# *Summary and Recommendations from Working Group 1: model uncertainty representations in convection-permitting / shorter lead-time / limited-area ensembles*

# Summary and Recommendations from Working Group 1: Model uncertainty representations in convection-permitting / shorter lead-time / limited-area ensembles

Co-chairs: *Bob Plant (Uni Reading) and Alan Hally (Met Éireann)*

Rapporteur: *Sarah-Jane Lock (ECMWF)*

Participants: *Maike Ahlgrimm (DWD), Marco Arpagaus (MeteoSwiss), Werner Bauer (Kingston Uni), Carlo Cafaro (UKMO), Ihan Chen (CWB, Taiwan), James Fannon (Met Éireann), Henrik Feddersen (DMI), Axelle Fleury (Météo France), Inger-Lise Frogner (Norwegian Met. Inst.), Christoph Gebhardt (DWD), Karoliina Hämäläinen (FMI), Maicon Hieronymus (JGU Mainz), Janne Kauhanen (FMI), Salah Kouhen (Uni Oxford), Patrick Kuntze (JGU Mainz), Humphrey Lean (UKMO), Ashu Mamgain (NCMRWF), Anne McCabe (UKMO), Marco Milan (UKMO), Jon Petch (UKMO), Matjaž Puh (LMU München), Christoph Schraff (DWD), Alison Stirling (UKMO), Kirsten Tempest (LMU München), Aristofanis Tsiringakis (KNMI), Florian Weidle (ZAMG), Meryl Wimmer (LMD), ZhiZhen Xu (Chinese Academy Met. Sci.)*

Working group 1 considered the treatment of model uncertainty (MU) in high-resolution ensembles, at grid spacings of order 1-5 km. These systems are often run for regional weather forecasting, perhaps over a single country, and for lead times of up to 5 days. Looking ahead, ECMWF's strategy seeks to deliver global medium-range ensemble forecasts with 3-4 km grid spacings by 2030. It is questionable for what grid spacing we should dispense with a deep convection parameterization, but it will be either switched off or damped in these systems, such that deep convection can be assumed to be dominated by explicit motions. One of the problems with limited-area ensemble systems at this scale is that spread depends not only on the modelling system itself but also on the variability inherited from the large-scale boundary conditions. There is often thought to be a lack of spread in our high-resolution EPS (ensemble prediction systems), but this could reflect a lack of diversity on larger scales. The relative importance of lateral-boundary diversity and the model uncertainty mechanisms is regime dependent. The lateral boundaries will generally be more important in midlatitude winter but less so for summertime convection in relatively weak synoptic flow.

Here we present a summary of our discussions and some recommendations for ECMWF and the wider community. Recommendations are written in italics and labelled as *[Rx]*.

## State of the art, and its issues

The main MU packages in use are stochastically perturbed parameterizations (SPP) and stochastically perturbed parameterization tendencies (SPPT) with variants of each (e.g., iSPPT, pSPPT). The existing approaches clearly do have value (e.g., SPP was able to address known weaknesses for fog in HarmonEPS). At ECMWF and elsewhere, there appears to be a preference for SPP at higher resolutions. This may be in part because without the deep convection parameterization then “there is less for SPPT to work on” and so perturbing parameters can be a more direct way to induce impacts on model behaviours.

An important issue at high resolution is to generate sufficient perturbations within the boundary layer. To some extent, this may reflect a greater focus on (near-)surface variables in the outputs from these models, but the issue also arises because of the role of the boundary layer in generating variability within the forecast as a whole. For example, boundary-layer fluctuations will disrupt the starting conditions for when and where explicit moist convection develops. In most implementations, SPPT is tapered within the boundary layer for reasons of numerical stability and over-dispersion. More favoured results at higher resolution have been obtained in systems where the tapering is not required (e.g., MeteoSwiss). In systems where tapering is necessary, the move towards SPP is partly motivated by providing additional boundary-layer variability. Another important motivation for SPP is physical consistency, in the sense that it retains the energy and moisture conservation of the unperturbed forecast.

Examples were described in which the search for sufficient near-surface spread encouraged the use of large perturbations that occasionally produce a physically unrealistic state. These included cases of excessive cold-pooling (with SPPT at MeteoSwiss), unrealistic suppression of vertical mixing (with SPPT in the IFS), and (with SPP in HarmonEPS) large perturbations of a mixing length parameter which can transform a nocturnal stable boundary-layer into a well-mixed state (see Figure 1; notice the areas of large mixed-layer depth in member 3 over France and southern Germany that coincide with large positive perturbations of the minimum mixing length scale).

*[R1] Climate model uncertainty studies have sought to identify regions of parameter space which a model should not be allowed to occupy to avoid unphysical results. WG1 recommends that this thinking should be considered for constraining MU in NWP.*

To do this will require the monitoring of simulations and the development of suitable methods to identify problematic perturbations. This should include soliciting input from forecasters/end-users, who can advise when individual forecasts are not credible. Furthermore, within the W2W[1] programme, a tangent linear approach is being developed which aims to select model parameters that could usefully be perturbed, and to identify the time and place where a perturbation would have significant impact.

[1] https://www.wavestoweather.de

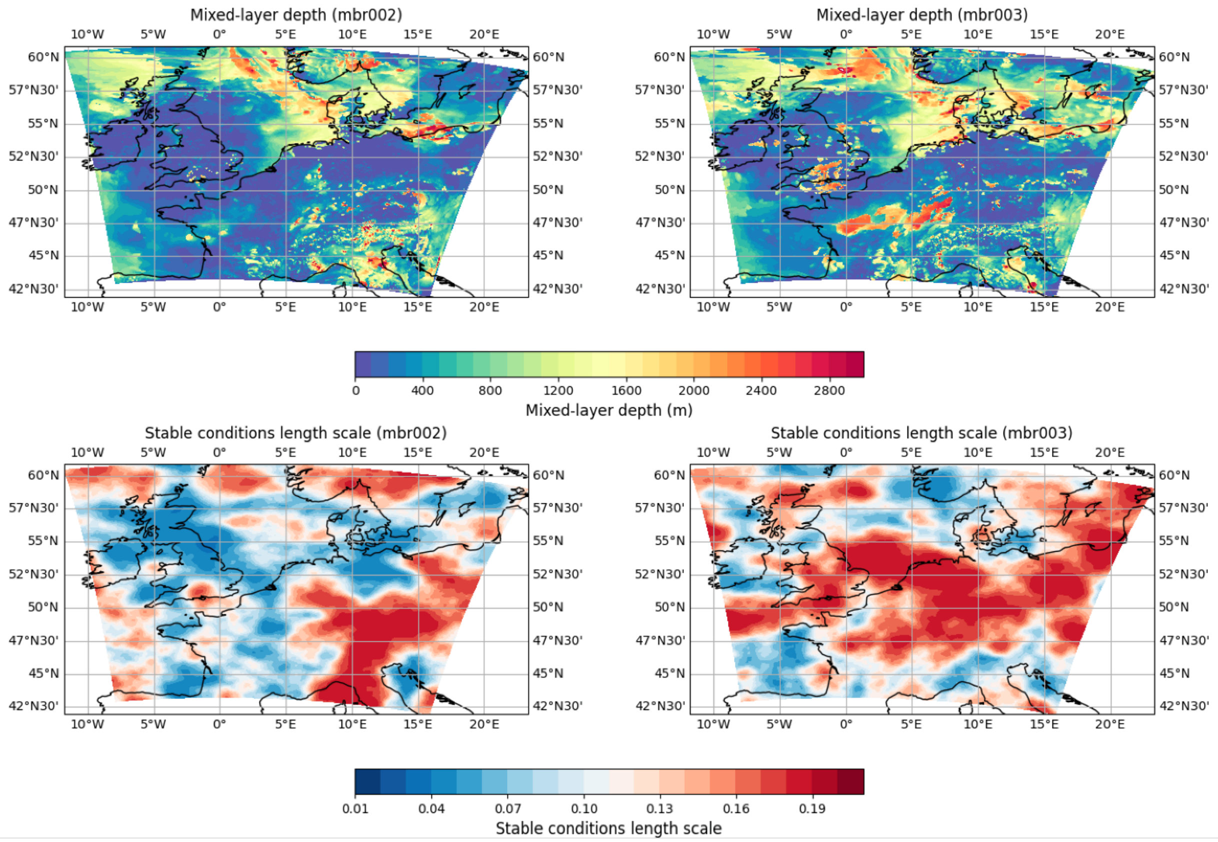


Figure 1: Modelled mixed-layer depth (in m) and stable condition length scale (SPP name RZC_H) for members 2 and 3 of the HarmonEPS at 0000 UTC 12 July 2020 (24-h forecast time) over the Netherlands domain ($\Delta x$ = 2.5km). Results are selected from a 7-day HarmonEPS experiment (06 – 13 July 2020), with 36-h forecasts starting at 0000 UTC every day, 3-h 3DVAR data assimilation for the control member (member 0) and SPP perturbations for 9 parameters, active in members 1-6. The default value for the stable condition length scale is 0.11 m. (Figure courtesy of A. Tsiringakis.)

There are physically-based schemes which aim to stochastically modify deterministic parameterizations at the process level to represent particular mechanisms that cause uncertainty. A shallow convection scheme (Sakradzija et al., 2016, *JAMES*) and two boundary layer turbulence schemes (SBL and PSP2) were discussed at the workshop.

*[R2] WG1 recommends that any stochastic shallow convection scheme should be developed and investigated in conjunction with a compatible stochastic boundary layer method.*

These methods were well regarded by the group in terms of their physical appeal, but there were concerns about whether they would generate sufficient spread sufficiently quickly in practice compared to more generic approaches like SPP and SPPT. The perturbations do not have an associated mesoscale pattern (like those used in SPP/SPPT) except that which is generated as they grow upscale from the near-grid scales. However, the process may not be rapid enough to induce enough spread in practice given the length of the forecast period and/or the residence time of the air mass in the limited-area domain. On the other hand, the workshop has shown examples where stochastic parameterizations are valued because they can improve skill, and so such methods can be worth pursuing even for traditionally deterministic applications.

The STOCHDP method (being developed at ECMWF) and the advection error element in the SPP version implemented in the Canadian global GEPS introduce perturbations to the semi-Lagrangian departure point interpolation procedure, to address MU associated with dynamical features. This has mainly been tested at lower resolutions so far.

*[R3] WG1 recommends that efforts be made to explore such dynamical perturbations at higher resolutions.*

They might prove more effective where fields are rougher and vary rapidly on short scales, especially in the vicinity of partially-resolved convection. However, it was also discussed within the group that care would have to be paid to any implementation of STOCHDP (or similar) in convection-permitting models due to the potential for instability issues, particularly as these models transition to single precision versions of their code.

The most effective approaches in high-resolution ensembles may not only depend on physical and dynamical uncertainties and their interactions, but also on the region and domain of interest. For example, the UK Met Office runs a limited-area model over Northern Africa and a smaller UKV model over the UK only. In the latter case, the smaller domain and stronger climatological flow means that the boundary conditions from the global model dominate the ensemble more quickly and so there is a greater need for the perturbations within the high-resolution system to act quickly.

*[R4] WG1 recommends that consideration be given to perturbing ensemble boundary conditions in order to develop convective-scale structures in limited-area domains more quickly.*

We are not aware of this having been tried so far but such an approach could help better represent the uncertainties related to the track of e.g., mid-Atlantic winter storms.

## Path to future developments

Model uncertainty representation at high resolution is a less mature topic than at lower resolution.

*[R5] WG1 recommends encouraging a diversification of methods and ideas in the field over the short-to-medium term, for the next 5 years or so*.

New avenues should be considered including SPDEs (for stochastic transport) and exploring uncertainties attributable to different aspects of the modelling system.

*[R6] Over the same period, WG1 recommends that the community seeks to extend its methods of assessment to encompass more bespoke approaches for higher resolution EPS, and to converge in identifying a subset of assessments that are particularly informative for these scales.*

It is currently very difficult to translate from the results for a handful of traditionally-reported metrics from one MU scheme in one high-resolution model into other settings. New assessment methods were considered in more detail by Working Group 3, and are under active development (e.g., SINFONY[2] project). WG1 emphasised the need for distinct approaches at high resolution to focus on storm scales and for more process-based assessments better targeted towards user concerns. Methods that assess the spatial scales of differences between ensemble members are also expected to be valuable, not least to disentangle the effects of perturbations due to a larger-scale parent model from those within the high-resolution system itself.

*[R7] WG1 recommends that, in the longer term (and following [R5] and [R6]), we seek to collaboratively assess a diverse set of candidate MU representations with a view towards convergence on the most effective representations for high resolution.*

More physically-based schemes are desirable but may not prove sufficient in isolation and combinations of approaches will likely be required. However, choices will have to be made with care and assessments will be needed of combinations of schemes and interactions between them. In order to be able to recognize overlaps between scheme designs:

*[R8] WG1 recommends that model uncertainty developers should be as explicit as possible in explaining the intended purpose of their method and any specific physical process that is targeted.*

This will help to develop awareness of whether combinations of schemes are valid combinations of distinct uncertainty sources or whether they may be “double counting”.

With SPP, there are many choices to be made about which variables to perturb, how much and with what inter-variable correlations. Some correlation structures may be desirable for physical consistency, such as where microphysical uncertainties ultimately derive from an unknown drop size distribution. With both SPP and SPPT, there are also parameters to be chosen in formulating pattern scales.

*[R9] WG1 recommends further exploring methods that can guide such choices in the ensemble context, including analysis of model tendencies, and perhaps applying machine learning.*

High-resolution ensembles are run for different reasons to their lower-resolution counterparts, particularly with a view towards high-impact weather such as fog or extreme wind or rainfall, and their downstream effects such as flooding. There was agreement that the different objectives can (indeed,

[2] http://www.dwd.de/EN/research/researchprogramme/sinfony_iafe/sinfony_en_node.html

should) affect the model uncertainty strategy. For example, it may motivate which parameters an SPP scheme targets in order to obtain sufficient spread in the metrics of key importance to forecast users.

*[R10] WG1 recommends further work to understand the needs of end-users for high-resolution systems, which should feed back into the strategies we choose to focus on.*

## Ways of working to facilitate future developments

We considered this at several levels:

### Within an organisation

*[R11] WG1 proposes that the stochastic aspects of parameterizations, and of the modelling systems more generally, should be considered a key part of the EPS. As such, they should be considered at every stage, including physics development, verification methods and communication with forecasters and end users.*

This requires more joined-up working within organisations. How scientists focussed on MU are perceived (and even where they are physically located!) within the organisation can be important. Several contributors felt that constructive steps were starting to happen. For example, UKMO is making big efforts to integrate EPS thinking into development and assessment. Also, some contributors from organisations which had developed SPP methods reflected that there was a need to work closely with model developers to build those methods, and that this had proved to be a positive process. Even the simple fact that the perturbed parameter settings are visible in the model code can concentrate minds and encourage collaboration.

### Operational centres and universities

It is generally necessary to take a longer-term view of the scientific aims in such collaborations since advances can take considerable time to fully feed through into operational practice. However, that longer perspective can also be seen as an advantage. It can, for example, encourage a more theoretical view (e.g., SPDEs) which might lead to larger changes over 10+ years and is complementary to the shorter- and medium-term processes of seeking improvements in existing methods and existing ways of thinking.

*[R12] WG1 recommends fostering collaborations between operational centres and universities, which should not necessarily be judged by whether there are immediate impacts on forecasting systems.*

The community more generally

It has been noted above that it can be difficult to learn from the experiences of others in testing MU representations at high resolutions because of differences in model domains, typical meteorological regimes and the situations that are of most pressing interest to local forecasters and users.

*[R13] WG1 recommends that centres share more about their operational acceptance criteria for adopting changes to MU representation, so that testing experiences can be better interpreted in the full context and MU schemes can have an operational focus from the earliest stages of their development.*

Testbed activities were discussed, where operational meteorologists, model developers and evaluation scientists work together, typically for several weeks, in assessing the day-to-day behaviour of the operational EPS system. These may take place internally or involving multiple forecasting centres (e.g., the ESSL Testbed[3]; the NOAA Hazardous Weather Testbed[4]). Such activities were felt to be valuable, and members of the group expressed desires to collaborate further.

*[R14] WG1 recommend that model uncertainty should be made a focus topic for some testbed events over the medium term, whether as bespoke events or as an occasional focus within ongoing activities.*

We recognize that a challenge with this strategy can be to maintain momentum since the greatest benefits may be realized only once interactions are embedded after several such events.

[3] https://www.essl.org/cms/essl-testbed/
[4] https://hwt.nssl.noaa.gov