Current and Emerging Developments in Subseasonal to Decadal Prediction


ABSTRACT: Weather and climate variations on subseasonal to decadal time scales can have enormous social, economic, and environmental impacts, making skillful predictions on these time scales a valuable tool for decision-makers. As such, there is a growing interest in the scientific, operational, and applications communities in developing forecasts to improve our foreknowledge of extreme events. On subseasonal to seasonal (S2S) time scales, these include high-impact meteorological events such as tropical cyclones, extratropical storms, floods, droughts, and heat and cold waves. On seasonal to decadal (S2D) time scales, while the focus broadly remains similar (e.g., on precipitation, surface and upper-ocean temperatures, and their effects on the probabilities of high-impact meteorological events), understanding the roles of internal variability and externally forced variability such as anthropogenic warming in forecasts also becomes important. The S2S and S2D communities share common scientific and technical challenges. These include forecast initialization and ensemble generation; initialization shock and drift; understanding the onset of model systematic errors; bias correction, calibration, and forecast quality assessment; model resolution; atmosphere–ocean coupling; sources and expectations for predictability; and linking research, operational forecasting, and end-user needs. In September 2018 a coordinated pair of international conferences, framed by the above challenges, was organized jointly by the World Climate Research Programme (WCRP) and the World Weather Research Programme (WWRP). These conferences surveyed the state of S2S and S2D prediction, ongoing research, and future needs, providing an ideal basis for synthesizing current and emerging developments in these areas that promise to enhance future operational services. This article provides such a synthesis.
Beyond the tremendous progress in weather forecasting witnessed in recent decades (Bauer et al. 2015), predictive capabilities have expanded, increasingly seamlessly, to encompass climate on subseasonal to decadal time scales (Fig. 1 and Kirtman et al. 2013). These advances have been enabled by better observations, data assimilation schemes, and models originating both from the weather prediction and long-term climate simulation communities, together with increased computational power supporting progressively higher resolution and larger ensembles that allow uncertainties to be better estimated and, in some cases, reduced.

International efforts under the auspices of the World Weather Research Programme (WWRP) and World Climate Research Programme (WCRP) have helped drive this progress through coordinated research to improve the accuracy and utilization of weather and climate predictions. Community research efforts under the WCRP led initially to climate predictions one to two seasons ahead becoming part of the World Meteorological Organization (WMO) operational infrastructure (Graham et al. 2011). More recently a joint WWRP and WCRP Subseasonal to Seasonal Prediction Project has started tackling the so-called weather–climate prediction desert from two weeks to a season (Robertson et al. 2018; Mariotti et al. 2018), aiming to underpin new WMO operations on those time scales (Vitart et al. 2017), and the NOAA-led Subseasonal Experiment (SubX) project has similar aims (Pegion et al. 2019). At longer ranges, WCRP-enabled research

Fig. 1. Schematic depiction of (bottom) temporal ranges and (top) sources of predictability for weather and climate prediction. The subseasonal range encompasses the S2S time scales and the seasonal and annual-to-decadal ranges encompass the S2D time scales that are considered in this paper. Longer multidecadal and centennial ranges derive predictability mainly from forcing scenarios rather than initial conditions, and are typically represented through climate projections originating from historical simulations begun in preindustrial times rather than predictions initialized from more recent observation-based climate states. Some important sources of predictability and approximate time scales over which they are most influential on surface climate are indicated in the top portion of the figure; acronyms are defined and associated phenomena are discussed in the main text.
has quantified predictability from a year to a decade, and corresponding WMO operational infrastructure for annual-to-decadal climate prediction is now in place (World Meteorological Organization 2018; Kushnir et al. 2019).

As each of these efforts has progressed it has become increasingly apparent that common challenges exist across predictive time scales. These include understanding and adequately representing in models processes that give rise to predictability in the Earth system, consisting of the physical climate system—atmosphere, ocean, land, and sea ice—together with associated biogeochemical cycling, especially of carbon (top part of Fig. 1); capturing and communicating inherent uncertainties caused by the chaotic nature of weather and climate; correcting for and reducing imperfections in models that may systematically degrade forecast quality; and providing forecast information in a form that is applicable to decision-making. At the same time, opportunities for usefully predicting elements of the Earth system beyond long-term means of standard meteorological variables, including land, ocean, and sea ice properties and risks of weather extremes, have come into focus. The ultimate collective endeavor is to improve the prediction of the spatial-temporal continuum connecting weather to climate through a coordinated, seamless, and integrated Earth system approach for the benefit of society.

In September 2018, international conferences1 on subseasonal to seasonal (S2S) prediction (encompassing forecast ranges from 2 weeks to a season) and seasonal to decadal (S2D) prediction (encompassing ranges longer than a season, up to a decade) together with crosscutting plenary sessions were convened jointly by WWRP and WCRP. This represented a confluence of research and operational climate prediction expertise and knowledge exchange across prediction time scales that was unprecedented in scope. Selected outcomes, organized by themes encompassing the challenges outlined above, are synthesized in this article.

Mechanisms of predictability

Subseasonal to seasonal. A major source of S2S predictability is the organization of tropical convection by the Madden–Julian oscillation (MJO) (Woolnough 2019), which is predicted skillfully by S2S project models up to 3–4 weeks ahead (Vitart 2017). The MJO has worldwide impacts that depend on its amplitude and phase, including modulation of tropical cyclone activity (Lee et al. 2018; Zhao et al. 2019) and extratropical phenomena such as the East Asian summer monsoon (Li et al. 2018). The associated tropical–extratropical teleconnections (Lin et al. 2019) impart S2S forecast skill for many of these extratropical phenomena including Euro-Atlantic weather regimes, position of the jet stream, atmospheric rivers (DeFlorio et al. 2019), and hail/tornado activity (Baggett et al. 2018). However, good representations of the basic state both in the tropics and in the extratropics, as well as tropical air–sea interactions and atmospheric convection (e.g., Yoo et al. 2015), are necessary for these teleconnections to be correctly simulated by general circulation models (Henderson et al. 2017).

S2S predictability also derives from the stratosphere through its relatively long time scales of variability2 and lagged influences on the troposphere (Kidston et al. 2015). Interactions between the stratosphere and the troposphere from the tropics to the extratropics thus provide a promising source of S2S prediction skill (Butler et al. 2019). For example, in the winter Northern Hemisphere stratosphere the climatological westerly polar vortex exhibits extremes in variability, including sudden stratospheric warmings (SSWs) that are driven largely by Rossby waves from the troposphere. SSWs have lagged impacts on sea level pressure, surface temperature, and precipitation,
including pronounced tendencies for cold anomalies over northern Eurasia and warm anomalies over northeastern North America (e.g., Sigmond et al. 2013). Initializing forecasts during extreme stratospheric events provides increases in prediction skill of surface climate in such regions up to 3–6 weeks later (Domeisen et al. 2020b). However, the predictability of specific extreme stratospheric events is limited, ranging from a few days to about 2 weeks (Fig. 2) for different SSWs (Karpechko 2018; Taguchi 2018; Domeisen et al. 2019), although models show evidence of underconfident forecasts in the stratosphere on S2S time scales (O’Reilly et al. 2019). Outstanding questions remain about the mechanisms of stratosphere–troposphere coupling processes, in particular on the causes, variability, and trends for the occurrence of SSW events (Ayarzagüena et al. 2018; Simpson et al. 2018) and why not all SSW events have similar downward effects (e.g., Garfinkel et al. 2013; Maycock and Hitchcock 2015). In addition, further research is needed to assess the degree to which prediction models capture both the stratospheric variability and coupling processes.

Among atmosphere–surface influences, land–atmosphere interactions have their greatest impact on subseasonal time scales in forecasts where land is initialized (Dirmeyer et al. 2018a), but also can contribute skill on weather prediction and multimonth time scales (Dirmeyer and Halder 2016, 2017). The most broadly impactful land attribute is soil moisture (Koster et al. 2004, 2016), but anomalies in soil temperature (Y. Zhang et al. 2019; Yang et al. 2019), snow cover (Jeong et al. 2013; Orsolini et al. 2013), and vegetation states (Williams et al. 2016) can all have significant impacts. A number of recent studies have focused on nonlocal impacts of land surface anomalies, showing, for example, that soil moisture anomalies can exert remote as well as local influences in boreal summer through driving of quasi-stationary Rossby waves and associated circulation anomalies (e.g., Teng et al. 2019; Wang et al. 2019). In addition, land surface and subsurface temperatures in spring may exert delayed downstream influences on precipitation (Xue et al. 2018), and evapotranspiration may remotely influence precipitation over land (Wei and Dirmeyer 2019).

Atmosphere–ocean interactions, fundamental for S2D predictability, can also be influential on S2S time scales. For example, submonthly prediction skills for precipitation and temperature are enhanced over certain land areas including parts of Australia, the Maritime Continent, and the contiguous United States when tropical sea surface temperature (SST) anomalies associated with El Niño–Southern Oscillation (ENSO) are present (Hudson et al. 2011; Li and Robertson 2015; DelSole et al. 2017). Extratropical SST anomalies also can impart S2S skill through teleconnections, as shown, for example, by McKinnon et al. (2016), who identified a SST anomaly pattern in the midlatitude North Pacific that tends to precede heat waves and rainfall deficits in the eastern United States by up to 50 days.

Fig. 2. Forecast probabilities of 13 SSWs that occurred on the indicated dates as a function of lead time, based on ensemble hindcasts from the ECMWF monthly forecasting system. Most of the SSWs are predicted between lead times of 8 and 12 days with a probability of 0.5–0.9, which is considerably larger than the average frequency of SSW occurrence (from Karpechko 2018).
Sea ice strongly influences surface fluxes and lower-atmospheric temperatures particularly in the marginal ice zone, and provides a source of S2S predictability for polar and possibly midlatitude regions (Chevallier et al. 2019). This motivates the development of S2S forecasts for sea ice, which thus far have shown significant, albeit region-dependent skill for predicting intraseasonal Arctic sea ice variability (Liu et al. 2018; Zampieri et al. 2018).

### Seasonal to decadal.

A primary general source of S2D atmospheric predictability is remote influences from a variety of teleconnections (e.g., Yuan et al. 2018; Ruprich-Robert et al. 2018; Beverley et al. 2019). Teleconnections associated with anomalous atmospheric circulation patterns arise from changes to the Walker circulation usually driven by anomalous zonal SST gradients (Cai et al. 2019), and changes to the Hadley circulation usually driven by anomalous meridional SST gradients, especially interhemispheric differences (Kang et al. 2018). These influences impact tropical cyclones and rainfall, whereas anomalous upper-level divergence due to tropical rainfall anomalies leads to Rossby waves that impact the extratropics (Scaife et al. 2017; O’Reilly et al. 2018). Besides giving rise to atmosphere–ocean interactions that alter the atmospheric circulation, SST anomalies can induce low-level temperature and moisture anomalies that are advected elsewhere by climatological winds (Dunstone et al. 2018).

S2D atmospheric predictability arising from teleconnections requires that SST anomalies be predictable. On seasonal time scales, tropical SST anomalies are dominated by ENSO (Yang et al. 2018), though there is some independent variability in the tropical Atlantic and Indian Oceans that also drives teleconnections (e.g., Nnamchi et al. 2015; Lim et al. 2017). The impacts of ENSO are sensitive to ENSO diversity (Capotondi et al. 2015), including the longitude at which maximum SST anomalies occur (Yeh et al. 2018; Patricola et al. 2018). ENSO SST anomalies are largely predictable out to a year particularly in winter and early spring (Barnston et al. 2019), whereas predictability may extend to two years for some La Niña events (DiNezio et al. 2017b), and to 1.5–2 years for certain El Niño events (Luo et al. 2008).

Decadal SST variability occurs in both the Atlantic and Pacific Oceans, often referred to as Atlantic multidecadal variability (AMV) and Pacific decadal variability (PDV), respectively (e.g., Kushnir et al. 2019). The causes of AMV are not fully understood, especially the relative roles of internal variability and external forcing from aerosols. However, AMV is modulated to some extent by the oceanic Atlantic meridional overturning circulation (Yeager and Robson 2017), which together with the North Atlantic subpolar gyre is influenced by deep ocean density anomalies particularly in the Labrador Sea (Robson et al. 2016); these influences contribute to the especially high multiyear predictability in the North Atlantic (Buckley et al. 2019). AMV couples to the Hadley circulation, affecting hurricanes and Sahel rainfall as illustrated in Fig. 3 (Sheen et al. 2017), and can initiate atmospheric Rossby waves with remote influences including temperatures in parts of China (Monerie et al. 2018). AMV can influence PDV (Ruprich-Robert et al. 2017), and vice versa. PDV may also be influenced by off-equatorial heat content anomalies in the western Pacific Ocean (Meehl et al. 2016). Decadal variability of deep convection in the Southern Ocean influences temperatures in that region, potentially explaining recent increases in Antarctic sea ice (L. Zhang et al. 2019).

S2D atmospheric predictability also arises from longer time-scale processes over land, mainly involving soil moisture (Chikamoto et al. 2017; Ardilouze et al. 2019) and vegetation (Weiss et al. 2014; Bellucci et al. 2015). These highlight the need for land surface initialization (Prodhomme et al. 2016a) and realistic vegetation models (Alessandri et al. 2017).

An additional source of S2D predictability is variations in radiative forcing, which provide significant skill on multiyear time scales (Smith et al. 2019). Much of this skill arises from changes in greenhouse gases, but anthropogenic aerosols may force decadal variations in AMV (Booth et al. 2012) and PDV (Smith et al. 2016; Takahashi and Watanabe 2016). Solar variability (Misios et al. 2019), volcanic eruptions (Ménégoz et al. 2018) including through their influence...
on ENSO (Khodri et al. 2017; Wang et al. 2018) and possibly AMV, and the North Atlantic Oscillation (NAO; Swingedouw et al. 2017) affect climate on seasonal to decadal time scales and are potentially important sources of predictability. However, the relative roles of external radiative forcing and internal variability (W. M. Kim et al. 2018) continue to be explored.

S2D predictability of carbon cycle and other biogeochemical Earth system variables stems from relatively long time scales of variation and predictable climatic influences. Assessing capabilities for predicting such variables is an emerging research area that is discussed further in the sidebar “Frontiers in Earth system prediction.”

**Time-scale interactions.** The quasi-biennial oscillation (QBO) is a downward-propagating ~28-month oscillation of easterly and westerly zonal jets in the tropical stratosphere, driven by upward equatorial waves from the troposphere (e.g., Kim and Chun 2015). In addition to having high predictability and some teleconnected influence on winter surface climate (e.g., Scaife et al. 2014a), the QBO modulates the amplitude, persistence, and rate of propagation of the boreal wintertime MJO (Fig. 4) through its impact on tropical convection via changes in static stability near the tropopause (Yoo and Son 2016; Nishimoto and Yoden 2017). MJO amplitude is better predicted at longer leads during the easterly phase of the QBO (Marshall et al. 2017), likely as a result of longer persistence of the MJO rather than its greater initial amplitude (Lim et al. 2019).

The modulation of SSW probability of occurrence by tropical sources of variability, such as the QBO, ENSO, or MJO, may extend probabilistic predictability of stratospheric variability to a few months or longer if these relationships can be adequately captured by prediction models (Garfinkel and Schwartz 2017; Garfinkel et al. 2018; Domeisen et al. 2019, 2020a).

There is increasing evidence of additional interactions between various sources of S2S and S2D predictability across time scales. One example is that seasonal time-scale variations in ENSO modulate the MJO (Chen et al. 2016) and its impact on the NAO (Lee et al. 2019) with
consequent influences on weather over remote regions. Another is that ENSO teleconnection to the extratropics has varied over multidecadal time scales spanning the past 100+ years (O’Reilly 2018), possibly modulating ability to predict the NAO (Weishiemer et al. 2019), although sampling variability can also give rise to such long-term changes in teleconnections (Yun and Timmermann 2018).

**Modeling issues**

**Subseasonal to seasonal.** Because S2S operational prediction is a relatively new enterprise, considerable efforts focusing on fundamental aspects of forecast system design are occurring at operational centers worldwide (Takaya 2019). One major emphasis consists of methods to represent the uncertainty in initial conditions (bred vector, singular vector, ensemble data assimilation) and model physics (stochastic physics; Leutbecher et al. 2017). In addition, configurations of real-time forecasts and hindcasts, including ensemble size, ensemble strategy (lagged ensemble with different initial times or burst ensemble with a common initial time), and hindcast period, impact forecast quality and ability to evaluate the performance of the hindcast (see the sidebar on “Hindcast and forecast quality assessment” for further discussion of this topic). Identifying suitable compromises and trade-offs in forecast system design is a challenge under practical constraints for operational activities (costs, priorities, timeliness) and demands further research.

From the modeling perspective, multiple operational centers are moving toward a unified, or “seamless” coupled forecast system that can be applied across time scales from days to seasons or longer. More S2S models are incorporating ocean and sea ice components, and becoming increasingly complex and complete in representing coupled processes in the Earth system. On the other hand, poor representation of model physics, in particular clouds (Morcrette et al. 2018), results in model drifts and biases in surface land and ocean temperatures, which is a long-standing modeling issue that can degrade the skill of S2S predictions (Vitart and Balmaseda 2017). Improvements in cloud parameterizations (Stan and Straus 2019) and in representing the diurnal cycle of the atmospheric boundary layers are crucial for advancing S2S modeling. The Earth system modeling approach poses another challenge to initialize the ocean and sea ice components with high accuracy; for example, there is a relatively large dispersion of initialized sea ice fields in current S2S models (Chevallier et al. 2017; Zampieri et al. 2018).

Another important S2S modeling issue is predicting the MJO, owing to its importance as a source of subseasonal predictability (H. Kim et al. 2018). Multimodel evaluations have shown that S2S models have difficulties in representing MJO propagation across the Maritime Continent. Process-oriented diagnostics (Maloney et al. 2019) have identified a
dry bias in the lower troposphere as one of the causes for the poor MJO propagation through weakening the horizontal moisture gradient over the Indian Ocean and western Pacific (Lim et al. 2018) and dampening the organization and propagation of the MJO. A recharge process whereby moisture builds up in the lower troposphere during the suppressed convection phase of the MJO, and that is key for MJO propagation around the Maritime Continent in boreal winter, is underrepresented in S2S models due to the dry bias (Kim 2017). Ocean coupling is another important process for the MJO (DeMott et al. 2015), and several studies have demonstrated that ocean coupling can improve MJO propagation and enhance predictive skill in models.

Poor vertical resolution, low model lid height, inadequate orographic and nonorographic gravity wave parameterizations, and biases in the tropospheric mean state (e.g., the location of stationary Rossby waves) could limit the predictive skill from stratosphere–troposphere coupling processes (Tripathi et al. 2015; Butler et al. 2016), but new generations of prediction systems have rapidly improved in many of these areas. Future model development could prioritize improved representation of orographic and nonorographic gravity wave drag and an internally generated QBO (Butchart et al. 2018). Better understanding of stratosphere–troposphere coupling processes and the role of the stratosphere on surface skill could be gained through case studies and stratospheric nudging experiments (Hansen et al. 2017). Improved observations of the stratosphere (e.g., aerosols and chemistry) and other climate components may improve S2S predictions. Finally, there is potential for modeling of stratospheric ozone chemistry, which provides surface temperature predictability on S2S time scales due to its influence on high-latitude stratospheric circulation anomalies together with their lagged surface impacts (Stone et al. 2019). Although that may currently be too resource-intensive due to the many species and reactions that must be modeled, emerging machine-learning techniques may provide pathways for incorporating chemistry-climate information into S2S forecasts (Nowack et al. 2018).

**Seasonal to decadal.** Modeling issues for S2D prediction naturally overlap with those for S2S prediction. However, the longer time scales of S2D prediction lead to a greater emphasis on representing slower climate variations such as ENSO and AMV, and greater attention to reducing model biases in the ocean that may take months to years to develop. Increased model resolution can reduce model biases as illustrated in Fig. 5 (Jia et al. 2015; Müller et al. 2018), and improve skill (Prodhomme et al. 2016b; Schuster et al. 2019; Infanti and Kirtman 2019), although the greater computational cost is not always justified (Scaife et al. 2019). More fundamental issues...
strategies involve analyzing/understanding model biases, before attempting to correct them a priori or a posteriori. Such analysis methods include comparing hindcasts with observations and multidecadal historical or other simulations to distill causation for model errors, such as in the tropical Pacific (Shonk et al. 2018) or Atlantic (Voldoire et al. 2019). Similarly, errors in modeled variability or teleconnection patterns can be characterized by examining their evolution with lead time. Model biases can be corrected both through simple methods such as statistical bias correction and anomaly coupling (Toniazzo and Koseki 2018), and more complex methods such as supermodeling, through which multiple models exchange information during a climate simulation (Shen et al. 2016).

Performance of S2D predictions is strongly tied to initialization of model components beyond the lower atmosphere. For example, stratospheric initial conditions are a source of

---

**Hindcast and forecast quality assessment (or “The unexamined life is not worth living”)**

Besides helping to inform decision-making, the careful assessment of forecast quality is critical to guiding forecasting improvements, but has many and varied considerations. Simply answering the question “Is this forecast better than that one?” is complicated, as the appropriate skill metric or means for comparison is not always obvious. Some recent studies have focused on newer statistical methods for comparing one forecast to another. One relatively simple approach is the random walk test (DelSole and Tippett 2016), illustrated in Fig. SB1. This method is applicable to a wide range of measures such as a score based on the Earth mover’s distance metric (Düsterhus 2020), while also being robust and fair in its discrimination.

The utility of forecast assessment can be illustrated through two very different applications of seasonal forecasts: sea ice and hurricanes. The assessment of seasonal sea ice forecasts is complicated by the low quality of sea ice observations, but nevertheless reveals that initializing sea ice thickness using observational datasets generally improves the prediction of Arctic sea ice extent and edges (Blockley and Peterson 2018). Comparison of multiannual forecasts of Atlantic hurricane activity obtained by direct tracking of storms in decadical hindcasts and through a hybrid approach combining predicted SSTs and observed statistical relations finds that each approach is skillful, especially hybrid forecasts based on a SST index for AMV (Caron et al. 2018).

A robust assessment of model performance should include the model’s simulation of climate modes and teleconnection patterns such as ENSO, MJO, and NAO, since they are major sources of predictability and errors representing their patterns or strength (e.g., Yang and DelSole 2012; Vitart 2017) can degrade skill in affected regions (Gleixner et al. 2017; Lu et al. 2017). In cases where modeled teleconnection patterns are imperfect, forecast skill may be improved by means of statistical methods that use model forecasts of relevant climate modes such as ENSO as predictors (e.g., Strazzo et al. 2019). It remains desirable, however, for models to improve so that their simulated teleconnection patterns are sufficiently realistic that such corrections are not needed.

---

**Fig. SB1.** Random walk test comparing monthly mean forecasts of the Niño-3.4 index for equatorial Pacific SST at a 2.5-month lead, between the multimodel mean (MMM) and individual models in the NMME. Counts (vertical axis) increase by 1 when the MMM squared error is smaller than that an individual model (MMM is more accurate) and decrease by 1 otherwise (individual model is more accurate), and are accumulated forward for all initial months and years (horizontal axis). Accumulated counts above or below the shaded region indicate skill differences according to the squared error metric that are significant with >95% confidence (MMM is more skillful above the shaded region, and individual model is more skillful below). Niño-3.4 anomalies are relative to 1982–98 climatological values, and span each month in 1999–2015 (from DelSole and Tippett 2016).
seasonal winter NAO skill (e.g., O’Reilly et al. 2019; Nie et al. 2019) as illustrated in Fig. 6, and ocean initial conditions are crucial for skillfully predicting ENSO (Balmaseda and Anderson 2009), as well as decadal variability in the subpolar North Atlantic (Yeager and Robson 2017; Borchert et al. 2018). However, initialization using full-field observational values can lead to initial shocks affecting skill (Kröger et al. 2018) and in such cases initialization combining observed anomalies with the model’s own climatology can be beneficial until underlying model errors can be reduced (Volpi et al. 2017). Basic initialization strategies continue to be an active research area particularly for decadal prediction (Brune et al. 2018), and methods extending to forecast runs such as the ensemble dispersion filter, which replaces the ensemble members with the ensemble mean every 3 months (Kadow et al. 2017) are also being explored. Comparisons that apply different initialization methods to the same model can yield valuable insights (Polkova et al. 2019); further issues specific to the initialization of the land, ocean, and sea ice components are considered in the next section.

Tackling these diverse and persistent modeling issues effectively will require sustained effort, as simple model-specific solutions may not cure the underlying problems. Ideally this should involve coordination between the S2S/S2D prediction, climate modeling, and data assimilation communities.

**Initialization issues**

**Atmosphere initialization.** Accurate atmospheric model initialization is a basic requirement for numerical weather prediction because atmospheric initial conditions are the primary source of predictability on time scales of less than a week or two (Fig. 1). It is enabled by sophisticated data assimilation systems that are the result of decades of advancement (Bauer et al. 2015). Subseasonal and seasonal prediction systems generally initialize their atmospheric components by such means, with the additional requirement that historical observations must be assimilated similarly to produce reanalyses that are used to initialize hindcasts.
Frontiers in Earth system prediction

New frontiers in S2D prediction have been enabled by Earth system models (ESMs; Flato 2011) that represent the carbon and other biogeochemical cycles in addition to the physical climate system. These frontiers include prediction of ocean and land carbon sinks and biogeochemistry and their important contribution to global carbon storage, as well as ecosystem services. The world’s oceans are a fundamental regulator of global carbon storage and variability. The strength of ocean carbon uptake, together with uptake of carbon by the land, determines the fraction of anthropogenic emissions remaining in the atmosphere, and hence modulates present and future warming. Observed global-mean ocean carbon uptake shows variability on decadal time scales that can be represented by ESMs in which physical climate variables are assimilated (H. Li et al. 2019).

ESM simulations indicate that internal variability of the ocean carbon uptake on decadal time scales is as large as the forced climate change trend (Li and Ilyina 2018), pointing to the potential importance and utility of decadal carbon cycle predictions. Decadal predictions from a number of ESMs are assessing the predictability of the ocean and land carbon sinks and other ocean tracers such as dissolved oxygen. These forecasts are part of the Decadal Climate Prediction Project (Boer et al. 2016) and international programs such as the World Climate Research Programme’s Grand Challenge on Carbon Feedbacks (Ilyina and Friedlingstein 2016). Initial results based on individual models have demonstrated potential predictive skill, assessed through verification against the assimilating reconstructions that provide initial conditions, for ocean carbon uptake in certain regions such as the North Atlantic reaching up to 7 or more years (Li et al. 2016; Lovenduski et al. 2019a), and skill in predicting actual variations estimated from observations (Fig. SB2) has been demonstrated (H. Li et al. 2019).

ESM-based studies also point to the drivers of this predictability. Air–sea CO$_2$ flux mainly varies due to pCO$_2$ changes in the ocean. While thermal influences on pCO$_2$ play a role in shorter-term predictability, predictability beyond 3 years is maintained mainly by nonthermal influences of ocean circulation and biological modification of surface dissolved inorganic carbon and alkalinity (H. Li et al. 2019; Lovenduski et al. 2019a). Investigations in progress are finding potential for multianual prediction of additional biogeochemical fields such as net primary productivity and interior dissolved oxygen concentrations. In addition, potential predictability and skill for terrestrial carbon uptake have also begun to be assessed, with promising initial results (Lovenduski et al. 2019b). These examples demonstrate that skillful multiyear prediction is likely achievable for biogeochemical and ecological Earth system components, and open prospects for the utilization of such information although significant challenges including the paucity of long-term observational data for initialization and verification will need to be overcome.

Fig. SB2. Temporal evolution and predictive skill of global CO2 flux into the ocean from the MPI-ESM-HR decadal prediction system. (a) Annual values of anomalous CO2 flux into the ocean from data-based estimates (SOM-FFN; gray) and MPI-ESM uninitialized simulations (blue), year 2 of initialized decadal predictions (red), and data-constrained assimilation run (black). Anomaly correlations and root-mean-square errors (in parentheses) verifying against SOM-FFN data are indicated. (b) Anomaly correlation skill for global CO2 flux into the ocean, verifying against SOM-FFN. The blue dot and dashed line show the uninitialized skill for which lead time is not relevant, and the red dots show the initialized skill for different forecast years, along with 90% confidence intervals and p values based on a bootstrap approach. (c) As in (b), but verifying against the assimilation run (after H. Li et al. 2019).
Because in situ and remotely sensed atmospheric observations are relatively dense there is generally good agreement between different reanalyses for the modern era implying relatively low uncertainty at heights below about 10 hPa, although temporal inconsistencies can result from changes in observing systems (Long et al. 2017). Because atmospheric initial conditions contribute less to predictability on multiannual time scales, some decadal prediction systems do not initialize the atmosphere (e.g., Yeager et al. 2018).

**Land initialization.** Climatically important land variables such as soil moisture and snow can be initialized by driving land surface models with observed atmospheric fields (e.g., Koster et al. 2009; Sospedra-Alfonso et al. 2016a) or, more directly, assimilation of land observations principally from satellites (Bilodeau et al. 2016; Muñoz-Sabater et al. 2019; Toure et al. 2018). Yet predictability from land surface states is being harvested only to the extent that land initial conditions and the relevant processes are represented realistically in forecast models (Koster et al. 2011; Ardlouze et al. 2017). Historically, land surface and atmospheric models are developed separately and their coupled behavior is not calibrated or validated (Dirmeyer et al. 2018b).

There are also observational limitations. In situ measurements of soil moisture are of varying quality and uneven distribution, and are not designed for real-time operational use (Dorigo et al. 2011). Satellite soil moisture monitoring (Entekhabi et al. 2010; Kerr et al. 2010), provides either very shallow or total-column measurements including groundwater (Li et al. 2012), and is subject to uncertainties caused by vegetation and other factors (Al-Yaari et al. 2017). By contrast, soil moisture in forecast models is mainly a gross reservoir for the surface water balance, and its variations do not represent all of the observed processes, particularly at subgrid scales. Therefore, model soil moisture is only a crude representation of reality, although it still contains useful information that can be largely consistent across different land models (Koster et al. 2009).

Climate forecasts can be improved by making high-quality land-state observations an operational priority for real-time reporting, and planning for long-term continuity in satellite monitoring (Balsamo et al. 2018). This includes vegetation, especially as its interannual variability and cycles of agricultural planting and harvest are not represented and can affect surface fluxes and predictions (Alessandri et al. 2017). In addition, realistic snow initialization can positively impact subseasonal predictions of surface temperatures (e.g., F. Li et al. 2019). Along with coupled land–atmosphere model development (Santanello et al. 2018), such efforts would facilitate improved predictions on weather to subseasonal time scales, as demonstrated by numerous forecast model-based sensitivity studies such as that of Koster et al. (2011).

**Ocean and sea ice initialization.** The importance of initializing the oceans stems from their relatively long thermal and dynamical time scales, through which they play an essential role in S2D climate predictability (Cassou et al. 2018). In addition, the oceans can influence S2S variability, for example, through air–sea interactions affecting the MJO (DeMott et al. 2015) and mesoscale eddy impacts on atmospheric circulation (Saravanan and Chang 2019). Predicting future ocean evolution, especially on S2D time scales, requires estimates of 3D ocean states for initialization. This in turn requires a data assimilation method (usually in conjunction with a dynamical model) to constrain ocean state estimates based on available observations. Similar considerations apply to state estimates of sea ice. Comparisons of different ocean and sea ice state estimates as in Fig. 7 can point to variables and regions for which they are most robust, as well as to where uncertainties are relatively large (Balmaseda et al. 2015; Chevallier et al. 2017). Observing system experiments in which certain observations are withheld have shown, for example, that data from tropical ocean moorings positively impact state estimates even when Argo float data are also available (Fujii et al. 2015).
Recent enhancements in observing capabilities are enabling improvements in ocean and sea ice state estimates, potentially leading to more accurate initial conditions and hence better forecasts. For example, assimilation of satellite measurements of sea surface salinity (SSS) leads to improvements in tropical Pacific ocean states and ENSO forecasts in experiments using an intermediate-complexity coupled model (Hackert et al. 2019), whereas assimilation of satellite-derived sea ice thickness (SIT) measurements has shown potential for improving sea ice forecasts in operational seasonal forecasting systems (Chen et al. 2017; Blockley and Peterson 2018). A major limitation is that these data sources have been available for less than a decade, whereas considerably longer hindcast periods are required for forecast post-processing and skill assessment, and temporal consistency of observational data used for initialization is required to avoid artificial biases between hindcasts and forecasts. Forecasts thus continue to be initialized typically without assimilation of SSS or SIT, from initial conditions that deviate appreciably from available observations especially for SIT (Uotila et al. 2019). This motivates alternative approaches for initializing SIT over multidecadal hindcast periods (Dirkson et al. 2017).

**Coupled data assimilation.** The atmosphere, land, ocean and sea ice components of climate prediction models have often been initialized individually, without coupling. However, such an approach does not make optimal use of observations, which may exert influences across the interfaces of the model components. In addition, physical inconsistencies between the separately initialized components may lead to rapid adjustments, or shocks. To overcome these limitations attention has increasingly turned toward developing coupled data assimilation methods that treat multiple components, such as atmosphere and ocean, simultaneously using observations from each (Penny and Hamill 2017). Such methods are termed weakly or strongly coupled (Penny et al. 2017). Weakly coupled methods apply assimilation independently to each model component within the coupled model, so that the components may exchange information across their interfaces. Such techniques have shown promise for reducing...
shocks (Mulholland et al. 2015), and have begun to be applied operationally (e.g., Browne et al. 2019). Strongly coupled methods apply assimilation to multiple model components in an integrated manner, so that observations assimilated in one component can directly influence others. Such methods remain experimental and thus far have been applied mainly in simplified models (e.g., Penny et al. 2019).

**Ensemble predictions and forecast information**

*Subseasonal to seasonal.* In contrast to ensemble weather forecasts, a consolidated verification strategy for S2S predictions is not yet established, and developing such a framework that encompasses important forecast attributes such as accuracy, association, discrimination, reliability, and resolution has thus emerged as a priority (Coelho et al. 2018). (Accuracy measures error, or distance between forecast and observed values; association measures strength of the linear relationship between forecast and observation as through temporal or spatial correlations; discrimination measures by how much forecasts differ given different outcomes; and reliability measures how well forecast probabilities correspond to observed frequencies of occurrence; resolution measures by how much outcomes differ given different forecast probabilities. Forecast quality encompasses all these attributes, whereas skill indicates quality relative to some benchmark such as persisted anomalies or climatological probabilities.) As for seasonal predictions, a purpose of S2S hindcasts is to provide a larger sample for more confident verification statistics than real-time forecasts because they cover more years. However, because S2S hindcasts are initialized using reanalysis and most often have a smaller ensemble size, their verification generally underestimates real-time forecast quality. Operational centers are encouraged to compute and monitor verification statistics based both on hindcasts and real-time forecasts.

As has been demonstrated for seasonal prediction, S2S multimodel ensembles (MMEs) generally outperform individual models (Vigaud et al. 2017; Pegion et al. 2019). Currently, the S2S and SubX MME projects are providing testbeds for research\(^3\) as well as a foundation for operational use (Vitart and Robertson 2019; Pegion et al. 2019). One focus for exploiting such datasets is developing calibration procedures, postprocessing steps that improve the properties of probabilistic forecasts, to enable S2S ensemble

---

\(^3\) Hindcast and near-real-time forecast data are available from S2S (at www.s2sprediction.net) and from SubX (at http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/).
forecasts to provide reliable probabilities for particular conditions occurring or thresholds being exceeded, especially for extreme events. The varied current choices among S2S project modeling systems for hindcast and near-real-time initialization dates, hindcast period and ensemble size is, however, limiting advances in developing multimodel calibration and combination procedures. In addition, the value of these datasets for research would be enhanced if more comprehensive stratospheric data were to be available across models.

S2S ensemble forecasts have shown promise in providing useful predictions and early warnings for high-impact climate and weather events including severe heat waves and cold spells, as well as regional probabilities of the occurrence of tropical storms as illustrated in Fig. 8 (Vitart and Robertson 2018). Examples include severe cold conditions over Europe associated with the negative phase of the NAO, whose useful predictability into week 3 is enhanced by tropical–extratropical teleconnections resulting from MJO activity (Ferranti et al. 2018), and atmospheric rivers, plumes of intense water vapor transport that often trigger weather and hydrologic extremes and are especially predictable at lead times of 3–5 weeks during certain MJO and QBO phase combinations (Baggett et al. 2017). While modest overall skill at ranges longer than a week has been found for S2S predictions of springtime Sahelian heat waves including measures of heat stress, such conditions following a strong El Niño were accurately forecast, pointing to the tropical Pacific as a source of predictability for extremes in that region (Batté et al. 2018).

A global precipitation hindcast quality assessment of the S2S prediction project models (Fig. 9) was performed by de Andrade et al. (2019). Subseasonal prediction quality is modulated by the MJO, QBO, and ENSO in the tropics, changes in large-scale flow in the extratropics, and stratospheric tropical and extratropical variability (Butler et al. 2019). It is therefore important to estimate the predictive skill of such events and identify their impacts on predictions of weather and weather extremes. Evaluating the conditional prediction quality associated with the key low-frequency variability modes is instrumental for better understanding S2S predictability mechanisms. For example, MJO predictive skill in the S2S MME ranges between 12 and 36 days and is affected both by the MJO amplitude and phase errors (Vitart 2017; Lim et al. 2018; H. Kim et al. 2018). Communicating these variations in forecast quality, including if the forecasts are no better than climatology, is extremely important as users with such knowledge can better

Fig. 9. Global averages of correlations between hindcast and observed precipitation anomalies over the 80°S–80°N latitudinal band for weeks 1–4 for S2S project models initialized from November to March, 1999–2009. (left) Hindcast quality assessment based on ensemble means using the full ensemble size for each model, as indicated in the figure legend. (right) Hindcast quality assessment based on ensemble means using three ensemble members for each model. The reduced spread of the lines shown in the right panel when ensemble sizes are identical compared to the spread of the lines shown in the left panel demonstrates the value of using larger ensembles for subseasonal precipitation forecasting (adapted from de Andrade et al. 2019).
utilize and benefit from the forecast information. Furthermore, capitalizing on “windows of opportunity” when skill is especially high increases the value of S2S forecasts (Mariotti et al. 2020), and motivates their frequent initialization (ideally daily).

Seasonal to decadal. Limited forecast quality in current S2D ensemble prediction systems motivates research initiatives that focus on extracting skillful and reliable information from the large amounts of forecast and hindcast data available to potential users.3

One emerging theme of such research is that S2D prediction systems sometimes underestimate the predictable signal (Eade et al. 2014; Scaife and Smith 2018). As a result, very large ensembles that effectively filter out unpredictable noise demonstrate higher skill in predicting phenomena such as the winter NAO (Scaife et al. 2014b; Dunstone et al. 2016) and seasonal to multiannual regional precipitation variations (Dunstone et al. 2018; Yeager et al. 2018) than was previously thought possible. While very large ensemble sizes hold value for isolating weak predictable signals, much smaller ensemble sizes are sufficient for skillful prediction of tropical SST, for which signal-to-noise ratios are much larger (Zhu et al. 2015). The causes of unrealistically low modeled predictable signals (sometimes called the “signal-to-noise paradox”) remain under investigation. Two hypotheses stemming from hindcast experiments are that winter NAO skill is enhanced by skillful prediction of a QBO teleconnection that is too weak in models (O’Reilly et al. 2019), and that transient eddy feedbacks are too weak in models (Scaife et al. 2019). Others based on simple models suggest that the NAO predictable signal is too weak because climate models switch between NAO regimes too rapidly (Strommen and Palmer 2019), or exhibit less persistent NAO variability than is observed (Zhang and Kirtman 2019).

In the case of the winter NAO, which is a key source of variability over the midlatitude North Atlantic and Europe, another approach to extract relevant information from overdispersive ensembles that leads to improved skill is to subsample ensemble members that are close to a “first guess” statistical prediction of the NAO (Dobrynin et al. 2018); subsampling has shown potential for improving European summer forecasts as well (Neddermann et al. 2019).

Estimating and realizing the predictability of key modes of variability is still a major challenge at S2D time scales. ENSO is considered one of the most predictable phenomena on multiseasonal time scales, but longer-range skill has been viewed as limited. However, multiyear ensemble predictions have shown evidence of skill in predicting long-lasting La Niña events that follow warm events up to 24 months ahead (DiNezio et al. 2017a; Luo et al. 2017). Challenges in the initialization of such longer time-scale predictions remain, as evidenced by multiyear predictions in which skill for SST and precipitation over land improves with lead time in some areas, suggesting that short-term adjustments following initialization may tend to degrade skill (Yeager et al. 2018).

Calibration of ensemble forecasts is a necessary step to reduce the areas for which S2D forecasts are unreliable and potentially misleading. Combinations of several forecasting systems such as the North American Multimodel Ensemble (NMME; Kirtman et al. 2014) are now routinely used to increase ensemble reliability and improve forecast skill. Several recent efforts have explored weighted multimodel calibration methods to combine ensembles from different models in order to improve probabilistic seasonal forecasts for temperature and precipitation anomalies as well as forecasts of extremes (Becker 2017). Calibration methods have also been developed for ensemble decadal hindcasts to adjust both the bias and ensemble spread with a parametric dependency on lead time and initialization time (Pasternack et al. 2018). Such methods are found to improve both the conditional bias and probabilistic skill of decadal hindcasts.

Seasonal hindcast data from the WCRP Climate System Historical Forecast Project (CHFP; Tompkins et al. 2017) are available online (at http://chfps.cima.fcen.uba.ar/access.php), and data from the North American Multimodel Ensemble (Kirtman et al. 2014) including real-time forecasts are also online (at https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/). Decadal hindcast data from the WCRP Coupled Model Intercomparison Project phases 5 and 6 are available online as well (via https://esgf-node.llnl.gov/projects/cmip5/ and https://esgf-node.llnl.gov/projects/cmip6/).
Climate forecasts for decision-making

Subseasonal to seasonal. Many decisions are made on the S2S forecasting time scale, which sits between weather forecasts and S2D climate outlooks; therefore, the continued development of S2S forecasts has the potential to benefit many sectors of society (Fig. 10). S2S forecasting is a rapidly maturing discipline, with emerging recognition for both the need and the potential use of forecasts on this time scale (White et al. 2017). While S2S forecasts are increasingly being used in government as well as a range of sectors including agriculture, energy, finance, health, and water resource management, more engagement between S2S forecasters and end users is needed to increase the wider awareness and uptake of S2S forecasts.

Although scientific knowledge gaps, computational capacity, and gaps in observations and modeling currently limit the use of S2S forecasts to some degree, by increasingly placing decision-makers at the forefront of S2S forecast development, an improved dialogue between S2S forecasters, developers and end users will accelerate the awareness and application of S2S forecasts to real-world decision-making.

To support the increased use of S2S forecasts for decision-making, the following recommendations were identified for action following the Boulder conference:

- A summary of existing stakeholder case studies is planned to be created to demonstrate past and ongoing “success stories,” and support better engagement with end users and stakeholders. As the S2S forecast needs and associated performance varies greatly between different sectors and users, the wider community is increasingly working together on the codesign and production of S2S predictions in order to better meet user needs. Several applications of S2S forecasts are now being developed in different disciplines, such as the EU-funded Subseasonal to Seasonal Climate Forecasting for Energy (S2S4E) project in the energy sector, a quasi-operational excess heat outlook system in the health sector (Lowe et al. 2016), and S2S hydrologic prediction in the water management sector. These efforts need to be catalogued and disseminated to guide further user-driven decision-support products, and to support the continued development of S2S forecast, verification metrics and related services.

- Systematically assessing the relative skill (or lack thereof) of forecasting a series of historical high-impact events, such as heat waves, extreme rainfall events, or wildfires, on the S2S time scale would be a useful way to help demonstrate the potential of S2S forecasts to decision-makers across multiple sectors. At present, such case studies are often ad hoc and typically not published in the wider

Fig. 10. Schematic illustration of relationships between an S2S forecast range of 10–30 days and other prediction time scales, including examples of actionable information that can enable decision-making by various sectors. Indicated actions are examples that are not exclusive to a particular forecast range (after White et al. 2017).
literature; however, a collaborative effort that brings together a set of demonstrable case studies, involving both forecasters and end users, would fill this gap. For example, a series of “tailored narratives,” or “storylines” (approaches that construct stories of plausible, non-probabilistic climatic futures that relate to a specific person or sector to counter perceived barriers; e.g., Hazeleger et al. 2015), may aid in the understanding of what S2S forecasts may deliver in the future.

• To support the codesign, uptake and use of S2S forecasts, S2Sapp.net is being established as a new network of researchers, modelers and practitioners—an “open to all” global community with a shared aim of exploring and promoting cross-sectoral services and applications of this new generation of forecasts from across government, academia, and the private sector.

**Seasonal to decadal.** Research efforts are assessing the value of S2D forecast information for many applications, and initiatives such as the WMO’s Global Seasonal Climate Update5 and Annual to Decadal Climate Update (Kushnir et al. 2019) are making such information more widely available. However, consultation with decision-makers is essential in order to tailor forecast information to the needs and expectations of users.

Fisheries management is one application for which S2D forecast information holds promise (Tommasi et al. 2017). This is due to reasonable skill for ocean prediction in regions of interest, coupled with strong influences of S2D climate variability on fish populations. Case studies employing fisheries management decision frameworks have shown that SST forecast information can potentially increase fishery yields while reducing the risk of population collapse from combined effects of environmental factors and overfishing. However, significant challenges remain for fully realizing this potential. These include a need for improved initialization and reduced model errors in key ocean regions such as the U.S. Northeast continental shelf, dynamical downscaling in cases where important spatial scales are not resolved by global models, and sufficiently accurate observational data for hindcast verification on these scales. In addition, incorporating biogeochemistry and marine ecosystem components into prediction systems will expand their potential capabilities, while posing additional verification challenges.

Another current focus of application-oriented research is water management. Global climate prediction models have been shown to have skill in predicting the next winter season’s snowpack throughout much of the western United States, where spring snowmelt is an essential water resource (Kapnick et al. 2018; Sospedra-Alfonso et al. 2016b). Because temperature influences snowmelt and runoff efficiency, skill in seasonal temperature forecasts can provide improved skill for seasonal water supply forecasts in this region (Lehner et al. 2017). Seasonal forecast skill has also been demonstrated for monsoon rainfall (e.g., Jain et al. 2019) and drought (Hao et al. 2018) with potential to inform water management decisions in many regions of the globe. Decadal forecasts potentially can meet planning horizon needs but currently are less familiar to water managers than seasonal forecasts and long-term climate projections. Efforts to apply decadal climate information for water management decisions have included assessing the role of decadal modes of variability, and using statistically downscaled decadal predictions as hydrological model inputs. Developing information that is credible and compatible with existing decision frameworks is an important consideration (Towler et al. 2018).

Additional sectors for which S2D forecasts are being assessed for decision-making include agriculture (Klemm and McPherson 2017), energy (demand and wind power generation, Clark et al. 2017; Lledó et al. 2019), tropical cyclone (Bergman et al. 2019) and coastal flooding (Widlansky et al. 2017) preparedness, Arctic marine transportation (Stephenson and Pincus 2018), wildfire risk (Turco et al. 2019), and food security (Funk et al. 2019).

---

Initiatives to develop and deliver climate forecast information range in scale from international, regional, and national (e.g., Marotzke et al. 2016), to individual users, all of which aim to provide forecast information having practical value for decision-makers. In all cases, it is crucially important that uncertainties are adequately quantified and conveyed in order to avoid any false sense of certainty and to build trust in forecast information providers, although sometimes this requires overcoming a preference of users for deterministic information. Additional considerations are that expectations of users need to be conditioned to generally modest levels of skill, but that this information can nonetheless be advantageous when applied consistently in the long term. The likelihood that climate forecast information gets used increases when efforts are made to build relationships with potential users, and dialogs are opened to enable forecast products to be codesigned (Kolstad et al. 2019).

Crosscutting issues in S2S and S2D prediction

Initialization shock and model error. Model biases are an endemic modeling issue that is common across S2S and S2D prediction time scales and influence all aspects of the prediction systems—complicating ingestion of assimilated observations, degrading skill, and necessitating postprocessing steps such as bias correction and calibration for product development and delivery.

Model biases begin to develop on fast time scales and lead to drifts from the climate represented by the initial conditions to that of a model’s biased equilibrium state. It has been extremely hard to understand the mechanisms behind these drifts, and further, pathways for their diagnosis are not clear although some progress is being made (Sanchez-Gomez et al. 2016; Shonk et al. 2018; Voldoire et al. 2019). Such difficulties arise due to nonlinear interaction between various physical processes that are parameterized, and because biases could be nonlocal in their origin. Long time scales before models’ equilibrium states are attained make understanding the causes of drifts even harder. The Boulder meeting recognized that the S2S–S2D prediction community needs to pay particular attention to developing pathways for understanding the onset of model biases and put together mechanisms (such as summer schools) to train the next generation of scientists with interest and expertise in climate modeling and model diagnostics.

Initialization shocks that arise from imbalances in initial states with respect to the formulation of the model and can be caused by limitations of observations and data assimilation as well as model biases were also recognized as a major issue, particularly in the context of decadal predictions. Initialization shocks result in the degradation of initial information that may be the primary source of predictability for the subsequent forecast. Even after considerable research and investment in decadal predictions it is still not clear what may be best approaches, such as between full field versus anomaly initialization, to retain predictive information in the initial state while minimizing the influence of initial shocks on the subsequent forecast. The continuing prominence of model drift and initial shocks as important issues reinforces a long held sentiment that these are outstanding problems that need to be studied more systematically if progress in translating inherent predictability into prediction skill is to be made.

S2S and S2D research interactions. The examples of interaction among modes of variability across S2S and S2D time scales noted earlier emphasize the fact that continued interaction and communication across the S2S and S2D research communities will be important to make progress. Furthering our understanding of time-scale interactions will require investments in process-level understanding of these phenomena and will not only benefit our understanding about their lower-frequency variations but will also contribute to improved process-level diagnostics of model simulations. Better understanding of time-scale interactions is likely to
require the use of a hierarchy of models, such as simple linear models to investigate the characteristics of tropical–extratropical interactions (including their influence on storm tracks), to diagnose possible causes for errors in their representation in complex GCMs (Dias et al. 2019).

Another aspect of research interactions across time scales is quantifying the fidelity of models in S2S and S2D prediction as well as projections of climate on longer time scales based on their simulation and prediction of shorter time-scale phenomena. The advantage of such an approach is that much larger samples for predictions of shorter time-scale phenomena are available, and an assessment of the reliability of such predictions can be used to build confidence in prediction on longer time scales. Theoretical basis for extrapolating the reliability of forecasts across different time scales may also require the use of a hierarchy of models (Weisheimer and Palmer 2014; Christensen and Berner 2019).

**Research and operations.** Postprocessing to improve forecast quality is an important area of research that bears directly on operational activities. Postprocessing is necessary because biases in forecasts can be as large as the predicted signal, and therefore require the use of bias correction and calibration techniques to adjust real-time predictions before their delivery to the users. These requirements are shared across subseasonal to decadal prediction time scales; however, because of different levels of experience (seasonal predictions having the longest history) the opportunity for cross-community interactions was recognized. Some aspects for postprocessing are specific to time scale, for example, bias correction for decadal predictions requires methods to account for the nonstationarity of climate, and research needs to develop such methods were stressed.

Necessity for postprocessing requires an extensive set of hindcasts to accompany real-time predictions. Because of limited resources, decisions about hindcast period, ensemble size, and forecast start dates are not straightforward and call for further guidance from the research community. Such questions about the operational infrastructure for long-range prediction systems, including ensemble generation techniques and recommendations for harmonizing hindcast and real-time forecast production, provide an opportunity to link operational and research communities that was highlighted during the conference.

Product development and communicating forecasts to the user community is also a common thread across the S2S and S2D communities. Communication of probabilistic forecast information (including confidence in the forecast based on past verifications) to users for their decision-making has been a challenge, and once again there is much to be gained from lessons learned from the experiences of different communities. Similar challenges and opportunities also exist in the context of product development that incorporate user needs based on an ongoing dialog from the very start of the process. In addition, users often wish to have information on smaller spatial scales than are represented in global climate models. For such applications either statistical or dynamical downscaling is possible and can be effective in reducing local climatological biases, although clear demonstrations that downscaling can improve the skill of climate predictions remain elusive (e.g., Manzanas et al. 2018).

In summary, research needs for further development of operational infrastructure, product generation, and communication of probabilistic forecasts were themes often repeated during the conference.

**Conclusions and the future of subseasonal to decadal prediction**

This paper has outlined many commonalities in the prediction of weather and climate across time scales and Earth system components, and through the value cycle from basic research to operational delivery.

Earth’s weather and climate is inherently chaotic and challenges the best currently available modeling capabilities. There remains, however, untapped skill, and realizing this skill
will require improvements on numerous fronts. These include fundamental understanding of finescale processes, leading toward their robust parameterization; accurately representing property exchanges across Earth system components through realistic coupling limiting systematic errors; sustained Earth observing systems and advanced data assimilation methods enabling balanced initial conditions that avoid shocks and mitigate model drifts; and innovative numerical and ensemble generation techniques to address model scalability issues. Additional important avenues toward improved services include development of probabilistic information for high-impact weather and climate events including unprecedented extremes, and optimal postprocessing and data fusion to add value to multimodel ensembles, among many others.

These challenges are broad but so are opportunities for steady progress, ranging from curiosity-driven science to the systematic model evaluation and improvement in a collaborative and open research/operational environment.

The joint WWRP–WCRP conferences in Boulder clearly demonstrated the benefit in bringing relevant stakeholders together to cross-fertilize their experience, knowledge, respective issues and working cultures, which will surely help frame a new and vibrant research portfolio, and inspire the next generation of science leaders to ensure that society has access to the best possible weather and climate prediction science.

Acknowledgments. The International Conferences on Subseasonal to Decadal Prediction on which this paper is based were sponsored by U.S. CLIVAR, NSF, UCAR, NCAR and its Climate and Global Dynamics Laboratory (CGD), NOAA's Climate Variability and Predictability (CVP) and Modeling, Analysis, Predictions and Projections (MAPP) Programs, Copernicus Climate Change Service, IPSL, and WWRP/WCRP's Subseasonal-to-Seasonal (S2S) Prediction Project.

References


multidecadal variability using the GFDL CM2.1 and NCAR CESM1 global coupled models. J. Climate, 30, 2785–2810, https://doi.org/10.1175/JCLI-D-16-0127.1.


