To: National Science and Technology Council's (NSTC's) Fast Track Action Committee for Earth System Predictability (ESP-FTAC)

RFI on Earth System Predictability Research and Development

This response comes from the following individuals on behalf of the **Ecological Forecasting Initiative** (https://ecoforecast.org/) [Emails were removed for the website version of this document]

- Michael Dietze
- Tess McCabe
- Jake Weltzin
- Alexey Shiklomanov
- Andrew Fox
- William K. Petry
- Melissa Kenney
- Gyami Shrestha
- Rebecca Chaplin-Kramer
- Ethan White
- Güray Hatipoğlu
- Jake Zwart
- Jeff Atkins
- Thilina Surasinghe
- Ge (Jeff) Pu
- Christa Torrens
- Kai Zhu
- Bruce E. Wilson
- Chris Jones
- Nico Franz
- Daijiang Li
- Ann Raiho
- Yang Song
- William Hammond
- Alex Young
- James Guinnip
- Tim Keitt
- Quinn Thomas
- Leah R. Johnson
- Eric R. Sokol
- Alexa Fredston
- Umesh Gupta
- Daniel Gianotti

1. Needs and benefits: Widespread environmental change is the defining challenge of our time, with impacts occurring across all levels of society from individuals to humanity as a whole[1,2]. We are facing continual, accelerating change, as demonstrated by the recent pandemic, wildfires, droughts, and coral bleaching. Our ability to understand, manage and conserve natural systems, and sustain economic growth in the 21st century, requires a capacity to anticipate changes in the Earth System at a scale and speed beyond what is currently possible.

Being able to predict changes in the **biotic** components of the Earth system is central to the predictability of the Earth system as a whole, the sustainable use of living resources, and the socio-economic benefits of improved prediction. In the National Academies "Next Generation Earth System Prediction" report, a top recommendation was to "Include More Components of the Earth System in S2S Forecast Models", and specifically to advance prediction capabilities for soil-state and seasonal vegetation growth and aquatic and marine ecosystems[3]. The important role of vegetation and soil moisture in controlling the exchange of energy, water, and momentum between the land surface and the atmosphere is relatively well understood over short timescales[4–6]. Over longer timescales Earth system predictability can be enhanced by more effective modeling of the slow evolution and long memory of the biosphere[7], exploiting land-atmosphere interactions to better predict radiative forcing and GHG changes associated with land use and land cover change. Improved modeling, parameterization, and initialization of ecosystem processes will increase and extend model prediction skill for events such as droughts, heat waves, floods, monsoons, and storm formation[8–12].

The biosphere is the life support system of the planet and contributes an estimated $145 trillion annually to human well-being; twice the global GDP[13]. Socio-economic benefits of improved prediction affect any sector, community, system, or industry that have a biotic component[14], for example: agriculture, forestry, fisheries, water resources management, tourism and recreation, disaster management, human health, energy, and infrastructure. Furthermore, it is clear that there are unmet management opportunities where improved prediction would lead to improved decision making[15–17]. For example, literature surveys found that 30-80% of natural resource and conservation management operations experienced unanticipated weather impacts, predominantly on sub-seasonal to interannual timescales[18]. Improved predictions, alone or combined with projections under alternative management scenarios, will allow society to anticipate challenges on decision-relevant timescales, to adapt to change, and to improve decisions at all scales, from individual citizens to organizations to nations.

Improving prediction within the biosphere is not only urgent, but it is also timely and achievable. The capacity to generate and improve predictions is fueled by recent advances in sensor technologies, satellite-based observation systems (e.g. NASA's Earth Observing System), genomic tools, community science initiatives, and shifts toward large-scale networked science (that leverage recent advances in physical, cyber and communications infrastructure) that now provide access to previously unimaginable volumes of near real-time environmental data. Standardized observatories, such as NSF's National Ecological Observatory Network (NEON) and Ocean Observatories Initiative (OOI), and long-term monitoring data (e.g. USGS, NOAA, USDA) are particularly valuable in these efforts, both directly and for connecting high-volume satellite data to process-level detail across scales. Increases in data volume, velocity, and openness are having a revolutionary impact on efforts to expand and formalize prediction for the biosphere.

Increasing data availability has enabled novel, iterative predictions for the biosphere with the potential to directly address the socio-ecological challenges of the 21st century[19]. By continually confronting short- to medium-term predictions with new observations, iterative prediction systems are a win-win that simultaneously provide more societally-relevant information on decision-relevant timescales while accelerating research and improving our fundamental understanding of predictability[20]. For example, real-time forecasting improves the process of research by forcing scientists to make specific, quantitative, and falsifiable predictions that are “pre-registered”[21] and can be rapidly validated against out-of-sample (future) data.

Beyond making useful predictions, our theoretical understanding of predictability has recently emerged as an important research area for the biosphere[22–24]. One key question asks which uncertainties dominate predictions on different timescales: how long do initial conditions matter, when do long-term climate scenario uncertainties take over, and which uncertainties (model structure, inherent stochasticity, environmental heterogeneity, parameter variability) dominate in between? The answers to these questions are likely to vary across systems and scales, and a comparative approach is necessary to understand the patterns to predictability in the biosphere[19]. Competing hypotheses consider the role of the physical environment, biotic interactions, biological traits, and evolutionary constraints. Similar questions exist about the transferability of forecast models across locations and study systems (e.g. how similar in structure and parameters are models for different harmful cyanobacteria blooms?). Understanding the patterns to predictability across the biosphere addresses discipline-spanning grand challenge questions in biology. Furthermore, it has a direct impact on practical science priorities by highlighting what monitoring data are most needed, when we need to model specific systems in more detail, or how we could structure more holistic/integrative models of general properties.

2. Gaps and barriers: Although there is much to be gleaned from the theory and approaches used to predict the physical Earth system, the biosphere poses unique challenges that may require different solutions. Other parts of the Earth system (atmosphere, ocean, etc.) have well-defined governing equations based on fluid mechanics and thermodynamics. The biosphere has these too, but adds complexities across a whole cascade of scales from genomes to cells to organisms to ecosystems. Just one of the estimated 8.7 million species on Earth[25] may contain gigabytes of information in its genome and we possess genomic data for a mere 0.04% of species. Furthermore, our ability to predict biological function at the cellular- or organismal-scale based on genomes is in its infancy and largely restricted to the bacteria. Because **we remain far from being able to scale up to processes at the ecosystem- and biosphere-scale based on first-principles**, we are strongly reliant on our (incomplete) understanding of emergent phenomena and empirical calibrations. This notorious messiness of biology does not preclude prediction, but leads to important challenges to how we generate predictions (a mix of semi-mechanistic, statistical, and machine learning approaches) and how we study predictability (comparative approaches, numerical simulation, uncertainty analysis, scaling theory)[24].

Ecological systems are tightly coupled to human systems and respond rapidly to human decisions. Furthermore, some predictions (e.g. fisheries) will have to account for the fact that humans will make different decisions depending on a forecast's output, and those decisions can change the outcome of the prediction (similar to what is seen in economic forecasts). This coupled representation of ecosystem dynamics, human action, response, and outcome illuminates the challenges of predicting natural-human system states and behaviors, but also opportunities to improve decision support and our understanding of human-natural system predictability. Predictions need to be better integrated into socio-economic benefit analyses such as ecosystem service valuation and value of information studies[19]. Earth system modeling is about putting the pieces together, and while predictability is complicated it can be improved by the coupling between the components. Models of this type must be capable of interrogating the feedback uncertainties and process dynamics, and provide "experimental worlds" where the impact of different policy interventions can be tested[26].

Finally, compared to one-off analyses, **iterative predictions have much higher requirements for reproducibility, robustness, computational efficiency, and automation**. Successful forecasting systems for the biosphere will require sustained funding for the development and maintenance of models, tools, cyberinfrastructure, and for the technical training of the next generation of scientists in such methods. As noted earlier, these predictions also require uncertainty estimation and propagation, which requires more sophisticated analytical techniques, is more computationally intensive, and produces larger and more complex results that are more difficult to store, distribute, and communicate. In addition to these general forecasting challenges, some additional cyberinfrastructure challenges are unique to ecological predictions. For one, the diversity of study systems and associated predictive models in ecology means that ecological data vary in type, spatial and temporal scale, format, and distribution method. This heterogeneity poses methodological challenges (e.g. how to model relationships between variables of different type and scale) and cyberinfrastructure challenges (e.g. processing pipelines; data formats).

3. Opportunities and activities:

Opportunity 1 - Identify dominant uncertainties through synthesis: One knowledge gap that is critical, but at the same time actionable over the next ten years, is to focus efforts on **improving our understanding of the relative contributions of different uncertainties to the predictability of the biosphere and coupled human-natural systems**. Early on, the uncertainty in numerical weather prediction (NWP) was shown to be driven primarily by the unstable, chaotic nature of the atmosphere, and therefore forecast improvement required increasing the volume and precision of observations of the present state (i.e., initial conditions)[27–29]. This theoretical insight into predictability had profound implications on the design of models, monitoring efforts, and data assimilation systems. By contrast, while the biosphere and coupled human-natural systems are replete with nonlinearities there are few definitive examples of chaos[30–32]. Therefore diagnosing the predictability of these systems requires that additional uncertainties be quantified and analyzed: drivers, model parameters, model structure, system heterogeneity (statistical random effects), and inherent stochasticity. Similar to NWP, which of these uncertainties dominate in which situations will have major impacts on how we monitor and model the biosphere, how we assimilate new information into our predictions, and how we make decisions under uncertainty. Although we can derive some basic expectations about the relative importance of each of these terms[24], much of our understanding of predictability needs to be empirical. For example, while improved atmospheric forecasts at the subseasonal-to-seasonal and interannual-to-decadal timescales would improve many biological forecasts, the observation that many biological (and societal) processes are time-integrating and possess substantial memory means that it is not an *a priori* given that the current skill of atmospheric models is insufficient to produce societally-useful and scientifically-interesting biological predictions. The unfortunate reality is that existing analyses of predictability are largely incomplete, especially when it comes to the large-scale processes in Earth system models, with only a subset of uncertainties being considered in any given study, and multiple uncertainties frequently being convoluted (e.g. parameters, model structure, initial conditions)[33–35]. **Complete uncertainty analyses** have recently become possible at small scales[36], but a robust partitioning of errors is labor-intensive and **needs to be scaled up**. This needs to be done not just for individual study systems, but also through databases and repositories - both virtual (data) and physical (samples) - that will allow for large-scale synthesis and comparative analyses of predictability across systems and scales. Such syntheses will tackle grand challenge scientific questions and have a revolutionary impact on our ability to understand, manage, and conserve the biosphere.

Opportunity 2 - Build Community Cyberinfrastructure: Although some ecological forecasts run within fully-coupled Earth system models, many more are simpler to implement offline. Currently, ecological forecast systems are developed independently, leading to large redundancies that increase the time, costs, and learning curve needed to develop and operate these forecasts. Shared, community cyberinfrastructure would **increase economies of scale and accessibility**, making it easier for different groups to deploy and manage forecasts for many different ecological systems[37]. Forecast pipelines should be built on **flexible, composable, and reproducible modules** that can be deployed across modern distributed systems, and adhere to **community-driven standards** and **unified development approaches** that cover cases across a wide range of biological subdisciplines. Forecast inputs and outputs should also strive to be **"FAIR"—Findable, Accessible, Interoperable, and Reusable**— for both human users[38] (e.g. help decision makers find and use forecasts) and "machine" users (e.g. automatically ingest new data)[39]. Key components include data ingest tools that can handle both data volume and heterogeneity, standards for forecast inputs, outputs, and metadata, public archives, and tools for data assimilation, uncertainty propagation, visualization, and dissemination. In some cases, tools and software for addressing these needs already exist[37], though creating and maintaining forecasting pipelines that use these components still requires specialized training. Other forecasting challenges—particularly those outlined in the previous sections—demand the development of new techniques and technologies that could be accelerated by a collective community platform. Importantly, the methodological and technological advances necessary for successful ecological forecasting systems may produce positive externalities in the same way that comparable advances in weather forecasting have found applications far beyond atmospheric science.

Opportunity 3 - Grow human and institutional capacity: One of the most critical, actionable opportunities to advance biosphere prediction is the **need to strengthen and sustain human and institutional capacity**. Although predictive approaches are routine in other Earth system sciences (e.g. NWP), they are relatively new to biology. Few undergraduates are exposed to the role of prediction and forecasting in research, resource management, policy-making and conservation of the biosphere. Few graduate programs offer explicit training in the concepts, tools, and techniques used by ecological forecasters, or in the management and decision science approaches that take advantage of this information, and there is a need to bring existing researchers and managers up to speed.

Another aspect of capacity building revolves around the critical **need to identify and engage stakeholders and end users** in the process of developing and applying predictions. Forecasts are only useful if they are used. One approach to maximize the utility of predictions is to employ co-production, the explicit partnering of producers and consumers through the life-cycle of forecast development. Co-production requires communication, trust and time, and relies on partnerships that span federal, state, and local agencies, academia, industry, tribes, NGOs and their stakeholder communities including the public.

There is also a critical **need for interagency coordination and collaboration**, similar to the U.S. Carbon Cycle Science Program/Carbon Cycle Interagency Working Group and its science community-led North American Carbon Program (NACP), to unite researchers, networks, and international government agencies around a central goal. Coordination minimizes the potential for unnecessary redundancy across agency efforts, helps identify and prioritize unmet needs, minimizes competition while maximizing complementarity, and improves the potential for successful technology transfer (e.g., from academia to decision-makers). Explicit consideration of coordination, roles and responsibilities provides a critical opportunity to address the current 'gulf' between research and operations. Many promising research-grade forecasts are not brought into operations because the agencies that support research frequently lack the mandate to operate forecasts. This is compounded by the dearth of support mechanisms (funding, coordination) to transfer forecasts to organizations (e.g., agencies, industry, or NGOs) that have operational authority and capabilities.

Big Idea: Establish a National Center The theoretical, computational, social, and organizational challenges to making biosphere predictions are deeply intertwined. Tackling these challenges simultaneously in a coordinated manner will achieve far greater return on investment than a piecemeal approach focused on individual components. We recommend[14] establishing a National Center focused on prediction and predictability in the biosphere. By **coordinating cyberinfrastructure, stakeholder engagement, basic natural and social science research, and operations** across agencies, academia, and industry we can achieve deep synergies and economies of scale that will drive rapid scientific advances and improved decision making.

Works Cited

1. IPCC. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press.
2. IPBES. *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES secretariat, Bonn, Germany.* (2019).
3. National Academies of Sciences, Engineering, and Medicine. *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts.* (2016).
4. Green, J. K. *et al.* Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nature Geoscience* **10**, 410–414 (2017).
5. Seneviratne, S. I. *et al.* Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews* **99**, 125–161 (2010).
6. Swann, A. L. S., Hoffman, F. M., Koven, C. D. & Randerson, J. T. Plant responses to increasing CO2 reduce estimates of climate impacts on drought severity. *PNAS* **113**, 10019–10024 (2016).
7. Guo, Z., Dirmeyer, P. A. & DelSole, T. Land surface impacts on subseasonal and seasonal predictability. *Geophysical Research Letters* **38**, (2011).
8. National Research Council. *Assessment of Intraseasonal to Interannual Climate Prediction and Predictability.* (2010).
9. Dirmeyer, P. A., Halder, S. & Bombardi, R. On the Harvest of Predictability From Land States in a Global Forecast Model. *Journal of Geophysical Research: Atmospheres* **123**, 13,111-13,127 (2018).
10. Halder, S., Dirmeyer, P. A., Marx, L. & Kinter, J. L. I. Impact of Land Surface Initialization and Land-Atmosphere Coupling on the Prediction of the Indian Summer Monsoon with the CFSv2. *Front. Environ. Sci.* **5**, (2018).
11. Koster, R. D. & Walker, G. K. Interactive Vegetation Phenology, Soil Moisture, and Monthly Temperature Forecasts. *J. Hydrometeor.* **16**, 1456–1465 (2015).
12. Roundy, J. K. & Wood, E. F. The Attribution of Land–Atmosphere Interactions on the Seasonal Predictability of Drought. *J. Hydrometeor.* **16**, 793–810 (2014).
13. Costanza, R. *et al.* Changes in the global value of ecosystem services. *Global Environmental Change* **26**, 152–158 (2014).
14. Chaplin-Kramer, R. *et al.* Global modeling of nature's contributions to people. *Science* **366**, 255–258 (2019).
15. Banger, K., Yuan, M., Wang, J., Nafziger, E. D. & Pittelkow, C. M. A Vision for Incorporating Environmental Effects into Nitrogen Management Decision Support Tools for U.S. Maize Production. *Front. Plant Sci.* **8**, (2017).
16. Bradford, J. B., Weltzin, J. F., McCormick, M. & Baron, J. *Ecological Forecasting: 21st century science for 21st century management*. 85 (2020).
17. Wu, W. *et al.* Ensemble flood forecasting: Current status and future opportunities. *WIREs Water* **7**, e1432 (2020).
18. Tulloch, A. I. T., Hagger, V. & Greenville, A. C. Ecological forecasts to inform near-term management of threats to biodiversity. *Global Change Biology* (2020).
19. Dietze, M. C. *et al.* Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc Natl Acad Sci U S A* **115**, 1424–1432 (2018).
20. Bradford, J. B., Betancourt, J. L., Butterfield, B. J., Munson, S. M. & Wood, T. E. Anticipatory

natural resource science and management for a changing future. *Frontiers in Ecology and the Environment* **16**, 295–303 (2018).

21. Nosek, B. A., Ebersole, C. R., DeHaven, A. C. & Mellor, D. T. The preregistration revolution. *Proc Natl Acad Sci USA* **115**, 2600–2606 (2018).
22. Petchey, O. L. *et al.* The ecological forecast horizon, and examples of its uses and determinants. *Ecol Lett* **18**, 597–611 (2015).
23. Dietze, M. C. *Ecological Forecasting*. (Princeton University Press, 2017).
24. Dietze, M. C. Prediction in ecology: a first-principles framework. *Ecological Applications* **112**, 6252–13 (2017).
25. Strain, D. 8.7 Million: A New Estimate for All the Complex Species on Earth. *Science* **333**, 1083–1083 (2011).
26. Holling, C. S. *Adaptive Environmental Assessment and Management*. (John Wiley & Sons, 1978).
27. Lorenz, E. N. *Deterministic Nonperiodic Flow*. *Journal of the Atmospheric Sciences* vol. 20 (1963).
28. Lorenz, E. N. *Three Approaches to Atmospheric Predictability*. *Bull Amer Meteor Soc* vol. 50 (1969).
29. Lorenz, E. N. *Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas? American Association for the Advancement of Science Annual Meeting Prog.* (1972).
30. Massoud, E. C. *et al.* Probing the limits of predictability: data assimilation of chaotic dynamics in complex food webs. *Ecology Letters* **21**, 93–103 (2018).
31. May, R. Simple mathematical models with very complicated dynamics. *Nature* **261**, 459–67 (1976).
32. Grenfell, B., Bjørnstad, O. N. & Finkenstadt, B. F. Dynamics Of Measles Epidemics : Scaling Noise, Determinism, And Predictability With The TSIR Model. *Ecological Monographs* **72**, 185–202 (2002).
33. Bonan, G. B. & Doney, S. C. Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science* **359**, (2018).
34. Lovenduski, N. S. & Bonan, G. B. Reducing uncertainty in projections of terrestrial carbon uptake. **12**, (2017).
35. Lovenduski, N. S., Bonan, G. B., Yeager, S. G., Lindsay, K. & Lombardozzi, D. L. High predictability of terrestrial carbon fluxes from an initialized decadal prediction system High predictability of terrestrial carbon fl uxes from an initialized decadal prediction system. (2019).
36. Raiho, A. *et al.* Determinants of Predictability in Multi-decadal Forest Community and Carbon Dynamics. *bioRxiv* 2020.05.05.079871 (2020) doi:10.1101/2020.05.05.079871.
37. Fer, I. *et al. Beyond Modeling: A Roadmap to Community Cyberinfrastructure for Ecological Data-Model Integration*. https://www.preprints.org/manuscript/202001.0176/v1 (2020) doi:10.20944/preprints202001.0176.v1.
38. Atkins, D. E. *et al. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. *National Science Foundation* 84 http://www.nsf.gov/od/oci/reports/atkins.pdf (2003).
39. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016).