Group 1

Andy, Antonietta, Steph, Kristen, Savinni,
Thomas, Greg, Robert, Luke, Tony, Tongtong

Breakout 1

Questions:

- What are the knowledge gaps regarding ocean/cryosphere predictions?
- What are the potential untapped sources of ocean predictability?
- Are there specific observational needs for improving multi-year predictions that are not met with the existing observational system?

Breakout 1

1. What are the knowledge gaps regarding ocean/cryosphere predictions?

We have insights in regions and quantities that are more predictable, but we are still **lacking sufficient understanding of the processes** underpinning that predictability (e.g., Atlantic subpolar gyre, or ocean BGC). BGC processes that are more predictable in regions that are nutrient-limited, pointing to an important role of stratification, upwelling, etc., but these connections needs to be clarified. Also, can BGC inform physics?

There is increasing evidence that **mesoscale processes** in the ocean and atmosphere may be important to accurately capture the atmospheric response to oceanic SST anomalies and its downstream response. A better understanding of these processes and the possibility to parameterize them seems important.

Statistical and empirical models provide a “coarse” representation of the ocean-atmosphere-cryosphere interactions, but they may implicitly capture the influence of small-scale processes, but these implicit relationships must be explained.

Breakout 1

2. What are the potential untapped sources of ocean predictability?

The ocean is a source of memory, which could be further exploited. But we need to know the main ”reservoirs” of memory, and the impact that they can have on the atmosphere.

Meltwater forcing from land ice could impact ocean circulation and stratification, but it is very difficult to observe and predict.

Breakout 1

3. Are there specific observational needs for improving multi-year predictions that are not met with the existing observational system?

Denser and more frequent observations in the ocean interior, especially from Argo floats, would be helpful.

Better resolved coupled meltwater-ocean dynamics

However, it is hard to answer this question in a general way. E.g., the TPOS has provided invaluable understanding of tropical Pacific dynamics, but there are unresolved questions that need dedicated process studies. They may be useful in other regions, e.g. KOE region.

Breakout 2

Questions:

- Who are the main targeted users of ocean/cryosphere predictions, and what aspects of these predictions are necessary to meet their needs?
- How are multi-year to decadal predictions currently used in the ocean/cryosphere?
- For what applications might downscaling of multi-year predictions be necessary (and is there potential for added value)?

Many users are interested in prediction of high-tide flooding, especially extreme events, which may be controlled by local factors (e.g., local winds, river discharge, etc.). For these problems downscaling can be important, although sources of predictability may come from the large-scale.

Compound extremes

Water temperature near ice margins and subsurface temperature under the ice can be important for predicting ice edge, and ice melt.

Breakout 3

Questions:

- What are the knowledge gaps regarding land/atmosphere predictions?
- What are the potential untapped sources of land/atmosphere predictability?
- Are there specific observational needs for improving multi-year predictions that are not met with the existing observational system?

Land-Atmosphere interactions are a large knowledge gap. Land surface is very heterogeneous and the influence of this diversity on the atmospheric boundary layer is poorly constrained. Observations are very limited.

Land use appears to be very influential in climate model simulations and projections, but there may be uncertainties in the parameterization of land-atmosphere interactions, as well as in the definition of climate change scenarios.

Need to better understand how large-scale dynamics (e.g. ENSO) control land surface-atmosphere predictability

Need to more explicitly consider the societal-relevance of predictability: from watershed management, to tropical storms, to monsoon dynamics → these all influence human safety, food security, and political stability.

Breakout 4

Questions:

- Who are the main targeted users of land/atmosphere predictions, and what aspects of these predictions are necessary to meet their needs?
  What are the potential applications of multi-year predictions? What operational capabilities are needed to fulfill them?
- How can multi-year predictions best be formatted for decision making? (e.g., choice of predicted variables, averaging intervals, deterministic vs probabilistic forecasts, skill evaluation, identifying forecasts of opportunity)

Humanitarian needs are a big driver of the need to have predictability/forecasts of land/atmosphere: main ones discussed are drought and farming/crop yields, especially in 'global South'.

Critical to consider consecutive and compound extremes, and how the role of memory plays into consecutive events (e.g. consecutive drought years).

Co-production with stakeholders and potential end-users is critical to build a 'useful' forecast.

Group 2

Breakout summary

Session 1: Ocean/crysophere processes/predictability

Knowledge gaps

- Best practices for ocean initialization (cost/performance tradeoff)
- Identification of prediction targets and best metrics to estimate them
- Identification of of what we can predict with skill now
- Impacts from resolving smaller scales: ocean eddies, coastal processes, finer topographic effects - how do these impact predictability?

Sources of year-to-year predictability [different from seasonal]

- Trends (anthropogenic and low frequency natural variability)
- Time varying aerosols
- Multiyear ENSO
- OHC in reemergence regions

Session 1: Ocean/crysophere processes/predictability

Promising avenues toward improving multi-year ocean/cryosphere predictions

- Increasing ocean resolution - critical for coastal regions, WBCs
- Nested grids or infrastructure like in MPAS where certain regions have higher resolution
- Adding ML onto dynamical systems for applications like uncertainty quantification
- Experimentation with empirical methods
- Experimentation with snow/ice initialization - improved predictions?

Session 2: Ocean/crysophere applications

Targeted users

- Fisheries - stock and distribution models, changes in preferred environments, tourism (sport fishing)
- Real estate - sea level impacts on coastal building, impacts on property value
- Indigeneous people - ocean harvesting for food, sea level impacts on land

Variables of interest:

- Sea level: aquaculture, insurance (asset liability), Army Corp of Engineers, Government (investment in coastal communities)
- Sea ice: Naval navigation, commercial tourism, oil and gas, transportation routes

To meet their needs…

- Uncertainty quantification and effective communication of these metrics
- Likely prefer information related to key thresholds

Session 2: Ocean/crysophere applications

Untapped applications

- Ecological disruptions - infrastructure development (wind farms), trends in T
- Human health - predictions of the potential for mosquito borne illness
- Aquaculture - require specific environment for healthy populations
- Agriculture - saltwater intrusion on water sources, crop production

How do we determine "useful" skill for multi-year predictions?

- Does it provide actionable information?
- Definition of "Useful" depends on the user (how risk averse are they)
- Our skill metrics may be less useful for decision makers
- Will require multiple iterations with stakeholders

Session 3: Land/atmosphere processes/predictability + forecast models

Knowledge gaps

- Land-atmosphere interaction feedbacks - representation in models
- Land initialization - What variables? Datasets?
- Will higher resolution or resolved microphysical processes improve skill?
- Deep soil moisture (limited obs) - impacts on predictability? Processes?
- Integrated effect of high frequencies onto the multiyear timescale?

Untapped sources of predictability

- Linkages with the ocean and related teleconnections
- Shorter term anthropogenic effects (e.g., land use changes)
- Transition phases of climate modes - difficult to predict
- Permafrost, slowly evolving cryosphere

Session 3: Land/atmosphere processes/predictability + forecast models

Best practices to build trustworthy ML forecast models and reliable numerical prediction models

- Empirical models - Why are they skillful? Can they help inform our understanding of the system?
- Train models on diverse data sources to maximize predictability
- Distinguishing between explainable (after the fact) versus interpretable (utilize established relationships between variables/processes) model approaches
- Give appropriate attention to issues like overfitting

Session 4: Land/atmosphere applications

Targeted users

- Agriculture, reinsurance, water resource management, recreational/winter sports, human health, international humanitarian efforts, fire risk

Needs for downscaling

- Soil moisture
- Ski industry (snow line, snow pack)
- Water on land (Coastal inundation → need high resolution coastlines)
- Urban climate (high resolution land use)
- Water resource management (high resolution topography, rivers)

Paths forward toward making timely multi-year predictions available for decision making?

- Stakeholder engagement, liaisons between research and stakeholder
- Coordinated efforts for downscaling
- Infrastructure to make data publicly available

Virtual Breakout Sessions Report

Summary Across all Virtual Breakout Sessions

Gaps in Understanding?

- Very large gap in understanding interactions between trend and natural variability
- Non-local in both space and time for atmosphere-land predictability remains a very open problem
- How does soil moisture memory vary with location and soil type?
- Need for progress in understanding why modeled QBO tropospheric teleconnections are too weak (high-top not the answer), + other predictable atmospheric signals (e.g. NAO) ↔ ‘signal to noise paradox’
- Long-time scale reanalysis efforts important, but noted some important historical events not captured (e.g., dust bowl)
- Initialization strategies need much additional research

What Makes Multi-Year Distinct?

- Understand that multi-year is distinct from seasonal in going beyond 1-year lead, but is a natural extension of the seasonal efforts
- Multi-year is distinct from decadal in terms of frequency of initialization and that monthly or seasonal averages are under consideration as opposed to annual means or multi-year means
- Months 13-15 first, and perhaps, the easiest target
- Data assimilation issues with deep ocean (density) that are not as challenging on the seasonal and multi-year time-scale compared with the decadal problem.
- Whether extending seasonal or decadal is a target phenomenon question

Untapped Sources of Predictability

- Ocean: improved process representation at high resolution
- Sea Ice: realistic sea ice thickness initialization + subgrid properties (rheology, melt ponds?)
- Atmosphere: QBO, too-weak predictable signals
- Land: soil moisture initialization, re-emergence

Observational needs

- Consistency in initialization across hindcast periods is a fundamental need
- But high coverage may only exist in recent years (e.g. Argo, SMOS)
- Need new reanalysis methods and/or strategies for ‘filling in’ more sparsely observed periods
  - Non-local covariances
  - How well can surface-only initialization constrain deep ocean?
  - Machine learning?
- Ocean BGC is still sparsely observed
- High-resolution models require high-res observations for initialisation and/or evaluation of the predictions. Such data is generally lacking.

Promising avenues toward improving predictions

- Models: reduced model biases (mean and variability), improved initialization techniques (reduced shock/drift), higher resolution
- Observations: see previous slide!
- Understanding: case studies of well (or poorly) predicted events may improve knowledge and modeling of key processes

Identifying Forecast Use/Need?

- Targeted users: Fisheries (marine heat waves), coastal flood risk (sea level variations), Antarctic/Arctic operations (sea ice prediction), water management and food security (drought)
  - Some Examples of Year-2 Use Already In Place (e.g., wind energy, reinsurance in China)
- Need to integrate users early in the process to ensure forecasts inform decision support needs. Cannot simply produce forecasts and throw them “over the fence” and hope someone uses them
  - Project support is vital for this (e.g., CSIRO/Hydro Tasmania, Copernicus case studies)
- There is tension between demonstrating some expectation for predictability and prediction skill and the recognition that forecasts only have value if they are used to inform decisions. Even forecasts that have low skill based on traditional metrics may have use for decisions.

Communicating forecasts

- Forecasts are fundamentally probabilistic. Though some users may prefer deterministic, shouldn’t assume this.
- Communicating in a format the user understands is key ↔ co-design
- Tercile probabilities are common, but probabilities for user-specified thresholds or events are more suited for decision making
- Streamlined info such as classification of skill (green ↔ more confidence), interactive maps can help engage users
- Information for more sophisticated users can be layered behind streamlined “front page”

Group 3
Emily, Abdullah, Marika, Elizabeth, Gian Giacomo, Ginny, Kimberly, Erik, William, Erin, William, Andy, Tongtong, Pedro

Knowledge gaps / untapped sources of predictability

- **Tapping new sources of predictability vs. tapping cascades of predictability.** How does predictability of one component propagates into predictability of another component?
    - Potential large impact from model biases, e.g. errors simulating teleconnection patterns.

- **Small scale processes:**
    - Interactions with small scale features, e.g. coastlines and mountains.
        - Effect of topography can be adressed via downscaling.
        - Detailed coastlines are important for sea-ice.

Predicting sequences of years is useful for:

- **Water management.** Depending on reservoir size.
- **Public health.** E.g. valley fever, influenced by life cycle of rodents)
- Wild fires
- **Perennial crops.** Orchards, grapes, cranberries. Avoid persistent drought to plant them.
- **Wildlife management.** A wet year with good food sets up conditions for higher/healthier populations.
- Budgeting and planning operations. Maintenance.
- Energy sector (21 day forecast for operations vs. 30 year projections for planning).
- Shipping companies make plans on multi-year timescales.
- **Observing system design.** Scientific planning and discovery. For instance, study of events like the blob.
- **Marine resource management.** They typically use “persistence” forecasts, but it is hard to change the way people do things.

Best practices

- **Forecast forensics.** Diagnostics of what goes wrong in failed predictions

- Maintaining predictability in the model development cycles.Consider Earth System predictability during the model development cycle.

- Downscaling
  - Stakeholders are confused about the quality of the different techniques for downscaling. How do you ensure or promote good downscaled products?
  - Need for best practices for downscaling variables. Make sure the verification or validation is available with the product.
  - Intercomparison project?

- Dangers of ML/AI:
  - Shortness of training period, overtraining, influence of climate change on training period.

- Deep engagement with stakeholders.

- Need for infrastructure for sharing forecast data.

Breakout questions - group 3

Session 1: Ocean/cryosphere processes/predictability

- What are the knowledge gaps regarding ocean/cryosphere predictions?
- What are the potential untapped sources of ocean predictability?
- What are the most promising avenues toward improving multi-year ocean/cryosphere predictions?

What are the gaps?

What we think is societally relevant that we know is predictable. What societally relevant information needs to be predicted?
Water managers want year to year information. The sequence of years for reservoirs.
Reclamation: water in the west depends on the snow pack. Multi-year averages are useful. But a sequence of events is important. Valley fever depends on the sequence of events.
Useful sequences can be based on terciles.
Wild fires also depend on seasonal sequences.
2-year forecasts are useful for budgeting, planning for the next year.

Early or late melt out.

Ice area/volume predictability, but timing is more important. Is timing easier? Harder? There is seasonal predictability of freeze up based on melt up from previous season - via the effect
Interactions with coastlines might be important - missing processes: land fast ice and higher resolution is needed. This is important for local communities.

Model errors still have big influence on impacts of sea-surface temperatures.

What is the role of noise in predictions?
How much of the skill is forced vs. how much is initial value?

Cascade of predictability
How does predictability of one component propagates into predictability of another component?

Tapping new sources of predictability vs. improving cascades of predictability?

- **Statistical improvements to existing model data,**
- **Forecast forensics: diagnostics of what goes wrong in failed predictions,**

What variables are needed for energy forecasts:

High spatial resolution needed.
Attribution of extremes in a changing climate is still important.
A few years down the road for energy production. Planning out to 30 years
21 day forecast for operations vs. 30 year projections for planning.
Power generation needs multi-year forecasts.

Honesty omn what is predictable.

Maintaining predictability in the model development cycles.
Routine initialized predictions are useful to asses predictability and prediction skill in the development cycle.

Add ESM predictability as one of the goals of the model development and assess it along the cycle.

Session 2: Ocean/cryosphere applications

- Who are the main targeted users of ocean/cryosphere predictions, and what aspects of these predictions are necessary to meet their needs?
- For what applications might downscaling of multi-year predictions be necessary (and is there potential for added value)?

- How do we determine "useful" skill for multi-year predictions?

Users of multi-year predictions:
Fisheries has a chart to define timescales.
Water management - only certain reservoirs have sufficient capacity to hold water for more than one year.
Shipping companies make plans on multi-year timescales.
Scientific planning and discovery of impact of events like the blob. Observing system design.

Marine resource management relies on "persistence" forecasts. It is hard to change the way people do things. We have to demonstrate that there is added value in doing things differently.
Nature paper (Mark Jacobs) showing that dynamical forecasts of heat waves are better than persistence.
We don't know how useful sea-ice forecasts really are.
Is the shipping industry making sea ice forecasts?
Shell might be using sea-ice forecasts.
NOAA is developing a climate-fisheries initiative, it includes a goal to work directly with stakeholders to provide them with actionable information. Stakeholder fatigue is a problem. Agencies want to do this in an organized way.
Users love terciles (terrestrial hydrology).

Is hard to determine usefulness because it depends on the user.

Certain coastal stakeholders are tired of being asked what ocean data they need.
NOAA is trying sustained stakeholder engagement to learn about their needs. This is a bit of a "trial and error" approach that avoids spending a lot of time and effort producing data that might not be used.

Downscaling is important for terrestrial hydrology & stream flow prediction - some watersheds are smaller than a grid point. Soil moisture might be less dependent on downscaling.

Downscaling is useful for fisheries. Any place that is spatially heterogeneous will require downscaling.

Statistical downscaling might solve most problems. But how about extreme events?

Downscaling is important for terrestrial hydrology & stream flow prediction - some watersheds are smaller than a grid point. Soil moisture might be less dependent on downscaling.

Session 3: Land/atmosphere processes/predictability + forecast models

- What are the knowledge gaps regarding land/atmosphere predictions?

We don't know how to compare simulated and satellite-measured soil moisture. Satellites measure a layer of soil that is thinner than the model's top most layer. A soil moisture satellite simulator?

- Should multi-year prediction be viewed as an extension of seasonal forecasting or should it be viewed as the shorter lead part of the decadal forecast problem, or is it something entirely different? How might this affect development of its operational prediction?

Is it useful to make a distinction between lead times? Should be thought of as a balance between initial-value vs. boundary-forced predictability. This boundary is starting to be explored in land/atmosphere predictability.

- What are the best practices to build trustworthy ML forecast models and reliable numerical prediction models? For example, how long a record is needed for both construction and independent verification of ML models, and how long a hindcast record is needed to evaluate and bias-correct numerical models?

Best practices for ML:
Do we train using observations?
ML danger for overtraining. Records used in training are too short to sample variability and extremes.
You are sampling noise and turning it into signal.
Concern about the effect of forced changes in training intervals.
Using ML for fixing model biases is better than applying it to "post process" or to correct predictions.
Regularization techniques in ML are underutilized and could address the issue of low sample size.

Can we use ML techniques to assess what physical models are missing?

Include prediction in the model development cycle. We could envision diagnostic packages that include prediction metrics - as long these metrics help understanding of physical processes.

Maintaining old, good-for-prediction versions of a climate model could be a solution. But it is a challenge for software engineering.

Session 4: Land/atmosphere applications

- Who are the main targeted users of land/atmosphere predictions, and what aspects of these predictions are necessary to meet their needs?

Public health. Things that cue on sequences of years. Valley fever depends on a sequence of wet and dry years (can drive food availability for rodents…)
Fire risk. Wet year followed by a dry year.
Wildlife management. A wet year with good food sets up conditions for higher/healthier populations.
Realities of having to plan a budget, maintenance of infrastructure more than a year ahead of time..
Agriculture.
Perennial crops. Orchards, grapes, cranberries. Avoid persistent drought to plant them.
Wildlife management. A wet year with good food sets up conditions for higher/healthier populations.

- For what applications might downscaling of multi-year predictions be necessary (and is there potential for added value)?

Water storage in reservoirs.
Some variables don't need to be downscaled - e.g. drought.

- Who should do the downscaling?

Stakeholders are confused about the quality of the different techniques for downscaling. How do you ensure or promote good downscaled products?

Need for best practices for downscaling variables. Make sure the verification or validation is available with the product.

- How can multi-year predictions best be formatted for decision making? (e.g., choice of predicted variables, averaging intervals, deterministic vs probabilistic forecasts, skill evaluation, identifying forecasts of opportunity)

Co-development / deep engagement with stakeholders.

Taylor et al. 202… showed that users prefer formats they are familiar with and are easy for them to understand.

Is there a vision to produce/distribute forecast data in a way that can be used by developers of climate applications (e.g. consultants, start ups).

Limited choices vs. options?

Learning from what works for seasonal prediction.