Breakout 3. Forecasting Methods Group A

Shelf vs Open Ocean

- Challenge: Data on shelf is sparse
- Granular funding structure to support coastal/shelf observing
  - Mostly coming from urgent local concerns
  - Socioeconomic impacts on shelf/coast are huge
- Simultaneous observations can lead to mechanistic understanding

Integrating observations and modeling

- Challenge: Need for national backbone to support this
  - Consolidation of individual modeling efforts
- NMFS perspective: required for “climate smart” management
  - Think about observations from application/state estimation perspective
- Workload can be shared internationally

Reanalysis Products:

- Complete field of interpolated data - good observations help
- Metadata consistent
- Should not be treated as observations

Breakout 3. Forecasting Methods Group B

1. **Computational limitations:**
   a. Need ensembles of ensembles to better characterize the uncertainty in coupled models
      i. How many members? ii. How much computational resources?
   b. Need not only HPC resources, but analysis time & storage/sharing of output

   Solution: need centralized HPC resources as well as data serving centers (IOOS?)
2. **Data limitations:**
   a. Need more data on continental shelf (between coast and ocean)
   b. Need uncertainty characterization of data, more mechanistic studies; agency barrier

   Solution: Funding opportunities are available for data synthesis efforts!

   Data sharing repositories (e.g., Figshare, Zenodo)

   Uniform data formatting (consistent variable names, units, etc.); funding agencies should expect 10-25% of project $/time to go toward data curation)
3. **Gaps in knowledge:**
   a. Along boundaries: Land/water boundary; sediment/water interface; coast/shelf boundary
   b. Lab/field experiments: more rates, rate formulations, multi-stressor experiments

   Solution: Can programs like LTERs help?
4. **Leverage modelling for prediction:**
   a. Comparison & combination between machine learning and dynamical models for mutual improvement
   b. Usage of machine learning to fill in gaps of observational data
   c. Collaboration with data scientists

Breakout 3. Forecasting Methods Group C

Q1 : Major challenges in coupling models

- High technical requirements - engineering challenge, need computer science expertise
- High computational resources needed for ensembles, may be too large to be feasible
  - Propagate PDFs with forecast lead time instead of running each ensemble member? Might not work if ecological forecast is climate path-dependent
- Understanding deterministic ecology model runs & effect of initial states and degree of non-linearities
  - Degree of 'realism'
- Lack of ecological data to make reasonable comparisons between the coupled models
- Hindcasts in climate models not available with ecological models, not obvious how to do
- Biological coupling may highlight errors in particular model outputs and underlying assumptions of physics (e.g., due to lack of resolution near coasts)
- Predictability assessment of component parts (climate, ecosystem) may not be the same as for coupled components (e.g., Peruvian coastal ecosystem more sensitive to Niño1.2 whereas Niño3.4 more predictable)

Q2: Methods for leveraging machine learning techniques

- Define best practices
  - ML needs to be "reliable" (predict probabilities correctly) and/or can it identify "forecasts of opportunity"
  - ML needs to be interpretable (to justify use of its forecasts), ideally represent dynamics
- Machine learning to improve sub-grid scale parameterization (partially being done on physical side but very new - not yet on biological side) and for resolving fine-scale mesoscale features (wcr eddies, meanders) in regional models used in downscaling approaches
- Machine learning algorithms to set boundary conditions in dynamical downscaling for fields you don't otherwise have

Breakout 3. Forecasting Methods Group B

Gaps, Opportunities, Strategies for combining mechanistic and machine learning methods

1. Development of hierarchical and interoperable models and testbeds as well as tools for analysis to more systematically understand model skill and challenges.
2. Investment is needed in community development of curated data sets for model calibration and validation.
3. Note: NASA Export program could be an opportunity for testing different model hierarchies and types, using consistent, quality controlled and extensive data sets.
4. Development of consistent machine learning approaches with well-characterized limitations for marine science application.
5. Need for development of user-friendly ecological forecasts with an emphasis on applications that users can compare with their direct observations (or some direct observations) in order to develop trust and familiarity. (E.g. use of waver forecasts by surfers)
6. Need for reporting of uncertainty and conditional skill in ecological forecasts (note that meteorologists have demonstrated that “naive” users make better decisions when provided with forecasts that include uncertainty.)
7. Machine learning approaches to integrating different types of measurements across spatial and temporal scales that differ widely.
8. Machine learning can also be used for uncertainty quantification, and forecast correction

Breakout 3. Forecasting Methods Group D

What are the major challenges in coupling models across scales and disciplines (e.g., open ocean vs. coastal regions, physics/biogeochemistry/fish)?

- Gaps in knowledge
- Data limitations
- Computational limitations

How can we leverage machine learning, empirical and mechanistic modeling for physical/ biogeochemical/ecological prediction?

- Use model hierarchies to construct appropriate test bed frameworks for models
- Process resolution, interoperability across components.
- Interoperable model
- Make models more accessible
    - User-friendly real-time forecasting tools
- Constrain uncertainty and express uncertainty in a manner intepretable to end users
- Make strong recommendations that program managers can actually implement
- Use newer ML approaches to make data and models more interoperable
- Observational hierarchies
    - Use ML to connect the dots and interpolate

Breakout 3. Forecasting Methods Group E - virtual

1. What are the major challenges in coupling models across scales and disciplines (e.g., open ocean vs. coastal regions, physics/biogeochemistry/fish)? **Please consider computational limitations, data limitations, and gaps in knowledge.*
   - Challenge in combining disparate data sources for single variable, as well as multiple
     - Need for standardized best practices in data collection and processing
   - Lack of access to global biogeochemical model output at the spatial and temporal scales needed for regional downscaling (e.g., issue for downscaling of IPCC5) and seasonal prediction (only sfc values for ocean state)
   - Need for multi-trophic level integrated datasets (& thus multi-trophic level observing systems)
   - Lack of understanding (mechanisms/processes) of where predictability comes from an ecological standpoint
   - Challenge in linking spatial and temporal scales of information (e.g., physics, BGC, ecological data) provided from one piece of coupled modeling system to the next
     - E.g., zooplankton is recorded as counts in the field but modelers need C / mg m3
   - Need systematic approach to aggregate data by system sensitivity / management question
   - Qualitative vs. quantitative needs of stakeholders
2. How can we leverage machine learning, empirical and mechanistic modeling for physical/ biogeochemical/ecological prediction?
   - Coupling probabilistic forecast
   - Working with knowledgeable computer scientists to optimize ML techniques
   - Beneficial when have disparate data
   - Use ML to replicate dynamical model behavior (i.e., reduced form model)