–2688.

Sabatier, T., Paci, A., Lac, C., Canut, G., Largeron, Y. & Masson, V. (2020b) Semi-idealized simulations of wintertime flows and pollutant transport in an Alpine valley: Origins of local circulations (Part I). *Quarterly Journal of the Royal Meteorological Society*, 146, 807–826.

Sakradzija, M. & Klocke, D. (2018) Physically constrained stochastic shallow convection in realistic kilometer-scale simulations. *Journal of Advances in Modeling Earth Systems*, 10, 2755–2776.

Sakradzija, M., Seifert, A. & Dipankar, A. (2016) A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, 8, 786–812.

Sakradzija, M., Seifert, A. & Heus, T. (2015) Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Processes in Geophysics*, 22, 65–85.

Schalkwijk, J., Jonker, H.J., Siebesma, A.P. & Van Meijgaard, E. (2015) Weather forecasting using GPU-based large-eddy simulations. *Bulletin of the American Meteorological Society*, 96, 715–723.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Di Girolamo, S. et al. (2020) Kilometer-scale climate models: Prospects and challenges. *Bulletin of the American Meteorological Society*, 101, E567–E587.

Schär, C., Leuenberger, D., Fuhrer, O., Lüthi, D. & Girard, C. (2002) A new terrain-following vertical coordinate formulation for atmospheric prediction models. *Monthly Weather Review*, 130, 2459–2480.

Schoetter, R., Caliot, C., Chung, T.-Y., Hogan, R.J. & Masson, V. (2023) Quantification of uncertainties of radiative transfer calculation in urban canopy models. *Boundary-Layer Meteorology*, 189(1), 103–138.

Schoetter, R., Kwok, Y.T., de Munck, C., Lau, K.K.L., Wong, W.K. & Masson, V. (2020) Multi-layer coupling between SURFEX-TEB-v9. 0 and Meso-NH-v5. 3 for modelling the urban climate of high-rise cities. *Geoscientific Model Development*, 13, 5609–5643.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A. et al. (2016) Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, 142, 1453–1472.

Schwartz, C.S. & Sobash, R.A. (2017) Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Monthly Weather Review*, 145, 3397–3418.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F. et al. (2011) The AROME-France convective-scale operational model. *Monthly Weather Review*, 139, 976–991.

Sgoff, C., Acevedo, W., Paschalidi, Z., Ulbrich, S., Bauernschubert, E., Kratzsch, T. et al. (2022) Assimilation of crowd-sourced surface observations over Germany in a regional weather prediction system. *Journal of the Royal Meteorological Society*, 148, 1752–1767. Available from: https://doi.org/10.1002/qj.4276

Shahtahmassebi, A.R., Li, C., Fan, Y., Wu, Y., Gan, M., Wang, K. et al. (2021) Remote sensing of urban green spaces: A review. *Urban Forestry & Urban Greening*, 57, 126946.

Sharma, K., Lee, J.C.K., Porson, A., Chandramouli, K., Roberts, N., Boyd, D. et al. (2023) Adaptive selection of members for convective-permitting regional ensemble prediction over the western Maritime Continent. *Frontiers in Environmental Science*, 11, 1281265.

Shaw, J. & Weller, H. (2016) Comparison of terrain-following and cut-cell grids using a nonhydrostatic model. *Monthly Weather Review*, 144, 2085–2099.

Short, C.J. & Petch, J. (2022) Reducing the spin-up of a regional NWP system without data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 148(745), 1623–1643.

Smith, D.K., Renfrew, I.A., Dorling, S.R., Price, J.D. & Boutle, I.A. (2021) Sub-km scale numerical weather prediction model simulations of radiation fog. *Quarterly Journal of the Royal Meteorological Society*, 147, 746–763.

Smolarkiewicz, P.K., Kühnlein, C. & Wedi, N.P. (2014) A consistent framework for discrete integrations of soundproof and compressible PDEs of atmospheric dynamics. *Journal of Computational Physics*, 263, 185–205.

Smolíková, P. & Vivoda, J. (2023) Stability properties of the constant coefficients semi-implicit time schemes solving non-filtering approximation of the fully compressible equations. *Monthly Weather Review*, 151(7), 1797–1819.

Stappers, R. & Barkmeijer, J. (2011) Optimal linearization trajectories for tangent linear models. *Quarterly Journal of the Royal Meteorological Society*, 138, 170–184.

Steppeler, J., Bitzer, H.-W., Minotte, M. & Bonaventura, L. (2002) Nonhydrostatic atmospheric modeling using a z-coordinate representation. *Monthly Weather Review*, 130, 2143–2149.

Stewart, I.D. & Oke, T.R. (2012) Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93, 1879–1900.

Sullivan, P.P. & Patton, E.G. (2011) The effect of mesh resolution on convective boundary layer statistics and structures generated by large-eddy simulation. *Journal of the Atmospheric Sciences*, 68, 2395–2415. Available from: http://journals.ametsoc.org/jas/article-pdf/68/10/2395/3526167/jas-d-10-05010_1.pdf

Surcel, M., Zawadzki, I. & Yau, M. (2015) A study on the scale dependence of the predictability of precipitation patterns. *Journal of the Atmospheric Sciences*, 72, 216–235.

Sützl, B.S., Rooney, G.G., Finnenkoetter, A., Bohnenstengel, S.I., Grimmond, S. & van Reeuwijk, M. (2021a) Distributed urban drag parametrization for sub-kilometre scale numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 147, 3940–3956.

Sützl, B.S., Rooney, G.G. & van Reeuwijk, M. (2021b) Drag distribution in idealized heterogeneous urban environments. *Boundary-Layer Meteorology*, 178, 225–248.

Tang, Y., Lean, H.W. & Bornemann, J. (2013) The benefits of the met office variable resolution NWP model for forecasting convection. *Meteorological Applications*, 20, 417–426.

Tóth, H. & Szintai, B. (2021) Assimilation of leaf area index and soil water index from satellite observations in a land surface model in Hungary. *Atmosphere*, 12, 944.

Tricht, K.V., Degerickx, J., Gilliams, S., Zanaga, D., Battude, M., Grosu, A. et al. (2023) Worldcereal: A dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping. *Earth System Science Data*, 15, 5491–5515. Available from: https://essd.copernicus.org/articles/15/5491/2023/

Valkonen, T., Stoll, P., Batrak, Y., Køltzow, M., Schneider, T., Stigter, E. et al. (2020) Evaluation of a sub-kilometre NWP system in an arctic fjord-valley system in winter. *Tellus A*, 72(1), 1–21.

van Stratum, B., Theeuwes, N., Barkmeijer, J., van Ulft, B. & Wijnant, I. (2022) A one-year-long evaluation of a wind-farm parameterization in HARMONIE-AROME. *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002947.

Van Weverberg, K., Morcrette, C.J., Boutle, I., Furtado, K. & Field, P.R. (2021) A bimodal diagnostic cloud fraction parameterization. Part I: Motivating analysis and scheme description. *Monthly Weather Review*, 149, 841–857.

Vannitsem, S., Bremnes, J.B., Demaeyer, J., Evans, G.R., Flowerdew, J., Hemri, S. et al. (2020) Statistical postprocessing for weather forecasts–review, challenges and avenues in a big data world. *Bulletin of the American Meteorological Society*, 102(3), E681–E699.

Varquez, A.C.G., Kiyomoto, S., Khanh, D.N. & Kanda, M. (2021) Global 1-km present and future hoy anthropogenic heat flux. *Scientific Data*, 8, 1–14.

Veerman, M.A., van Stratum, B.J.H. & van Heerwaarden, C.C. (2022) A case study of cumulus convection over land in cloud-resolving simulations with a coupled ray tracer. *Geophysical Research Letters*, 49, e2022GL100808. Available from: https://doi.org/10.1029/2022gl100808

Verrelle, A., Ricard, D. & Lac, C. (2017) Evaluation and improvement of turbulence parameterization inside deep convective clouds at kilometer-scale resolution. *Monthly Weather Review*, 145, 3947–3967. Available from: https://journals.ametsoc.org/view/journals/mwre/145/10/mwr-d-16-0404.1.xml

Vié, B., Nuissier, O. & Ducrocq, V. (2011) Cloud-resolving ensemble simulations of mediterranean heavy precipitating events: Uncertainty on initial conditions and lateral boundary conditions. *Monthly Weather Review*, 139, 403–423. Available from: https://journals.ametsoc.org/view/journals/mwre/139/2/2010mwr3487.1.xml

Villefranque, N., Fournier, R., Couvreux, F., Blanco, S., Cornet, C., Eymet, V. et al. (2019) A path-tracing Monte Carlo library for 3-D radiative transfer in highly resolved cloudy atmospheres. *Journal of Advances in Modeling Earth Systems*, 11, 2449–2473.

Vionnet, V., Marsh, C.B., Menounos, B., Gascoin, S., Wayand, N.E., Shea, J. et al. (2021) Multi-scale snowdrift-permitting modelling of mountain snowpack. *The Cryosphere*, 15, 743–769.

Vivoda, J., Smolíková, P. & Simarro, J. (2018) Finite elements used in the vertical discretization of the fully compressible core of the ALADIN system. *Monthly Weather Review*, 146, 3293–3310.

Voitus, F., Bénard, P., Kühnlein, C. & Wedi, N.P. (2019) Semi-implicit integration of the unified equations in a mass-based coordinate: model formulation and numerical testing. *Quarterly Journal of the Royal Meteorological Society*, 145, 3387–3408.

Vosper, S., Carter, E., Lean, H., Lock, A., Clark, P. & Webster, S. (2013) High resolution modelling of valley cold pools. *Atmospheric Science Letters*, 14, 193–199.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S. et al. (2013) The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*, 22, 1–29.

Walsh, E., Bessardon, G., Gleeson, E. & Ulmas, P. (2021) Using machine learning to produce a very high resolution land-cover map for Ireland. *Advances in Science and Research*, 18, 65–87.

Walters, D., Baran, A.J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J. et al. (2019) The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geoscientific Model Development*, 12, 1909–1963.

Wessel, B. (2016) TanDEM-X ground segment–DEM products specification document. file:///C:/Users/lewis.blunn/Downloads/TD-GS-PS-0021_DEM-Product-Specification_v3.1.pdf.

Weyn, J.A., Durran, D.R., Caruana, R. & Cresswell-Clay, N. (2021) Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, 13, e2021MS002502.

Wilhelmson, R. & Ogura, Y. (1972) The pressure perturbation and the numerical modeling of a cloud. *Journal of Atmospheric Sciences*, 29, 1295–1307.

Wu, Y., Teufel, B., Sushama, L., Belair, S. & Sun, L. (2021) Deep learning-based super-resolution climate simulator-emulator framework for urban heat studies. *Geophysical Research Letters*, 48, e2021GL094737.

Wyngaard, J. (2004) Toward numerical modeling in the "terra incognita", 61, 1816–1826.

Yu, T., Sützl, B.S. & van Reeuwijk, M. (2022) Urban neighbourhood classification and multi-scale heterogeneity analysis of Greater London. *Environment and Planning B: Urban Analytics and City Science*, 2050(6), 1534–1558.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G. et al. (2022) ESA WorldCover 10 m 2021 v200. https://doi.org/10.5281/zenodo.7254221

Zängl, G. (2012) Extending the numerical stability limit of terrain-following coordinate models over steep slopes. *Monthly Weather Review*, 140, 3722–3733.

Zhang, F., Snyder, C. & Rotunno, R. (2003) Effects of moist convection on mesoscale predictability. *Journal of the Atmospheric Sciences*, 60, 1173–1185.

Zink, M., Bachmann, M., Brautigam, B., Fritz, T., Hajnsek, I., Moreira, A. et al. (2014) TanDEM-X: The new global DEM takes shape. *IEEE Geoscience and Remote Sensing Magazine*, 2, 8–23.

How to cite this article: Lean, H.W., Theeuwes, N.E., Baldauf, M., Barkmeijer, J., Bessardon, G., Blunn, L. *et al.* (2024) The hectometric modelling challenge: Gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, 1–38. Available from: https://doi.org/10.1002/qj.4858