Report of the 25th Met Office Scientific Advisory Committee Meeting (18-22nd January 2021)

Response from the Met Office Chief Scientist in red

We thank MOSAC for their contribution to a very productive meeting. This year, because of COVID restrictions, the meeting was conducted virtually via Microsoft Teams, with committee members joining from across the globe across many different time zones. We would like to thank the committee for their help in making a success of the meeting – particularly those who worked in unsociable hours to do so.

The MOSAC report is both detailed and insightful. The following section summarises the main points, which are made in more detail within the Annex. Responses from Met Office are therefore provided solely in the Annex in order to avoid duplication.

Summary

The 25th MOSAC meeting was held the 18-22 January 2020. MOSAC has appreciated that many of the topics included in the meeting agenda were aligned with the suggestions of MOSAC from last year. A notable point is that Applied Science is continuing its significant growth, as we have seen over the past years. This brings important public and private engagement questions on the Met Office strategy regarding what it wants to develop/provide versus what it would prefer the commercial sector to do, and what might be done in collaboration with the commercial sector to extend the reach of Met Office capabilities. Not many national meteorological services in the world are as well positioned than the Met Office to thrive in these interesting times. MOSAC is confident that the Met Office prediction systems and the delivery of its Research and Innovation Strategy is fit for purpose and one of the critical factors for the Met Office to be in such good position to provide a world class competitive service.

In the following we provide the highlights of the detailed MOSAC report in the Annex. The associated recommendations and questions are referred between brackets (e.g., NGMS-R1).

Science Communication

The Met Office has a high profile in the media, with government and with the public. There has been a lot of good work in communication in recent years and we should acknowledge initiatives like infographics, the naming of winter storms, enhanced work with the Science Media Centre and the Schools programme. MOSAC thinks there are ongoing questions on communication on climate topics and there is always more than can be done here, regarding: i) the use of the 'authoritative voice' and when and where that is appropriate (SC-R1); and ii) the value of engagement and communication activities for researchers (SC-R2 and 3).

Numerical Prediction R&D

MOSAC praises the recent numerical prediction system advances. In terms of operations, it was good to see beneficial use made of COSMIC-2 and SPIRE data. The changes for the trial atmosphere-ocean coupled forecast system have been rather swift as an example, but overall, there is still more testing to go (WS-R1) and planning to do (WS-R2). MOSAC welcomes the efforts to involve operational meteorologists in evaluating these new numerical prediction systems through testbeds (OP-R1 and 2). Communication of configurations and

evaluations should be a priority. Newsletters or blogs may be helpful to complement the (slower) scientific journal articles (FS-R6).

Next Generation Modelling Systems (NGMS) is progressing according to plan. MOSAC finds evidence of sound management methods and practices for NGMS. Risks, including those associated with timely convergence of development paths towards operations and the computational performance, are being tracked and plans adjusted as need arises. Current computational performance of the global model in NGMS is improving but it remains too expensive for operations by several factors. MOSAC is assured that efforts by Met Office personnel and Science and Technology Facilities Council partners will continue to narrow the gap, but there are still many challenges (NGMS-R1 and 2).

MOSAC praises the creation of the Joint Centre of Excellence in Environmental Intelligence (JCEEI). This constitutes a major milestone for the Met Office. The JCEEI already managed to develop partnerships with some of the leaders in the field. Its 'bottom-up' approach via selected pilot projects, initiated under the strategic action on fusing simulation with data science, is an excellent starting point. These projects are expected to expand, multiply, and inform on standards and lessons learnt for new projects (DS-R1 and 5).

Toward Numerical Environmental and Weather-to-climate Prediction (NEWP). Environmental modelling has a long history at the UKMO. The recent years' acceleration toward coupled systems is made possible by this long and strong base in numerous fields and emerging experts that operate at the interface of these newly coupled systems. The increased complexity brought upon by the growing number of coupled systems and related testing, including data assimilation (DA-R4), bring new challenges like performance verification for multiple customers or generation of ensemble members. These challenges can be tackled through different approaches like using simplified versions of the numerical prediction systems. (CM-R1, 2, 3, CM-Q1 and 2)

The future of NEWP lies in ensemble prediction, but more effort is needed at all time scales. Research and development on the production of NEWP ensembles has largely slowed down during NGMS development, but there will be a major need for this work immediately after the rollout, including more *quantitative* evidence based on numerical experiments. During this window of opportunity, where the expert personnel, tools and computing capacity are available, there should be a plan to address how to proceed forward for weather and climate ensemble prediction, including ocean (FS-R5, CS-R1, ENS-R1 and 2).

You cannot manage or improve what you do not measure. MOSAC welcomes the focus on probabilistic evaluation, in keeping with the strategy to use ensembles at all time scales. It also recognizes the value of demonstrating to the Met Office Board the improvement in model accuracy over time using a KPI suitably time-aggregated to account for interannual variability. Nevertheless, international comparison of scores is useful for all centres so comparative approaches should continue to be used, including air quality and ocean. Although collaboration is at the heart of our science, this healthy competition also keeps us all striving for better (NWP-R1, 2, NWP-Q1 and DS-R4).

Low-hanging fruits to improve forecasts with enhanced post-processing. In any of the world leading centres, condensing the endless numbers of model runs into readily accessible and useful information for a variety of users is a challenge. In general, the Met Office appears to continue to deliver and innovate on that front. MOSAC recognizes that IMPROVER is a real success story, adding value to forecasts and that the close link between usual public forecasts and industry products is beneficial for both groups. We are pleased with blending efforts and look forward to the impact of the inclusion of other data (e.g., ECMWF ensemble). Additional efforts in the post-processing area could give significant improvements to forecasts in the short-medium term. More work on machine learning and consideration of more

reforecasts for training purposes could be a cost-effective investment (WS-R3, PP-R1, 2, 3, 4, 5, 6, DS-R2 and 4).

Toward scale-aware model physics for numerical weather and climate prediction. The new model physics (e.g., convection and cloud) developed at the Met Office and in the PARACON project is an excellent example of innovation addressing key model deficiencies, but the plans for bringing this research in operation are still unclear. Moving forward, the process needs to step up transfer of results to exploit the time remaining within the PARACON funding period (FS-R1, MDP-R1, 2, 3 and 4).

The 100m scale modelling project is very ambitious. It includes fundamental science into model development gained through field campaigns and observations, exploitation of ensembles which is very new at that scale but mandatory in a weather forecasting perspective and learning through testbed applications for a wide range of phenomena. It will probably be very difficult to do it all together and to draw definitive conclusions about the added value of 100-m scale modelling for operational forecasting, let's say, relative to the 300-m scale modelling. Focusing first on the urban scale, taking the opportunities of the Paris Olympic Games WMO Research Demonstration Project in 2024 and the associated international field campaign in 2022, would be a welcome focus. (100SM-R1, 2, 3 and 100SM-Q1).

Decision making needs to be seamless. MOSAC encourages the application of weather-related tools such as Decider to decision making in the climate context to bridge the weather/climate interface, including in the seasonal/decadal timeframes (AS-R2).

Data Assimilation and Observations R&D

MOSAC supports the adoption of the Joint Effort for Data assimilation Integration (JEDI) community infrastructure. While technical training as part of adopting JEDI may constitute a steep learning ramp for some scientists (DS-R3), this investment can also bring new opportunities, such as tighter integration with machine learning frameworks. Many challenges need to be tackled before JEDI delivers its full potential at the Met Office (DA-R1).

More data assimilation early in model development is needed. MOSAC is pleased to see that the need for early testing with data assimilation is recognised in the timelines for model development and looks forward to seeing examples of how it has impact in practice (FS-R2, DA-R2 and 3).

Novel observation techniques: shaking up old habits. The four themes highlighted in the paper on novel observation techniques (enhanced observations for model initialisation, enhanced use of observations for nowcasting, custom hardware development, and novel strategies for measurement campaigns targeting model improvements) are good. There are many challenges in this endeavour, but nothing unreasonable (NOT-R1, 2, 3, 4, 5, and NOT-Q1).

Excellent utilisation of observations but beware of COVID-19 impact. MOSAC is pleased to see again several examples of the contributions of observations in scientific analysis and model evaluation. It is very positive to hear that OBR contributed flights as part of the Clean Air programme to measure air quality during COVID lockdown in the UK – a huge opportunity for research due to lowered pollution emissions. However, COVID has seriously delayed experimental activity for the MO and academic partners, most of which is time-limited due to funding. The knock-on impact of these delays should be considered carefully over the coming years, in terms of science and staff resources. This further encourages a more strategic approach in fostering links between OBR and external partners (e.g., UK Universities, NCAS, and international partners). (FS-R3, 4 and FS-Q1)

Applied Science and Observations R&D into a new Programmes Directorate could bring benefits. Technological change (future mobility, big data, machine learning) provides new opportunities for Met Office science and services in partnership with other players. The new capabilities demonstrated by case studies showed are quite encouraging (AS-R1). ML for observation processing is a promising approach and might be exploited to make greater use of third-party observations in Met Office business and user applications (DA-R1 and 5). Collaboration with universities could be a good and low-cost approach to further develop this capability.

Concluding Remarks

This year the MOSAC meeting was done through video-conference tools with pre-recorded presentations due to the worldwide COVID-19 lockdown. Overall, this was a more intense meeting than usual where the live Teams interactions were focusing on verbal and chat Q&A. Kudos and many thanks to the technical and administrative staff, this meeting was a real success, exhausting but very effective. Finally, MOSAC would like to thank the Met Office leadership, speakers, carousel and poster presenters for their hard and excellent work and willingness to discuss in collegial manner all aspects of their work. Indeed, this was an interesting, fruitful, and well-organized meeting.

Gilbert Brunet (Chair) on behalf of the Met Office Scientific Advisory Committee: Thomas Auligné, Janet Barlow, Natacha Bernier, Andy Brown, George Craig, Véronique Ducrocq, Beth Ebert, John Michalakes, and Ian Renfrew.

Annex: Recommendations and Questions

Foundation Science

The Met Office possesses a unique capability among national weather services for fundamental research. This, in turn, provides a unique opportunity to rapidly turn innovative ideas into operational applications. Numerous examples of this process were presented in the report and at the meeting.

FS-R1: The new convection and cloud microphysics parameterizations, CoMorph and CASIM are excellent examples of the innovation to application process, addressing key model deficiencies, both in specific phenomena like the MJO and in overall forecast scores. At the same time, the new schemes provide a path to unify the model physics across scales from climate prediction to regional NWP. The PARACON project, where CoMorph was developed, has the potential to be a flagship example, but the plans for bringing in the academic work are still vague. Moving forward, the process needs to step up transfer of results to exploit the time remaining within the PARACON funding period.

Please see the response to MPD-R1.

FS-R2: An important concern from MOSAC24 was the need to include data assimilation early in model development work to avoid surprises late in the testing process. We are pleased to see this need clearly recognized in the timelines for Data Assimilation and NGMS. However, so far we have seen relatively few examples of this type of testing having been done in practice (perhaps CoMorph), so there is still more work to be done.

We agree that early testing of model changes with data assimilation is important and have modified our model development process to reflect this. Every global model change which has more than a trivial impact now has a DA trial run before it can leave

the Research Cycle and be submitted for package testing. During the meeting the committee just happened to only be shown the CoMorph results.

We are pleased to see again several examples of the contributions of observations in scientific analysis and model evaluation. The use of Cardington measurements during 2018 drought to evaluate JULES and provide specific guidance for future development of the land surface model is a good example of the Met Office's capability to respond rapidly to specific meteorological events that challenge the model.

FS-R3*:* It is very positive to hear that OBR contributed flights as part of the Clean Air programme to measure air quality during COVID lockdown in the UK – a huge opportunity for research due to lowered pollution emissions. However, COVID has seriously delayed experimental activity for the MO and academic partners, most of which is time-limited due to funding. Certain field campaigns may be under threat or scaled back, e.g., PICASSO and Terra Maris, which would test CASIM, etc, further. The knock-on impact of these delays should be considered carefully over the coming years, in terms of science and staff resources.

We agree. Updating campaign plans in light of COVID developments has been an ever-present theme over the last year and looks set to remain so for the near future. We will continue to work closely with partners to actively review plans and recognise that some degree of prioritisation may be needed to ensure that our limited resources are directed to the most pressing science areas.

FS-Q1*:* An internal OBR Science Advisory Group was set up last year to try and join up activities across Foundation Science and make sure OBR and APP (for example) are pulling in the same direction. MOSAC welcomed this initiative last year. Has this group been able to function well in the past year? Given the COVID impacts, it has a critical role to play in identifying changes in priorities.

The OBR SAG has now met five times since October 2018. These meetings have largely focussed on reviewing future proposals for field campaigns – ten in total and several which have now either been undertaken (e.g. SOFOG3D, EUREC4A), or form part of future plans (e.g. ACAO, LIAISE). To a lesser extent, the group has also reviewed progress with the pull through of results from recently-completed campaigns (e.g. CLARIFY and MACSSIMIZE) and advised on strategy for ongoing long-term OBR research endeavours (e.g. development of capability for snow surface emissivity modelling).

Over the last year, the SAG has primarily focussed on review of three new field project proposals, which now form part of our future plans for 2022-2024 – WESCON, CCREST and Paris BL urban work. The SAG has not been involved significantly in operational decision making related to COVID. In part this is by design, as the SAG is intended primarily to be an advisory rather than management group. It is also the case that we have not yet had to prioritise between Met Office funded campaigns with a view to stopping, rather than delaying, activities. It is very possible that this could change in the coming year, in which case the SAG will play an important role.

FS-R4: Reflecting recommendation MOC-R1 in last year's MOSAC report, we further encourage a more strategic approach in fostering links between OBR and external partners (e.g., UK Universities, NCAS, and international partners), with an emphasis on regular two-way communication beyond dissemination at e.g., MOSAC, Royal Met Society conferences, etc. A further example is the DCMEX and Southern Ocean Clouds projects are part of the NERC CloudSense programme (Uncertainty in Climate Sensitivity due to Clouds) where field work is planned over New Mexico and Southern Ocean.

Though we did not highlight progress in this area directly during the meeting, we have taken some steps over the last year to enhance the way in which we engage partners in the development of OBR research projects, particularly field campaigns.

One example of this is the CCREST project, which aims to deploy the FAAM aircraft for cirrus radiative transfer studies on the 2024 timescale. This project started as a research proposal from Met Office that was considered by the OBR SAG in 2020. OBR then facilitated a workshop to introduce the project to potential collaborators. This was well attended with representatives from both UK academia and several international institutes Following the workshop, a white paper for the project has been written, providing a basis for growing collaboration in this project.

This model for growing collaborative involvement in Met Office-initiated campaigns is also being extended to other projects. For example, we will be holding a similar workshop for the WESCON convection project in April 2021.

In addition to developing collaboration around Met Office-initiated work, OBR continues to collaborate significantly in research activities led by other organisations. Currently this includes involvement in projects such as Terra Maris, MPhase and urban boundary layer observations. The number of possibilities for contributing to collaborative projects of this nature generally outweigh our capacity significantly and as such prioritisation is needed.

FS-R5: Research and development on the production of NWP ensembles has largely paused during NGMS development, but there will be a major need for this work immediately after the rollout. During this window of opportunity, where the expert personnel, tools and computing capacity are available, there should be a plan to address topics including stochastic parameterization, ensemble size vs. resolution, ensemble data assimilation, perturbations environmental modelling components including the ocean, spread-skill relations, and calibration.

We intend to return to the development of stochastic physics and its impact on ensemble performance following completion of the physics porting to NGMS. As part of the planning for the next supercomputer this year we shall also be reviewing the trade-off between ensemble size, resolution, and model complexity.

FS-R6: The changes for the trial coupled forecast system have been rather swift, progressing through GC2 to GC4 in 3-4 years, with GC5 on the horizon. Communication of configurations and evaluations should be a priority. Newsletters or blogs may be helpful to complement the (slower) GC configuration journal articles.

The updates to GC science versions are made against a timetable agreed and communicated with a wide range of partners/users/stakeholders and upgrades to GC science can vary in complexity and don't necessarily involve all parts of the coupled system (GC3 to GC4 had no significant changes to ocean or sea-ice components).

However, given the importance of this activity, we agree that the final GC versions, and more generally the GC Programme as a whole, would benefit from increased visibility and communications (e.g. UM user newsletter, science seminars with MOAP). This is something that the GC Programme Board is actively considering and intends to make progress on during this coming year.

<u>Weather Science</u>

Despite COVID, Weather Science has continued to make progress this year. In terms of operations, it was good to see beneficial use made of COSMIC-2 and SPIRE data, and PS44 gave incremental improvements in performance. Looking ahead, good progress is being made towards coupled NWP, albeit with some residual uncertainties as to exactly what version will be implemented (GC4 or GC3.2). GC4 gives better results, but some stability problems have been identified in aspects of the full NWP testing which are only undertaken relatively late in the development cycle. The links to operational forecasters also continue to be developed through work on R2O and O2R (e.g., ideas for testbeds).

MOSAC has previously advised that more data assimilation testing should be performed earlier in the research cycle and is pleased to see that the Met Office has been active in this direction. However, given some past challenges in transferring GA developments to operations, it suggests that this should continue to be an area of focus. Additionally, as recognized by the Met Office, the new enhanced focus on ensembles and the planned introduction of coupled NWP (both of which are supported by MOSAC) bring additional challenges to the testing process (need for early, affordable indication of NWP performance of coupled ensemble system).

WS-R1: Continue to review end-to end testing process (from research cycle to operations) for NWP developments in the light of lessons learned over last few years and new issues that will be brought to the fore by the move to coupled NWP.

We agree that continual review of the testing process is important. The end-to-end testing process for NWP relies on close coordination between Foundation and Weather Science. Model development work is governed by formal programmes (e.g. the GC programme) on which the Hd R2O sits as the Weather Science senior user and the implementation process relies on input and support from Foundation Science. There is also representation on major implementation projects (e.g. Hd GMED is a senior supplier on the coupled NWP project). The model development and R2O processes all adopt a continual improvement approach, in which we challenge ourselves to develop and deliver more efficiently and robustly and learn lessons from previous issues and successes. This includes reviews of what changes are made as part of, or outside of a parallel suite.

The new HPC should present a huge opportunity for the Met Office to improve NWP performance. At the same time though, there could be challenges in continuing to improve the performance of the operational models while at the same time dealing with the migration to the new supercomputer and with increasing efforts being diverted to the development of the NGMS system. For this reason, it is suggested that the Met Office flesh out further the roadmap for operational NWP for the next few years, identifying the priority upgrades and ensuring that there is enough resource (and process flexibility e.g., relationship between GC and NWP plans? Aeolus outside full parallel suite?) to enable them to be implemented.

WS-R2: Clarify roadmap for upgrades to operational NWP models for next few years.

We agree we must address this. The short-term plan focusses on the implementation of global coupled NWP in PS45, followed by the HPC port to be completed by Autumn 2022. In parallel, we are developing a longer-term roadmap of potential upgrades to both realise the benefits laid out in the supercomputer business case and pull through capability delivered by the NGMS programme. This roadmap will be further developed over the coming year.

We recognise both the need to maintain research compute capacity and the possibility of a potential 1.5x cost increase from the implementation of the NGMS model. This means we will need to continually monitor, coordinate, and adjust these plans to make appropriate use of our HPC resource.

The committee considers that there may be relatively 'low-hanging fruit' in terms of improving accuracy or forecasts to users through continuing and enhancing the progress being made in post-processing (e.g., operational developments reported last year, progress with IMPROVER). These could be particularly valuable if migration efforts result in a relative lull in raw NWP improvements. It suggests further consideration of whether additional efforts in the post-processing area could give significant improvements to forecasts in the short-medium term. For example, this could include deciding to make this area a particular focus of the enhanced efforts on machine learning and consideration of whether production of more reforecasts (allowing improved post-processing) could be a cost-effective investment. While at the end of the spectrum involving creation of a UM-based reanalysis this could be lot of work, it seems plausible that cheaper options might still be useful. For example, a global reforecast can be initialized from a reanalysis such as ECMWF's. A regional reanalysis can then be performed locally over the UK driven with the global reforecast.

WS-R3: Review opportunities to further increase added-value delivered by post-processing (e.g., machine learning and reforecasts).

Post-processing is an important aspect of the Met Office science programme. IMPROVER represents the initial implementation of several years of R&D investment in probabilistic post processing and this is critically important if we are to derive value from increased use of ensemble prediction. In the short term IMPROVER will feed our automated Web and App forecasts. The next steps for this work will be demand-led from the customers, but our solutions will look to advance and improve our methods. This will be done within the context of the Met Office Common Data Platform programme. Machine learning is very much in the mix of methods that will be used. However, we remain cautious about the level of improvement to be had from "quick wins" in this area.

A review of the opportunities to be had from a reforecasting activity does seem timely, given the likely benefits for post processing. We will consider this over the coming year.

<u>Climate Science</u>

Overall, the Hadley Centre seem to be striding forward on many fronts and carrying out excellent work. Ensembles provide a common methodological link across weather and climate timescales. Links between the regional and seasonal groups did not come across as that strong.

CS-R1: We would encourage the development of links between seasonal and regional ensemble groups (e.g., seasonal uses reforecasts routinely, regional uses neighbourhood statistics).

We are discussing and exchanging methodologies across timescales for global and regional predictions in our cross-office Ensembles Working Group. We would not wish to embark upon dynamical downscaling of seasonal forecasts as the jury is out on whether this adds enough additional skill to justify the considerable technical effort. However, we are discussing using the first days/weeks of our seasonal hindcasts for

analysis of ensembles on weather timescales and are also considering creating an explicit ensemble weather reforecast.

Applied Science and Observations

The decision to merge Applied Science and Observations R&D into a new Programmes Directorate could bring benefits. Technological change (future mobility, big data, machine learning) provides new opportunities for Met Office science and services in partnership with other players. The new capabilities demonstrated for three case studies shown (CAVs, environmental public health, radar QC) are quite encouraging.

AS-R1: The Programmes Directorate should seek opportunities to bring access to commercial datasets/opportunistic sensor data as part of mutual relationships.

ML for observation processing (e.g., radar QC) is a promising approach and might be exploited to make greater use of third-party observations in Met Office business (DA, situational awareness, calibration/validation of models & forecasts) and user applications. Collaboration with universities could be a good and low-cost approach to further develop this capability.

We are keen to engage externally to access commercial/opportunistic sensor data and would welcome further advice from the committee in this area. The use of observational data to support sub-km modelling is likely to be an increasingly important challenge for NMHSs in the coming years. We believe we will need to work closely with others to identify, evaluate and deliver new types of observations to enable the benefits of these models to be realised.

We can see a big opportunity for ML for observations processing (especially in terms of exploiting new 3rd party observations) and agree that should give this topic a greater focus. Collaboration with universities (and other external partners) offers great potential here and we will seek opportunities to do more in this space.

In addition to emerging new sources of meteorological observations, working with industry users provides an opportunity to access data related to their operations (as noted in PP-R4) which, in combination with new machine learning techniques, can help us to take the step from hazard to impact. We will continue to take advantage of opportunities to gain access to and develop understanding of these types of data as they arise, however, these can be resource intensive activities so will be balanced with utilising more traditional observations sources and developing our underpinning science. While we are already co-supervising PhD students in these areas, we appreciate the suggestions above and will explore whether further collaboration with universities can help us to achieve this balance.

AS-R2: MOSAC encourages the application of weather-related tools such as Decider to decision making in the climate context (e.g., using UKCP) to bridge the weather/climate interface, including in the seasonal/decadal timeframes.

Thank you for this suggestion. We are actively looking at further options to reapply tools on new timescales and hope to provide examples to MOSAC next year.

Applied Science is a significant growth area, as we have seen over the past years. MOSAC would like to understand better the Met Office strategy regarding what it wants to develop/provide versus what it would prefer the commercial sector to do (consultancies,

routine services), and what might be done in collaboration with the commercial sector to extend the reach of Met Office capabilities.

MOSAC ask a big question here. The Met Office Executive and Board are actively discussing this question. We expect that Met Office will focus on a few sectors where we have scientific and stakeholder knowledge and then find ways to multiply efforts into other sectors, perhaps through partnership, and perhaps also through better delivery of data.

Update on NGMS Programme

MOSAC observes that NGMS is progressing according to plan and finds evidence of sound management methods and practices (P3M/PRINCE2) for a project of the size (50 FTE) of NGMS. Risks, including those associated with timely convergence of development paths towards operations (DA and GC) and the computational performance of LFRic, are being tracked and plans adjusted as need arises.

MOSAC commends the Met Office for its response to MOSAC 24 (NGMS-R1) in establishing specific acceptance criteria for computational performance of LFRic. Measured against these criteria, current computational performance of the LFRic global model in NGMS is improving but remains too expensive for operations by several factors. MOSAC is assured that efforts by Met Office personnel and STFC partners will continue to narrow the gap through, for example, a new Helmholtz operator and hybridization of the solver, and optimization of the physics interface through transposition of memory layouts and eventually through refactorization.

NGMS-R1: MOSAC urges that these computational efforts remain a primary focus and would like another report on progress with respect to acceptance criteria for the next year's review. MOSAC requests the report include computational profiles of performance (dynamics and physics), time series showing the pace of improvement in overall performance and scaling projected forward, and, if warranted, discussion of performance-related contingency measures and their potential impact on NGMS fidelity or coverage.

We agree with, and welcome, this recommendation and will provide MOSAC with such a report for next year's meeting. Recognizing the importance of this aspect, we will include computational assessment of the new atmosphere model as one of three elements in next year's NGMS corporate Key Performance Indicator.

NGMS-R2: As a primarily scientific panel, MOSAC uncovered no shortcomings in HPC or software engineering proficiency, methods, technology, practices, etc.; however, and considering the importance of these to the success of NGMS, we recommend that the Met Office consider instituting jointly with STFC a more formal technical external review process to assess and report on the key aspects of LFRic related to computational performance and software quality. This should be done in a way that minimizes impact to project timelines and with the least possible burden to technical staff. Suggested topics for technical review include an assessment of the now decade-long Psyclone domain-specific language development effort and the DSL's effectiveness mapping high-level scientific code for optimum efficiency on the Met Office's investments in supercomputing hardware up to and beyond HPC2.

We understand that MOSAC view is wide, and therefore limits to the depth that can be explored during the annual meeting. We are though reluctant to instigate a further review panel, particularly noting that we are implementing the strategy and design that was a significant output of the 5-year GungHo project, a joint project with NERC and

STFC. But we do recognize the importance of what MOSAC is recommending. Therefore, we will explore ideas for how we might respond to the recommendation.

Application of ensembles across all timescales for developing services

The committee notes and supports the enhanced focus on ensembles in the Met Office strategy. The paper and presentation nicely highlighted the current state of play, with application of ensembles for NWP, seasonal and decadal projections and, through perturbed parameter ensembles, to climate projections (e.g., UKCP). It was noted that the applications on different timescales have developed to some extent independently, and hence there is quite a range of approaches used (e.g., perturbation methods, use or otherwise of reforecasts). MOSAC supports the stated desire to try to bring together these slightly disparate efforts and regards the cross-application work that clearly went into the production of the paper as a useful first step in this direction. It was less clear of the extent to which there was firm plan beyond that (e.g., how will work in this area be taken forwards, what are priority next steps?), but we imagine that this will become clearer over time.

ENS-R1: Clarify ambitions/ priorities for next steps with cross-cutting ensemble work.

We are developing a plan of activities for ensembles R&D across the office and across timescales following this year's analysis of the state of play for ensembles in weather, seasonal to decadal predictions and climate projections. We would be happy to report on progress at a future meeting.

On NWP timescales, evidence was given of over-confidence in the ensemble (e.g., case studies for tropical cyclones and high impact UK weather). However, to put these case-studies into a broader context, MOSAC very strongly encourages systematic creation and bringing together of more *quantitative* evidence. For example, data was shown demonstrating that near surface temperatures in the UK ensemble were over-confident/under-spread (which is not particularly surprising given experiences in other systems). However, a finding that the global ensemble was significantly over-confident synoptically across multiple cases would be somewhat surprising (and hopefully addressable given sufficient attention). Hence, the MOSAC considers that the situation is very likely more nuanced than a general statement that the NWP ensembles are over-confident (e.g., situation likely different global synoptic to UK near surface; conceivable, but not proven, that extreme cases might be more of an issue). Further understanding of the current situation would seem to be a pre-requisite to improving it.

We note the need for more comprehensive analysis going beyond case studies and are considering the feasibility of a larger set of reforecasts. We also note the suggestion to use the existing seasonal system to carry out initial analyses.

ENS-R2: Quantify across multiple cases the situation regarding reported NWP ensemble over-confidence (spread-error; bias). This should be done separately for global and UK models, and separately for upper-air (synoptic) measures and near-surface variables.

Evaluation of our global and regional ensemble systems, both in operations and in their development, is a priority and we are considering producing a larger set of reforecasts (see below). This includes both the monitoring but also the understanding of the results. A new team is being established in R2O that will have responsibility for monitoring and understanding the performance of our NWP systems including the ensemble.

The case studies presented in the meeting were intended to complement the existing objective evaluation so that we can more strongly link our understanding from objective evaluation with the perception of our end users. It is likely that case studies, or subjective assessment of spread by end users, provide measures of only very selective samples compared to the objective statistical measurement of spread as used in trials and verification. Testbeds may provide an opportunity to compare the two in a statistically meaningful way.

ENS-R3: Given increased focus on ensembles, ensure that assessment of impact of any proposed upgrades on spread-error relationship is routinely carried out as part of testing process (i.e., ensemble performance is not just CRPS).

For NWP timescales, the assessment of spread/error is already an important part of our assessment of trials and operational upgrades. For global NWP, this is built into our evaluation suite software and the same will soon be true for regional NWP, which previously relied on stand-alone verification scripts.

For seasonal forecasting we routinely calculate a suite of both deterministic and probabilistic scores over a 23-year set of reforecasts, this includes reliability diagrams and scores to measure spread/error and hence determine any over or under-confidence. Nevertheless, we note that in almost all cases on these timescales, the probabilistic scores vary systematically with the deterministic scores. We also note that spread-error ratios are unable to detect the unusual under confidence and associated problems with weak signal-to-noise ratios found in our extratropical predictions, so while we calculate them, they have limited use in this context.

MOSAC has also suggested consideration (PP-R2, WS-R3, CS-R1 and ENS-R2) of the value of creating ensemble reforecasts on NWP timescales as a potential way of improving post-processing and ultimately the accuracy of forecast products. Given that lower resolution global forecasts and reforecasts on NWP timescales are already available from the seasonal system, some analyses of these might give some indication of the potential value for NWP post-processing of different sized reforecast sets.

Thank you for this suggestion, it is being seriously considered and we will report back at the next meeting.

In addition to the quantitative work described above, it is important to continue work to ensure that ensemble forecasts are useful and used as part of the operational forecasting process. As part of this, feedback from operations will be crucial, and MOSAC welcomes the efforts on O2R (e.g., testbeds).

<u>Data Sciences – the opportunity, the plan, and a progress</u>

MOSAC praises the creation of the Joint Centre of Excellence in Environmental Intelligence (JCEEI), which constitutes a major milestone for the Met Office. The JCEEI already managed to develop partnerships with some of the leaders in the field. Its ‘bottom-up’ approach via selected pilot projects, is an excellent starting point. These projects are expected to expand, multiply, and inform on standards and lessons learnt for new projects.

DS-R1: This approach should later be complemented with a ‘top-down’ approach to identify critical use cases of Data Sciences across the entire Met Office value chain and develop a

comprehensive strategy. The new cross-cutting initiative should be a good vehicle for this goal.

Building on the Research and Innovation Strategy we plan to develop a strategy for ML that develops the 'Fusing simulations with data sciences' theme and capitalizes on the gains made in the Strategic Action on Data Sciences. This strategy will cover the entire Met Office value chain.

DS-R2: Short-term, IMPROVER is perceived as the low hanging fruit for demonstrating the added value of AI/ML and should therefore be promoted as a flagship. Longer-term the project can expand to other uses such as systematic model error in DA (DA-R5).

We agree that IMPROVER should be an important focus for ML development (see also WS-R3). Further development of the project will depend on budget and be informed by the priorities of sponsors.

DS-R3: We encourage continued effort to develop internal staff skills more broadly via training and strong community. Expanding on scientist-engineer and external partnerships is also desirable.

We agree and welcome MOSAC's encouragement.

DS-R4: It is important to build an ecosystem that includes easy access to large training and validation datasets, an integrated suite of software tools for data analytics, and "data-proximate" computing capabilities. Initial successes of platform agnostic implementations on cloud and on-prem HPC (e.g., Pangeo-as-a-service) should be used to develop a common data platform. This platform will find benefits beyond AI/ML applications, such as the development of refined metrics for model evaluation and diagnostics.

Development of an ecosystem to enable ML (and Data Sciences) to flourish within the Met Office will in part be informed by its new Common Data Platform (a Met Office Strategic Action) and the new HPC provider. We agree that the benefits of a ML considerate ecosystem will stretch beyond Data Sciences.

DS-R5: To solidify plans for data sciences, the Met Office should quickly identify risks and opportunities associated with its new HPC procurement.

While the risks and opportunities associated with the new HPC procurement are well understood, documented, and managed, links to Data Sciences are perhaps less so. We thank MOSAC for this timely recommendation.

<u>Science Communication</u>

The Met Office is in a privileged position in having a high profile in the media, with government and with the public. There has been a lot of good work in communication in recent years and we should acknowledge initiatives like the Parallel Suite infographics; the naming of winter storms (in partnership with Met Éireann and KNMI); enhanced work with the Science Media Centre; the Schools programme; as being terrific and welcome.

MOSAC think there are ongoing questions on communication on climate topics and there is always more than can be done here. MOSAC guess the SRG will focus more on this

element. There was an interesting discussion during the session about the use of the 'authoritative voice' and when and where that is appropriate.

SC-R1: It is important that the Met Office is aware of the concerns around this issue and uses its authority in the right times and places (e.g., during UK hazard warnings) and does not give in to lecturing. This may be something to include in Met Office science engagement training.

In the weather and hazards, there is a project underway with Cabinet Office to clarify roles, responsibilities and having clearly designated Authorities around different warnings. We will also be reviewing how the "Authoritative Voice" term should be used across weather and climate science and services as it may vary in different situations. We would always aim to speak with authority.

There was a question in discussion about the value of engagement and communication activities for early career researchers. There is a useful discussion to be had on this topic, especially as MOSAC think attitudes are changing within science. A greater emphasis is now placed on excellent communication of one's research, and engagement activities are now being valued more explicitly in recruitment and career progression, in a trend that is being driven by social media and the expansion of what academics and researchers are 'supposed to do' as part of their jobs, e.g., the REF-led impact agenda. The Met Office should be conscious of this trend within the broader scientific community.

The skills framework we use to guide our scientists' career development explicitly includes communication and engagement as a core skill. When staff put themselves forward for progression to more senior roles, emphasis is placed on outreach and engagement within their evidence. To support the development of these skills, a wide range of in-house training programmes are provided, alongside opportunities for external training. We encourage our early career staff to participate in outreach competitions (including STEM for Britain and Soapbox Scientists) and we celebrate the successes of those staff who are nominated or who win awards for their outreach and engagement work.

Across Science, there are a number of initiatives, which although not specifically focused on Early Career Scientists, are pitched at a level that allows them to be involved. These include science writing workshops, aimed at both scientific papers and creating writing. There are also opportunities put on by networks such as Women in Climate. For climate scientists interested in communication, the Climate Science directorate runs a group with the Communications Team to provide guidance and support around communication skills to give scientists the confidence to engage in media opportunities.

SC-R2: Developing science literacy, analytical minds, and educating is critical to the Met Office's mission. We commend the MO efforts and engagement with schools, parents, and the general public. This should continue.

We thank MOSAC for their encouragement. There are plans to expand the education content strands as well as the reach of the network of engaged schools using our schools programme resources. During the extensive period of home schooling through 2020 and 2021, many parents have made use of our resources as we saw from increased engagement on our dedicated twitter feed and web pages. We plan to do more, especially in the lead up to COP26.

Communication is clearly taken very seriously at the Met Office and the video clearly showed the many ways in which Met Office weather and climate science is shared with the public. It did not discuss the science of communication.

Under the SPF Climate Resilience project, the Met Office and University of Leeds have researched the 'communications of uncertainty'. This work reviews empirical studies from the cognitive, psychological, and behavioural sciences exploring responses to climate information and its visualisation. It provides best practice evidence-based design principles for communicating climate information and associated uncertainties in more accessible and transparent ways for informing urgently required climate policy decisions. We have plans to make these readily available to Science staff to allow the recommendations to be adopted to improve communications.

In addition we are working on the concept of Storylines, narratives of climate change that are a more qualitative representation of the climate model data. The research introduces a typology of four reasons for using storylines to represent uncertainty in physical aspects of climate change.

SC-Q1: For example, what social science is done and used to support improvements to weather and climate communication?

Corporately, we regularly carry out research which is used to improve our weather communications. For example:

- **"Name our Storms" is based on evidence which supports the benefits of naming storms in terms of increasing reach, influence, and authority of severe weather advice.**
- **Evidence is gathered from users via ad hoc post-event NSWWS surveys and PWS Annual Public Perception Survey (PPS).**
- **Research on Storm Doris (Charlton-Perez, Vukadinovic Greetham and Hemingway, 2019) provides insight into the behavioural and communications science of relevance to warnings.**
- **Research was carried out among responders about awareness and action during the heat events in 2019 & 2020, the results of which inform future campaigns.**
- **We carry out user testing across our digital platforms (Desktop, Mobile and App) to inform all improvements. We also have plans this year to improve perceptions of accuracy and understanding of our forecasts by looking at the way users understand and interpret our weather symbols and percentage chance of precipitations. Results of the research will inform changes which will benefit users across our digital platforms.**

For climate science:

- **We have incorporated climate related questions in the Public Perception Survey, including asking people how likely they think different severe weather elements are to be connected to climate change.**
- **Our Climate Tone of Voice and Positioning work with an external agency (presented at MOSAC) has helped us to ensure the needs of stakeholders are considered.**
- **We identified need from all audiences- especially media to understand the likelihood that climate change is impacting the severity or frequency of a**

weather event. We are working with media stakeholders to agree attribution statements and integrate within communications and products.

- **The Communications Team are also working with a Met Office Socio-Meteorologist to look at social science at the Met Office. We want to understand how we can incorporate positive messaging around climate change and how the public interpret and interact with our climate content e.g. how they can use Future Weather Forecasts to better understand climate projections.**

Data Assimilation

MOSAC supports the adoption of the Joint Effort for Data assimilation Integration (JEDI) community infrastructure, and the development of JOPA and JADA in the JEDI environment. While training in C++ as part of adopting JEDI may constitute a steep learning ramp for some scientists, this investment can also bring new opportunities, such as tighter integration with machine learning frameworks.

DA-R1: JOPA short-term plan on engineering activities to reproduce OPS capabilities is well thought out. MOSAC recommends the elaboration of a longer-term plan for JOPA highlighting the main risks and areas of growth beyond current capabilities, such as the use of innovative observations from Internet of Things (NOT-R2) and the integration of machine learning in the observation processing chain (PP-R3).

The key priority of NG-OPS is to deliver equivalent capability to what we have now within the new JOPA framework. However, we do want to continue to advance our capabilities and ensure any systems we have allow us to effectively adopt new observations and make use of ML techniques. Currently we have no explicit plans to make sure of ML and as we get closer to delivering NG-OPS we will begin to develop our longer-term plans in this area. Planning is likely to be at least a year from now for these longer-term activities given the key priorities we have now.

DA-R2: NGMS project management should continue to recognize the need to include DA early in model development work (FS-R2), especially considering the proposed revised timeline for integration of DA-NGMS into the parallel suite. This will require a plan for model-DA interface that addresses interfacing PSyclone-Atlas data structures for 4DVar.

Discussions are underway between the LFRic and JADA infrastructure teams as to the best way of doing this. We are confident that an effective interface will be available for 4DVar, before the linear and adjoint models are finished.

DA-R3: MOSAC recalls the conversation last year about regional convective-allowing scale DA degrading scores and is happy to see an operational fix in the works. A rapid update cycle global DA, proposed to improve boundary conditions for convective scale DA, could bring other valuable applications for situational awareness outside the UK.

This remains in our plans for a future upgrade to global deterministic NWP, after porting to our next HPC. It is also possible that a more rapidly updating global ensemble could also provide benefit, although more research is required in this area.

DA-R4: Coupled DA is a complex research problem, with a variety of scientific options and significant uncertainties now. For next year's review, MOSAC would like to see a roadmap for the coupled DA work that aligns with coupled modeling activities and plans.

Short-term goals will be to include an ocean ensemble in the coupled ensemble system, and to develop JADA. Medium-term aspirations are to make the 1DVar in JOPA fully coupled (this also addresses DA-R1) and to develop DA for all Earth system components using JEDI. We will monitor other research ideas to improve coupling but will want to establish current developments before diverting effort.

DA-R5: The handling of systematic model error in DA could particularly benefit from the Met Office's goal of 'fusing simulations with data science'. The DA team should develop a connection with Data Sciences and Post-processing groups focusing on AI/ML parameterization of model error covariances for use in DA.

Within JEDI framework there are ideas and opportunities for ML activities.

100m Scale Modelling

The 100m scale modelling R&I Flagship project is one of the four flagship projects of the "Path to high resolution" theme. The paper clearly identifies the key issues at this resolution (e.g., representation of convection, urban surfaces, cost, size of the domain), founded on a strong experience in 100m model simulations.

The project is very ambitious, as it includes fundamental science into model development gained through field campaigns and observations, exploitation of ensembles which is very new at that scale but mandatory in a weather forecasting perspective and learning through testbed applications for a wide range of phenomena. It will probably be very difficult to do it all together and to draw definitive conclusions about the added value of 100-m scale modelling for operational forecasting.

100SM-R1: In view of the broad scope, MOSAC recommend time scales of the project be specified, with few objectives for a short-term flagship and longer-term objectives to be addressed in a second stage.

We agree on the need for clear time scales and detailed activity plans to support the 100m Scale Modelling project and are actively progressing this. We aim to be in position to share these more specific plans and continue further discussion to support alignment with UK and international collaborators within the next 2-3 months. Priorities for the first phase (with focus on demonstrating benefits on HPC1) include developing/testing key parametrizations, understanding the resolution/cost trade-offs for different applications, engaging with planned observation campaigns (see below), delivery of a UK based 100m resolution research system for case study evaluations, and defining a clear upgrade path for the 300m London model.

Focusing first on the urban scale, taking the opportunities of the Paris Olympic Games WMO Research Demonstration Project in 2024 and the associated international field campaign in 2022, would be a welcome focus for the short-term flagship. Coming up with an urban-focused strategy for exploitation of opportunistic sensing, and potential need for novel observation development would be a good outcome, as urban areas are potentially data-rich.

The 2022 observation campaign associated with the Paris RDP and 2023 WesCon Convection field campaigns provide a clear focus for model development and evaluation. We welcome the encouragement to also develop a broader urban-focussed strategy for observations, and commit to working across Science, Applied Science and

Observations R&D to develop this within the 100m Scale Modelling scope, and engage with external groups and expertise where possible (see also response to AS-R1).

Running 100-m simulations in an ensemble mode and over a domain with a sufficient size for spin-up is a huge challenge, in terms of computer time, storage and exploitation. Variable resolution at the boundaries of the domain does not seem to be a durable solution to cope with the too-small inner domain unless a scale-aware turbulence parameterization is developed for the "grey-zone" variable resolution domain. On demand system is a reasonable solution to the problem of cost, if the 100-m ensemble prediction can run in a time consistent with real time operational use.

A key motivation for developing the 100m Scale Modelling flagship project is to focus effort on assessing the appropriate design choices, model costs and performance benefits of systems run across a range of sub-km model resolutions and targeting a range of applications. It is important to note that the spin-up issue depends a great deal on the particular application – being less serious for more locally forced situations (e.g. stable BL fog). In the short term at least, we envisage variable resolution approaches to be a cost-effective approach to reducing the influence of lateral boundaries, and we note that development of turbulence parameterization targeting grey zone issues will continue in collaboration with UK academia under the 100m Scale project.

As noted under Operational Perspectives, the recent Winter Resilience testbed activity provided motivation to run the 300m resolution ensemble on demand for several cases focussed on pre-agreed local area domains. This was successfully run exploiting part of the UK allocation of ECMWF HPC resources.

100SM-R2: MOSAC strongly encourages exploration of alternative methods to take full advantage of next-generation computing architectures, such as machine learning based methods to enlarge the size of the 100-m ensemble or to reduce the cost of physical parameterizations. Research on this topic would also be beneficial to coarser resolution ensemble NWP.

The opportunities and challenges presented by machine learning will be coordinated under the Data Science theme (e.g. see Data Sciences discussion, and PP-R1), but we agree that this application could be an interesting use case. This is explicitly included within the scope of the 100m Scale Modelling project, and we have already been engaged in discussions with UM Partners with a view to future proposals in this area.

100SM-Q1: On the shorter time, are there any plans to increase the domain size and/or vertical resolution of the London model at 300m resolution, given the observed problems for convection and for fog (e.g., LANFEX)?

Establishing a clear upgrade path for the currently 300m resolution London model is an identified early priority within the 100m Scale Modelling project, with both domain size and vertical resolution under consideration. Our recent experience assessing 300m ensemble performance linked to the SOFOG field campaign in central France provides additional evidence that supports this activity.

Measuring the added value of very-high resolution forecast is challenging, given the limited observational basis on those scales and the limited number of cases of any given weather type that will be simulated. It would be probably difficult to identify clear benefits from 100-m simulation results over those less costly, at 200-m or 300-m resolution or even at coarser resolutions. The involvement of O2R and testbeds at early stage is potentially very valuable to help identify the possible added value of such high-resolution system for operational weather forecast and stakeholders.

100SM-R3: MOSAC recommend establishing a systematic evaluation plan for testbeds, including hypotheses regarding expected benefits and quantitative measures to demonstrate that benefits are realized, focusing as proposed on high-impact weather we may expect benefit of a higher resolution (fog, urban heat island during heat waves, etc.).

Measuring the added value of different resolution, complexity and ensemble design choices is a fundamental evaluation challenge for all model development, and a key motivation for the 100m Scale Modelling project. We will ensure this work benefits from related 'Path to High Resolution' projects looking across scales with similar evaluation challenges to assess the value of increased resolution (as well as related computational challenges). Testbeds provide one opportunity to focus evaluation research, and we have made early progress this winter working across RMED, R2O and O2R (see OP-R1). We agree that systematic evaluation plans are required for testbeds, but also across the breadth of 100m Scale Modelling research in order to build a more comprehensive evidence base on the cost/benefits of sub-km systems.

Novel Observation Techniques

The four themes highlighted in the paper (enhanced observations for model initialisation, enhanced use of observations for nowcasting, custom hardware development, and novel strategies for measurement campaigns targeting model improvements) are good. Our response to other questions is:

i) Expanding requirements for observations

NOT-R1: 3rd party observations are currently used but there is potential for much greater use, particularly of observations collected by local governments and industries. Most of the measurement campaigns involved aircraft and cloud processes but considering the 100m modelling it's also very important to focus effort on urban street-level and boundary layer observations.

We agree with this recommendation and are keen to explore ways of doing more in this area through collaboration. From an Obs R&D perspective it's important to note that the vast majority of our resources are dedicated to routine lifecycle support for our current observational networks. We are looking hard at opportunities to increase our level of support for 3rd party/novel observations but, working within our anticipated resources, we are expecting to be reliant on establishing high quality collaborations to build a pipeline of prospective observations (currently at low TRL) to enable observational support for benefits realisation for sub-km modelling.

NOT-R2: MOSAC paper 25.13 (Next Generation Post Processing) discussed observation networks based on the "Internet of Things." Are there active or potential connections there – or other separate Internet of Things activities that are part of MO's Novel Observation Techniques efforts?

We agree there are potential future connections. We are working to put in place the infrastructure (communications, hardware, and processing) to enable IoT approaches in surface observations. We hope to work more closely with PP teams (and others) and get benefits from different connections that people have to widen our chances of finding new opportunities in this field to work together.

NOT-R3: Drones and UAVs are going to be a potentially huge area due to commercial interest in drone delivery, amongst other areas. In the future, is it conceivable that a similar system to Mode-S or CAV could exist, with data sharing from on-board sensors? Given existing good

engagement in the use of UAVs for atmospheric research (response to MOC-R2 in last year's report noted), this could be an opportunity.

We completely agree with this suggestion and are actively pursuing opportunities from UAVs/Drones. We hope to leverage strong relationships between Applied Science and Observations R&D to identify and develop future opportunities. In particular, our Aviation Science, Defence Science and Industry Consultancy teams have links with users that are likely to widen data access for the benefit of the wider met community

ii) Collaboration on observations within the UK Testbeds

NOT-R4: For optimal use in forecasting testbeds, field campaign observations would need to be analysed and available in near real time. This might be difficult when universities (i.e., PhD students with limited experience) are involved so would require significant Met Office technical preparation. Related to Weather Science, are there plans to progress the use and improvement of impact observations (e.g., damage, disruption, etc.)? This would be relevant to testbeds and verifying warnings derived from models, and partnership would be essential.

We agree that in order to make use of university provided observations in near real time the Met Office might need to provide assistance, although this would be less of an issue when e.g. collaborating with NERC Centres and facilities. We do see impact observations as useful, in particular in the development of impact-based forecasts, but the collection of such observations would be an area for collaboration rather than a focus area for the Met Office.

iii) Stimulate academic interest in low technology readiness level (TRL) observations:

NOT-R5: MOSAC feels it is a positive move for the Met Office to want greater engagement with the academic community around development of new observation techniques and instrumentation. Instrumentation development needs a potentially lower level of investment but over a longer timescale than the average research project. The MO should explore with UKRI (the umbrella UK research funding organisation, covering research and innovation) funding mechanisms to assist this. There is expertise beyond the usual NERC-funded communities (which include NCAS and other Centres) – for instance, EPSRC (engineering and physical science) and STFC (science and technical facilities).

We thank MOSAC for the suggestions and the encouragement to explore this.

NOT-Q1 A focus on UK collaboration on low TRL observations initially is practical, but obviously international engagement is needed to better leverage respective limited resources. Can the WMO provide a useful forum (e.g., Instruments and Methods of Observation Programme (IMOP), which fosters development of new instruments)?

Thank you for this suggestion. We entirely agree that exploring international opportunities is important and that WMO and EUMETNET could provide useful forums. We are also keen to explore whether the WMO Paris experiment might enable us to establish/improve links with other NMHSs who are working in this area.

<u>Next Generation Post Processing</u>

In any of the world leading centres, condensing the endless numbers of model runs into readily accessible and useful information for a variety of users is a challenge. In general, the UKMO appears to continue to deliver and innovate on that front. MOSAC recognizes that IMPROVER is a real success story, adding value to forecasts and that the close link between usual (PWS)

forecasts and industry products is beneficial for both groups. We are pleased with blending efforts and look forward to the impact of the inclusion of other data (e.g., ECMWF ensemble).

PP-R1: MOSAC recommends continuing to invest to support the expansion of machine learning and reap rapid gains.

We will continue to carry out research involving the role of machine learning in post processing, collaborating where appropriate. While significant gains may ultimately come from methods such as Machine Learning (ML), it is important to recognise that the work is not without challenge, so these may not necessarily be rapid! Some problems are of course more amenable to ML than others. Sustained improvement in the NWP model itself, remains critically important.

PP-R2: MOSAC recommends that reforecasts, especially for ensembles, be used to rapidly increase the datasets available for development and evaluation. Similarly, very high-resolution models could be used to generate pseudo-observation training sets.

We thank MOSAC for this recommendation and are now planning to hold a workshop to review whether Met Office should routinely carry out reforecasts, considering the range of benefits and computer cost. Outcomes of this will available before the next MOSAC.

PP-R3: With a rapidly growing number of non-traditional observations coming online, ML of this additional data can support post-processing but can also be used to begin exploring quality control of these new observations in preparation for their future consideration for high resolution DA.

In many cases DA is the most effective means of automated quality control of observations, but other ML methods may also be effective where there are ways to train it to recognise poor quality observations. These types of data may indeed be useful in post-processing but the first and most challenging requirement of any ML technique is quality control of the data used. For PP we ideally need a really good high-resolution analysis to use for calibration or verification, which can also be used to initiate nowcasts, and work is planned in the new Nowcasting team to develop these analyses.

PP-R4: Some variables are particularly difficult to observe (e.g., snow, visibility). As a result, reliable data sets for ML or evaluation of the accurateness and usefulness of the forecast are not available. We encourage the exploration, in partnership with authorities and industry non-conventional metrics such as reduction of accidents following alerts.

Availability of sufficient quality-controlled observational data is indeed critical for ML techniques to avoid over-fitting or producing dangerously misleading results. Industry observations may be valuable in supplementing conventional Obs, especially in bespoke industry services. "Observations" of the type suggested carry many additional challenges. They are useful for evaluation of impact-based services, particularly for assessing forecast value rather than just accuracy, but are affected by many confounding complexities. For the specific example given this would include human decision-making (response to alerts) and the lack of a control observation (how many accidents would have occurred in the absence of an alert). We will continue to take advantage of opportunities to gain access to and develop understanding of these types of data as they arise.

PR-Q1: Two calibration approaches are used – Ensemble MOS for temperature and wind and Reliability Calibration (RC) for cloud and precipitation. Has EMOS been shown to outperform RC for temperature and wind speed? Could RC be used to calibrate all variables, given its advantages of being non-parametric and of reduced computational cost?

EMOS is parametric, whereas RC is not. So far, we have applied the different techniques for the variables we think they are best suited to. Temperature errors are approximately Gaussian so well-suited to a parametric approach; precipitation and cloud are highly non-Gaussian in different ways so better handled by RC. We have built an infrastructure that allows different methods to be introduced and compared. We do want to compare EMOS with RC to see if only one is needed. These methods can be used as a baseline from which to compare further enhancements to either or new methods such as ML approaches

PR-R5: MOSAC encourages UKMO to pursue the creation of calibrated ensemble members (e.g., using ensemble copula coupling or some other method(s)) and to continue to seek improved input data of use to both public and industry post-processing, thus continuing to bring the two strands of post-processing closer together.

There are various demands and requirements in terms of processing our model output to be of optimal value for a wide range of users. We must consider the demand and consider science and technical aspects in any solutions we take forward. However, from a scientific perspective we do recognise the potential value of calibrated ensemble members and that ensemble copula coupling may be a good method to generate these.

Coupled Modelling

Environmental modelling has a long history at the UKMO. The recent years' acceleration toward coupled systems is made possible by this long and strong base in numerous fields and emerging experts that operate at the interface of these newly coupled systems.

CM-R1: The UKMO raised concerns regarding the increased complexity brought upon by the growing number of coupled systems and related testing. The MOSAC pointed out that other centres are facing similar challenges. The identification of tests that can be done using varying resolution can offer means of reducing the costs of testing. MOSAC notes that over the coming years, it could be interesting to share such knowledge as it is developed.

We thank the committee for this suggestion. Once the global coupled NWP system is operational (early 2022), we plan to use the pre-operational trial results to inform a new investigation into the sensitivity of our early testing framework to model-DA interactions. Following's the committee's recommendation, we will consider expanding that investigation to cover the resolution of test configurations. We will certainly look to share experience with other coupled model prediction centres and made a start on this with a joint Met Office-ECMWF coupled NWP workshop in early March 2021.

CM-R2: As the complexity of systems continues to increase, improvements may come with degradations in other parts of the system. This creates challenges for verification. It is not clear how the final judgement on what is or is not acceptable is taken. MOSAC cautions that establishing some clear limits on what acceptable degradations could be or similarly what is an improvement that warrants degradations could help simplify the acceptance process.

We agree that this is a challenge, and so through the GC Programme we are working to establish more complete acceptance criteria with all stakeholders. We note that while

the move towards more coupled NWP adds challenges to establishing the balance of performance changes across different parts of the system, there are also opportunities to take a more whole-system approach in these assessments, with potentially more visibility of the impact of changes in one component on others than across current systems.

CM-Q1: It was mentioned that the sea-ice systems are coupled to the atmosphere through the land surface model for flux exchange. Does this provide an advantage for snow modelling on ice?

Potentially it could do in future but would require a merge of the snow schemes in the sea-ice and land models. The sea-ice model snow scheme includes marine-specific processes such as flooding of snow by sea water. At present the role of the land surface model (JULES) is limited to surface exchange.

CM.R3: The UKMO is putting a lot of resources toward ensemble modelling. The role and emphasis on the development of perturbations for environmental systems is less clear. Additional information would be welcomed.

Research in this area was not strongly highlighted to MOSAC this year, but there has been considerable activity beyond atmosphere-focussed ensemble development. For example, under the WCSSP India project, and linking to UK academia, we are developing a new global ocean ensemble capability in which ocean spread is generated both through different atmospheric forcing and perturbations to ocean observations. This has been successfully implemented within a hybrid En3DVAR capability using NEMOVAR. While the first operational implementation of coupled NWP will use the same SST perturbation schemes used in MOGREPS-G, the NEMO ensemble has been shown to generate more representative ocean spread and we would plan to implement this in future coupled model upgrades.

Activity to explore similar methods to develop a regional ocean ensemble is under way, linked to the Regional Environmental Prediction (REP) flagship project. The Coupled Modelling paper to MOSAC also highlighted initial developments to establish a regional coupled ensemble capability under REP, as the basis for implementing a research-mode multi-hazard demonstrator on the new HPC.

The Met Office has also run operational storm surge and wave model ensemble systems, in which perturbations are derived through the spread in forcing developed from MOGREPS-G, for some time. Practical use of these systems for coastal flood forecasting has illustrated the need to ensure relationships are maintained between ensemble members in each model due to the co-dependence of the surge and wave conditions on both weather and tide. This need would be more simply met in a coupled framework, for which ongoing developments of WAVEWATCH III and NEMO for coupling will provide the enabling technology.

CM-Q2: There are numerous coupled systems at all resolutions and spanning several model versions. MOSAC is concerned about the strain on resources of this version hopping. Does UKMO has plan to rationalise this situation?

Supporting the wide range of resolutions is a challenge, particularly for the ocean, and we regularly review whether we have adequate resource to support all resolutions. On the other hand, the proliferation of versions arises from the need for 'branch' configurations (e.g. GC3.1 for climate, GC3.2 for NWP & seasonal) to allow operational systems to make minor science changes according to their specific needs. One goal of

the GC Programme is to make 'trunk' configurations (GC3.0, GC4.0) perform well for multiple applications, in order to minimise the need for branch configurations.

GC4 has been a success in this regard, outperforming each of the branch configurations in their respective application metrics. The frequency of new trunk configurations has been approximately 2 years in the past, and we expect to maintain this frequency going forward (though with a slightly longer wait for the first NGMS configuration GC6). We have found that this frequency represents a reasonable balance between development resources and the timescale for pull-through of new physics.

Operational Perspective – research to operation and operation to research (Winter resilience)

MOSAC welcomes the efforts on O2R to involve operational meteorologists through testbeds at early stage of the development process. This could be potentially very valuable and would improve the R2O/O2R process.

The committee will be looking forward to seeing examples of how this process has led to improvements in the system and lessons learned from this first testbed this winter at the next MOSAC. This will help to define a clear positioning of the testbeds with respect to the objective and subjective assessments performed within the R2O and how they should interfere with the whole NWP strategy.

OP-R1: Define clear scientific and operational forecasting objectives at the beginning of the testbed.

The main purpose of the proto-winter testbed is to learn about how to run testbeds. Its main output will be a report on lessons learned but we also hope for some outcomes that advance our understanding of our modelling systems and how they are used. As we move to testbeds for which learning about our systems, processes and tools are the key output then we absolutely must define clear objectives across a range of areas that include model performance, operational methods, tools etc.

Regarding the 300m ensemble testbed, this should lead to profitable interactions for a wider use of ensemble forecasts. It is not clear however how other environmental systems could benefit from this ensemble prediction, as well as how stakeholders could be involved without products tailored to their needs.

OP-R2: Focus the evaluations on the added value for a few high impact weather events and/or a few stakeholders to help identify when such systems should be considered.

We agree with this suggestion.

Model Physics Developments

The highlights in Model Physics Development this year certainly include the performance of the new convection and cloud schemes CoMorph and CASIM in a wide range of testing. An important aspect of CoMorph is that it serves as a framework for building in many future improvements (cold pools, stochastic perturbations, etc.). CASIM is also developing well and shows potential to unify the different microphysical schemes used on different scales. The improvements in computational efficiency are particularly notable, since this is often the factor blocking the introduction of more sophisticated microphysical parameterizations in operations. We also welcome the range of activities exploring the potential of machine learning in parameterization development.

MPD-R1: Develop a plan for incorporating and evaluating PARACON results from academic partners - this is potentially a flagship project for close MetO-academic collaboration, but the proof will be in the implementation of new ideas in the MetO models.

For MOSAC25 we focussed on the operational development of CoMorph, as significant advances were made this year. We propose that at MOSAC26 we will present the work of the wider ParaCon programme, highlighting the academic contributions to CoMorph from process research, progress and insights from the novel parametrisation approaches, and the plans for pull-through and legacy over the longer term.

MPD-R2: ML development should focus on low-hanging fruit (the first two levels of the hierarchy that was presented), building expertise while working on projects that are likely to have an impact.

We agree and this is very much in our plans for the next 2-3 years.

MPD-R3: Following up on last year's report, we are concerned about how well the three significant new parameterizations schemes (CoMorph, CASIM and large-scale cloud) will work together for the model and for a variety of weather system events. The TestBed approach could play an important part in the evaluation.

We feel the challenge here is less about these major developments ultimately working together, and more to do with current differences between global and convective-scale model science (and between tropical and mid-latitude convective-scale configurations), along with NGMS timeline pressures preventing a single large implementation of all three new schemes at the same time. Consequently, the development path needs to deliver acceptable performance of each scheme in the different current control systems whilst also targeting the long-term goal of all the new schemes ultimately working together to unify the different configurations.

As always, a wide variety of testing will be undertaken in each system, but we agree that a testbed approach could form a valuable part of this.

MPD-R4: The pipeline for current Atmospheric Processes and Parameterizations (APP) team developments to become operational requires meeting several cut-off dates for different model configurations, and as a result can be very slow, with years passing between initial testing and acceptance in a parallel suite. Is the Met Office content with the timelines or can something be done to speed things up?

Whilst the timeline can appear slow, a large degree of testing is required in order for model changes to be properly evaluated (e.g. in last few years the addition of model-DA tests and idealised climate sensitivity tests to the research cycle). For, say, an atmosphere model change which currently passes the standard tests to show its individual performance is acceptable, a considerable amount of further testing and potentially tuning is required when it is packaged together with other model physics changes. These will include tests in the coupled NWP system, higher resolution DA trials and, following feedback from MOSAC (which we agree with), ensemble tests. Experience has shown that this process takes a good year (and longer if issues arise). The new configuration is then subject to wide-ranging evaluation including metrics and diagnostics not in standard tools and subjective forecaster assessment of the results. In the case of GC5, completion of this step would put us in the middle of the port to the new HPC, hence the need for a later parallel suite. Lessons learned from each time we go through this cycle tend to point to the need to do more, rather than less, testing. Whilst we do always look to streamline the process, this approach does ensure our

testing is robust and that we have high confidence in the configurations delivered to operations.

NWP Evaluation and Accuracy

MOSAC welcomes the focus on probabilistic evaluation, in keeping with the strategy to use ensembles at all time scales. It also recognizes the value of demonstrating to the Met Office Board the improvement in model accuracy over time using a KPI suitably time-aggregated to account for interannual variability. Nevertheless, international comparison of scores is useful for all centres so comparative approaches should continue to be used. Although collaboration is at the heart of our science, this healthy competition also keeps us all striving for better.

NWP-R1: MOSAC recommends that the MO not only continues to compare its ensemble against other sharing centres but also seeks to encourage and include ensemble producing centres that are not yet contributing their data.

We certainly plan to continue to monitor the performance of our model in a wide range of ways including the current approach that compares against other centres. This wide-ranging monitoring is important because it helps us to prioritise and focus on continuous improvement. The paper presented to MOSAC described our new KPI measures specifically, responding to concerns raised by the Met Office Board over the importance of demonstrating improvements over time and that it, in order to be consistent with our strategy, the KPI measure should focus on ensemble capability.

NWP-R2: Regarding new components for ocean and air quality, candidate metrics should be measured and archived so that, once agreed, the time series underpinning the KPIs can be as long as possible and intercompared with other international producers.

We agree with this recommendation.

MOSAC agrees that measuring the fraction of component scores that improve over a period of years, from amongst an agreed basket of components, is more informative than the composite NWP score, assuming the components are shown using a red-amber-green matrix or similar. It is also more robust as one outlier variable can have a strong impact on a composite score. Most people would want to see the time series as well, as suggested in the paper.

NWP-Q1: The NWP KPIs for global (using analysis truth) and convective-scale (HiRA neighbourhood methods using site truth) models are different, reflecting the different purposes of the global and convective-scale models. Given that forecasters use both MOGREPS-G and MOGREPS-UK in the medium range (e.g., days 4-5) do you present them with any routine comparisons of global and convective-scale model performance using HiRA?

We do provide comparison of deterministic UKV and Global models over the UK using HiRA, but not currently the ensembles. HiRA was designed as a tool for verifying convective-scale models, with use of neighbourhoods not nearly so necessary or relevant on the global resolution models. This was discussed in the original HiRA reports, but we have not invested further effort in applying to global models. In the future, as we move towards convective-scale global modelling, this should be revisited.