BAMS Article

Explaining and Predicting Earth System Change

A World Climate Research Programme Call to Action

Kirsten L. Findell, Rowan Sutton, Nico Caltabiano, Anca Brookshaw, Patrick Heimbach, Masahide Kimoto, Scott Osprey, Doug Smith, James S. Risbey, Zhuo Wang, Lijing Cheng, Leandro B. Diaz, Markus G. Donat, Michael Ek, June-Yi Lee, Shoshiro Minobe, Matilde Rusticucci, Frederic Vitart, and Lin Wang

ABSTRACT: The World Climate Research Programme (WCRP) envisions a world "that uses sound, relevant, and timely climate science to ensure a more resilient present and sustainable future for humankind." This bold vision requires the climate science community to provide actionable scientific information that meets the evolving needs of societies all over the world. To realize its vision, WCRP has created five Lighthouse Activities to generate international commitment and support to tackle some of the most pressing challenges in climate science today. The overarching goal of the Lighthouse Activity on Explaining and Predicting Earth System Change is to develop an integrated capability to understand, attribute, and predict annual to decadal changes in the Earth system, including capabilities for early warning of potential high impact changes and events. This article provides an overview of both the scientific challenges that must be addressed, and the research and other activities required to achieve this goal. The work is organized in three thematic areas: (i) monitoring and modeling Earth system change; (ii) integrated attribution, prediction, and projection; and (iii) assessment of current and future hazards. Also discussed are the benefits that the new capability will deliver. These include improved capabilities for early warning of impactful changes in the Earth system, more reliable assessments of meteorological hazard risks, and quantitative attribution statements to support the Global Annual to Decadal Climate Update and State of the Climate reports issued by the World Meteorological Organization.

KEYWORDS: Climate prediction; Climate models; Ensembles; Decadal variability; Climate services

https://doi.org/10.1175/BAMS-D-21-0280.1
Corresponding author: Kirsten L. Findell, kirsten.findell@noaa.gov
In final form 7 September 2022

AFFILIATIONS: Findell—NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey; **Sutton**—National Centre for Atmospheric Science and University of Reading, Reading, United Kingdom; **Caltabiano**—WCRP Secretariat, World Meteorological Organization, Geneva, Switzerland; **Brookshaw and Vitart**—European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom; **Heimbach**—University of Texas at Austin, Austin, Texas; **Kimoto**—National Institute for Environmental Studies, Tsukuba, Japan; **Osprey**—National Centre for Atmospheric Science, Leeds, and University of Oxford, Oxford, United Kingdom; **Smith**—Met Office Hadley Centre, Exeter, United Kingdom; **Risbey**—Oceans and Atmosphere, Commonwealth Scientific and Industrial Research Organisation, Hobart, Tasmania, Australia; **Z. Wang**—University of Illinois at Urbana–Champaign, Urbana, Illinois; **Cheng and L. Wang**—Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China; **Diaz**—Departamento de Ciencias de la Atmósfera y los Océanos, Facultad de Ciencias Exactas y Naturales, and Centro de Investigaciones del Mar y la Atmósfera, Universidad de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and Instituto Franco-Argentino para el Estudio del Clima y sus Impactos, Buenos Aires, Argentina; **Donat**—Barcelona Supercomputing Center, and Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain; **Ek**—National Center for Atmospheric Research, Boulder, Colorado; **Lee**—Research Center for Climate Sciences, Pusan National University, and Center for Climate Physics, Institute for Basic Science, Busan, South Korea; **Minobe**—Hokkaido University, Hokkaido, Japan; **Rusticucci**—Departamento de Ciencias de la Atmósfera y los Océanos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and Instituto Franco-Argentino para el Estudio del Clima y sus Impactos, Buenos Aires, Argentina

The formulation of robust policies for mitigation of, and adaptation to, climate change requires quantitative understanding of how and why specific changes are unfolding in the Earth system, and what might happen in the future. Quantitative explanation of observed changes—through robust process-based detection and attribution—is also fundamental to specification of confidence in climate assessments, predictions, and projections. However, the capacity to deliver these capabilities is very limited, particularly for the annual to decadal (A2D) time scales that lie between the time scales of days and seasons—the focus of numerical weather prediction (NWP) and seasonal forecasting—and the multidecadal-to-century time scales that are the primary focus of climate projection efforts. The World Climate Research Programme (WCRP) Lighthouse Activity (LHA) on Explaining and Predicting Earth System Change (EPESC) is intended to address this need. We adopt the nomenclature A2D to define our time scales of interest to be consistent with the community working with initialized predictions at subseasonal to seasonal (S2S), seasonal to interannual (S2I), and seasonal to decadal (S2D) time scales (Meehl et al. 2021).

Given the current nonstationarity of the climate system and the limited sampling of extreme events in our global observational records, climate statistics and probabilities of hazards and extremes based on past observations are no longer adequate for infrastructure or disaster planning (Milly et al. 2008). Indeed, in a changing climate, understanding the development and precursors of extreme events, attributing causal factors, and determining the impacts of background conditions on the likelihood of event occurrence is crucial (Stott et al. 2011). Actionable predictions and risk assessments require full appraisal of all the relevant uncertainties, including those stemming from uncertainties in observational records, from forcings and climate responses, from internal variability, from climate model structural differences, and from interactions between each of these sources of uncertainty (Hawkins and Sutton 2009;

Frölicher et al. 2016; Lovenduski et al. 2016; Marzeion et al. 2020; Lehner et al. 2020; Aschwanden et al. 2021). Understanding and quantifying these uncertainties is particularly challenging for small regions and A2D time scales, yet information about these spatial and temporal scales is needed to inform adaptation.

Decadal time scales were targeted by the WCRP Grand Challenge on Near-Term Climate Prediction (Kushnir et al. 2019), where the authors highlighted the dual dependence on natural climate variability and anthropogenically imposed climate change. As an outgrowth of this large-scale international effort, multiannual forecasts are now routinely issued by the World Meteorological Organization (WMO) Lead Centre for Annual to Decadal Climate Prediction and in the WMO Global Annual to Decadal Climate Update (GADCU; Hermanson et al. 2022). However, improved understanding and attribution of predicted signals is needed to gain further confidence in the forecasts and to gain insight on how to improve these forecasts. In addition, the WCRP Grand Challenge on Weather and Climate Extremes (Zhang et al. 2014) targeted the improved understanding of climate-related hazards. This was organized around four overarching themes, to document, understand, simulate, and attribute such extremes. This LHA aims to build on the earlier efforts of the now-complete Grand Challenges by first establishing and applying attribution methodologies to help explain A2D changes in the climate system and their influence on hazards (including extremes), while also evaluating the requirements needed to fully observe and model these changes. Additional effort will be directed toward defining the elements of an operational capability that integrates attribution and prediction methods to better understand and predict climate hazards on A2D time scales. Outputs of these efforts will enhance the value of A2D climate forecasts issued by WMO.

The overarching objective of the WCRP Lighthouse Activity on EPESC is *to design, and take major steps toward delivery of, an integrated capability for quantitative observation, explanation, early warning and prediction of Earth system change on global and regional scales, with a focus on annual to decadal time scales*.

On global to regional and A2D scales, changes in oceanic and atmospheric circulation and their consequent impacts are of particular interest because of their importance in shaping hazards, and because current capabilities to explain and predict changes in circulation are particularly limited. Some examples of changes of interest include the rapid warming of the North Atlantic Ocean that occurred in the 1990s (e.g., Robson et al. 2012; Yang et al. 2016; Cheng et al. 2017; Yeager 2020), weakening of the North Atlantic subpolar gyre (Häkkinen and Rhines 2004; Piecuch et al. 2017), changes in the phase of the interdecadal Pacific oscillation (e.g., Thoma et al. 2015; Meehl et al. 2016), persistent marine heatwaves such as in the North Pacific during 2013–16 (e.g., Di Lorenzo and Mantua 2016; Oliver et al. 2018), persistent droughts such as in the Sahel during the 1970s and 1980s (e.g., Held et al. 2005), and the apparent slowdown in global mean surface temperature rise that was observed in the 2000s (e.g., England et al. 2014; Fyfe et al. 2016). This last example is a particularly fitting case study of how natural decadal variability on top of long-term trends can combine to produce a long-lasting signal that can capture both research and public attention (Fyfe et al. 2016; Risbey et al. 2018).

This LHA is concerned both with events that have A2D duration and also with understanding how regional and larger-scale changes (e.g., broad atmospheric or oceanic circulation changes) on these time scales influence the characteristics of hazards (e.g., severe convective storms, tropical and extratropical cyclones, atmospheric rivers, terrestrial and marine heat waves, wildfires) occurring on shorter space and time scales. Examples of A2D variability influencing hazards can be found in the impact of Atlantic multidecadal variability on tropical cyclones in the Caribbean basin (Goldenberg et al. 2001) or of El Niño–Southern Oscillation (ENSO) on droughts in the United States (e.g., Trenberth et al. 1988; Schubert et al. 2009; Findell and Delworth 2010) or on fire weather in Australia (Squire et al. 2021) and their secondary impacts (Damany-Pearce et al. 2022). Physical predictions on A2D time scales can be useful even for marine biological forecasting (Minobe et al. 2022).

Given the breadth of the targeted goals, we have found it useful to organize the scientific challenges and opportunities around three major themes with associated working groups:

- theme 1: monitoring and modeling Earth system change;
- theme 2: integrated attribution, prediction, and projection of Earth system change; and
- theme 3: assessment of current and future hazards.

Figure 1 provides an overview of the three scientific themes, how they interact, and how they will deliver benefits to society. Expertise in many areas relevant to these themes is found in many of WCRP's core projects and other Lighthouse Activities. Active communication between EPESC and other WCRP entities is crucial to the success of this endeavor.

Three cross-cutting dimensions connect the work of the LHA's three thematic elements. First, the development of a capability to observe, explain, and predict changes in the Earth system requires the *tight integration of observations and models*, including characterization and quantification of uncertainties. Comprehensive model calibration and evaluation of model skill each require observational datasets that capture the phenomena of interest, but also computational frameworks for achieving rigorous model calibration. Just as observations can be used to confront models, calibrate model parameters, and determine model skill, models can be leveraged as tools to inform the design of efficient, targeted observing systems (e.g., Fujii et al. 2019; Cheng and Zhu 2016). We envision an interactive workflow between model and observing system improvement, as both represent incomplete yet

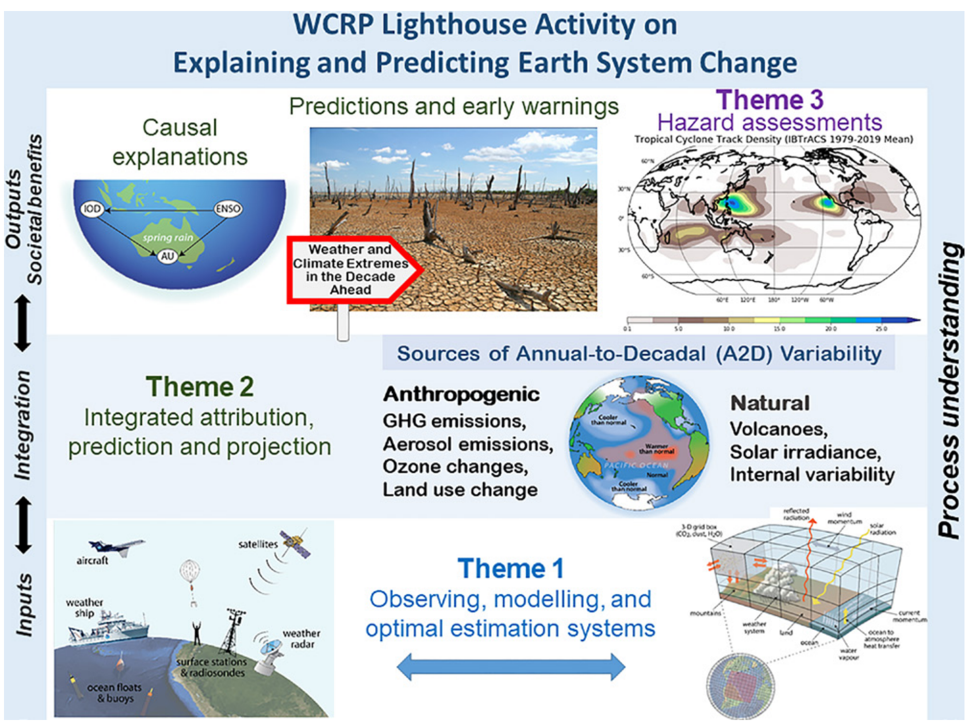

Fig. 1. Key elements of the Lighthouse Activity. The bottom layer shows the importance of coordinated observational and modeling efforts serving as key tools and inputs to the integrated attribution, prediction, and projection efforts in the middle layer. Both of these layers feed into the outputs and societal benefits displayed in the top layer: causal explanations, predictions and early warnings, and hazard assessments. Arrows along the left side indicate that outputs (themes 2 and 3) and integration (theme 2) can feedback to improve the inputs (theme 1). Fundamental physical process understanding runs through all aspects of the Lighthouse Activity. [Causal explanations figure following Kretschmer et al. (2021).]

complementary knowledge bases. Similarly, identification of causal factors and processes leading to large-scale climate regime shifts or changes in regional hazard risk require integrated usage of both observations and modeling systems.

Second, initial steps to develop a capability to observe, explain, and predict Earth system change will focus on a few (two or three) compelling *case studies* targeting climate "events" that have occurred in recent decades, such as the examples given above. Through these case studies, we seek to develop a systematic approach across all three themes to identify causal factors shaping these events, to assess the potential for predictions of the events themselves, to investigate opportunities for observations targeted at realizing predictability of the events, and, where relevant, to determine the impact of the event on hazard likelihoods.

Finally, we envisage that *large ensembles of single-forcing experiments* will inform the activities at the heart of each of the themes. These are essential to characterize the responses to different forcing factors, thereby informing observing system design (theme 1), providing quantitative process-based attribution (theme 2), and improving our understanding of the drivers of changing hazard frequencies and intensities (theme 3).

Theme 1: Monitoring and modeling Earth system change

Key research questions.

1) What are the observational and modeling requirements to measure, explain, and predict changes in the Earth system on A2D and regional to global scales?
2) How can we most effectively combine observations and models to quantify, explain, and predict changes in the Earth system on A2D and regional to global scales?
3) Which enhanced observations will offer the greatest improvements in predictive and explanatory skill, and where should those enhancements be targeted?

The Global Climate Observing System (GCOS) has developed over many decades, with a steady expansion of the spatial coverage and physical quantities recorded, punctuated by major advances in funding allocations and/or observing technologies (e.g., GCOS 2021). For example, though satellite observations of environmental quantities began in the late 1950s, the beginning of the satellite era is commonly recognized as 1979, when microwave measurements were included in NOAA weather satellites, enabling measurement of tropospheric temperature (Thorne et al. 2010) and sea ice cover in polar regions (Parkinson 2019). Similar expansions occurred with the beginning of the era of satellite altimetry for monitoring sea level change in 1992/93 (Fu et al. 2019). Observations of upper ocean temperature for climate monitoring relied on hydrographic measurements from research vessels including buckets and moorings (sea surface), Nansen bottles, mechanical bathythermographs (MBTs) (upper 250 m), and expendable bathythermographs (XBTs) (upper 700 m) since the late-nineteenth century (Abraham et al. 2013). During the 1990s, increased subsurface measurement coverage of the global ocean using high-quality hydrographic sections was achieved as part of the World Ocean Circulation Experiment (WOCE; Ganachaud and Wunsch 2000). Subsequently, since around 2005, the Argo array began to dominate ocean observations, measuring the upper 2,000 m at unprecedented resolution (Johnson et al. 2022). Plans for deep ocean Argo measurements are well-developed (Roemmich et al. 2019), and would provide improved capabilities for model initialization and verification for A2D understanding and applications (Meehl et al. 2021).

Simultaneous development of global climate modeling capabilities over more than 60 years has seen similar gradual improvements in complexity, resolution, and skill, with occasional step changes in both theoretical understanding and computational capacity (e.g., Forster 2017; Manabe and Broccoli 2020; Balaji 2021). These largely separate (though interdependent) efforts have covered enormous ground and helped the climate science community substantiate the unequivocal human influence on climate (IPCC 2021; Hegerl 2022). However, for the near- and long-term climate-related challenges the world now faces, tighter integration between the global climate observing system and the climate modeling community is necessary to address several interrelated obstacles. The joint consideration of observation and modeling challenges (Fig. 1, bottom panel) provides a conceptual framework for identifying major gaps and opportunities for progress in observing, monitoring, and modeling Earth system variability and change.

Fundamental to this tighter integration is the need to better understand the observational and modeling requirements to measure, explain and predict changes in the Earth system on A2D and global to regional scales, as well as the current limitations on these capabilities. These requirements and limitations are certainly case specific, but through the initial case studies discussed above, we aim to develop a systematic methodology that can be applied subsequently to assess a wider set of events and address a number of scientific questions. Foremost among these, How early were these events recognized as significant and how well were they monitored by different elements of GCOS? Additionally, we will need to assess how well models, analysis, and reanalyses represented these events. Case studies will also prove useful for determining how well observations constrained the underlying metrics (e.g., regional versus global ocean heat content anomalies; global mean values as small residuals of large regional variations; climate anomalies at the margins of the polar ice sheets), and if current observations allow for a mechanistic understanding of the propagation or evolution of relevant anomalies. In particular, these case studies will be testbeds to determine if observations were sufficient to provide coverage of "upstream" or precursor processes that led to the events of interest. Many potential case studies are active areas of research and may require additional investigation to determine which measurable quantities are the most relevant upstream indicators. With this in mind, we seek to develop methods that could inform quantitative observing system design, targeting scales relevant to EPESC goals, and help objectively determine *what climate indices are to be measured? What measurements constrain such indices? Where, and over what time horizons should they be measured? How many observations are sufficient? What are optimal combinations of different observing networks (satellite and in situ)?* The observing networks that operate under GCOS and the Global Ocean Observing System (GOOS 2020) can play a major role in this approach. For such case studies, the value of these networks could be assessed, major gaps (as well as potential redundancies) identified, and observational requirements formulated.

While addressing observational limitations on Earth system understanding, we can also tackle persistent Earth system model and reanalysis biases through the use of comprehensive estimation methods that bring modeling and (re-)analysis closer together and lead to better usage of the diverse, heterogenous observing networks underlying the in situ ocean, terrestrial, and atmospheric networks, in addition to satellite capabilities. This necessarily touches on the need to harness and improve data assimilation (DA) efforts, viewed more broadly as parameter estimation or inference methods, and objective analysis procedures. This opportunity for a "DA for climate" initiative is being approached in partnership with the Digital Earths Lighthouse Activity. This collaborative effort allows for a focus on major climate-specific needs (e.g., initial condition estimation versus model parameter calibration) and issues that might not receive much attention in other data assimilation applications (e.g., conservation laws and other physical constraints that are key on climate time scales, but not of primary concern in NWP). In addition, the exploration of synergies between data assimilation and machine learning concepts will be beneficial (e.g., Schneider et al. 2017; Abarbanel et al. 2018; Ham et al. 2019; Gordon et al. 2021).

A parallel issue relates to the data available for assimilation. We currently have both sparse observational sampling of various elements of the Earth system, and an underutilization of the wide array of observational data which are collected. Underobserved variables allow for errors to be hidden during assimilation of observed quantities (e.g., altering soil moisture when assimilating near-surface temperature and humidity; Mahfouf et al. 2009). Model calibration efforts could benefit from expanded use of GCOS and GOOS observations, and at the same time, optimization techniques could be harnessed to identify regions of Earth where enhanced observations could offer substantive improvements in predictive and explanatory skill (e.g., Hakim et al. 2020) or reduce uncertainty in chosen climate indices (e.g., Loose and Heimbach 2021). All of this must include novel approaches for dealing with the combined stream of uncertainties from observations and models.

Another important objective is the development of calibration and uncertainty quantification (UQ) strategies for existing or emerging models, observations, and methodologies (e.g., data assimilation). Quantifying uncertainties of global and regional changes in relevant climate metrics, based either on observations, models, or synthesis/data assimilation products, remains a great challenge, in part because of the computational complexity of the underlying problem (e.g., Oden et al. 2010, for a general perspective). Inspired by examples described in Bui-Thanh et al. (2012), Kalmikov and Heimbach (2014), Schneider et al. (2017), Loose et al. (2020), and Aschwanden et al. (2021), an activity should develop frameworks and workflows that will account jointly for uncertainties in observations (instrument error, representation error, sampling, etc.), models (parametric errors, structural model errors, etc.), and assimilation strategies (to the extent that they exist) into comprehensive uncertainty propagation flows that seek to combine these error sources and, for example, propagate them onto specific target metrics relevant to climate diagnostics. Ensemble methods used in the climate modeling community would benefit from such a systematic approach, both at the point of ensemble generation and—even more so—when using observations to “constrain” the ensemble, or to reduce uncertainty in these calculations. Coordination with efforts that focus on developing and building communities for novel modeling approaches, such as the Digital Earths LHA, will be crucial.

Theme 2: Integrated attribution, prediction, and projection of Earth system change

Key research questions.

1) How can we best identify and attribute the drivers of changes in the Earth system on global to regional and A2D scales?
2) What are the requirements for an operational integrated attribution and prediction capability focused on global to regional and A2D scales to provide early warnings to inform decision making?

On A2D time scales, climate is influenced by many factors, including internal variability and external forcing from greenhouse gases, aerosols, ozone, solar variations, volcanic eruptions, and land-use changes (Cassou et al. 2018; Kushnir et al. 2019; Merryfield et al. 2020). Climate model simulations are essential to disentangle the relative roles of these different factors. Promising results demonstrate that there is initial state skill extending into A2D time scales (Meehl et al. 2021, and references therein). At the same time, other evidence shows that even if most of the A2D time scale skill is coming from external forcing, the initial state of the climate system can substantively impact the forcing trajectory (e.g., Bordbar et al. 2019). However, climate models are imperfect, with issues of model bias and drift posing challenges for A2D predictions (Meehl et al. 2022). Additionally, currently available simulations do not take into account the latest estimates of, and uncertainties in, the various radiative forcings. Developing a prototype operational attribution capability therefore requires two initial stages:

1) Critical assessment of the ability of models to simulate the full range of relevant internal variability and responses to radiative forcings. A key outcome of this stage will be recommended strategies to eliminate, reduce, or adjust for model errors.
2) Operationalization of attribution simulations using the latest estimates of radiative forcings and uncertainties, and application of corrections diagnosed in stage 1.

Taken at face value, large ensemble historical simulations suggest a dominant role for irreducible internal variability in regional climate change on decadal time scales (Deser et al. 2020). However, there is mounting evidence that climate models may underestimate atmospheric circulation signals in subseasonal (Domeisen et al. 2020; Charlton-Perez et al. 2019), seasonal (Eade et al. 2014; Scaife et al. 2014; Baker et al. 2018; Lee and Ha 2015), interannual (Dunstone et al. 2016), and decadal (Athanasiadis et al. 2020; Smith et al. 2020) predictions, and in historical simulations (Lee et al. 2014; Zhang and Kirtman 2019; Sévellec and Drijfhout 2019; Klavans et al. 2021; Zhang et al. 2021). This error is especially clear in the North Atlantic, although there is some ongoing debate about the potential role of nonstationarity and sampling issues (Christiansen et al. 2022; Weisheimer et al. 2019). On decadal time scales it appears to occur in most regions where there is skill for atmospheric circulation (Smith et al. 2019, 2020). Figure 2 illustrates this error for decadal predictions of the North Atlantic Oscillation (NAO). The left panel shows that the ensemble mean has little signal and high uncertainty. However, there is high correlation (0.79) between the forecast ensemble mean and the observations such that ensemble mean forecasts scaled to match the observed variance more closely follow the observed changes over this period (Fig. 2, right panel). The mismatch between the high correlation and small signal of the ensemble mean occurs because the models underestimate the predictable signal by an order of magnitude. Attribution of A2D changes in climate is therefore complicated by the possibility that models may not properly represent the relative roles of internal variability and external factors (Scaife and Smith 2018), and due to the difficulties of providing robust statistical verification for decadal forecasts (Christiansen et al. 2022; Weisheimer et al. 2019).

Understanding the causes of the signal-to-noise problem, and improving models so that the problem does not arise, are major long-term challenges. In the meantime, given the current landscape of model capability, the proposed way forward for this LHA is to diagnose the response to individual forcing factors from the mean of large ensembles, and then to assess their relative roles (and their additivity) by scaling to reconstruct the observed historical

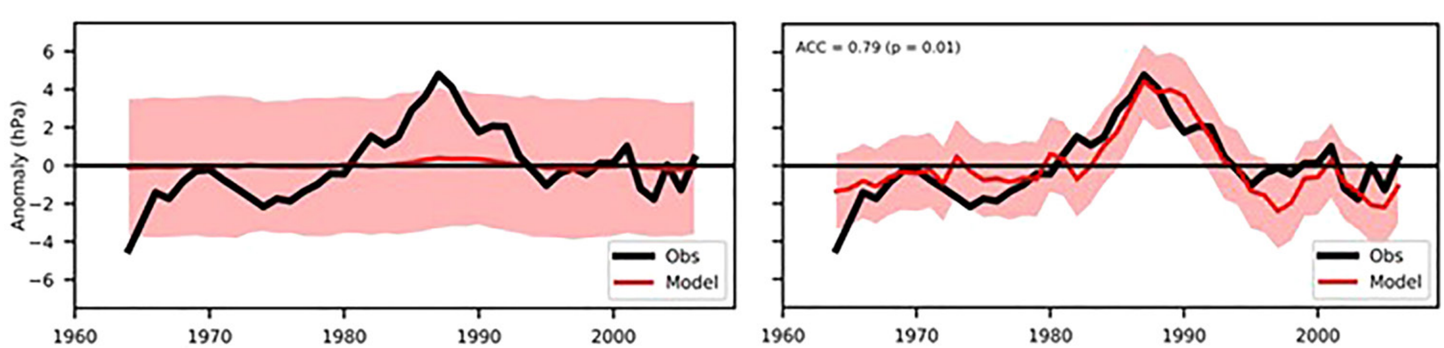

Fig. 2. Decadal predictions of the NAO. (left) Observed (black) and model forecast (years 2–9; red) 8-yr running mean boreal winter NAO index (hPa). The red curve shows the ensemble mean; the red shading shows the 5%–95% confidence interval diagnosed from the individual members. (right) As in left panel, but the ensemble mean has been adjusted to have the same variance as the observations and the confidence interval has been diagnosed from the errors. Adapted from Smith et al. (2020).

record and treating the residual as internal variability. Such single forcing experiments have been proposed by the Detection and Attribution Model Intercomparison Project (DAMIP; Gillett et al. 2016). However, these experiments are generally low priority and modeling centers have either not completed them, or produced only a few ensemble members. *Hence, a key objective of theme 2 is to develop large ensembles of single forcing historical simulations* (LESFs). A LESF Model Intercomparison Project (LESFMIP) based on this initiative is detailed in Smith et al. (2022), with tier 1 experiments detailed in Table 1 below. Multiple modeling centers have already committed to producing these large ensembles. Additional centers are welcome and would not be required to contribute all of the experiments if the computational demands are too high. These LESFMIP experiments are expected to provide information on model behavior that will feed back on the model development process. Ideally, these experiments will also inform the theme 1 activities designed to identify high-priority regions for expansion of observational networks, though the complexity of this task should not be underestimated.

Analysis of the LESFs will provide scaling factors for the different forcings (Table 1) and hence corrections for the model simulations. A key part of the analysis of LESFMIP will be to exploit differences between the models to diagnose the real-world situation. Hence, multimodel simulations are essential, though understanding the causes of model errors and developing emergent constraints will be a significant challenge. This will likely involve detailed analysis of recent case studies, assessment of observational and forcing uncertainties alongside model biases, and exploration of possible nonlinear interactions between the responses to different forcings. Initial analysis will likely focus on explaining A2D changes in sea surface temperatures (SSTs) in the Atlantic, Pacific, and Indian Oceans, with a goal of providing initial attribution statements to upcoming WMO reports on the State of Climate and GADCU (Hermanson et al. 2022). Subsequent efforts will go beyond this analysis of SST changes and focus on their associated impacts (e.g., tropical and extratropical cyclones, droughts, wildfires, marine heatwaves).

In order for an operational system of attribution simulations to produce measures of the relative importance of different forcing factors in the observed changes shortly after their occurrence, real-time estimates of individual forcing factors, together with their observational uncertainties, will be required. Theme 2 will therefore seek to identify annually updated sources of forcing information. There will also be a need for research to explore how results from near-real-time attribution can be used to constrain and improve decadal predictions. This could involve, for example, exploring the sensitivity of predictions to the modification or exclusion of individual forcing factors, or sampling large ensembles to match recent observations (Sparrow et al. 2018; Mahmood et al. 2022). Overall, this new integrated approach to attribution and prediction promises to provide a step change in our

Table 1. Large ensemble single forcing experiments. All experiments to cover the time period 1850–2020, and then to be extended with real-time estimates of radiative forcings. All experiments are part of DAMIP (Gillett et al. 2016) except for hist-LU. Target ensemble size is 50 members for all simulations, with a minimum of 10 members. Tier 1 experiments are listed here; see Smith et al. (2022) for further details of experiments and analysis plan.

Experiment name	Description
hist-GHG	Well-mixed greenhouse gas–only historical simulations (WMGHGs)
hist-aer	Anthropogenic-aerosol-only historical simulations (BC, OC, SO_2, SO_4, NO_x, NH_3, CO, NMVOC)
hist-sol	Solar-only historical simulations (solar irradiance)
hist-volc	Volcanic-only historical simulations (stratospheric aerosol)
hist-totalO3	Ozone-only historical simulations (stratospheric and tropospheric ozone)
hist-lu	Historical simulations with only land-use changes

understanding of the drivers of A2D climate changes and in our ability to provide early warnings for decision-making.

Theme 3: Assessment of current and future hazards

Key research questions.

1) How do internal variability and external forcings influence the characteristics and occurrence of meteorological hazards on A2D scales in different regions?
2) How can we use observations, models, and process understanding to deliver robust assessments of current and future hazards for specific regions and hazard classes?

Climate hazards and disasters are increasingly costly to human lives and livelihoods, with the best estimates for 2021 alone indicating roughly 10,500 lives lost and $343 billion (U.S. dollars) in worldwide economic losses (Aon 2022). Given the enormity of those numbers, improved understanding of the causal factors influencing a wide range of meteorological hazards and improved predictions of such hazards merits substantial investments in climate science. As such, a key goal of this LHA is to better understand, quantify, and predict changes in the characteristics and likelihoods of regional weather and climate hazards on A2D scales, taking into account nonstationarity and multidecadal variability, particularly on large scales like those associated with ENSO or the Atlantic meridional overturning circulation (AMOC). Hazards of interest include tropical and extratropical cyclones, droughts, floods, heatwaves, wildfires, and cold air outbreaks. The theme 3 portion of Fig. 1 (upper-right frame) indicates that tropical cyclone (TC) frequency is a critically important quantity for this LHA to assess, in part because of the documented regional dependence of TCs on A2D variability (e.g., ENSO; Lin et al. 2021; the Atlantic meridional overturning circulation; Dunstone et al. 2011; Smith et al. 2010). Figure 3 provides an additional example of a relevant hazard metric assessing the changing risk of fire weather days across the globe. We aim to quantify the current likelihood of specific weather and climate hazards, as well as changes in weather and climate hazards on A2D scales. A key component of quantifying those changes must be improved understanding of the processes connecting changes in hazards to natural and anthropogenic drivers of climate variability and change. Each climate hazard brings its own particular requirements

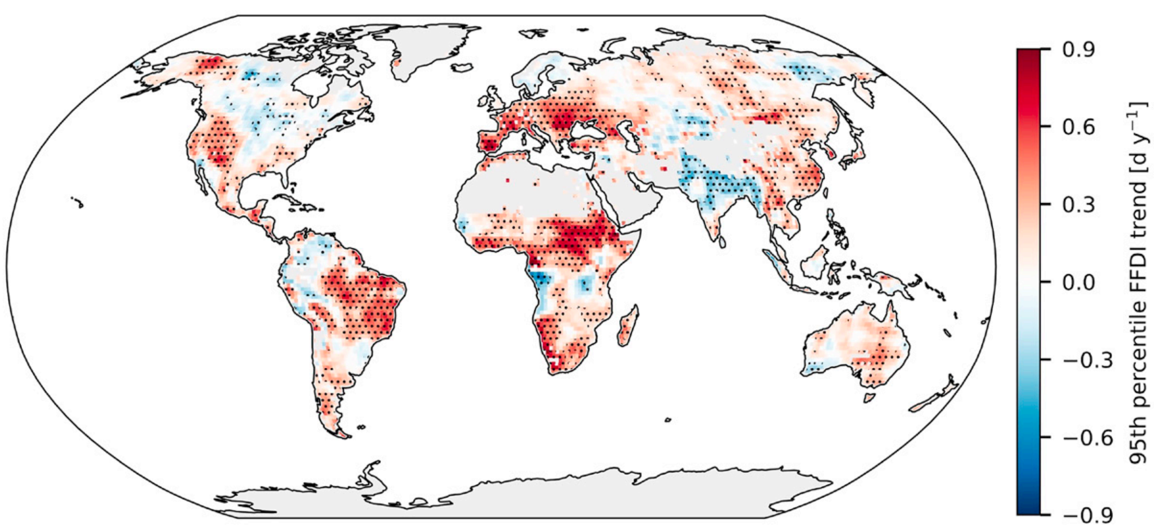

Fig. 3. Trend in the number of days per year between 1970 and 2020 for which the forest fire danger index (FFDI) exceeds the climatological 95th percentile, adapted from Richardson et al. (2022). Based on data from the reanalysis product JRA-55 (Japan Meteorological Agency 2013). Shading is the Theil–Sen slope. Stippling indicates a statistically significant result from the Mann–Kendall trend test, treated for multiple testing and autocorrelation. See Richardson et al. (2022) for details.

in terms of the strength of observational data and modeling underpinning current understanding, and in terms of the complexity of the hazard in human systems. As such, dedicated analysis is required for each type of climate hazard. Our initial efforts will be focused on TCs because of their high salience for climate impacts (Fig. 1, theme 3 highlight, top-right corner).

There are many knowledge gaps impeding the quantification of the impacts of natural and anthropogenic drivers on hazards. This is in part due to the limited length and limited spatial coverage of reliable observational records and the relatively rare occurrences of many types of hazards. These factors make it challenging to identify statistically significant trends and to distinguish internal variability from responses to external forcings. Our current capability to explain hazard changes is also limited due to a lack of process understanding about how drivers of large-scale changes may affect hazards, making this a key focal point of this theme. One example is the debate on whether Arctic warming impacts midlatitude extreme weather (e.g., Barnes and Screen 2015; Blackport et al. 2019). In general, hazards or weather/climate extremes are often regarded as the tail of the distribution of a climate variable, and strong observational or theoretical constraints do not exist for most types of hazards.

Our current capability to predict and project hazard changes is limited due to several factors, many of which have already been touched upon in discussion of the other themes. First, many types of hazards are related to mesoscale, or even convective-scale processes (such as TCs and tornadoes), or are closely tied to the coupling between different components of the climate system (such as the role of land–atmosphere interactions in droughts or heat waves), neither of which are adequately represented in most global models. Additionally, even if a model skillfully predicts the mean regional change, there are often large biases in the regional distribution of hazards, leading to deficiencies in capturing the impacts of climate variability or changes in these hazards. Model biases, sparse observations, and technical difficulties also present major challenges to the comprehensive calibration and balanced initialization of coupled prediction systems (theme 1). Furthermore, as discussed above, large intermodel spread exists in predicted and projected climate changes on the regional scale, due to differences in model formulation (e.g., physics parameterizations and resolution), and signals related to anthropogenic forcing are often weak compared to internal variability (theme 2). Finally, computationally demanding large ensembles are required to adequately sample rare events. Use of the LESFs discussed in theme 2 will allow us to improve our ability to quantify current and future risk of hazards, and attribute changes in hazard risk to internal or external climate drivers. These efforts in combination will enable us to assess the predictability and uncertainty of changes in hazard risk.

Additionally, a range of experimental designs will be useful to make progress with the challenges in this theme, including hindcast datasets to quantify current risk (e.g., Thompson et al. 2017; Squire et al. 2021); coupled single-forcing experiments (as described in theme 2); large ensemble atmospheric general circulation model (GCM) experiments, possibly including regional downscaling (e.g., Mizuta et al. 2017; Imada et al. 2020); and targeted nudging and/or pacemaker experiments (e.g., Kosaka and Xie 2013; Watanabe et al. 2014). While atmospheric GCM experiments cannot fully address predictability and fixed SST experiments can provide biased estimates of changes in climate extremes (Fischer et al. 2018), they can nonetheless be helpful to provide mechanistic understanding of the natural and anthropogenic contributions to changing hazards, such as the regional risk of heavy precipitation (Imada et al. 2020).

A gap also exists between the research and user communities regarding A2D prediction products that are useful, usable, beneficial, and feasible to produce, especially at regional scales. For example, although research has demonstrated the predictability of basinwide statistics of TCs (e.g., Smith et al. 2010; Dunstone et al. 2011; Caron et al. 2018), users are often more interested in *landfalling* TC statistics. With increasing computing power, improved climate models, and a better understanding of A2D predictability sources, skillful predictions of landfalling TC statistics are achievable over some basins (Chang and Wang 2020). In addition, there is likely a middle ground, and codesign between researchers and users is needed for it to be identified. These overlapping challenges highlight the benefit of combining the proposed activities under the broad umbrella of this LHA.

Conclusions

The WCRP Strategic Plan (WCRP JSC 2019) for the coming decade highlights four scientific objectives, three of which relate directly to the objectives of this Lighthouse Activity. *Fundamental understanding of the climate system* (objective 1) and *prediction of the near-term evolution of the climate system* (objective 2) are at the heart of this LHA's effort to explain and predict annual to decadal-scale Earth system change. Furthermore, this LHA will ensure that advances in fundamental understanding of Earth system change are targeted to meet the needs of decision-makers facing climate-related risks and opportunities. Societal benefits to be delivered by this LHA include early warning of significant global- and regional-scale changes in the climate system, and quantification of current and future hazard risk on regional scales. The benefits of this new actionable information will be enhanced through codevelopment with diverse stakeholders (e.g., governments, businesses, public), and thereby offer a major contribution to WCRP's efforts to *bridge climate science and society* (objective 4).

WCRP's new Lighthouse Activities constitute a bold effort to tackle some of the most persistent and difficult issues in climate science today. These efforts will require close collaboration with many different groups within WCRP and beyond to undertake a full, integrated assessment of our observational and modeling capabilities that cross component (ocean, atmosphere, land, ice) and disciplinary boundaries, and help push forward the capabilities of explanation, prediction, and uncertainty quantification for annual to decadal time scales. Successfully addressing these issues calls for collaboration and coordination of climate scientists around the world and will require support from funding agencies that is commensurate with the magnitude of the task at hand.

Acknowledgments. DS was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra and by the European Commission Horizon 2020 CONSTRAIN project (GA 820829). SO acknowledges support from the U.K. Natural Environment Research Council and funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement 101003469. Thanks to Cathy Raphael and Jiacheng Ye for contributions to Fig. 1, and to Doug Richardson for contributing Fig. 3. Thanks, too, to Tom Delworth and Feiyu Lu for helpful and insightful reviews of the manuscript. This article has been published with the financial support of the World Meteorological Organization (WMO). The opinions, findings, interpretations, and conclusions expressed in this article are those of the authors and do not purport to reflect the opinions of WMO or its members.

Data availability statement. No datasets were generated or analyzed during the current study.

References

Abarbanel, H. D. I., P. J. Rozdeba, and S. Shirman, 2018: Machine learning: Deepest learning as statistical data assimilation problems. *Neural Comput.*, **30**, 2025–2055, https://doi.org/10.1162/neco_a_01094.

Abraham, J., and Coauthors, 2013: A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Rev. Geophys.*, **51**, 450–483, https://doi.org/10.1002/rog.20022.

Aon, 2022: 2021 weather, climate and catastrophe insight. Aon Rep., 53 pp., www.aon.com/weather-climate-catastrophe/index.html.

Aschwanden, A., T. C. Bartholomaus, D. J. Brinkerhoff, and M. Truffer, 2021: A roadmap towards credible projections of ice sheet contribution to sea level. *Cryosphere*, **15**, 5705–5715, https://doi.org/10.5194/tc-15-5705-2021.

Athanasiadis, P. J., S. Yeager, Y.-O. Kwon, A. Bellucci, D. W. Smith, and S. Tibaldi, 2020: Decadal predictability of North Atlantic blocking and the NAO. *npj Climate Atmos. Sci.*, **3**, 20, https://doi.org/10.1038/s41612-020-0120-6.

Baker, L. H., L. C. Shaffrey, R. T. Sutton, A. Weisheimer, and A. A. Scaife, 2018: An intercomparison of skill and overconfidence/underconfidence of the wintertime North Atlantic Oscillation in multi-model seasonal forecasts. *Geophys. Res. Lett.*, **45**, 7808–7817, https://doi.org/10.1029/2018GL078838.

Balaji, V., 2021: Climbing down Charney's ladder: Machine learning and the post-Dennard era of computational climate science. *Philos. Trans. Roy. Soc.*, **A379**, 20200085, https://doi.org/10.1098/rsta.2020.0085.

Barnes, E. A., and J. A. Screen, 2015: The impact of Arctic warming on the mid-latitude jet-stream: Can it? Has it? Will it? *Wiley Interdiscip. Rev.: Climate Change*, **6**, 277–286, https://doi.org/10.1002/wcc.337.

Blackport, R., J. A. Screen, K. van der Wiel, and R. Bintanja, 2019: Minimal influence of reduced Arctic sea ice on coincident cold winters in mid-latitudes. *Nat. Climate Change*, **9**, 694–704, https://doi.org/10.1038/s41558-019-0551-4.

Bordbar, M. H., and Coauthors, 2019: Uncertainty in near-term global surface warming linked to tropical Pacific climate variability. *Nat. Commun.*, **10**, 1990, https://doi.org/10.1038/s41467-019-09761-2.

Bui-Thanh, T., C. Burstedde, O. Ghattas, J. Martin, G. Stadler, and L. C. Wilcox, 2012: Extreme-scale UQ for Bayesian inverse problems governed by PDEs. *Proc. Int. Conf. on High Performance Computing, Networking, Storage and Analysis*, Salt Lake City, UT, IEEE, https://doi.org/10.1109/SC.2012.56.

Caron, L.-P., C. G. Jones, A. Dobbin, J. Imbers, L. Lledó, and G. A. Vecchi, 2018: How skillful are the multiannual forecasts of Atlantic hurricane activity? *Bull. Amer. Meteor. Soc.*, **99**, 403–413, https://doi.org/10.1175/BAMS-D-17-0025.1.

Cassou, C., Y. Kushnir, E. Hawkins, A. Pirani, F. Kucharski, I. Kang, and N. Caltabiano, 2018: Decadal climate variability and predictability: Challenges and opportunities. *Bull. Amer. Meteor. Soc.*, **99**, 479–490, https://doi.org/10.1175/BAMS-D-16-0286.1.

Chang, C., and Z. Wang, 2020: Multiyear hybrid prediction of Atlantic tropical cyclone activity and the predictability sources. *J. Climate*, **33**, 2263–2279, https://doi.org/10.1175/JCLI-D-19-0475.1.

Charlton-Perez, A. J., J. Bröcker, T. N. Stockdale, and S. Johnson, 2019: When and where do ECMWF seasonal forecast systems exhibit anomalously low signal-to-noise ratio? *Quart. J. Roy. Meteor. Soc.*, **145**, 3466–3478, https://doi.org/10.1002/qj.3631.

Cheng, L., and J. Zhu, 2016: Benefits of CMIP5 multimodel ensemble in reconstructing historical ocean subsurface temperature variations. *J. Climate*, **29**, 5393–5416, https://doi.org/10.1175/JCLI-D-15-0730.1.

——, K. E. Trenberth, J. T. Fasullo, T. Boyer, J. Abraham, and J. Zhu, 2017: Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.*, **3**, e1601545, https://doi.org/10.1126/sciadv.1601545.

Christiansen, B., S. Yang, and D. Matte, 2022: The forced response and decadal predictability of the North Atlantic Oscillation: Nonstationary and fragile skills. *J. Climate*, **35**, 5869–5882, https://doi.org/10.1175/JCLI-D-21-0807.1.

Damany-Pearce, L., B. Johnson, A. Wells, M. Osborne, J. Allan, C. Belcher, A. Jones, and J. Haywood, 2022: Australian wildfires cause the largest stratospheric warming since Pinatubo and extends the lifetime of the Antarctic ozone hole. *Sci. Rep.*, **12**, 12665, https://doi.org/10.1038/s41598-022-15794-3.

Deser, C., and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Climate Change*, **10**, 277–286, https://doi.org/10.1038/s41558-020-0731-2.

Di Lorenzo, E., and N. Mantua, 2016: Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nat. Climate Change*, **6**, 1042–1047, https://doi.org/10.1038/nclimate3082.

Domeisen, D. I., and Coauthors, 2020: The role of the stratosphere in subseasonal to seasonal prediction: 1. Predictability of the stratosphere. *J. Geophys. Res. Atmos.*, **125**, e2019JD030920, https://doi.org/10.1029/2019JD030920.

Dunstone, N. J., D. M. Smith, and R. Eade, 2011: Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. *Geophys. Res. Lett.*, **38**, L1470, https://doi.org/10.1029/2011GL047949.

——, ——, A. A. Scaife, L. Hermanson, R. Eade, N. Robinson, M. Andrews, and J. Knight, 2016: Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nat. Geosci.*, **9**, 809–814, https://doi.org/10.1038/ngeo2824.

Eade, R., D. M. Smith, A. A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do seasonal to decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, **41**, 5620–5628, https://doi.org/10.1002/2014GL061146.

England, M., and Coauthors, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Climate Change*, **4**, 222–227, https://doi.org/10.1038/nclimate2106.

Findell, K. L., and T. L. Delworth, 2010: Impact of common sea surface temperature anomalies on global drought and pluvial frequency. *J. Climate*, **23**, 485–503, https://doi.org/10.1175/2009JCLI3153.1.

Fischer, E. M., U. Beyerle, C. F. Schleussner, A. D. King, and R. Knutti, 2018: Biased estimates of changes in climate extremes from prescribed SST simulations. *Geophys. Res. Lett.*, **45**, 8500–8509, https://doi.org/10.1029/2018GL079176.

Forster, P., 2017: Half a century of robust climate models. *Nature*, **545**, 296–297, https://doi.org/10.1038/545296a.

Frölicher, T. L., K. B. Rodgers, C. A. Stock, and W. W. L. Cheung, 2016: Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors. *Global Biogeochem. Cycles*, **30**, 1224–1243, https://doi.org/10.1002/2015GB005338.

Fu, L.-L., T. Lee, W. T. Liu, and R. Kwok, 2019: 50 years of satellite remote sensing of the ocean. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, *Meteor. Monogr.*, No. 59, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0010.1.

Fujii, Y., and Coauthors, 2019: Observing system evaluation based on ocean data assimilation and prediction systems: On-going challenges and a future vision for designing and supporting ocean observational networks. *Front. Mar. Sci.*, **6**, 417, https://doi.org/10.3389/fmars.2019.00417.

Fyfe, J. C., and Coauthors, 2016: Making sense of the early-2000s warming slowdown. *Nat. Climate Change*, **6**, 224–228, https://doi.org/10.1038/nclimate2938.

Ganachaud, A., and C. Wunsch, 2000: Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature*, **408**, 453–457, https://doi.org/10.1038/35044048.

GCOS, 2021: The status of the Global Climate Observing System 2021: The GCOS status report (GCOS-240). World Meteorological Organization Rep., 384 pp., https://gcos.wmo.int/en/publications/gcos-status-report-2021.

Gillett, N. P., and Coauthors, 2016: The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6. *Geosci. Model Dev.*, **9**, 3685–3697, https://doi.org/10.5194/gmd-9-3685-2016.

Goldenberg, S. B., C. W. Landsea, A. M. Mestas-Nuñez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, **293**, 474–479, https://doi.org/10.1126/science.1060040.

GOOS, 2020: A roadmap for the implementation of the Global Ocean Observing System 2030 strategy. GOOS Rep. 249, 40 pp., www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=26687.

Gordon, E. M., E. A. Barnes, and J. W. Hurrell, 2021: Oceanic harbingers of Pacific decadal oscillation predictability in CESM2 detected by neural networks. *Geophys. Res. Lett.*, **48**, e2021GL095392, https://doi.org/10.1029/2021GL095392.

Hakim, G. J., K. A. Bumbaco, R. Tardif, and J. G. Powers, 2020: Optimal network design applied to monitoring and forecasting surface temperature in Antarctica. *Mon. Wea. Rev.*, **148**, 857–873, https://doi.org/10.1175/MWR-D-19-0103.1.

Häkkinen, S., and P. B. Rhines, 2004: Decline of subpolar North Atlantic circulation during the 1990s. *Science*, **304**, 555–559, https://doi.org/10.1126/science.1094917.

Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, https://doi.org/10.1038/s41586-019-1559-7.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1.

Hegerl, G. C., 2022: Climate change is physics. *Commun. Earth Environ.*, **3**, 14, https://doi.org/10.1038/s43247-022-00342-8.

Held, I. M., T. L. Delworth, J. Lu, K. L. Findell, and T. R. Knutson, 2005: Simulation of Sahel drought in the 20th and 21st centuries. *Proc. Natl. Acad. Sci. USA*, **102**, 17 891–17 896, https://doi.org/10.1073/pnas.0509057102.

Hermanson, L., and Coauthors, 2022: WMO global annual to decadal climate update: A prediction for 2021–2025. *Bull. Amer. Meteor. Soc.*, **103**, E1117–E1129, https://doi.org/10.1175/BAMS-D-20-0311.1.

Imada, Y., H. Kawase, M. Watanabe, M. Arai, and I. Takayabu, 2020: Advanced event attribution for the regional heavy rainfall events. *npj Climate Atmos. Sci.*, **3**, 37, https://doi.org/10.1038/s41612-020-00141-y.

IPCC, 2021: *Climate Change 2021: The Physical Science Basis.* V. Masson-Delmotte et al., Eds., Cambridge University Press, 2409 pp.

Japan Meteorological Agency, 2013: JRA-55: Japanese 55-year Reanalysis, daily 3-hourly and 6-hourly data. UCAR, accessed 14 July 2020, https://doi.org/10.5065/D6HH6H41.

Johnson, G. C., and Coauthors, 2022: Argo—Two decades: Global oceanography, revolutionized. *Annu. Rev. Mar. Sci.*, **14**, 379–403, https://doi.org/10.1146/annurev-marine-022521-102008.

Kalmikov, A. G., and P. Heimbach, 2014: A Hessian-based method for uncertainty quantification in global ocean state estimation. *SIAM J. Sci. Comput.*, **36**, S267–S295, https://doi.org/10.1137/130925311.

Klavans, J. M., M. A. Cane, A. C. Clement, and L. N. Murphy, 2021: NAO predictability from external forcing in the late 20th century. *npj Climate Atmos. Sci.*, **4**, 22, https://doi.org/10.1038/s41612-021-00177-8.

Kosaka, Y., and S.-P. Xie, 2013: Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**, 403–407, https://doi.org/10.1038/nature12534.

Kretschmer, M., S. V. Adams, A. Arribas, R. Prudden, N. Robinson, E. Saggioro, and T. G. Shepherd, 2021: Quantifying causal pathways of teleconnections. *Bull. Amer. Meteor. Soc.*, **102**, E2247–E2263, https://doi.org/10.1175/BAMS-D-20-0117.1.

Kushnir, Y., and Coauthors, 2019: Towards operational predictions of the near-term climate. *Nat. Climate Change*, **9**, 94–101, https://doi.org/10.1038/s41558-018-0359-7.

Lee, J.-Y., and K.-J. Ha, 2015: Understanding of interdecadal changes in variability and predictability of the Northern Hemisphere summer tropical–extratropical teleconnection. *J. Climate*, **28**, 8634–8647, https://doi.org/10.1175/JCLI-D-15-0154.1.

——, B. Wang, K.-H. Seo, J.-S. Kug, Y.-S. Choi, Y. Kosaka, and K.-J. Ha, 2014: Future changes of Northern Hemisphere summer tropical–extratropical teleconnection in CMIP5 models. *J. Climate*, **27**, 3643–3664, https://doi.org/10.1175/JCLI-D-13-00261.1.

Lehner, F., C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti, and E. Hawkins, 2020: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dyn.*, **11**, 491–508, https://doi.org/10.5194/esd-11-491-2020.

Lin, I.-I., and Coauthors, 2021: ENSO and tropical cyclones. *El Niño Southern Oscillation in a Changing Climate*, M. J. McPhaden, A. Santoso, and W. Cai, Eds., Amer. Geophys. Union, 377–408, https://doi.org/10.1002/9781119548164.ch17.

Loose, N., and P. Heimbach, 2021: Leveraging uncertainty quantification to design ocean climate observing systems. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002386, https://doi.org/10.1029/2020MS002386.

——, ——, H. R. Pillar, and K. H. Nisancioglu, 2020: Quantifying dynamical proxy potential through shared adjustment physics in the North Atlantic. *J. Geophys. Res. Oceans*, **125**, e2020JC016112, https://doi.org/10.1029/2020JC016112.

Lovenduski, N. S., G. A. McKinley, A. R. Fay, K. Lindsay, and M. C. Long, 2016: Partitioning uncertainty in ocean carbon uptake projections: Internal variability, emission scenario, and model structure. *Global Biogeochem. Cycles*, **30**, 1276–1287, https://doi.org/10.1002/2016GB005426.

Mahfouf, J.-F., K. Bergaoui, C. Draper, F. Bouyssel, F. Taillefer, and L. Taseva, 2009: A comparison of two off-line soil analysis schemes for assimilation of screen level observations. *J. Geophys. Res.*, **114**, D08105, https://doi.org/10.1029/2008JD011077.

Mahmood, R., M. G. Donat, P. Ortega, F. J. Doblas-Reyes, C. Delgado-Torres, M. Samsó, and P.-A. Bretonnière, 2022: Constraining low-frequency variability in climate projections to predict climate on decadal to multi-decadal timescales—A poor man's initialized prediction system. *Earth Syst. Dyn.*, **13**, 1437–1450, https://doi.org/10.5194/esd-13-1437-2022.

Manabe, S., and A. J. Broccoli, 2020: *Beyond Global Warming: How Numerical Models Revealed the Secrets of Climate Change.* Princeton University Press, 216 pp.

Marzeion, B., and Coauthors, 2020: Partitioning the uncertainty of ensemble projections of global glacier mass change. *Earth's Future*, **8**, e2019EF001470, https://doi.org/10.1029/2019EF001470.

Meehl, G. A., A. Hu, and H. Teng, 2016: Initialized decadal prediction for transition to positive phase of the interdecadal Pacific oscillation. *Nat. Commun.*, **7**, 11718, https://doi.org/10.1038/ncomms11718.

——, and Coauthors, 2021: Initialized Earth system prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.*, **2**, 340–357, https://doi.org/10.1038/s43017-021-00155-x.

——, H. Teng, D. Smith, S. Yeager, W. Merryfield, F. Doblas-Reyes, and A. A. Glanville, 2022: The effects of bias, drift, and trends in calculating anomalies for evaluating skill of seasonal-to-decadal initialized climate predictions. *Climate Dyn.*, **59**, 3373–3389, https://doi.org/10.1007/s00382-022-06272-7.

Merryfield, W. J., and Coauthors, 2020: Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, **101**, E869–E896, https://doi.org/10.1175/BAMS-D-19-0037.1.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, **319**, 573–574, https://doi.org/10.1126/science.1151915.

Minobe, S., A. Capotondi, M. G. Jacox, M. Nonaka, and R. R. Rykaczewski, 2022: Toward regional marine ecological forecasting using global climate model predictions from subseasonal to decadal timescales: Bottlenecks and recommendations. *Front. Mar. Sci.*, **9**, 855965, https://doi.org/10.3389/fmars.2022.855965.

Mizuta, R., and Coauthors, 2017: Over 5000 years of ensemble future climate simulations by 60 km global and 20 km regional atmospheric models. *Bull. Amer. Meteor. Soc.*, **98**, 1383–1398, https://doi.org/10.1175/BAMS-D-16-0099.1.

Oden, T., R. Moser, and O. Ghattas, 2010: Computer predictions with quantified uncertainty, Part I. *SIAM News*, Vol. 43, No. 9, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1–3, https://archive.siam.org/pdf/news/1842.pdf.

Oliver, E. C. J., and Coauthors, 2018: Longer and more frequent marine heatwaves over the past century. *Nat. Commun.*, **9**, 1324, https://doi.org/10.1038/s41467-018-03732-9.

Parkinson, C. L., 2019: A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far exceeding the rates seen in the Arctic. *Proc. Natl. Acad. Sci. USA*, **116**, 14 414–14 423, https://doi.org/10.1073/pnas.1906556116.

Piecuch, C. G., R. M. Ponte, C. M. Little, M. W. Buckley, and I. Fukumori, 2017: Mechanisms underlying recent decadal changes in subpolar North Atlantic Ocean heat content. *J. Geophys. Res. Oceans*, **122**, 7181–7197, https://doi.org/10.1002/2017JC012845.

Richardson, D., A. S. Black, D. Irving, R. J. Matear, D. P. Monselesan, J. S. Risbey, D. T. Squire, and C. R. Tozer, 2022: Global increase in wildfire potential from compound fire weather and drought. *npj Climate Atmos. Sci.*, **5**, 23, https://doi.org/10.1038/s41612-022-00248-4.

Risbey, J. S., S. Lewandowsky, K. Cowtan, N. Oreskes, S. Rahmstorf, A. Jokimaki, and G. Foster, 2018: A fluctuation in surface temperature in historical context: Reassessment and retrospective on the evidence. *Environ. Res. Lett.*, **13**, 123008, https://doi.org/10.1088/1748-9326/aaf342.

Robson, J., R. Sutton, K. Lohmann, D. Smith, and M. D. Palmer, 2012: Causes of the rapid warming of the North Atlantic Ocean in the mid-1990s. *J. Climate*, **25**, 4116–4134, https://doi.org/10.1175/JCLI-D-11-00443.1.

Roemmich, D., and Coauthors, 2019: On the future of Argo: A global, full-depth, multi-disciplinary array. *Front. Mar. Sci.*, **6**, 439, https://doi.org/10.3389/fmars.2019.00439.

Scaife, A. A., and D. M. Smith, 2018: A signal-to-noise paradox in climate science. *npj Climate Atmos. Sci.*, **1**, 28, https://doi.org/10.1038/s41612-018-0038-4.

——, and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, **41**, 2514–2519, https://doi.org/10.1002/2014GL059637.

Schneider, T., S. Lan, A. Stuart, and J. Teixeira, 2017: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys. Res. Lett.*, **44**, 12 396–12 417, https://doi.org/10.1002/2017GL076101.

Schubert, S. D., and Coauthors, 2009: A U.S. CLIVAR project to assess and compare the responses of global climate models to drought-related SST forcing patterns: Overview and results. *J. Climate*, **22**, 5251–5272, https://doi.org/10.1175/2009JCLI3060.1.

Sévellec, F., and S. S. Drijfhout, 2019: The signal-to-noise paradox for interannual surface atmospheric temperature predictions. *Geophys. Res. Lett.*, **46**, 9031–9041, https://doi.org/10.1029/2019GL083855.

Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife, 2010: Skilful multi-year predictions of Atlantic hurricane frequency. *Nat. Geosci.*, **3**, 846–849, https://doi.org/10.1038/ngeo1004.

——, and Coauthors, 2019: Robust skill of decadal climate predictions. *npj Climate Atmos. Sci*., **2**, 13, https://doi.org/10.1038/s41612-019-0071-y.

——, and Coauthors, 2020: North Atlantic climate far more predictable than models imply. *Nature*, **583**, 796–800, https://doi.org/10.1038/s41586-020-2525-0.

——, and Coauthors, 2022: Attribution of multi-annual to decadal changes in the climate system: The Large Ensemble Single Forcing Model Intercomparison Project (LESFMIP). *Front. Climate*, **4**, 955414, https://doi.org/10.3389/fclim.2022.955414.

Sparrow, S., and Coauthors, 2018: Finding ocean states that are consistent with observations from a perturbed physics parameter ensemble. *J. Climate*, **31**, 4639–4656, https://doi.org/10.1175/JCLI-D-17-0514.1.

Squire, D., and Coauthors, 2021: Likelihood of unprecedented drought and fire weather during Australia's 2019 megafires. *npj Climate Atmos. Sci.*, **4**, 64, https://doi.org/10.1038/s41612-021-00220-8.

Stott, P. A., N. Christidis, and R. A. Betts, 2011: Changing return periods of weather-related impacts: The attribution challenge. *Climatic Change*, **109**, 263–268, https://doi.org/10.1007/s10584-011-0265-8.

Thoma, M., R. J. Greatbatch, C. Kadow, and R. Gerdes, 2015: Decadal hindcasts initialized using observed surface wind stress: Evaluation and prediction out to 2024. *Geophys. Res. Lett.*, **42**, 6454–6461, https://doi.org/10.1002/2015GL064833.

Thompson, V., N. J. Dunstone, A. A. Scaife, D. M. Smith, J. M. Slingo, S. Brown, and S. E. Belcher, 2017: High risk of unprecedented UK rainfall in the current climate. *Nat. Commun*., **8**, 107, https://doi.org/10.1038/s41467-017-00275-3.

Thorne, P. W., J. R. Lanzante, T. C. Peterson, D. J. Seidel, and K. P. Shine, 2010: Tropospheric temperature trends: History of an ongoing controversy. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 66–88, https://doi.org/10.1002/wcc.80.

Trenberth, K., G. Branstator, and P. Arkin, 1988: Origins of the 1988 North American drought. *Science*, **242**, 1640–1645, https://doi.org/10.1126/science.242.4886.1640.

Watanabe, M., H. Shiogama, H. Tatebe, M. Hayashi, M. Ishii, and M. Kimoto, 2014: Contribution of natural decadal variability to global-warming acceleration and hiatus. *Nat. Climate Change*, **4**, 893–897, https://doi.org/10.1038/nclimate2355.

WCRP JSC, 2019: World Climate Research Programme Strategic Plan 2019–2028. WCRP Publ. 1/2019, 28 pp.

Weisheimer, A., D. Decremer, D. MacLeod, C. O'Reilly, T. N. Stockdale, S. Johnson, and T. N. Palmer, 2019: How confident are predictability estimates of the winter North Atlantic Oscillation? *Quart. J. Roy. Meteor. Soc*., **145** (Suppl. 1), 140–159, https://doi.org/10.1002/qj.3446.

Yang, C., S. Masina, A. Bellucci, and A. Storto, 2016: The rapid warming of the North Atlantic Ocean in the mid-1990s in an eddy-permitting ocean reanalysis (1982–2013). *J. Climate*, **29**, 5417–5430, https://doi.org/10.1175/JCLI-D-15-0438.1.

Yeager, S., 2020: The abyssal origins of North Atlantic decadal predictability. *Climate Dyn.*, **55**, 2253–2271, https://doi.org/10.1007/s00382-020-05382-4.

Zhang, W., and B. Kirtman, 2019: Understanding the signal-to-noise paradox with a simple Markov model. *Geophys. Res. Lett.*, **46**, 13 308–13 317, https://doi.org/10.1029/2019GL085159.

——, ——, L. Siqueira, A. Clement, and J. Xia, 2021: Understanding the signal-to-noise paradox in decadal climate predictability from CMIP5 and an eddying global coupled model. *Climate Dyn.*, **56**, 2895–2913, https://doi.org/10.1007/s00382-020-05621-8.

Zhang, X., G. Hegerl, S. Seneviratne, R. Stewart, F. Zwiers, and L. Alexander, 2014: WCRP Grand Challenge: Understanding and predicting weather and climate extremes. WCRP White Paper, 10 pp., www.wcrp-climate.org/images/documents/grand_challenges/GC_Extremes_v2.pdf.