Report of the 27th Met Office Scientific Advisory Committee Meeting (16-20th January 2023)

Response from the Met Office Chief Scientist in red

The 27th Met Office Scientific Advisory Committee (MOSAC) hybrid virtual meeting was held 16-20th January 2023. The MOSAC has conducted a comprehensive review of some of the Met Office's scientific research programmes contributing to innovations in weather, climate, and oceanographic services. The goal of the review was to assess the quality and relevance of the R&D and its ability to support the Met Office's services and advice to sponsors.

MOSAC's review found that the Met Office's research programs are of a high quality and are well-aligned with the needs of its sponsors and the global meteorological and climate communities. The review also identified areas for improvement and made recommendations for enhancing the scientific basis of the Met Office's services and advice. In the following we provide the highlights of these recommendations, and to find more background in the detailed report (see the Annex) we provide their references explicitly (e.g., FC-R1).

MOSAC was pleased with the increase in the number of female speakers.

However, there is still room for improvement in terms of gender balance, both among the membership of committees and in terms of the number of women among the Met Office fellows (CSTO-R1). MOSAC would like more information next year on the Met Office's diversity and inclusion plan, including a demographic analysis (CSTO-R2). MOSAC is intrigued by the h-index in Figure 3 of Annex II of MOSAC supporting documents and would like to understand why publications seem to be less impactful as the number of publications has increased (CSTO-Q1).

Our corporate and Science & Technology Directorate Equality, Diversity and Inclusion action plans have a range of activities to improve diversity, including: i) targeting greater diversity of all types in senior roles ii) ensuring we consider diversity where we have discretion in assigning people to internal and external committees and iii) working with our staff to support those less likely to put themselves forward for such opportunities. The Met Office publish an annual EDI report that is available from our public website. We recommend MOSAC read this and if they still have specific question, our EDI domain experts that drive our activities can respond. The Met Office Board reviews our EDI progress and we also have external reviews of EDI through professional review bodies – currently we work with, and are accredited by, Investors in Diversity.

We apologize if the plots have caused confusion: the flattening of h-index is an artifact of what is being plotted. The historical time series of h-index for the Met Office is presented in figure 1. The Met Office h-index is increasing, providing evidence that our publications continue to have very high impact.

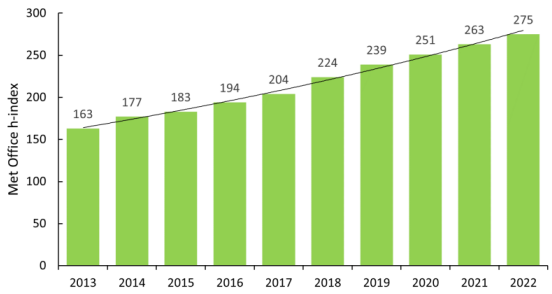


Figure 1. Change in Met Office h-index, a measure of impact of publications from 2013 - 2022

MOSAC is happy with the overall progress made by Weather and Foundation Sciences.

The focus on ensembles in Weather and Foundation Sciences and cross-office efforts to better exploit them were welcomed. The merging of IT teams and their relocation in Foundation Science also appears to be working well. MOSAC welcomes the appointment of Matt Hort, Richard Lawrence, Cath Senior and Jon Taylor as Principal Fellows. MOSAC would like to learn more about the future focus of the climate program and has some questions about the plans for PRISM. MOSAC would like to have next year a presentation on observational activities and gaps/priorities (FC-R3). MOSAC would also like to see more information on space weather products and the primary clients (SP-R1). Good progress has been made with the next generation processing and assimilation of observations (NG-PAO) and the JEDI work, although there is still work to be done in terms of interfacing with next generation model LFRic (SP-R2).

We thank MOSAC for these very positive remarks and look forward to presenting on the topics at next year's meeting. The point about NGMS and JEDI is addressed in detail under SP-R2.

The Met Office is commended for its efforts in addressing recommendations arising from last year's special-topic workshops on LFRic performance.

The efforts include increasing software engineering resources, appointing a new fellow to lead the NG-OPT project, and improving test coverage. MOSAC would like the Met Office to provide a report on its plans and timelines for addressing the panel's recommendations and highlights the importance of addressing all aspects of NGMS performance (NGMS-R1). The focus of the NG-OPT project should be coordinated with existing efforts for algorithmic/numerical optimization (NGMS-R2). MOSAC suggests that the Met Office consider open sourcing more generic code to benefit from community contributions (NGMS-R3). The appointment of a fellow to head the NG-INT project is welcomed and MOSAC is looking forward to hearing about the project's integration successes and challenges (NGMS-R4).

We thank MOSAC for their time and advice on the special-topic workshop. We have shared with the committee the recommendations from the May 2022 workshop and our responses. A further workshop has been arranged for mid-June 2023. We completely agree with the point on coordinating code optimisation with algorithm optimisation, and one of the work packages of the NG-OPT addresses this point. Many components of NGMS have been developed under the BSD-3 open licence, which has been welcomed by the communities using these models and we hope that it will encourage greater uptake of, and contributions to, these codes. We note that there are other components of the system developed by other organisations, and outside our control.

Observation-based Research

MOSAC congratulates OBR on their achievements despite the challenges posed by the pandemic. MOSAC appreciates the observation-based research done to support process understanding and model improvement and encourages prioritizing activities for maximum impact. MOSAC recommends early engagement with NERC-funded scientists and models in the upcoming WesCon campaign and maximizing its potential to evaluate models and parameterizations (FC-R2). OBR should consider its role in rallying the academic community for a potential tropical observational campaign (K-R4).

We are grateful to MOSAC for their positive comments regarding OBR. Our current priority is to complete implementation of global model enhancements that we know will improve tropical performance. Should we identify the need for further observations in the tropics, we would seek to work with partners to explore how these could be facilitated. We agree that any such activity

would be a major undertaking, requiring involvement across the UK community and with international partners.

The computational power conundrum

MOSAC expresses appreciation to the Met Office for seeking its input into their pioneering research programme. The questions raised in the paper are considered interesting and raise additional questions. For example, the importance of finding the balance between complexity, data, computational power, and ensemble size/resolution is an important question (PR-Q1).

We thank MOSAC for these insights. We recall that the pre-MOSAC workshop in 2022 discussed the issue of the balance of computational resources. It was very clear from that workshop that there is no simple single answer to this balance. Further discussion on this important question is provided in the detailed response.

Data Science: not a quiet revolution

MOSAC is excited about the potential for machine learning to improve forecasting performance in operations but would like to understand better the motivation behind using it (DS-R1). MOSAC is concerned about the potential loss of understanding of the underlying mechanisms (DS-R1) and wants to see more consideration of a split between science and operations numerical prediction systems (DS-R2). MOSAC is interested in seeing the Met Office's use of data science techniques in analysing vast amounts of weather data (DS-R3) and in future developments using CLIMAR (DS-R4). MOSAC is concerned about the lack of plans for re-forecast and reanalyses and what information will be used to train emulators (DS-R5). MOSAC commends the Met Office's machine learning plan.

We thank MOSAC for the rich discussion on data science at the meeting and in the report. We shall keep asking the question why deploy data science – as the response will be different in application of data science in different parts of the value chain. We agree that the use of machine learning (ML) should not be at the expense of physical understanding. For this reason, uncertainty and trust are core to our Data Science Framework. We shall keep MOSAC abreast of developments. A report on our review of re-forecasts was presented to MOSAC 26.14, and due to current priorities of NGMS and porting the new supercomputer, we will not develop reanalyses or reforecasts in the short term. Nevertheless, as our plans for data science become more ambitious, we shall keep this decision under constant review.

K-Scale to Urban Scale: toward where people live

MOSAC was impressed with the progress made in using a fleet of global, big domain LAMs and tropical channels runs. The primary challenge is to show the better representation of the smaller scales leading to a better evolution of the larger scales and finding the best metrics and diagnostics to use. The trailblazer was considered potentially useful, but MOSAC recommended further thinking on its purpose and where it lies on the spectrum between the experimental and operational ends (K-R1). MOSAC expressed concerns about the lack of sufficient parallelism to make global km-scale fast enough for operational NWP (K-R2). The full benefit of high-resolution global NWP will require advances for K-scale data assimilation (K-R3). MOSAC also wondered if the Met Office could make further efforts to galvanize the academic community around another field campaign in the tropics, such as Maritime Continent and/or Africa (K-R4).

We share these concerns about the trailblazer systems, and we shall establish and agree principles for the use of the data. We also agree that global k-scale is not achievable through traditional approaches and are investigating a range of innovations within NGMS and through data sciences. Similarly, data assimilation for k-scale indeed requires innovations and there are many ideas in the community. As a first step we are considering using ensemble Kalman filters.

Finally, we are currently finalising improvements to the global model to improve tropical performance, and once this is completed we may look to initiate a field campaign.

Ensemble First

MOSAC welcomes the new initiative on exploiting ensembles and noted the ambitious five work packages involved in the project (EE-R1). MOSAC recognizes that changing the culture from a deterministic to a probabilistic perspective will be a challenging aspect and suggests working with a professional change manager to achieve successful adoption of probabilistic thinking (EE-R2). MOSAC recognizes the use-case decision tree as a useful tool and emphasizes the importance of appropriate evaluation and verification approaches for each of the use cases (EE-R3). MOSAC suggests that efforts be made to consolidate, where possible, independently developed ensemble applications to reduce maintenance overhead and offer opportunities for re-use (EE-R4). Finally, MOSAC encourages the Met Office to make greater use of multi-model ensembles and to include their use in IMPROVER (EE-R5).

We appreciate the suggestion on the culture change required to embed ensembles and the strategic action on exploiting ensembles has just appointed a dedicated change manager. We thank MOSAC for the remarks on evaluation and verification and will explore a range of options. Similarly, we are grateful for the points made around consolidating applications. We are grateful for the points made around multi-model ensembles: we intend to include ECMWF data in IMPROVER post processing in the near future.

MOSAC has welcomed the new methodology to measure the NWP ensemble performance, but there is more to do for assessing the added value of post-processing (NV-R1). MOSAC recommends that the Met Office and to continue efforts to quantify the impact of observation uncertainty (NV-R2). MOSAC has also suggested that the Met Office revisit the KPI to monitor the value of high-resolution modeling (NV-R3) and to verify ensembles across all lead times (NV-R4). MOSAC emphasizes the importance of using appropriate metrics to measure the accuracy and utility of ensemble forecasts when used as storylines, which may require new metric development (NR-5).

Thank you for the comments around demonstrating the value add of IMPROVER over raw NWP, we are developing plans for how to do that. On observation uncertainty, the new verification system being developed in NGMS will have this capability. We appreciate the discussion on developing storylines from the ensembles and agree that measuring the accuracy and values of this approach deserves consideration.

Physics and model development

MOSAC congratulates the Met Office on the positive developments in the past year. However, MOSAC has concerns about the agility of model development and suggests developing criteria for including changes between cycles (FC-R4). Regarding the Unified Physics strategy, MOSAC recognizes the motivation behind it but raises questions about how the interaction between schemes for different model resolutions will be managed (PD-R2). MOSAC also emphasizes the importance of conforming with LFRic model code structures and interfaces and encourages close coordination between physics development and LFRic software engineering teams (PD-R3). MOSAC would like to see further testing to ensure the science performance remains stable and robust after coupling UM physics into LFRic (PD-R4).

We too have been concerned about the agility of model development: as the operational production systems become more complex, and the applications more numerous upgrades

become more difficult to manage. The operational upgrade process already includes the ability to make changes outside of a parallel suite and examples are given in the detailed response below. We are pleased that MOSAC recognises Met Office strength in physical process research.

The need for R2O testing and refinement of parameterizations was emphasized (FC-R1). MOSAC was excited about the full slate of field campaigns and encouraged the Met Office to clearly define its goals and allocate resources for data analysis (FC-R2). MOSAC mentioned the importance of including waves in coupled systems (FC-Q1).

The R2O team have a close working relationship with the parametrization and dynamics development teams which facilitates testing. We thank MOSAC for this positive feedback on the plans for field campaigns. We are actively investigating how to port the wave model to new computer architectures as described in the detailed response.

Science to Service: the technical challenges

MOSAC is keen to hear more about the progress of innovations from Science & Technology to operations, including the technical challenges in integrating scientific codes into the operations workflow and the move to the cloud. It is also interested to have more detail on the process of adapting science codes for customer consumption, and the use of tools for developing customer front ends (S2S-R1). Still the actual topics discussed demonstrated the high value of the Met Office's work in society's responses to difficult weather conditions. As an example, ensemble was discussed in relation to address extreme precipitation and the Met Office is encouraged to consider other high-impact variables in presenting probabilistic data to different customers (S2S-R2).

We are grateful to MOSAC for the positive comments on the science to service presentations. The detailed response below provides more information on how code is moved from research systems into production systems.

Warnings at the heart of the Met Office Services

MOSAC applauds the Met Office for its professional and comprehensive warnings and engagement with government bodies and costumers (CS-R1) during the case study events (e.g., winter storms, summer heat wave). It encourages the Met Office to continue to increase the lead time of warnings. MOSAC raises questions about false warnings (HIE-Q1) and the consideration of ethics in human interventions (HIE-Q2). MOSAC is pleased with the performance of MOGREPS-G and post-processing tools in improving the timing and accuracy of the National Severe Weather Warning Service. It welcomes the creation of the "socio-meteorologist" position and would like to hear more about how this is shaping weather and climate services (HIE-R1). MOSAC is interested in the evaluation of red warning events throughout the end-to-end value chain (HIE-Q3).

We thank MOSAC for these very positive comments. The Met Office has a subjective assessment process that verifies all issued amber and red warnings, including situations where such a warning may have been appropriate, but none was issued. The performance is evaluated by our sponsors. On the ethics and human behaviour, we are forging links with the academic community through our MOAP partners.

The future at the Met Office

The MOSAC has made several recommendations in its report to the Met Office regarding various future capabilities (PFC-R1 to PFC-R6).

MOSAC was impressed with the quality of the next generation of scientists and the presentations they made. The presentation on the Met Office Data Ethics Framework for Environmental Data was seen as especially important and it was pleasing to see that the Met Office recognizes the importance of handling data ethically. The presentation on interannual vs decadal response to the Pacific Decadal Variability was well received and MOSAC recommends a closer look at the relationship between the North Atlantic Oscillation and the Madden-Julian Oscillation (ECS-R1). The presentation on trends in global marine heatwaves and cold spells was interesting and of potential high interest to the blue economy. MOSAC encourages continued efforts in developing forecasts for fisheries applications. The presentation on the development of a regional LFRic model demonstrated the expertise in developing and testing numerical methods. The presentation on using storylines to engage new communities in climate change discussions was seen as energetic and engaging, and MOSAC appreciated the use of arts and community engagement in helping people understand climate change in a personal and non-scientific way.

We are grateful to MOSAC for the rich discussion on future capabilities and encouraging us to think beyond the Microsoft Generation 2 supercomputer. Responses to the specific points raised are below. Finally, we thank MOSAC for the very positive comments on the presentations from the early- to mid-career staff.

Concluding Remarks

The MOSAC concludes that the Met Office's research programmes are well-positioned to continue to provide valuable contributions to the advancement of weather, climate, and oceanographic science and services. The Met Office is encouraged to implement the recommendations made in the report to further enhance its scientific research programmes. MOSAC also has put forward human resource recommendations such as increasing gender balance in some aspects of the Met Office activities and to provide more information on the diversity and inclusion plan.

MOSAC was pleased with the progress made by LFRic, the next generation of post-processing tools and assimilation of observations. MOSAC was impressed with the use of machine learning to improve forecasting performance but concerned about the potential loss of understanding of underlying mechanisms. MOSAC is interested in seeing the Met Office's use of data science techniques and its plans for re-forecast and reanalyses. MOSAC is keen to hear about progresses in the methodology that brings innovations to operations, especially in the increasingly complex numerical Earth-system prediction landscape, and the technical challenges in integrating scientific codes into the operational workflow.

Kudos and many thanks to the technical and administrative staff. This meeting was well organised and run efficiently.

Finally, MOSAC would like to thank the Met Office leadership, speakers, and poster presenters for their great work. We found the meeting intellectually challenging and demanding as usual, but the openness and collegiality of the participants make it very pleasing and rewarding.

Gilbert Brunet (Chair) on behalf of the Met Office Scientific Advisory Committee: Natacha Bernier, Andy Brown, George Craig, Beth Ebert, Owain Kenway, John Michalakes and Ian Renfrew.

Annex: Recommendations and Questions

Chief of Science and Technology Overview

The Chief Scientist overview has permitted to have a good view of all the programmes and feedbacks are provided in the following short reports for each of the papers. MOSAC finds very useful the introduction of three-time horizons (A, B and C) for the purpose of planning and the introduction of the Met Office Science Progress session. MOSAC was pleased also to see this year the significant increase in the number of female speakers. This sends a strong message to the women scientific cohort and looks very well for the future of the Met Office gender balance. But there is still lot of work to do on both this and other aspects of diversity especially in fellow/senior roles.

The following items are related to topics not covered specifically in the papers.

CSTO-R1 MOSAC has taken note of the impressive list of committee members provided in Annex IV. In terms of membership there is some work to do on the gender balance especially given the importance of participation to committees for career progression and acquiring scientific management experience. Some senior scientists have numerous membership responsibilities, it would be good to consider, when it is justified, if mid-career scientists could take some of these responsibilities with in mind an improvement in gender balance.

Our corporate and Science/Technology Directorate EDI action plans have a range of activities in them that aim to improve diversity in all aspects of our work, and this includes representation on committees. The list provided in Annex IV is for a mixture of internal and external committees. For internal committees the membership is often defined by the role and for external committees we sometimes have little influence over their choices. However, we do recognise the issue and so aim to address it via some specific routes. Examples include: i) working towards getting better diversity of all types in our senior roles ii) ensuring we consider diversity where we have discretion in assigning people to committees and iii) working with our staff to support those less likely to put themselves forward for such opportunities. All these areas are activities we have running and through our corporate and Science Directorate EDI actions plans we aim to make continued progress in this area.

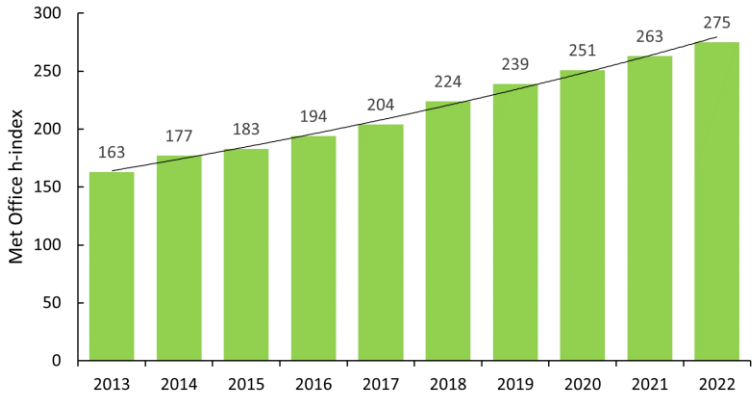
CSTO-R2 MOSAC would appreciate next year to have more information on the Met Office diversity and inclusion plan (including a demographic analysis).

The Met Office publish an annual EDI report that is available from our public website. We recommend MOSAC read this and if they still have specific question, our EDI domain experts that drive our activities can respond. The Met Office Board reviews our EDI progress and we also have external reviews of EDI through professional review bodies – currently we work with, and are credited by, Investors in Diversity.

CSTO-Q1 This is food for thought. MOSAC is intrigued by Figure 3 in Annex II. What is the interpretation of the h-index seeming to asymptote over the last 5 years for all institutions? It must mean that publications are getting less impactful as at the same time the number of publications per year has accelerated in the last 3 decades (see Figure 1). Are weather and climate sciences generating less and less important scientific results that attract the attention of the community?

Have we more papers, but with less quality or more and more specialized papers with smaller readerships?

The h-index presented in Annex II is calculated for the date range of 1981-2021, and is equal to 275 over this range. The curves in Figures 2 and 3 then show how papers published each year contribute to this value; more recent papers will have fewer citations than older papers, and for this reason this measure will tend to asymptote for the most recent years. This is also true for all centres, as shown in Figure 3. This is not the same as a historical time series of h-index and we apologize if the plots in Annex II have caused confusion. We are not able to easily reconstruct a h-index time series for other organizations, but we can do this for the Met Office. This is presented below. It is clear that the Met Office h-index has increased year-on-year and there is no evidence that the papers are becoming less impactful (by this measure at least).



Science Progress

MOSAC was pleased to see LFRic global working and giving reasonable scientific results and the response to the recommendation relative to computational performance (both workshop and subsequent progress), although it encourages continued focus on what Next Generation Modelling Systems (NGMS) (and Gen2) will be able to deliver in terms of performance for lower resolution climate. It was also good to learn of the recent good progress with the regional version now doing first runs.

The increased focus in Science on ensembles was welcomed, as were the cross-office efforts (reported later) to better exploit them. The merging of previously separate IT teams and locating in FS also appeared well justified and to be working satisfactorily.

The climate programme will be considered in more detail by the SRG. Nevertheless, MOSAC would be interested to learn more about the thinking on the evolution of the focus of the climate programme over the coming years.

Of the major change in the programmes, the plans (or even requirements) for PRISM were rather less clear to the committee, although it appreciates that this is still a work in progress.

SP-R1 Space weather has been mentioned over the past few years. MOSAC would welcome additional information. For example, what kind of space weather products are being prepared, what is their lead time, who are the primary clients and what is the desired lead time and foreseen product growth in 5-10 years to address costumer needs.

Since moving to 24/7 operations in 2014 the predominant Met Office customers for Space Weather are government, energy sector, defence, satellite operators and more recently aviation since the ICAO introduction of global space weather advisories. To date, the majority of the modelling capability utilised has been ported from NOAA Space Weather Prediction Centre, but the UKRI funded SWIMMR programme, will introduce UK models over the next few years to replace or supplement the existing US models. Currently we issue daily probabilistic forecasts of geomagnetic storms, solar flares, high energy proton events and high energy electron events up to 4 days ahead. The developments, listed below, will enable us to enhance the time discrimination and reduce the uncertainty of these forecasts while at the same time enabling us to provide more focused products that better meet our end user needs. In several areas we are developing machine-learning based approaches.

- Maximising utility and exploitation of the newly implemented SWIMMR models (Radiation Belt, atmospheric radiation, Ground Electric Field, coupled Thermospheric-ionospheric models and enhancements to CME/solar wind modelling)
- Developing capability to exploit the UK led ESA Vigil mission imagery & data, to improve geomagnetic storm forecasting , 2029/30 launch anticipated
- Improving thermospheric density modelling to support space traffic management
- Supporting military interests with enhanced ionospheric modelling and space weather impacts in the Arctic region
- Enhancing non-Earth space weather modelling to support human lunar missions

An extended version of the Unified Model with an upper boundary of 150 km has also been developed and tested although further work is required to ensure system before it could be made operational.

We would be happy to present progress at a future MOSAC meeting.

SP-R2 Good progress has been made with the Next generation processing and assimilation of observations (NG-PAO). On the wider JEDI work it was good to learn of continued progress (and successful working JCSDA), although there is inevitably further to go and more to be done in terms of interfacing with LFRic. It would be good to hear more next year.

We thank the panel for their recognition of the progress in this area and the hard work of all those involved in getting us to this point. The interfacing of JEDI with LFRic is indeed a priority and we expect to make progress on this over the coming year. We will be happy to share progress next year.

Foundational Capability

FC-R1 The strength and breadth of Met Office research into fundamental physical processes in the atmosphere continues to be a unique strength. MOSAC was pleased to see that the roadmap for Foundation Science continues to be developed. The use of 3 timeframes to structure the activities appears to be very useful to organize planning. It is important that R2O testing, and refinement of parameterizations is part of the planning process right from the outset.

We are pleased that MOSAC recognises the strength we have in physical process research. Maintaining this continues to be an important consideration for our science programme. On the topic of testing refinements to parametrization schemes, the R2O team have a close working relationship with the parametrization and dynamics development teams and the developers of the GC and RAL science configurations. This relies on a combination of informal working relationships

at both management and working level and formal process such as R2O representation on the governance bodies for the configuration development and delivery. We are also developing other ways to join the ends of the delivery chain at early stages of delivery such as the assessment of output from new parametrization schemes through O2R evaluations and testbeds.

FC-R2 It is very exciting to see the observation group is recovering from the COVID pause with a full slate of field campaigns. MOSAC recognizes the need to free time and resources to analyse the glut of data being produced and encourages this aspect to be clearly addressed in the planning of any campaign. The links with university researchers for WesCon, including the role of the campaign in the new Turbulent Processes Programme, are very welcome. Given the number and variety of partners, MOSAC encourages the Met Office to clearly define its own scientific goals in this campaign.

We thank MOSAC for this positive feedback. The WesCon campaign was born from scientific needs identified under the Met Office urban-scale modelling strategy. The funding of university researchers to join the observational effort is extremely welcome and will add greatly to what we can achieve, both during the campaign and in subsequent scientific work. The first integration workshop between all NERC-funded projects and the Met Office was held recently at the Met Office. It was a very useful first step in sharing and establishing priorities between observational groups and soliciting input from a wide range of data users to help fine-tune details of the deployment strategy. We anticipate significant further work and coordination in this domain in the run up to the summer campaign.

FC-R3 We welcome the appointment of Jon Taylor as Principal Fellow to take synergistic view of the observational activities across the Met Office. We would be very interested to have a presentation from him next year, giving an overview of the activities, and the gaps and priorities that he has identified.

Jon Taylor will be happy to provide such a presentation and is looking forward to the opportunity.

FC-R4 The roll-out of GC5 and RAL3 are major achievements, particularly given the background of urgent NGMS work. We continue to have concerns about the agility of the model development process and the response to problems that appear shortly before the cut-off dates. Recent examples include the late discovery of problems with tropical cyclone prediction in GC5 and the late correction of problems with the CoMorph convection scheme. MOSAC is pleased to note that the process has been modified in recent years to include NEWP system testing, including data assimilation, early in the development cycle. It might be useful in future to developing criteria for including more changes between cycles, and more closely defining the role that trailblazer models can play in this process.

The operational upgrade process already includes the ability to make changes outside of a parallel suite, although for model physics changes, we use this only for fixes to unacceptable performance issues that would not have a significant impact on other aspects of model performance. In addition, we have previously pulled larger (but not major) changes into an operational system as part of a parallel suite (but outside of the GC/RAL process) when these were known to significantly improve performance or a known operational issue at a time that we were not ready to implement a full GC/RAL release. Recent examples include the introduction of the multi-grid dynamical solver and an improved relationship between wind-speed and ocean surface drag. For the major changes that the committee have discussed above, however, the model development teams have started holding a list of model developments that are reaching

the levels of maturity where they could be considered for use in operational upgrades between GC/RAL releases, if they show evidence of enough benefit to justify an upgrade. For large or fundamental changes (such as CoMorph) we see the trailblazers as the most promising initial route for operational implementation, as they will allow us to gain useful feedback into model development whilst providing early benefit to users of our data.

The tropical cyclone intensity problems were a feature of GC4, discovered during operational implementation. These have largely been remedied in GC5, and additional test cases at high resolution were introduced into the model development cycle specifically to tackle this issue and prevent future degradation.

FC-Q1 The paper and talks have shown the importance of including waves in coupled systems at the regional scale. Including waves in ESM can also be relevant at the global scale (e.g., wave-ice coupling affects the location of the marginal ice zone and modulate both the formation and melt of ice). In 27.2 there was mention of the challenges of preparing WW3 to leverage future HPC capacity. What are the plans and impacts moving forward?

Our work to assess WW3 deployment on compute architectures has covered CPU based machines and near-future GPU based technology. For CPU architectures, the main benefits for our wave models are to be gained from using existing WW3 strategies to reduce memory use per node and increase processor count. The Met Office will likely run its existing operational wave models using WW3s hybrid MPI-OMP facility from Parallel Suite 47 (2024), and is developing and reviewing a new multi-grid Spherical Multiple Cell grid (SMC) global wave model for operational deployment on Horizon B timescales. The SMC multi-grid capability should now be available for wider use through the development branch of NOAA's WW3 github repository.

Achieving WW3 compatibility with GPUs is proving a hard task, although we have recently (post MOSAC) achieved some breakthroughs that enable us to run source term and (SMC) propagation components of the code on a GPU with performance equivalent to 2-3 CPU sockets. The price of getting this performance is a significant level of code restructuring, meaning that there will need to be a significant dialogue with the WW3 developer community to decide if these code updates could be practically implemented within the existing code structure. It is not fully clear yet whether the restructuring would also benefit a CPU only deployment. Our plan is to further analyse the system and provide details of our findings to the WW3 developer group during 2023, with the intention of achieving a group consensus on a way forward for the model.

Pioneering Research

We thank the MO for trusting MOSAC enough to seek our input into their programme. The questions in the paper are very interesting and raise additional questions.

PR-Q1 For example, what has the highest impact on prediction (i.e., order 1)? What is associated with the largest risks (e.g., of pushing the system toward a tipping point)? How computationally expensive are each component of the ESM? What has considerable feedback but must be optimized to make the cost manageable? Can AI help with some costly parts? Where can AI best be used so it provides significant added value and allows sufficient freedom for new model balances to develop?

This comes back to finding the right balance between complexity, data, compute power and ensemble size/resolution. To come to decisions over where to place resources (both computational and people) one really needs first to know the questions being asked.

We thank MOSAC for these insights and agree that the answer to the balance of resources depends on the questions being addressed. We also recall the issue of the balance of computational resources was the focus of the pre-MOSAC workshop in 2022. It was very clear from that workshop that there is no simple single answer to this balance. For instance, spatial resolution requirements are linked to the physics we are trying to capture and the bias introduced by not having sufficient resolution. As an example, work on high resolution global modelling shows how higher spatial resolution can lead to important changes in circulation and precipitation over Europe (e.g. Moreno-Chamarro et al., 2022). Similarly, ensemble size requirements follow from consideration of the required signal to noise ratio on seasonal timescales, and also what magnitude of return period we might want to examine directly with physically based model output. The level of complexity needed is determined by the question being addressed and consideration of the likely size of relevant earth system feedbacks. For instance, the nitrogen cycle being included has an impact on carbon uptake by the land, and also on carbon release from permafrost (as described in the MOSAC paper). Thus, our approach is to keep open the option to flex the balance between resolution, ensemble size and complexity according to user and scientific questions.

In the MOSAC presentation we included a description of the structured approach we take to decide on developing and including new complexity within the earth system model. We also retain within UKESM climate work the ability to run a faster version of the model where appropriate. For weather prediction we are prioritising the use of the ensemble (Ensembles R&I theme and a strategic action on Ensembles Exploitation) and for near-term climate prediction on months to years ahead we are prioritising ensemble size as this has been demonstrated to give more benefit than, say, doubling the current resolution. We are however investigating the role of coupled ozone for seasonal forecasting as this has also been shown to deliver potential benefits and we are investigating higher resolution to address the signal to noise paradox.

We continue to research the effect of greater resolution (as described in our approach to k-scale modelling) and of additional environmental complexity (e.g. making the fire model more complex) to inform the decisions on experimental design and model set-up for e.g. future CMIP experiments. In the capturing complexity theme we are considering how we can make more use of computationally fast emulators (e.g. connecting together models like FAIR with a simplified spatial treatment and link to JULES) and AI approaches for time intensive schemes. Additionally, we have been discussing within the Met Office and in the international km-scale community (e.g. at the WCRP storm-resolving workshop in Boulder) the relative (compared to the Atmosphere or Ocean) immaturity of work on ultra-high resolution land modelling. At these scales, many new aspects of hydrology could potentially be included (parameterized) and its clear we will need to work to prioritise the most important processes and what needs to be included that will have global upscale impacts. This will have implications for complexity/resolution choices for particular climate applications.

Constraints on our future ability to deploy suitable balances of resolution/ensemble size and environmental complexity in our major projects, including UKCP and CMIP, are taken into account in development of our future modelling systems (NGMS), for instance by having representation from the major projects that will use the ESM in future.

As the MOSAC question asks about the relative timing of major model components we include some statistics here. For UKESM, UKESM1 (N96Orca1) runs approximately 4 yr/d on 37 nodes. Of this, the ocean (Orca1) with MEDUSA uses only 14% of the total resources and so is

inexpensive compared to the N96 atmosphere. In the atmosphere, the physical core (dynamical core, radiation, convection, cloud microphysics, JULES etc) is about 50% of the total atmosphere runtime while aerosols are about 17%, and chemistry 33%. There is also some evidence that we could get a significant speed-up by running the chemistry code on GPUs. At N216 Orca025 the ocean and MEDUSA become more expensive than the N216 atmosphere.

PR-Q2 In the Air Quality section, the grid scale detail appears to come across in the visualisation. Is the map only of additional detail or of a forecast? Is there lateral advection? How are high-resolution details imbedded into the 2.2km grid?

The images show output from a series of model predictions. The gridded model outputs were shown from two separate runs of AQUM, one with a horizontal grid length of 11 km and the other with 2.2 km. The final image, showing roads, is generated by combining the output from a street model (ADMS-Local, developed by CERC) and the regional AQUM (2.2km model run) data. Both models are run separately. ADMS-Local is run using the same meteorology and the same emissions both on the grid and also remapped to the roads. The outputs are combined at each output time step through a process that ensures that emissions/concentrations are not double counted. The result is a consistent combined prediction of time-evolving regional and local air quality. The approach is outlined in Hood et al 2018 (https://doi.org/10.5194/acp-18-11221-2018).

Science to Services

From the title of this talk, MOSAC was hoping for a detailed discussion of how methods progress from the science side of the organisation to operations and how this will change in future, particularly the following:

S2S-R1 MOSAC is keen to hear about the technical challenges faced in "productionising" scientific codes written originally for researchers and integrating them into the operations workflow. We are keen to hear how the planned move to the Cloud is improving (or making harder) these efforts with the associated changes to ways of working and developing software for a flexible, software defined environment. MOSAC is also interested in more detail in the work that goes on in adapting science codes to produce outputs that are consumed by the Met Office's customers. For example, we are interested in whether there is a common toolkit used to develop customer front ends or whether a more bespoke approach is taken for each project. Do all customers consume data through a web (or other GUI) front end, or do some customers prefer to consume raw or filtered data output and pass that into their own applications? We would therefore appreciate more updated detail on this important process at a future MOSAC.

In terms of productionising codes originally written for researchers, the approach taken depends on several factors including the complexity of the code, the architecture of the system it is developed and implemented in and the Non-Functional Requirements of operations (NFRs, e.g. latency or resilience). For code that is simple, going into a very different system that it was developed in or going into a system with demanding NFRs then often we take the opportunity to refactor the code, with scientists involved in testing to ensure that the refactored code delivers the expected results. For code that is complex, going into a similar system that it was developed of that already performs well against NFRs, then we are more likely to implement this as-is, following an appropriate level of code review.

It is early days for cloud-based systems, but we already have some cloud-based systems (e.g. in the observations) and others that are imminent (SPICE). We are working through the changes in

working practices, which currently centre around understanding the extent to which cloud compute costs are exposed to, and manged, by individuals rather than more centrally.

The work to develop customer front ends is beyond the scope of Science: it sits with colleagues in other areas of the Met Office. Customers consume data through a variety of means including both web/GUI and direct data feeds into their applications. Customer needs are assessed by colleagues in other parts of the business with the Science role being to produce and assure the scientific integrity of the data. For customers paying us for a product or service we also provide scientific assurance about how the data should (and shouldn't) be used and we verify the data quality.

Although the topics discussed were not the technical detail focus expected they demonstrated the high value of the Met Office's work in society's responses to difficult weather conditions over the last year, from extreme heat to extreme precipitation. The technical background on the domain-specific applied topics (for example railway engineering) was appreciated and this made these parts of the talks much more understandable than might otherwise be the case.

S2S-R2 The use of ensemble to address extreme precipitation is important and relevant. We encourage the Met Office to continue to explore such avenues for other types of impacts.

We are also interested in the challenge of how ensemble/probabilistic data is presented to different customers and commend the Met Office for its clear consideration of this problem.

We thank MOSAC for this commendation and hope to highlight advances in Ensemble Exploitation in future MOSAC meetings.

Data Science: Plan and Progress Update

MOSAC is excited by the potential for machine learning derived emulators to improve the performance (both accuracy and computations) of the very time-limited forecasting part of Operations.

DS-R1 MOSAC feels from the presentations, however, that the actual motivation behind using machine learning in simulation workflows is unclear and believes that it needs to be better understood why we are doing this. The focus on emulators seems to be around speed in which case improved time to solution needs to be demonstrated. MOSAC understands that currently this is limited due to the architecture of the present machine and of the Gen 1 system (which are CPU-based) and hopes that some of the development work during Gen 1 will include testing performance at scale on GPU resources either in the Cloud or on systems run by Met Office partners as preparation for Gen 2. MOSAC is keen that any use of machine learning on the science side does not decrease the understanding of the underlying mechanisms.

We agree that the use of machine learning (ML) should not be at the expense of our understanding of the mechanisms/systems we are modelling. We are pursuing a 'fusing data science and simulation' approach to get the best of each approach and this includes understandability. This is also why uncertainty and trust are core to our Data Science Framework. Whilst there is a risk with ML techniques that insight could be lost, explainable AI, causal techniques and other methods can actually lead to discovery of new understanding of how systems behave, and not just reproduce our existing understanding.

An increase in the computational speed of models with integrated ML components requires further investigation and needs to be informed by our Gen 1 plans but also needs to inform our Gen 2 plans. Speed is pursued not just to reduce compute cost but also because this may allow more

complexity to be represented, or for example for certain parameterization schemes to be run more frequently or at a higher fidelity, leading to improved accuracy. However, increased speed is not the only application of intra-model machine learning.

The creation of Intra-Model Machine Learning team within the parametrizations group embeds these considerations into the heart of our model production 'engine', in the team where the core physical understanding resides, and can push forward with these challenges.

DS-R2 MOSAC would like to see more consideration of a potential split in numerical prediction systems between science (physical models) and operations (where machine learning methods might dominate) and the consequent loss of the "unified" part of the Unified Model.

We acknowledge that this is an important consideration, although it is currently unclear to what extent the adoption of machine learning for operational prediction will give rise to such a "loss of seamlessness".

DS-R3 MOSAC notes the Met Office's plans to use machine learning and other Data Science techniques to analyse and derive meaning from the vast quantities of weather (and related) data sets that it collects from various sources and is interested in seeing more of this in future.

Similarly, when data science techniques are used in post-processing, this is an exciting area of research, and we look forward to future developments from the Met Office.

In all these areas, the Met Office is at the forefront of this research, and we believe that both positive and negative results are valuable learning for the whole community.

We thank you for the feedback and look forward to sharing our ongoing results in future.

DS-R4 The use of CLIMAR seems remarkably interesting. MOSAC is interested in whether there are plans to use or develop open-source data so that countries with less developed climate and forecasting infrastructure (and who might bear great impact) can benefit from this work?

Thank you for the feedback. CLIMAR is aligned with wider activities including our R&I Strategy theme Hazard to Decision Making and our vision for Digital Twins. We agree these themes have potentially very positive impacts in a wide range of areas including countries with less developed climate and forecasting infrastructure and as our strategy in these areas develops, these themes will be increasingly used in our engagements both nationally and internationally for positive impact.

MOSAC commends the MO on their excellent machine learning, training, and development plan.

DS-R5 MOSAC are concerned that plans around re-forecast and reanalyses appear to no longer be considered. At the same time, there are talks around the increased usage of emulators. What information is the MO planning to use to train emulators?

A report on our strategic review regarding reforecasts was presented to MOSAC in 2022 (MOSAC 26.14). In this report we stated that, beyond the areas of seasonal and decadal prediction, there were no plans to develop reanalyses or reforecasts in the short term, but that we would investigate the benefits of making an archive of operational forecasts available. For example, operational high-resolution UK ensemble data are currently provided to the EUMETNET archive for post-processing research and plans are under development to make data available to external users. Nevertheless, as our plans for data science become more ambitious, we shall keep this decision under constant review.

K-scale to Urban Scale

MOSAC was impressed by the progress. On the global scale, the use of a fleet of global, big domain LAMs and tropical channels runs seems appropriate as the best way to make progress with experimentation, noting the potentially very large computational costs. It agrees that a primary challenge is to show evidence of the better representation of the smaller scales leading to a better evolution of the larger scales (and indeed finding the best metrics and diagnostics to use as part of this).

K-R1 The trailblazer sounds like a potentially interesting and useful way in between traditional experimental and fully operational work. However, MOSAC recommends further thinking on exactly its purpose and how it is pitched and used, particularly where on the spectrum it lies between the experimental and operational ends. This may be particularly important if there is to be a deterministic 5km trailblazer, as it could significantly cut across other efforts (technical and cultural) to push ensembles to the heart of everything.

We share these concerns, and we will work with potential users and those that prepare and provide our operational data feeds to determine how best we can use these applications without undermining the function of our flagship ensemble system. It will be important to establish and agree clear principles for the usage of the data feeds from these trailblazer systems.

K-R2 Looking further ahead, MOSAC is concerned that there will not be sufficient parallelism to make global km-scale fast enough to be useable for operational NWP. Hence (with a Gen3 time horizon) MOSAC encourages the Met Office to be thinking about (1) creative ways to make things affordable (e.g., work referred to on making different parts of the model run at different resolutions; other algorithmic developments; possibly variable resolution; emulators etc.), and (2) opportunities to add value using k-scale that do not require operational simulation rates; for example, reanalysis for ML training, nature runs, process studies, etc.

Ultimately to gain full benefit of high-resolution global NWP will require advances in the resolution (and approach?) for data assimilation (although there may still be some easy gains noting that current DA resolution is a long way behind model resolution and so could be increased while still being well short of convective scales). A further concern is data assimilation for short-range regional prediction, where there are fundamental scientific uncertainties about the algorithms and use of observations.

We agree that there are significant challenges to making optimal use of future architectures and getting to global k-scale for operational NWP. We will not achieve this through traditional approaches and will need to explore new avenues. While the current focus remains on getting LFRic ready (scientifically and computationally) for our Gen1 supercomputer, we are also considering the longer term Horizon C timescale. Specifically:

i. A number of technical innovations have been built into LFRic alongside the target dynamical core configurations. These include; the ability to run parts of the model (including IO) on different nested grids, potentially enabling more targeted deployment of computational resource; the underpinning infrastructure for higher-order finite elements, providing the potential for more intense compute on GPUs and the introduction of finite-element transport; capability for different operator-splitting strategies in the semi-implicit timestep, providing the potential for more coarse-grained parallelism in the algorithm. The Dynamics Research team will also continue to explore new algorithmic developments alongside changing hardware drivers.
ii. While Met Office resources are focussed on Gen1 developments, we are collaborating with academia and UM partners on a range of other 'blue sky' activities. This is particularly

the case through the ExCALIBUR programme with projects including Parallel-In-Time, task parallelism, IO and other cross-cutting themes (i.e., applicable to Weather and Climate, but also other industrial problems) which leverage a broader field of expertise.

iii. While the extent to which machine learning will help with this problem is currently unclear, it is important that we pay attention to the very rapid pace of development in this area. Some components of the model may be well suited to emulation, and for example our new Intra-Model Machine Learning team in the Atmospheric Processes and Parametrizations group is exploring emulators for physics schemes. We are also looking more broadly at all elements of the NWP process to understand where machine learning can provide an advantage. This analysis will inform our future plans.

We also agree that km-scale simulations can provide a rich resource outside of operational NWP. Making good use of datasets is a key aim under the Path to High Resolution (P2HR) R&I Strategy theme and specifically within the K-Scale project, where we will ensure that data from new global and large domain km-scale simulations can be explored by academic groups, for example to advance process understanding. We also anticipate these simulations will be a rich resource for Data Science exploitation. Fundamentally, at this stage, P2HR remains more focussed on exploring the value of increasing resolution (and cost) relative to the existing benchmarks and alternative routes to increasing predictability at lower cost (e.g. ensembles, complexity etc).

K-R3 K-scale data assimilation might be a useful topic for a workshop at some point.

As a first step we are considering our approach to regional km-scale DA within JEDI. This will be based on Ensemble Kalman Filters (EnKF) to start with, focusing on the gap between observation-based nowcasting and the MOGREPS-UK regional ensemble (i.e. initialisation of 0 to 12h forecast). The flavour of EnKF is not yet decided. We agree that a k-scale DA workshop could be very useful, and suggest that this would focus on exploring DA systems/methods which take into account non-linear processes and non-Gaussian distributions, appropriate for convective scales.

In terms of the sub-km urban scale work, it was encouraging to see first evidence of cases where evidence could be seen of added value from the very high resolution (e.g., more organized convection bands). However, it also seems appropriate that the Met Office also plans to compare with 1.5 km, including postprocessing, as a benchmark noting the large costs of the model.

We agree that as we continue to develop our sub-km modelling approach it will be important to fully understand its value vs post-processed, coarser resolution, model predictions

The WesCon field campaign, in collaboration with academic partners, seems like a good opportunity to gain relevant observational data. In order to maximize the value obtained, MOSAC encourages continued discussion around exactly what WesCon measurements would best help address the primary modelling questions (see FC-R2).

Please see the response to FC-R2 on the WesCon plan.

K-R4 MOSAC also wonders whether the Met Office could make further efforts to galvanize the academic community around another attempt to have a further field campaign in the tropics (e.g., Maritime Continent and/or Africa). Given current logistical difficulties the challenges are substantial but looking on a say ~5-year timeframe, there is perhaps a role for the Met Office in supporting and encouraging academic community, e.g., via a workshop, townhall meeting at an appropriate time.

We agree that there may well be a role for the Met Office to take here in the longer term. Our strategy is however to focus first on completing the implementation of global model enhancements

that we know will impact tropical performance significantly – particularly the introduction of the CoMorph convection scheme. We will then be in a position to assess the systematic errors in these systems, and in turn identify priorities for further fieldwork. Should we identify the need for further observations in the tropics, we would seek to work with partners to explore how these could be facilitated. We agree that any such activity would be a major undertaking, requiring involvement across the UK community and with international partners.

In addition, we are increasingly well positioned through our international collaborations (e.g. under the WCSSP programme, and via the UM Partnership) to benefit from routes to access data from long-term observational research facilities supported in at least some tropical areas, and we can explore making more of these opportunities.

NGMS: Technical Integration and Optimisation Projects

MOSAC commends the Met Office for convening last year's special-topic workshops on LFRic performance and for its ongoing and vigorous responses to the recommendations arising from those workshops. These include increasing (and protecting) dedicated software engineering resources for LFRic performance optimization, appointing a new Met Office fellow to head the newly created NG-OPT project, and improving test coverage for tracking gains for NWP and climate configurations. Considering plans to move climate simulation away from the UM by the end of the decade, MOSAC agrees with SRG on the vital importance of addressing all aspects of NGMS performance, especially per-node performance needed for lower-resolution/higher-complexity climate applications that are less able to scale to large numbers of processors.

NGMS-R1 MOSAC encourages continued follow-up activities between the Met Office and the LFRic performance panel through the remainder of its term. MOSAC requests that Met Office provide a report next year that includes detailed plans and timelines for addressing each of the panel's recommendations from the May 2022 workshop report and other follow-up recommendations. MOSAC requests summaries of the technical information that have been provided to the panel, along with the panel's responses, on (1) performance of LFRic on the suite of NWP and climate configurations highlighting remaining gaps in meeting performance objectives, (2) NG-OPT's accomplishments improving per-node computational and memory-system performance, including tools and methodologies, and (3) other accomplishments or challenges arising for improving LFRic performance and scalability.

We thank MOSAC for this encouragement and thank all members of the panel for their valuable time and expert advice. We are happy to say that the next follow-up meeting of the expert panel has now been arranged for mid-June 2023. We have already shared with the committee the panel's original recommendations from the May 2022 workshop and our responses to those. But we will be happy to report on any further recommendations and comments of the panel, as well as our progress on addressing those points and improving the model's computational performance.

NGMS-R2 The distinct focus of the NG-OPT project on *computational* optimization is on the mark; however, it is important to avoid separation from existing efforts to improve performance through *algorithmic/numerical* optimization. To ensure as holistic an approach as possible, MOSAC urges continued close coordination between these highly interrelated computational and algorithmic optimization efforts, with regular two-way exchange of information on latest performance and efficiency measurements, on code and data restructurings, and on the interplay with algorithmic/numeric updates. As noted elsewhere in this report, NG-OPT should also coordinate

closely on best software and performance engineering practice with other groups developing software destined for NGMS, such as those working on Unified Physics.

We completely agree with these points and, as a result, one of the work packages of the optimisation project (NG-OPT) is specifically about improving the algorithmic/numerical performance. This work package will be led and resourced by people from the Dynamics Research group. With regard to the final sentence of this recommendation, although we agree with the sentiment, we have to be mindful that NG-OPT has limited resources with very high priority objectives. Our preferred approach is, in time, to follow a similar approach to the UM, whereby performance aspects are part of the business-as-usual model development process, e.g., with specific optimisation/performance reviews and testing. In the shorter term several members of the LFRic infrastructure project have expertise in the development of HPC applications and they will work closely with NG-OPT in ensuring that performance engineering is part of the development process.

NGMS-R3 MOSAC notes that there is considerable software development effort going across a very large suite of often bespoke software. While we understand there are political challenges with open sourcing models, we would be keen for the Met Office to consider where more generic infrastructural or library code could be open sourced and distributed to the wider community. Hence the Met Office could make good use of code contributions and bug fixes from the community. A good example where more engagement could be carried out would be around Fab, which the Met Office has published under an open-source license on Github and Pypi. Fab solves problems that are generally applicable to large Fortran code bases but presents some issues for usability and accuracy of its documentation and so would benefit from wider community contributions.

The GungHo dynamical core, the LFRic infrastructure, the CASIM microphysics and the SOCRATES radiation schemes have all been developed under the BSD-3 open licence. Furthermore, about 18 months ago, and as reported to MOSAC two years' ago, a new general licensing policy for next generation modelling system components was agreed by the Met Office Executive. This policy allows such components to be developed and released under an open licence arrangement, such as the BSD-3 one. This process has been successfully applied to both the UK Chemistry and Aerosol code (UKCA) and the Joint UK Land Environment Simulator code (JULES). These developments have been welcomed by the communities using these models and we hope that it will encourage greater uptake of, and contributions to, these codes.

Regarding Fab specifically, we do not understand the comment about usability and accuracy of documentation, but we are keen to address any issues and so would welcome more context from the committee on this. However, we do agree that our decision to release it under open licence on Github will allow a wider community to contribute. It is also worth noting that we also have short term plans to deploy Fab as a build system for the Unified Model. This adoption process will undoubtedly contribute to improving usability and accuracy by providing an application case with a large user base.

NGMS-R4 The appointment of a Met Office fellow to head the new NG-INT project to ensure that technical critical paths that cross through the project internally and through external partnerships (XIOS, JEDI) is welcome and timely. We look forward to hearing details on successes and challenges for smooth NGMS project integration at the next meeting.

We will be happy to provide the requested details at the next meeting.

High Impact Events

MOSAC was impressed by the professional and comprehensive nature of the warnings and engagement with government bodies in the case study events highlighted (e.g., the winter storms, summer heat wave). The work here is interesting, especially the push toward as early as possible warnings and the considerations around social behaviour. We encourage the MO to continue to advance and increase the 70% baseline KPI target, in addition to increasing the lead time from 1h to several hours, for as many warning as possible. One question raised was:

HIE-Q1 Are there KPI's on false warnings?

The Met Office has a subjective assessment process that verifies all issued amber and red warnings (and also situations where such a warning may have been appropriate but none was issued). These warnings are assessed in terms of 'impact level', 'area affected' and 'validity time' and the scores result in a rating of very poor, poor, good or excellent guidance. The 'very poor' category includes both false alarms and missed warnings. There is a target for at least 80% of assessed warnings to provide 'good' or 'excellent' guidance, and for no more than 20% of warnings to score 'very poor', these scores based on a 36-month rolling average.

HIE-Q2 In the presentation, human intervention based on expected behavioural response (people flocking to the coast) and risks associated with this response were considered and resulted in the extent of the zone of an alert being increased.

What partners is the MO working with and how are ethics considered in this sort of intervention?

MOSAC is pleased to see the good performance of MOGREPS-G and post-processing tools (DECIDER and IMPROVER) in improving the timing and accuracy of the National Severe Weather Warning Service, including the extra lead-time provided by IMPROVER for extreme temperatures.

The issuing of severe weather warnings is a collaborative effort, with the civil contingency team (among others) providing insight and recommendations to the chief meteorologist. Civil Contingencies Advisors (CCAs) have regional insight into the variety of impact response which inform National Severe Weather Warning Service (NSWWS) decisions in real-time (e.g. via a dedicated Teams chat). In the case of the Summer 2022 heatwave the professional views of the UK Health Security Agency and Cabinet Office were sought before the warnings were issued.

In slower time, we are working with our Met Office Academic Partnership and other academic partners to inform background information relating to behavioural response. The Public Weather Service also commissions behavioural insights organisations to conduct specific research. We also conduct routine public surveys following the issue of amber and red warnings to measure awareness and response to warnings.

HIE-R1 MOSAC welcomes the creation of the "socio-meteorologist" position who has skills both in meteorology and behavioural psychology. We would be interested to hear in a future meeting how this new in-house capability is shaping weather and climate services. And also, how this position or group plans to interact with the large academic community (e.g. at Leeds, in the Tyndall Centre for Climate Change Research, etc.).

We can provide an update at a future meeting. In the meantime we are building a network incorporating a variety of disciplines, expertise and skills, with a particular focus on social and behavioural science. We have excellent links with the University of Leeds (via

MOAP), especially their centre for decision research, but aim to develop stronger social science links with other universities. We also aim to learn what international organisations are doing in this space, e.g. in the US and Australia, where we have a specific social science collaboration with the Bureau of Meteorology.

HIE-Q3 MOSAC was interested in the evaluation of red warning events, not only from the modelling and post-processing perspective, but right through the end-to-end value chain – how is this done?

The assessment of weather warnings is discussed in our response to HIE-Q1. The purpose of the new Weather and Climate Extremes and Impacts Team is to empower people to deal with extreme weather and remain safe and resilient in our changing world. The work of this transdisciplinary team spans the value chain, with National Severe Weather Warning Service (NSWWS) as a key focus. The team will consider real-time climate change attribution, contextualised advice, post-event analysis with a human-centred approach, and greater and more effective collaboration between science and services with a buddy system for researchers and operational meteorologists. Although our verification process does not differentiate amber warnings from red warnings, the Weather and Climate Extremes and Impacts Team will, by definition, be particularly interested in events in the red warning category.

NWP Verification

With the move to ensembles only, MOSAC is happy to see KPIs of NWP performance being measured using CRPS for the global ensemble and HiRA for the regional ensemble, including a CLIPER reference (rather than random chance) for the latter.

NV-R1 From a scientific perspective it is useful to see the value-add of IMPROVER over a raw NWP baseline, including the increase in performance for each step in the processing chain. From a service perspective, IMPROVER was shown to be more reliable and valuable than Best Data. Improved consistency was mentioned and shown on a map but not shown quantitatively. Messaging around this value-add from IMPROVER could be a useful part of the “ensemble first” journey within the Met Office. Further, it could be useful to develop/implement an appropriate metric for forecast spatial consistency so that this benefit can be measured.

Thank you for the comments around demonstrating the value add of IMPROVER over raw NWP. We agree that it would be helpful to quantify improvements in consistency. Our plan is that the verification using additional, non-Met Office surface observations such as those from WOW, should help to quantitatively assess the improvements in spatial consistency. This, together with the knowledge of how calibration is performed by the two post-processing systems, should demonstrate the improvements in consistency from IMPROVER.

NV-R2 There is increasing evidence that the consideration of observation uncertainty can significantly impact the interpretation of verification results (see, for example, the ECMWF work of Ben Bouallègue). This is of increasing importance with the availability of new unconventional data sources such as WOW that provide large amounts of low-quality data, but also the increasing use of ensemble products that produce distributions and uncertainties. A key problem is characterizing observation errors, including new instruments such as autonomous vehicle sensors, but also representativity issues (e.g., rain gauge vs. radar). MOSAC recommends that

the Met Office continue investigating and quantifying the effects of observation uncertainty on verification results and their interpretation.

We thank MOSAC for this recommendation. We agree that this is worthy of further consideration. With respect to verification, our current system (and our next generation, NG-Ver, system) has the capability to include observational uncertainty. In the short term we intend to focus on assessing the sensitivity of the verification to error estimates, with input from our colleagues in the Satellite Applications and Nowcasting and Surface and Satellite Assimilation teams to see if we can apply their knowledge of observational uncertainty quantification . In the longer term we recognise that we may need to refine these error estimates for use in verification.

Regarding the crowd-sourced WOW observations, our nowcasting team have made significant efforts into characterising the errors in these observations and we will bring this knowledge into our verification work using WOW (initially in the form of advanced quality control). Finally, in our "Sensor Assurance Framework" project with NPL, we are simultaneously measuring meteorological parameters and the response of autonomous vehicle perception sensors to explicitly explore both observation errors and representativity issues. This will provide a well-curated dataset to explore issues around observational uncertainty and we are keen to use this across relevant applications.

NV-R3 The KPI to monitor the value-add of the regional model over the global model was discontinued a few years ago. As models continue to move to finer scales with greater complexity, it would be useful to revisit this question and monitor the value of the high-resolution modelling investment using an appropriate KPI.

Although the value added by the regional model over the global model is no longer a part of the forecast accuracy KPI, the relative performance is still routinely monitored and is used in decision making as appropriate. As we begin to introduce higher resolution regional models (urban scale) it will of course also be important to also compare the forecast skill of such models with the current generation of km-scale regional models. In doing so, however, it is important to recognise the differing requirements and applications of models at different resolutions e.g. the use of higher resolution to capture local extremes.

NV-R4 Ensembles provide considerable value at long lead times. MOSAC would like to see verification across all lead times.

Our forecast verification is routinely carried out at lead times that span the full forecast range available from the models. We would be happy to present some longer lead time verification at a future MOSAC meeting and would also note that we are planning to expand our forecast accuracy KPI next year to include measures of seasonal forecast performance.

NV-R5 Ensembles are being used for storylines as well as probabilities, but the NWP verification metrics seem to only address the latter. Metrics appropriate for measuring the accuracy and utility of ensemble forecasts when used as storylines (as opposed to probabilities) should be used. This may require new metric development.

This is an excellent suggestion and we will investigate if this can be considered as part of our Exploiting Ensembles Strategic action. Our current expectation is that this will require a significant subjective component, although we recognise the value of objective metrics where they capture the important considerations.

Observation-based Research

MOSAC wants to congratulate OBR on their achievements in what has been a hectic time post-covid, with a confluence of field campaigns and a lot of logistical pressures. Participation in several campaigns in last year and first use of the airborne Ice Nuclei Counter are real highlights. MOSAC appreciates the wide variety of observation-based research to support process understanding and model improvement. We encourage the prioritisation of activities to deprecate the less useful ones and allow greater time for analysis of data from the most important campaigns.

The upcoming WesCon campaign has the potential to be very influential on planned turbulent processes science and the recently funded NERC-MO programme of research. It is vital that the Met Office OBR & APP be pro-active and engage with the NERC-funded scientists to optimise the experimental design deployed on all the observational platforms and thus get the most out of the new capabilities (e.g., using the FAAM's turbulence observations to benchmark and corroborate the EDR from the NERC radars). MOSAC recommends early and meaningful engagement on WesCon and with NERC-funded modelling staff to ensure that there is an understanding of what will be measured and how it can be used to evaluate the various models and parameterizations (see FC-R2). OBR should consider what role it plays in galvanising the academic community around a potential tropical observational campaign (see K-R4).

We thank MOSAC for these remarks and have responded to FC-R2 and K-R4 in the appropriate sections of this report.

Physics Development

PD-R1 There have been several positive developments in the last year, including the inclusion of CASIM in RAL3, contributing to the unification of the tropical and midlatitude regional model configuration, and showing substantial improvements in convective cell properties. On the other hand, the omission of CoMorph from GC5 is disappointing, and might be an example of a model change that might be introduced between the main development versions (see FC-R4).

We agree that the omission of CoMorph from GC5 is disappointing. We are planning to implement Unified Physics (UP) components (particularly CoMorph) in the 5km global model Trailblazer as they become available. A key goal that will drive the choice of science configuration for the trailblazers is that the schemes used should be a step towards the scale-aware physics strategy that is being pursued by the UP project.

MOSAC congratulates the Met Office team on the initiation of the Turbulent Processes Programme. This continues and strengthens the cooperation with university research groups doing research on physical processes that was successfully established in the ParaCon programme.

PD-R2 The plans for models at 10 km, 5 km and 300 m resolution will be a major challenge for parameterization development, since these are all "gray zone" for some subset of physical processes. The motivation for the Unified Physics strategy is strong, but there is not sufficient detail at this stage to comment on the plans in detail. A practical concern is whether the number of different applications and model configurations where UP will be used will make it very difficult to get the package accepted. On the other hand, if UP consists of several individual parameterization upgrades to include scale-awareness, how will the interaction between schemes

for different model resolutions be managed? Will the 5 km trailblazer configuration be a useful tool in the development? We would encourage the preparation of a plan addressing these questions across the 5- and 10-year time horizons.

The committee is correct that achieving acceptable performance metrics for all the applications, when the priorities for different user groups will vary, is the biggest challenge. We do have experience of achieving this across timescales for global and regional configurations, and the need for pragmatism in applications whilst following a common trunk is likely to be required. UP will be developed in a whole range of test environments including traditional global (AMIP, case studies, DA trials) and regional (various mid latitude and tropical domains) configurations, but also the 5km trailblazer, K-scale simulations, sub-km tests, WesCon simulations, etc. The strategy is very much focussed on accurate simulation of processes and interactions initially, working from the highest resolution towards coarser resolutions. Inevitably, tuning will then be required to achieve acceptable performance on all the necessary user metrics, but this may also be where machine learning approaches can help.

PD-R3 MOSAC believes that software development for new and upgraded model physics packages should conform with LFRic model code structures and interfaces earlier rather than later. At the same time, MOSAC understands the need for physics code to remain compatible with UM through the end of the decade. Especially about writing performant code, MOSAC encourages continued close coordination on LFRic software best practices between physics development and LFRic software engineering teams.

We fully agree with these points and will migrate UP work into LFRic as soon as the necessary infrastructure is in place to work in that environment. It is important to note that much of the new code is already ported and it is hoped that the majority of what is remaining will be developed and ported before any code fork. If this can be achieved, it will minimise risk from development on two separate code trunks. The Atmospheric Processes and Parametrizations team works very closely with Dynamics Research (who led the Gungho development), with one member of staff sitting in both groups to help expedite communication and interaction. As such, optimal development of physics codes for LFRic from a computational point of view, and plans around future model timestep structure and interfacing in general are routinely discussed.

PD-R4 Related to PD-R3, there was little in this year's papers on the science performance following the coupling of UM physics into LFRic. The impression was given that this has gone well and there are no significant issues. Is that the case? MOSAC would be interested to see if further testing at different resolutions and with different science configurations also remains stable and robust.

MOSAC is correct that so far there are no significant issues, but there is a lot still to do. The goal for proto-GA and proto-RA was to deliver something that worked and looked realistic - this has been achieved. The work of the GC5-LFRic and RA3-LFRic projects over the coming ~2 years is to dig into the detailed scientific performance of the models to establish whether they are acceptable for operational implementation (and provide any tuning etc that is required to achieve that). This work will also consider performance across the range of resolutions, including operational (10km and 5km) global and sub-km regional configurations.

Consultancy Support

CS-R1 MOSAC is impressed with the continued level of engagement with costumers. The work showcased in this piece is relevant and achievable thanks to the solid scientific capacities that

provide the necessary information for this work to be possible. We encourage the MO to continue to stress that such work is made possible through leveraging of outcomes of the research and development efforts.

We thank MOSAC for these observations. A strong focus continues to be on pulling science through to applications on all the timescales we consider. To help us understand the impact of our work the Met Office has a Strategic Action on Evidencing our Value which will develop case studies that demonstrate how our exceptional science, technology and operations have benefitted users. We are also working on providing more feedback on what we learn from customers to inform our underpinning science.

Ensemble Exploitation

MOSAC welcomes the new strategic initiative on exploiting ensembles. It is astonishing that only 5% of the Met Office's automated products make use of ensemble data.

EE-R1 The work packages on underpinning research on ensemble use; ensemble development; developing tools, processes and people; engaging and supporting users; and communicating work and thought leadership span the necessary efforts needed to better exploit ensembles. The challenges of ensemble exploitation seem unlikely to be solved in two years, but we look forward to good progress in this direction. The experts named on this project bring very valuable scientific, operational, and business expertise. It will be important for them to be given adequate head space to focus their ideas and efforts on this initiative. Given the importance of ensemble prediction to the Met Office and other major centres, and excellent work being done in some, the Met Office should continue to seek international collaboration across the five work packages.

Whilst the focus of this strategic action is to fully exploit and extract maximum value from the Met Office NWP-based ensemble systems, so that they underpin all our services and support users and customers in their decision-making, we recognise the importance of engaging with external international key players. We are proactively exploring collaborative opportunities with major National Met Services, as well as multi-model approaches. Greater exploitation of ECWMF ensemble data is particularly important in this regard (including bringing ECMWF data into IMPROVER as part of our extension to 14-day forecasts), especially given the upcoming significant upgrades to the IFS.

EE-R2 The most challenging aspect of the ensemble exploitation initiative is likely to be changing the culture from a deterministic to a probabilistic/ensemble perspective. Excellent change management will be required. MOSAC suggests working with a professional change manager to achieve the successful adoption of probabilistic/ensemble thinking, especially amongst operational staff.

A dedicated change manager has just been appointed and linked to the comments above we recognise cultural change takes time, so realistic timescales are being considered for those aspects.

EE-R3 MOSAC noted the helpful classification of ensemble use cases into three types for ensemble statistics and three that describe the use of one or more member. The use-case decision tree looks like a useful tool. It will be important to apply appropriate evaluation and verification approaches for each of these use cases – this should be part of the ensemble exploitation framework.

We thank MOSAC for these remarks. Evaluation and verification approaches are essential and this links to NV-R5 and our response to that recommendation. As highlighted, we will need to use a range of subjective evaluation and verification with research into developing objective measures that capture the key value that is added with the various methods.

EE-R4 The Met Office already exploits ensembles for a number of applications such as coastal flooding risk, ensemble-driven planning products, and forecast compression for the military, as well as scenarios in DECIDER. Applications that have been developed quite independently for different customers may be able to be generated using common processing chains (some of which may be incorporated in IMPROVER) which would reduce the maintenance overhead and offer opportunities for re-use. To streamline the generation of ensemble-based products, efforts should be made to consolidate, where possible, independently developed ensemble application software.

Defining and developing common post-processing data streams and common tools is a key focus for the Met Office and the relevant teams are involved in this discussion. It is a complex area involving many teams with a significant role for our business architects. We are actively investigating these questions in developing further the national capability concept and, and the future scope of the PRISM platform.

EE-R5 The ECMWF and other ensembles provide global guidance that can be exploited in combination with MOGREPS to enhance the ensemble distribution, especially near the extremes. MOSAC encourages the Met Office to make greater use of multi-model ensembles in its products, including greater focus on their inclusion in IMPROVER.

We fully recognise the value of utilising ensemble predictions from other forecasting centres, especially ECMWF, and our operational meteorologists are already using this information in some aspects of their work. The near-term focus of our Ensemble Exploitation work has been around the utility of ensembles from Met Office systems, but we agree we should not lose sight of the different value multi-centre ensembles provide. As noted under EE-R1 we intend to include ECMWF data in IMPROVER post processing in the near future.

Plans for Future Capabilities

MOSAC commends the Met Office on its planning built around various infrastructure upgrade. Throughout this report, MOSAC has commented on future capabilities from the perspective of projects presented to us. In this part of the document, we thus only focus on a few items not discussed thus far.

PFC-R1 MOSAC recognizes that it is always difficult to see too far into the future and there is an associated risk of getting set too early on a path forward. MOSAC nonetheless suggest an increase in the thinking of what GEN3 could bring and the early identification of some of plausible major scientific and technical challenges that lay ahead.

Much of our recent focus has understandably been on Horizon B and how we will exploit the Gen1 supercomputer, but we are increasingly now considering the transition to Gen2 and our key goals for Horizon C, including global k-scale modelling, the roles of data science and the development of environmental digital twins. These goals will undoubtedly extend beyond Gen2, but as MOSAC recognizes, it is difficult to plan so far ahead, particularly given the pace of development of potentially disruptive technology changes. Nevertheless, this is important and we will be keen to continue to share our thinking at future meetings.

PFC-R2 As the Met Office presents planning for successively more capable parallel suites, we would like to see estimates for required computational and data resources to show no danger of exceedance or curtailment of targeted capability.

The planning of the NWP roadmap included estimates of what will be affordable computationally on one "quadrant" of the Gen1 supercomputer, and thus includes consideration for appropriate balance between research and operations. We generally aim for an R:O ratio of ~4 on average over the lifetime of a supercomputer and the NWP and climate plans have taken this into account from the start (including in the building of the initial science case which underpinned the business case).

In addition, we recognize it is too early to estimate costs of Gen 2 systems until this is better defined in the coming couple of years.

PFC-R3 Leading NMHS are all facing challenges around increasing complexity, explosion of products and services, and so on. Stepping through rigorous procedures to ensure robustness and skill, are increasingly weighing on the agility and ability to rapidly pull innovations through to operations. We encourage the Met Office to reflect on its procedures and exchange with other leading centres to optimize the Met Office's balance between agility and rigour.

We last shared our experience on this through a series of discussions with ECMWF, NOAA and several European Met Services roughly 4 to 5 years ago following the formation of the R2O team in Weather Science. We agree that we should continue to reflect on our approaches and revitalise these discussions, which is something we started during the R2O conference as part of the recent AMS meeting in Denver.

PFC-R4 With an archive migration of 60PiB lasting some 1.5 year, and associated costs (energy and money), MOSAC strongly encourages the Met Office to have an in-depth reflection around data stewardship of the future (e.g., archive period, curation, expiration, cross over points between benefits of archiving vs rerunning, etc).

The recent re-organisation of Science IT allows more focus on data management with a single group, Science IT Data Management and Processing. This group will implement more effective and efficient management of our data in the Met Office. The group set data strategy for effective use of data resources across platforms. One key aim for this new team is to have nearly all data categorized and covered by set policy that governs where it should be stored, and for how long. The early phases of this have begun with a goal to save time and money during the MASS migration.

PFC-R5 MOSAC agrees it seems reasonable to continue to explore the usefulness of an on-demand ensemble system at 300m. MOSAC would like more information on ongoing tests and when the Met Office expects that the national ensemble system could reach a similar resolution? MOSAC would like to know what a plausible next step might then be? We encourage the Met Office to reflect on the perturbation of ensembles at various scales and on metrics to evaluate ensembles, especially those at the highest resolution.

Initial experiments with a relocatable 300m grid-length modelling system are discussed in a recent paper (Bain et al., 2022). Before progressing to further tests of a relocatable ensemble, however, we need to assess the performance of the 300m forecasts further for some additional weather regimes where we would expect the enhanced resolution to add value (e.g. orographically-enhanced rainfall). We also need to decide how to measure the performance of these expensive ensemble simulations compared to intelligent post-processing of the coarser grid length ensembles (e.g. using high-resolution orography information to infer the horizontal distribution of the rain/snow boundary at the surface).

PFC-R6 The Met Office is rightfully considering the potentials that ML/AI can bring at every step of their value chain. We encourage the Met Office to reflect on data necessary to train emulators to ensure that strategies around other programmes, such as that around reforecast, and reanalysis are aligned.

The question of what data we have for calibration and training should be continually reassessed. We agree that training data for machine learning approaches is one use case that may lead to a change in our strategy. We will consider this as we continue to implement the Data Science Framework and assess our machine learning priorities. As explained under DS-R5 in our review of reforecasting in 2021 the main priority was to make more past data available from our forecasts.

Early-mid Career Scientists

MOSAC was impressed by the quality of the early-mid career scientists.

We are pleased that MOSAC was impressed by the quality of these presentations. As always, we feel that this session is an important highlight of the meeting, along with the poster session. Both are a tribute to the exceptional talent of our early-to-mid-career scientists and scientific software engineers.

- Introducing Met Office Data Ethics Framework for Environmental Data (Helen Bartlett)
  MOSAC sees this as an especially important topic in future, addressing a challenge that many organisations are struggling with. MOSAC is pleased that the Met Office understands that this is a challenge for all data handling, not just limited to new, "trendy" fields like machine learning. It is also gratifying that the Met Office is making use of links to other expert organisations like the Alan Turing Institute to make sure that they adopt best practice when handling data. This was an excellent and engaging presentation presenting an important piece of work.

- Interannual vs Decadal Response to the Pacific Decadal Variability (Melissa Seabrook)
  This was a great in-depth analysis of the relationship of the Pacific decadal variability and the predictability of the North Atlantic Oscillation.
  ECS-R1 We recommend that the Met Office take a closer look at the well documented and observed relationship between the NAO and MJO (and its convection and radiative processes) in relation with the radiative impacts identified in this study.

  We thank MOSAC for this suggestion. We are indeed interested in this topic and so even though sub-seasonal (monthly) forecasting is not currently a priority for the Met Office, we will try to build the MJO-NAO relationship into our long-range forecasting research plans.

- Trends in Global Marine Heatwaves and Cold Spells (Robert Peal and Mark Worsfold)
  The presentation on heatwaves and cold spells in the ocean was quite interesting and potentially of high interest to the blue economy. Figures were well chosen, and the level of complexity well-balanced given the varied breath of expertise of the audience. The topic was clearly introduced and resulting impacts on local ecosystems were discussed using two extreme events. We encourage continued efforts to explore the feasibility of developing forecasts for fisheries application.

- Development of a regional LFRic model (Christine Johnson).

The development of a regional LFRic model was a well-done demonstration of savoir-faire (e.g., Big Brother numerical experiments) in developing and testing numerical methods with clear steps for the forward plan. This is an important leap for LFRic.

- Fishing For an Audience: Using Storylines to Engage New Communities in Climate Change Discussions (James Pope et al.)
  The presentation on communicating about climate change was energetic and engaging. It was great to see ways in which the arts and deep community engagement are being used to help people understand and interact with climate change in a more personal and non-scientific way.