<u>EMC RECOMMENDATION ON NEXT STEPS FOR BSIERP FUNDING DECISIONS</u>
(September 10, 2007)

The EMC recommends that this review document be transmitted to Dr. DeMaster, with the request that all the embedded numbered questions be addressed in writing by the appropriate persons or groups of persons, who will identify themselves in the responses.

[November 2, 2007]. Our answers were composed by **Kerim Aydin**, **Nick Bond**, **Franz Mueter**, **Andre Punt**, **Al Hermann**, **Georgina Gibson**, **Alan Haynie**, **Ken Coyle**, **Marc Mangel**, and **Jim Ianelli**. The primary author(s) of each answer are listed with each answer. Document tables display **budget** (Table 1), **timeline** (Table 2), **predictions** (Table 3), and **competing models** (Table 4).

In addition, we provide supplementary material for readers desiring more information: **revised modeling proposal** (Appendix 1), **overall BSIERP study plan** (as funded, Appendix 2), and **project-specific descriptions** (Appendix 3 (“zip” file that includes all project descriptions except for the patch dynamics and LTK projects)). Project-specific descriptions are referenced in this document by their project title and label (e.g., Seasonal bioenergetics, O2.24); in Appendix 3, the project-specific descriptions are referenced by their project label (e.g., O2.24).

OVERVIEW

The hoped for product from BSIERP is a capability for predicting the fisheries-relevant responses of some selected upper trophic level populations in the Bering Sea, conditional on selected probable scenarios of anticipated climate/oceanography changes, over a time horizon of a decade or so. The NPRB funding is for a 6-year program.

The intention of the NPRB in this endeavor is to set a new standard for coordination of the many subprojects and for statistical rigor. The questions posed here by the EMC are motivated by those issues.

PRESENT STATUS

The field component of the program has essentially been launched. Pending are decisions about the NPRB funded modeling and database components, and about possible expansion of the retrospective analysis component. The tentative allocations are $2.5 million for modeling (ie $417K/yr) and $810,000 for data management (ie $135K/yr). Roughly $300K has been committed to retrospective analysis. Another retrospective statistical analysis was funded by NPRB last year, is underway now, and could be reporting some results soon. These dollar amounts give some indication of the number of FTEs that might be committed to the tasks, which in turn will constrain the realistic scope.

DATA MANAGEMENT COMPONENT

At this point, we do not have a substantive work plan for the data management component. A director for that component has been selected by the BTU team.

The retrospective data analysis is obviously dependent on the availability of historical data. Given the nature of interannual variation and spatial heterogeneity, the lion's share of the data input to the models (both for initialization and especially for parameter estimation) will have to be existing (historical) data, which need to be assembled into a common database that can be accessed by the entire program. New data obtained by the program, which also need to be added to the common database when they are obtained, will serve primarily for validation and for very targeted rectification of critical, but limited, data gaps for model initialization and parameter estimation.

So, the work plan for the data management component must give a high priority to assembling, and making accessible, quality-controlled and adequately documented versions of all the pertinent historical data that will be used for the modeling and for the retrospective analyses. This, in fact, is a very large task (which is on top of the need to cope with the new data streams). Note that, for the purposes of this program, it generally is not sufficient to "point" to other databases. Those kinds of links are not always readily followed, and, more disconcertingly, the distributed data bases may be subject to multiple versions and moving-target revisions, which could create obstacles to reproducibility and validation within the program or in subsequent efforts. The pertinent historical data may span the past 50 years or so, and will doubtless involve a few obscure sources (eg perhaps the Japanese surveys, and surveys that were of short duration) not under direct institutional control of participants in BSIERP.

> ***[Q1]*** *We need a work plan for the data management component consistent with these priorities. This should include, for review by the EMC, a list of the data sets that will be sought, and a time table for posting them. The list should be cross indexed to data uses in the various modeling subprojects and the retrospective statistical analyses, and then the list should be verified by the PIs of those subprojects.*

A1 (Coyle): PIs have identified data sets, the timetable for posting them and project linkages. This information was requested from each PI through two methods: an on-line questionnaire and specific project descriptions. Principal investigators completed an on-line questionnaire located at http://www.eol.ucar.edu/cgi-bin/best/data_quest. This data form identifies project linkages by specifically requesting that the PI identify data sets their project depends on. Principal investigators also completed project–specific descriptions including sections on research projects, research links and research reporting. Planned data sets are described in the research projects section, the project linkages are described in the research links section and the timing is described in the research reporting section. This information is summarized in BSIERP Study Plan Figure 1 (data and model links) and BSIERP Study Plan Tables 3 (data products) and 4 (timelines) (Appendix 2, also displayed at the end of this document). The specific information (before summarization) is listed in the project-specific descriptions (Appendix 3).

The next step is to verify the specific information listed in the project-specific descriptions. This step is scheduled to occur during December 2007 through February 2008. The review period is lengthy because list verification will be an iterative process involving the modeling PIs, the data collection PIs and the BSIERP Data Manager (Ken Coyle). The data manager, in collaboration with the PIs, will generate data tables for holding the incoming retrospective, field and laboratory data. The PI will provide a template data set to the data manager, which will be used to generate the data table. Each data set sent to the data manager will be accompanied by a metadata document describing the data in detail. The data tables and metadata will be accessible from the BSIERP-BEST website which will serve as a "handshake" to ensure that the data collectors know what data is needed and the modelers know what data is coming.

The database for retrospective, field and modeling data will be accessible from the BSIERP-BEST web site. Data management will be completed using the AOOS (Alaska Ocean Observing System) web site and facilities.

The data manager will not post data until the PI has checked and verified the data. Upon receipt of the data, the data managers will check the data for missing values and scan the data for outliers. Computer programs will be written to scan incoming data files and plot or map the data. If outliers are found, the data manager will consult with the PI to ascertain the validity of the data, and either return the data to the PI for correction or enter corrections, depending on the volume of material. Species names will be checked for spelling using the NODC species listings and NODC species codes will be added to each species designation to facilitate sorting by phylum, class, order and family.

The data will be stored in relational database systems (PostgreSQL or MySQL). The relational database systems provide for flexible and rapid sorting and selection of data and files. The storage of files and database structures currently has an 8 Tb capacity. We will purchase an additional 8 Tb raid array and computer system for dedicated storage, serving and visualization of BSIERP data. Data will be backed to hard drives, DVD and CD. Relevant oceanographic data sets will be automatically sent to National Oceanographic Data Center (NODC) to provide additional off-site backup. An additional 100 Tb of storage will be provided by the Arctic Regional Super Computing center, primarily for storage of ROMS modeling data. We anticipate very large storage needs for ROMS physical and lower trophic level data. The actual size of these modeling data sets will exceed the capacity of the database system and will be served separately. ROMS data will be stored in netcdf format in a tape library in ARSC (Arctic Regional Super Computing Center) and will be made available to researchers on request. Live access to some modeling data may be possible as the project progresses, depending on the spatial resolution of the requests. Images and movies of model output provided by the modelers will be accessible from the BSIERP-BEST web site.

The data will be served through a web portal. The front page, which will serve as the entry point to BSIERP/NPRB data and project information, will be designed by NPRB personnel and implemented by the data managers. A Project Browser metadata application will allow discovery of BSIERP data sets. Where possible, web based visualization or sql (structured query language) options will be developed so end users can quickly visualize and select data for downloading. The database managers will participate in PI meetings and data synthesis activities to insure that data will be provided to end users in a format consistent with their requirements.

ONGOING RETROSPECTIVE ANALYSIS COMPONENT

> ***[Q2]*** *We need an interim report on the progress of the retrospective statistical analysis funded by the NPRB last year.*

A2 (Mueter): Two ongoing NPRB studies are relevant to this modeling effort. NPRB project number 605 is currently examining spatial patterns in temperature and prey availability to model potential growth of juvenile Pacific cod, but is not specifically conducting a retrospective analysis of cod in relation to climate trends. Interim reports on this project are available on the NPRB web site. A retrospective analysis of species-habitat associations of yellowfin sole, arrowtooth flounder, and Alaska Plaice (NPRB project 709) is using a generalized modeling approach to identify habitat characteristics that both determine and drive interannual variability in the spatial distribution of these species. No progress report is available at this point, as this work is just getting underway, but as two of the project PIs (Ciannelli, Bailey) also are BSEIRP PIs, it is expected that the development of the retrospective analysis will continue collaboratively with the BSIERP modeling effort. Among other products, the project will provide predictions of the spatial distribution of arrowtooth flounder under different environmental conditions. A comparison between these predicted patterns and spatial predictions for arrowtooth flounder from the BSIERP models can help either confirm or refute the hypotheses and mechanisms underlying the BSIERP models. The retrospective component of BSIERP (Fish, birds and mammals, O3.30) complements these analyses by focusing on interannual variability in productivity and abundance at the scale of the ecosystem.

GENERAL PERSPECTIVE ON THE BERING SEA ECOSYSTEM

<u>Climate</u>

The climate drivers of the Bering Sea ecosystem include the known global warming trend (fairly steady air temperature global average rise of about 0.4C over the past 50 yr), large-spatial-scale strong deviation events in the air pressure and sea surface temperature with persistence times of about a year and typical return times in the 2-7 yr range (El Nino), and phenomena with longer typical persistence times such as large-scale alternating patterns of sea surface temperature (PDO) and atmospheric pressure (Arctic Oscillation, Aleutian Low) which also correlate with other climate variables.

In combination, these are associated with recent sea surface temperature increases in the North Pacific and Bering Sea far in excess of the average global warming trend.

The climate indices underwent a discernible shift, in the mid 1970's, to predominantly warm phase PDO, strong Aleutian Low, more frequent El Nino, followed a decade or so later by a persistent excursion to far positive phase of the Arctic Oscillation. It is not clear whether that pattern is continuing now. Many of the climate indices are known to be somewhat correlated among themselves, and it is not necessarily decided which are "cause" and which are "effect," and whether the mid 1970's "shift" was a distinct departure from the underlying long-term dynamics, and whether other ocean-atmosphere phenomena such as ocean circulation patterns (North Pacific Gyre) also have an important role.

There has been a global long term trend in atmospheric CO2 build up, which is thought to play a role in the long term global warming. Accompanying the CO2 trend has been a trend in ocean acidity.

Physical Oceanography in Relation to Primary Productivity

The Bering Sea is influenced by, among other factors, prevailing northward currents through the Aleutian passes, which, with strong tides, are responsible for some localized very high productivity near some passes, and large scale eddies and upwelling along the slope break, which are responsible for a recognized Green Belt of productivity there. The ice edge itself affects upwelling, and ice melt forms the seasonal cold pool. One dramatic physical correlate of the recent warming (modulated also by winds and ocean currents) is the diminishing extent of winter sea ice in the Bering Sea and the earlier timing of ice melt.

Timing of ice melt in the Eastern Bering Sea affects the generation of the spring ice-edge bloom in an apparently qualitative way. Different phytoplankton groups have different responses to different ranges of temperature, nutrients, and light. Early ice melt is associated with a later spring bloom of a different character than the ice-associated bloom, and in conjunction with the warmer water temperature this promotes a food chain which is distinct from that fed by an early spring bloom.

Summer storms and wind patterns affect vertical mixing with consequences for the continuing summer productivity and development of a fall bloom.

Changes in ice, and changes in precipitation, including precipitation over land, will affect salinity.

Biology

Some biological time series, for the North Pacific, Gulf of Alaska, Aleutian Islands, and Bering Sea, most notably productivity of a number of fisheries, when analyzed pre- and post-1976, show considerable differences. However, it is not necessarily clear when the shifts took place in the respective biological variables, or the extent to which they were attributable to the shift in the variables reflected in the climate indices.

It is known, from longer term statistical analysis of fisheries elsewhere that production in some fisheries shows a correlation with climate indices, or with simple environmental variables such as temperature. It is also known from tree ring analysis that long tree-growth histories show some time series properties reminiscent of the climate indices (but the instrumental record of climate indices is much shorter than the tree ring histories).

Specifically in the Bering Sea, there have been indications of major biological changes at essentially all trophic levels in the past 4 decades. Not all these changes were coincident with the mid 1970s shift. Some of the finer resolution monitoring only began after the mid 1970's.

Among the phytoplankton, there have been occasional large scale coccolithophorid blooms, and changes in the size distributions of diatoms, and possibly changes in primary production overall. The ice-melt-timing switches in the spring bloom may be associated with the prevalence of diatoms and flagellates respectively.

Among the zooplankton, in the last few decades, decreases in the larger copepods and possibly euphausiids, and increases in chaetognaths, and the rise and fall of jellyfish have been noted.

In the northern Bering Sea benthos, decreases in bivalves and amphipods have been noted.

Among fish populations, large changes in turbot, arrowtooth flounder, rock sole, and walleye pollock have been noted.

There have been die offs of sea birds, and large population declines in fur seals and Steller sea lions, and signs of severe nutritional stress in gray whales.

It has long been noted, generally, that there is very large interannual variation in year class strength in the best-monitored fish populations. It is suspected that this is the result of match and mismatch in time and space of critical conditions for early life stage growth and survival. The critical conditions controlling year class strength for the important Bering Sea fisheries probably have not yet been identified.

PERSPECTIVE QUESTIONS FOR THE MODELING AND RETROSPECTIVE ANALYSIS TEAMS

> ***[Q3]*** *Which elements of the above "general perspective" story, and elaborations or alternatives thereof, are believed by the modelers to be relevant to the predictions of their modeling?*

A3 (Aydin): We answer this question in our answer to question 5, where the relevant elements of"general perspective" story are specifically identified. In general, our BSIERP hypotheses (BSIERP Study Plan, Section D.4 (Appendix 2)) frame the "general perspective" story and provide an integrated framework for the BSIERP study. Likewise, our proposed modeling approach provides an integrated framework to examine this "general perspective" story of the southeastern Bering Sea.

> ***[Q4]*** *How will the relevance be assessed, element by element, in the retrospective data analysis, and which data will be employed in that analysis?*

A4 (Mueter): The retrospective analysis will summarize the existing correlations between climate indices and biological variables, and extend the correlation to the finer details of the hindcast model outputs (e.g., produce correlations on a finer scale). Much individual correlation work between climate variables and specific populations has been published to date (e.g between PDO and salmon). The retrospective analysis will improve our understanding of the global correlations between climate and biology and improve our predictive capabilities by looking at correlations in intermediate quantities (e.g., “mechanistic link variables,” such as sea surface temperature in a particular domain in the Bering Sea; see Q16) that explain the link between the global indices and production. This will give us an “envelope” approach for determining what range and frequency of global index variation will still allow each correlation to hold; for example, if the current conditions correlate the PDO with fish production through a bottom-up mechanism, will a changing frequency of the PDO lead to a switch to top-down control which would cause the correlation to break down?

[Q5] *How will the relevant features of the story be represented in the models?*

A3 and A5 (Aydin): For *climate*, we will be using the IPCC downscaled climate scenario model output (hindcast and forecast) as the basis for driving biology. For the North Pacific, these IPCC models have been evaluated by a working group of the PICES Physical Oceanographic Committee for North Pacific relevance, specifically by examining how each model’s hindcasts fit statistically observed patterns such as the PDO. We will focus climate downscaling on the models which perform the best for these indices, including Bering Sea specific indices such as cold pool extent. Thus, the climate story will be represented in the models through the IPCC scenarios which best represent the local features of the ecosystem.

For *oceanography*, the ROMS regional oceanography model will be able to resolve winds, temperature, current, ice, flow through passes, and large-scale eddies in such a way that they can either directly affect the zooplankton or upper-trophic level populations. The NPZ component is specifically designed to be fit to available data and thus the output can assess the relative importance of each controlling variable in affecting primary production; this is not known at the outset of the model, but rather is a result.

For *biology*, we can examine the elements point by point as follows:

Some biological time series, for the North Pacific, Gulf of Alaska, Aleutian Islands, and Bering Sea, most notably productivity of a number of fisheries, when analyzed pre- and post-1976, show considerable differences. However, it is not necessarily clear when the shifts took place in the respective biological variables, or the extent to which they were attributable to the shift in the variables reflected in the climate indices.

“Shifts” as they occur will be emergent properties of more local or regional scale mechanisms; statistically significant biological shifts may or may not line up (with appropriate life-history lags) with global indexed climate shifts. The fitting methods for the vertically-integrated model to be used both within regime and between regimes, as discussed in Q20, will be used to identify the mechanism(s) behind any observed correlation between climate indices and production, and thus indicate when and why each “shift” took place. Of particular interest will be determining or confronting hypotheses concerning the controlling physical and lower trophic level processes that occur on a seasonal scale (e.g., bloom timing) and which likely give rise to the observed regime-scale patterns, including observed fishery production.

It is known, from longer term statistical analysis of fisheries elsewhere that production in some fisheries shows a correlation with climate indices, or with simple environmental variables such as temperature. It is also known from tree ring analysis that long tree-growth histories show some time series properties reminiscent of the climate indices (but the instrumental record of climate indices is much shorter than the tree ring histories).

Long-term (500+year) statistical properties of the vertically-integrated models, using climatology from random draws of hindcast conditions, may be used to examine whether there are long term “switch” frequencies, e.g., similar to dramatic fluctuations in sardines or anchovies reported in the world’s upwelling regions. The draws will explore the possibility that a biological model calibrated to current conditions contains “intrinsic variability” which may cause wide-scale fluctuations. As there are limited data on long-term production (e.g., sediment cores) for the Bering Sea, this analysis will be limited to statistical properties of the resulting model rather than examining fits to past paleontological records. Still, such long-term analyses will be useful in discussing the “natural range of variation” which is a reasonable benchmark for management purposes.

Specifically in the Bering Sea, there have been indications of major biological changes at essentially all trophic levels in the past 4 decades. Not all these changes were coincident with the mid 1970s shift. Some of the finer resolution monitoring only began after the mid 1970's.

Among the phytoplankton, there have been occasional large scale coccolithophorid blooms, and changes in the size distributions of diatoms, and possibly changes in primary production overall. The ice-melt-timing switches in the spring bloom may be associated with the prevalence of diatoms and flagellates respectively.

The NPZ model will have sufficient resolution in plankton size groups to explore the conditions which lead to these blooms, and reconstruct whether long-term changes are plausible from a modeling standpoint.

Among the zooplankton, in the last few decades, decreases in the larger copepods and possibly euphausiids, and increases in chaetognaths, and the rise and fall of jellyfish have been noted.

Copepods, euphausiids, and jellyfish will also be included explicitly in the NPZ and FEAST modeling. Of particular note with reference to euphausiids is their importance in late summer/early fall prey supply for forage fish. Chaetognaths and other large predatory zooplankton are not specifically represented in the current study (they will be represented by a constant predation mortality term) primarily due to lack of life-history, rate, and distribution data.

In the northern Bering Sea benthos, decreases in bivalves and amphipods have been noted.

The NPZ model will include a “benthic accumulation” box which will provide hindcast and forecast reconstructions of the energy available to the benthos. The “benthic accumulation” box will represent the total biomass of the benthic community (total secondary production, both infaunal and epifaunal) but not its species composition. Species pooling for the benthic community was chosen; modeling funding was limited and the commercially and ecologically more important pelagic community was prioritized.

Among fish populations, large changes in turbot, arrowtooth flounder, rock sole, and walleye pollock have been noted.

Among these species, the vertically-integrated model will examine walleye pollock, Pacific cod, and arrowtooth flounder, while the NPZ model will, as mentioned above, produce indices of overall benthic production which can be correlated with benthic feeders such as rock sole. Finally, the correlative/biomass dynamics model will examine all of these species for empirical interactions between their fluctuations and conditions.

There have been die offs of sea birds, and large population declines in fur seals and Steller sea lions, and signs of severe nutritional stress in gray whales.

The specific issues of nutritional stress for birds and mammals, especially for central place foragers, will be assessed both through the correlative analysis and inference made from the hindcast prey fields calculated through FEAST. In particular, we will calculate, both from FEAST output and through the behavioral foraging model, changes in overall foraging success, including energetic traveling costs to major prey concentrations. Of particular interest will be the search for thresholds; relationships between hindcast prey availability and extreme die offs, mortality, or nutritional stress which may indicate the threshold levels of prey (both in time and space) necessary for survival. Overall, birds and mammals provide sensitive indicators of poor prey conditions, and one project goal is to use hindcasts to formalize these indicators into direct measures of prey availability.

It has long been noted, generally, that there is very large interannual variation in year class strength in the best-monitored fish populations. It is suspected that this is the result of match and mismatch in time and space of critical conditions for early life stage growth and survival. The critical conditions controlling year class strength for the important Bering Sea fisheries probably have not yet been identified.

The seasonal cycle, both of temporal production and spatial production (e.g., moving blooms with ice-edge), is a particular subject of interest for the modeling study. In particular, we will focus effort on hindcasting and fitting shifts in seasonal timing in conjunction with the match/mismatch hypotheses. By modeling the growth and location of young-of-the-year pollock, the FEAST model will allow exploration of the importance of bloom timing, match/mismatch, and critical time/critical period effects on survival. **This latter element emphasizes a novel element of FEAST; the two-way coupling between large zooplankton (euphausiids) and pollock.** In particular, in the summer/fall, continued prey availability through large zooplankton and pollock (cannibalism) may play a critical role in determining winter survival. Matching growth and survival patterns to late season data, and modeling scenarios of change in both the spring and fall transitions, will provide one focus for hypothesis testing with the vertically-integrated models.

> ***[Q6]*** *How will adherence of the models to the known history of the relevant features be scored in the data assimilation and hindcasting?*

A6 (Aydin): We will score the model estimates based on their ability to represent population trajectories (running average) and variability comparing predictions and observations of BSIERP focal species (euphausiids, pollock, cod, arrowtooth flounder and the suite of forage species). Our scoring will measure focal species effects because of their ecological and economic importance. Each "feature" described in the data story or hypothesis will be associated with a set of confirmatory or rejection indexes (e.g., if flow to benthos is limiting, an index of total benthic production correlated with flow measures from the model will be assessed). For time series indices and data, and for specific indexes developed for spatial prediction (see "specific predictions" section of [Q9] answer, we will withhold specific years of data (either in sequence or drawn randomly) and assess the fit of the model to the indices calculated from the data using a likelihood criteria for scoring (appropriate probability distributions will differ from index to index, and will be assessed as part of the data assimilation process).

"Scoring" will be based on a table of likelihood components describing the model's fit to each aspect or index. The use of this scoring table in determining "truth" will be judgment-based but formal; for specific issues in tradeoffs of goodness-of fit, the comments on whole objective function optimization vs. optimizing individual parts of the objective function are found in the discussion of fitting FEAST found in [Q19]; the general aspect of examining multiple model parameter minimizations as described in the answer to [Q19] for FEAST are applicable to all models.

Following the evaluation techniques developed by PICES WG-20 for IPCC climate models, we will place specific emphasis on "regime and index" matching. For climate models for example, IPCC predictions are downscaled into index trends such as the Pacific Decadal Oscillation. Here, we will downscale the spatial modeling results into predicted recruitment variability by climate regime, including measures of autocorrelation resulting from the hypothesized interplay of top-down and bottom-up forcing through EOF and spectral analysis (e.g., are we getting the spatial and temporal frequencies correct?). For the actual scoring to be used, please see the model-specific answers to [Q15, Q17, and particularly Q19].

> ***[Q7]*** *How good a retrospective fit to the relevant features of this story is necessary for the integrated model to be deemed reasonably effective for predictive purposes?*

A7 (Aydin, Punt): The criteria for "sufficient goodness of fit" depend on prediction use. The breadth of BSIERP modeling provides several prediction types and scales, so several confirmatory criteria are reasonable.

- Operational recruitment forecasting (2-5 year time horizon) of focal species will be assessed for +/-30% difference from observed recruitment when challenged by withholding recent years from data fitting.
- There are several model-produced quantities that may be incorporated into single-species assessments, such as prediction of natural mortality trends for walleye pollock. The prediction skill for whether such results should be used in single-species model assessments will be whether the magnitude of the predicted change is greater than the standard error of residuals in mortality produced by the current stock assessment. This criterion was chosen because it is expected that these mortality estimates will not used for including year-to-year variation in mortality ("white noise") in stock assessments, but rather will be used to calculate trends or regime effects in long-term productivity which lie outside the range of currently assessed interannual variability.
- For longer-term (10-50 year time horizon) forecasting of ecosystem state, we will expect measures of relative change from the current state to hold in trend direction across a range of parameter values (where the range is either derived from posterior distributions or from sensitivity analyses). We will assess forecasting ability based on whether abundance estimates of focal species remain within +/-30% range across a set of parameter draws (e.g., within the simulation testing). The predicted "state of the ecosystem" may differ from recent years, including differences in frequencies of extreme climate events and regime changes.
- From the strict hypothesis standpoint (e.g., elements of the EMC story), we will use posterior distributions from the appropriate models to produce 95% probability limits for changes due to specific effects; this will serve to reject hypotheses which may be spurious results of time series calculations. We will present each relevant BSIERP hypothesis and each component of the EMC "ecosystem story" with the prior time-series analysis or conceptual modeling from the literature which led to that hypothesis. Then, each model will be scored on whether it "fits" the hypothesis. If a model has a poor fit to each hypothesis, two possibilities arise. The first is that the model contains insufficient detail or functional form to capture that aspect of the ecosystem. The second is that the story itself contains contradictions which come out in the modeling. Evaluating the "pieces which fit" in conjunction with the "models that fit" cannot be performed by strict likelihood criteria. Rather, it will lead to a broader discussion (including discussion with data providers and non-BSIERP experts) to evaluate the meaning of a model which fits poorly even when likelihood components are maximized.
- For development of long-term future predictions, and with regimes in particular, there is a tradeoff between tight fitting and flexibility. The ensemble of IPCC projections strongly suggests that the climate in the next 50+ years will enter uncharted territory, with monotonic increasing trends overpowering decadal variability (e.g., known regimes) within the next 20-30 years. Within the IPCC suite of models, each model has different trade-offs in features: in hindcasting, one model one may correctly capture ice retreat while another model may better duplicate the PDO. We do not have completely independent models in the BSIERP effort as the IPCC effort

has produced. Still, it is likely that different local minima in the fitting criteria for a "total" model score will fit different partial scores in a better manner [see Q19 for fitting criteria details for the vertically integrated models]. This is especially likely for story components which suggest how rules change across regimes. Each of these scores, in some sense, will represent the goodness of fit to a specific aspect of the "story".

- In some cases, parameters may be challenging to estimate, but the management regime may be robust to error in the parameter estimates. For long-term robustness testing of management regimes (e.g., in the Management Strategy Evaluation), we will focus on evaluating robustness of already established management models (including their uncertainty estimation routines) to the "plausible envelope of alternative ecosystem states" arising from the operating model (FEAST) (see "Measures of Uncertainty" discussion in answer to [Q9] for the definition of "plausible envelopes."

This process of determining fit quality will occur throughout the project, through workshops, symposia, and potentially meetings with the EMC. In general, measures of uncertainty will be assessed separately for each model component and for the vertically-integrated model as a whole.

With increased funding (see Q17 answer for funding gap description), one end product will be an "ensemble forecast" of the parameter sets which fit a reasonable selection of the story components well, as well as some expected winnowing and potential rejection of hypotheses or story components.

> ***[Q8]*** *How will the test of fit deal with the anticipated excess of free parameters in the integrated model?*

A8 (Aydin): For the vertically integrated models, the addition of spatial data through acoustics and trawl surveys will actually provide quite a wealth of data, although the independence of neighboring data points will need to be assessed through correlative analysis to determine effective sample sizes. A likely approach to ensuring appropriate sample sizes, while avoiding overcorrelated data, is to develop "pattern indices" such as centroids of distribution which are measurable from both data and models (see "specific predictions" list in [Q9] answer for a larger list of pattern indices which are calculated independently from models and data.

The addition of spatial data to a certain extent limits or replaces parameters rather than requiring large numbers of new parameters; for example, complex functional response parameters are replaced with bioenergetics parameters (taken from the BSIERP bioenergetics study (O2.24) and thus independently verified) and with simple movement rules. A greater difficulty than limited numbers of data points may actually be lack of contrast in the data which exists for projecting across regimes. For example, as most data was collected after 1982, there is limited simultaneous fitting to contrasting regimes. Our method for dealing with this is answered in [Q20].

Still, a step-wise fitting procedure will be required as described in Q19; parameters which are poorly estimated by fitting will be constrained by prior probability distributions. For constrained parameters, sensitivity analyses will be performed and reported.

For the competing models (time series, behavioral foraging, and correlative models) our primary approach will be to reduce model complexity so that the number of data points is at least 3X the number of parameters, implying that our modeling approach favors statistical estimation and representation of uncertainty in model forecasts over model complexity. In addition, as much as practicable, we will input externally estimated parameters with uncertainty (priors) (e.g., stock-recruitment parameters for pollock).

__[Q9]__ Exactly what fishery-relevant endpoints are the vertically integrated models committing to predict as their bottom line, and how will measures of confidence be attached to those predictions?

A9 (Aydin, Punt, Bond): *Specific predictions*

Overall, models generate terabytes of "data", so developing the appropriate metrics of biomass and communities (e.g., the "state" predictions) is an iterative process between modelers, field researchers, and the broad user community (see "general design principles for prediction development", below). From the outset, we envision the following specific predictions:

- Oceanographic conditions (temperature, salinity, current fields) on a 10-km (entire Bering Sea) or 3-km (southeastern Bering Sea) resolution.
- Time trajectories and 10-km or 3-km spatial distribution of the plankton (NPZ) community.
- Spatial distribution and production of walleye pollock (biomass, recruitment, growth). Emphasis will be placed on shifts in frequency (e.g., increase in variability, change in frequency of good vs. bad recruitment years, increased sensitivity to "regimes").
- Spatial distribution and production of euphausiids and other forage species. Emphasis will be placed on shifts in frequency (e.g., increase in variability, change in frequency of good vs. bad recruitment years, increased sensitivity to "regimes").
- Production rates of Pacific cod and arrowtooth flounder, estimated by overlaying predicted prey fields and temperature, and the results of the BSIERP seasonal bioenergetics study (O2.24), to map predicted "dynamic growth habitat" (growth potential in space) of these species. Recruitment scenarios for arrowtooth flounder and Pacific cod, predicted from the retrospective/biomass dynamics component of the modeling. Pacific cod predictions will be integrated into FEAST at a level of detail similar to that of pollock, based on a study requested in the NPRB 2008 RFP, if data from this study is available to modelers in a timely manner.
- Prediction of the expected impact of increasing cetacean populations on forage fish and therefore on fisheries.
- Downscaled indices of fine-scale patchiness on a scale relevant to predictions from the BSIERP patch dynamics study (O4.62).
- Indicators of community foraging success of central-place foragers (birds, mammals, and local human communities) calculated through measurable condition indicators including breeding success and economic output, and based on threshold analysis and shifting centers of distribution.
- Overall energy (total secondary production) of the benthos.
- Fisheries locations and concentrations, total economic value, and spatial economic profitability (based on production and distribution/travel costs). Prediction of these impacts on individual communities, focusing on LTK communities. Example metrics are travel distance to winter and fall pollock fisheries, pollock concentration (size and location of area that encompasses 80 and 95% of stock (centroid), a measure of how much effort will be necessary to reach quota), and pollock fishery value.

It is expected that several of these quantities can be analyzed further by other (non-BSIERP) researchers; for example, the prediction of oceanographic conditions can be used along with developed correlations to improve single-species stock assessments; if this occurs through independent collaborations, such results will be included as part of the "prediction".

The time target for long-term predictions will be approximately **50 years** in the future, described by Cox and Stephenson (2007) as a "sweet spot" of lowest uncertainty between initial condition uncertainty and scenario uncertainty in IPCC climate scenarios. However, in the management strategy evaluation (MSE) component, each individual prediction listed above will be scored at 5-year, 10-year, 20-year, and 50-year prediction potential compared to current stock assessment or recruitment predictions which have been produced by AFSC or other researchers (see MSE component for methodology). The five- and ten- year predictions in particular will be tested by validation through withholding data within and across regimes to measure fit to data (see answers to EMC questions on testing across regimes).

Measures of uncertainty

All of the individual models will produce estimates of uncertainty (confidence intervals or posterior probability distributions) for model predictions and these shall be reviewed. For formal evaluation by reviewing fisheries scientists, we will report relative probability density functions and cumulative distribution functions where feasible with available computing power; variance estimation will require multiple runs with differing weighting of data streams and re-estimation to determine the sensitivity of the model to data variance. Density functions will be compared between models, to explore the consequences of admitting additional uncertainty.

For spatial fitting of plankton and fish, we have defined *a priori* a set of "spatial indicators" (e.g., centroids of distribution, spread, and orientation/shape (e.g., North/South) of distribution) to describe goodness of fit to data. This "pattern-based" approach offers a direct method of aggregating 10km and 3km models to 37km resolution survey design (between-transect distance for acoustic surveys (O2.26), between-station distance for bottom trawl surveys (O2.25). These pattern indicators are scientifically responsive to hypotheses (e.g., mean foraging distance for a central placed forager) and management (e.g., trip distance for fishers). An analysis of the data itself will be performed to assess relative independence of these derived indicators.

In models for which formal Bayesian analysis is infeasible, "posterior probability distributions" will be represented by "envelopes of plausibility" derived from ensemble/Monte Carlo runs of each model, where the metrics for plausibility will be developed in conjunction with review by BSIERP field scientists and interviews with stakeholders/local communities (e.g., see "Delphi method" described in the "Interaction with Stakeholders" component in the revised proposal).

With additional funding (see Q17 answer for research gap), model predictions also will be compared in a blended forecast similar to that produced by the Intergovernmental Panel on Climate Change (IPCC) (see MSE component, M.55).

General design principles in prediction development

The suite of BSIERP models will provide predictions of Bering Sea future ecosystem production, spatial species distribution and inferences on broad changes in biological community structure (including human communities), and changing variability and uncertainty derived from scenarios of global climate change. The **specific predictions** will be indexed in time and space and production and distribution of walleye pollock and other forage species will be a focal point. Arising from the forage-centric focus, we will predict the relative production (based on prey supply) of Pacific cod and arrowtooth flounder populations. Furthermore, through the analysis of central-place foraging thresholds, we will predict changes in productivity of northern fur seals and seabirds, and thus the health and vitality of the human communities dependant on them. Pacific cod predictions will be "upgraded" to a full stock prediction if data resulting from the upcoming cod-focused work already solicited in the 2008 NPRB RFP can be included in a timely manner.

Another specific and rather novel product will be the development of indices of fine scale prey patchiness, which will be derived by related model-derived indicators on the 3-km scale (storminess, water movement, prey densities) to finer scale field studies undertaken by the patch dynamics study (O4.62). This fine scale patchiness may be an important element in local predator/prey interactions. Because it is computational infeasible to model on this fine a scale, such indices may offer the best hope for predicting the foraging success and thus viability of marine mammals and birds. We shall ask if 50-year climate change scenarios will dramatically increase the variability and scale of these indices, and thus foraging opportunities and resulting population trajectories.

We will estimate several parameters of relevance to fishermen, fishery managers and fishery economists, as well as the uncertainty (posterior probability distributions or ensemble envelopes) associated with these parameters. These quantities will be represented both as absolute and relative (referenced to present) values in case the models correctly represent trend, but not scale.

If the research gap identified by the EMC ("… MSE … seems a little thin …") is filled (see Q17 answer for our approach for filling this research gap), we will produce an "ensemble" set of predictions from these models on the common quantities (time series in biomass and recruitment) that they predict. The ensemble modeling will include the past back to 1960, so that it crosses known regimes to evaluate the tradeoff between simplicity of model and successfully crossing regime shifts.

With additional funding (see Q17 answer for funding gap description), the probabilistic nature of model forecasts will be communicated using novel indicators of direct relevance to stakeholders (e.g., NPRB/PICES workshop; Kruse et al. 2006). For example, uncertainty can be shown as frequencies of poor catch generated through Monte Carlo simulations; a 20-year "drought" of reduced pollock catch could be expected to occur much more often in high fishing than in low fishing scenarios. Indicators will be expressed in relative (percent change due to policy or long-term climate) rather than absolute terms (expected catches).

MODELING COMPONENT STATUS

The BTU proposal, the unsolicited early August addendum from the BTU team, and the Bond proposal from the NSF BEST program constitute what we have in writing, so far, about the BTU plans for modeling. The Bond NSF proposal has been funded. The rest is held in abeyance.

Physical and NPZ Modeling

The Bond proposal covers the physical and NPZ portions of the ecosystem. These will be modeled in a common coupled model in a common grid. The physical portion will be driven by a suite of climate scenarios selected from the IPCC forecasts and downscaled to mesh with the physical and NPZ model grid. Though the decision has already been made to fund this proposal from BEST, the EMC needs to include this project in its review of the BTU modeling, since the BTU upper trophic level portions will be integrating with this project for their NPZ and physical driving. The FEAST model component of upper trophic level BTU is clearly intended to integrate directly with the Bond proposal physical NPZ model in a 2-way coupling.

> ***[Q10]*** *We need a discussion to verify that the specific phytoplankton and zooplankton compartments (size and taxon categories) which will be modeled explicitly in the NPZ model are sufficient for driving the dynamics of the respective upper trophic level compartments in the FEAST model.*

A10 (Gibson, Bond): The NPZ model will represent two size classes of phytoplankton, two size classes of microzooplankton, two size classes of copepods, and a euphausiid compartment. In general, this resolution is appropriate for predicting forage fish or walleye pollock diets and production (based on data from AFSC groundfish food habits database and surface trawl (O2.23, BASIS surveys,). This resolution is sufficient to examine most of the ecosystem control hypotheses put forth as part of the EMC "ecosystem story" and the BSIERP hypotheses (BSIERP Study Plan, Appendix 2). Specifically, the current working hypotheses suggest that differences in production rates between copepod genera (*Pseudocalanus* versus *Neocalanus*) may influence production, and that changes in the microzooplankton community resulting from ice-edge blooms and stratification may alter energy flow. These groups are distinctly and spatially modeled in the oceanographic/NPZ component.

The limiting factor in linking NPZ and fish foraging models, in terms of taxonomic and especially size class resolution, is the historical resolution of prey in fish stomachs (food habits data). In particular, zooplankton species have been identified to copepods or euphausiids, but not in general to a finer scale (e.g., copepod species) with any degree of consistency. The diet collections associated with BSIERP (e.g., Functional foraging response, O2.16) will resolve this to a certain extent by using increased taxonomic resolution; meanwhile, overall the species planned for the NPZ model, including multiple size classes of copepods in particular, will provide the maximum resolution which can be reasonably modeled for fish.

One aspect which will need further consideration (e.g., the possible inclusion of dynamic "additional" mortality terms on zooplankton) will be the effects of large, predatory zooplankton other than euphausiids, such as chaetognaths and pteropods. These taxa can not be dynamically modeled explicitly with a reasonable degree of accuracy because life history and rate parameters are lacking. A range of mortality values including plausible trends will be included in ensemble runs. The importance of including dynamically-modeled predatory zooplankton will be judged by examining residuals of NPZ fitting results during hindcasting. Measuring their production rates in relation to oceanographic conditions is a potential area for future research (i.e., a current research gap).

N. Bond will be responsible for overall project management of the climate portion of the modeling; A. Hermann will be responsible for the ocean circulation portion of the modeling and interfacing with the biological modeling; G. Gibson will be responsible for the NPZ portion.

The main investigation of uncertainty in this physical/NPZ modeling project will be carried out by quantifying the variability in response to: (1) between-climate-model differences in the IPCC runs, and (2) between trajectory differences driven by stochastic variation within-trajectory to represent inherently stochastic (mainly physical) processes, and (3) sampling different fixed, but uncertain, rate parameters (mainly biological) "from a prescribed range."

Components (1) and (2) of this uncertainty analysis are straightforward and meaningful. The reasonableness and informativeness of component (3) will rest entirely on the choice of the "prescribed" ranges, for what is otherwise just a sensitivity analysis. Ideally, the retrospective fitting of the model to data should reveal the actual range of uncertainty in the estimated parameters, and this range in turn should be explored in quantifying the consequences of the parameter uncertainty for the uncertainty in predictions. The proposal is not clear about the formalities of their procedure for estimating rate parameters: "...data will be used for tuning of internal parameters to achieve the most dynamically consistent nowcast of present conditions."

> ***[Q11]*** *The Bond modeling team needs to articulate a statistical framework for their plans for estimating model rate parameters (especially the biological) so that the consequences of parameter uncertainty can be represented realistically in the planned sensitivity analyses.*

A11 (Hermann, Bond): A series of modeling experiments will being carried out to assess the effects of parameter uncertainty. These experiments primarily will involve the NPZ modeling component, because that component requires tuning many not well known parameters. The simulations have been, and will continue to be mostly in 1-D, to allow a sufficient number of realizations to compile meaningful statistics. This approach will establish the sensitivity of modeled outputs (such as zooplankton concentrations) due to prescribed parameters (such as phytoplankton growth rates). The ranges of values for the latter parameters are based on previously published results, and, to the extent it exists, data specific to the Bering Sea. The results will be analyzed with relatively simple statistical methods such as multiple linear regression.

The proposal suggests, but does not quite commit to, a "useful" test to at least indicate the prospects for future predictive power. They propose to test whether "tuning" the model to pre-1977 data will result in a model which successfully hindcasts the actual 1976/1977 "regime shift" when driven by the actual climate trajectory. This is an interesting idea which deserves more detailed development. It would also be worth asking the upper-trophic-level modeling teams to commit to the same test.

> ***[Q12]*** *We need a more detailed work plan and specifications for the test of the ability of a model that was fit to pre-1977 data to correctly hindcast the ecosystem changes observed in the mid 1970s.*

A12 (Bond): The changes in the physical oceanographic environment and upper trophic level species during the 70's are reasonably well documented (Livingston et al 1999; Schumacher et al 2003). On the other hand, phytoplankton and zooplankton data are scarce for the pre-regime shift period in the SE Bering Sea. As a result, the ROMS-NPZ model ensemble will depend on the planned coupling to models developed by the UTL modeling team for species such as pollock and cod, for which there is data. Further, cross-regime model validation will depend on comparison of observed and predicted upper trophic level data, rather than the lower trophic level data where there is little data.

In addition, the NPZ model initially will be tuned with data from more recent cold period for which observational data is available. Ideally, a correctly tuned model ensemble would be able to capture the ecosystem response to a shift in climate conditions. The experiment outlined in Q13 will give us confidence that the model is up to this task. Finally, NPZ model output pre and post regime shift may provide useful insights into why such large changes in upper trophic level biomass were observed during this period.

The proposal states, as the first of its two guiding hypotheses, a version of the OCH: that the timing of sea ice melt operates as a switch controlling the timing and nature of the spring bloom and the food chain that it supports. This also features in the "conceptual model" of the overall BTU proposal to NPRB. It would be telling to construct a cross-validation test around this specific phenomenon also, which may not be entirely isomorphic to the mid 1970s "regime shift."

> ***[Q13]*** *We need a discussion for testing the ability of the model to correctly hindcast the circulation and NPZ of early-ice-melt years when fit to (as recent as available) late-ice-melt years, and vice versa. Except for some long-lived zooplankton, most of the dynamics in the NPZ modeling are within-year effects, so it should be feasible to calibrate by-year to a subset of not necessarily consecutive years. If the modelers think this is a good idea, they should give us a work plan. If they think this is a bad idea, they should tell us why, and propose an alternative.*

A13 (Gibson): Most biological functions describing behavior of the biological model component are modified with temperature dependent rates. The temperature functions employed follow those commonly used in the modeling literature. This approach will be essential to successfully simulate a changing ecosystem with change in climatic regime. Gibson is presently testing the ability of a 1-D version of our Bering Sea NPZ model to correctly hindcast the LTL ecosystem dynamics of the relatively warm (early-ice-melt) 2004 following tuning to observational data from the relatively cool (late-ice-melt) year of

1999. This experiment is being used to determine success of parameter tuning for the biological model. Ideally, a 'correctly' tuned biological model will be able to simulate the key features of the observed ecosystem dynamics in both early-ice-melt and late-ice-melt situations with a single parameter set. Following successful completion of the 1-D testing, these experiments will be extended to the fully 3-D model simulations of 1999 and 2004.

The Bond proposal mentions use of anticipated field data from the Hopcoft BEST project for finer time and space resolution of NPZ, but NSF did not fund the Hopcroft proposal. The Bond proposal also states the importance of new data bearing on biological rate parameters, especially temperature dependence, but is not specific about who will provide those data needs.

> ***[Q14]*** *We need a more thorough discussion of the priority data needs for calibrating and validating the NPZ model, and the extent to which these will be satisfied by the funded field program.*

A14 (Gibson): The biological portion of the ROMS-NPZ model is being tuned for application to the Bering Sea, and hence requires data on plankton biomass and rates. Phytoplankton primary production is a fundamental measurement as its magnitude plays an important role in driving the overall dynamics of the ecosystem. Biomass measurements are necessary for constraining the model. Plankton process rates i.e., grazing, growth etc. are generally highly temperature dependent. This modeling effort would be greatly enhanced by additional data on lower trophic level process rates for species found in the study area. Rate information for both the spring and summer would help fine tune the model and ensure that it produces a correct simulation for the correct reasons. It is important to note that while these rate measurements will be highly valuable, particularly in light of a warming ocean, they should not come at the expense of biomass observations which are needed for assessing model accuracy and utility.

In the absence of any new biomass and rate measurements, we will be forced to rely on historical data for model tuning and validation and will experiment with rates that fall with in the range of values published in the literature. There are a number of historical data types that could be used for this purpose. For example; surface chlorophyll from the SeaWiFS and MODIS satellites, shipboard and mooring observations collected through PROBES (1978-1982), Bering Sea FOCI (1991-present) and the Inner Front Project (1999 and 2004) provide a picture of the general horizontal and vertical distribution of primary and secondary production along with limited process rate information on nutrient uptake, and growth and grazing rates of the plankton in the SE Bering sea. Without new data, we will estimate the probable errors due to uncertainties in rates and other processes based on the results of the 1-D experiments currently underway.

Some lower trophic observations will be made in the southeast Bering Sea through the BEST program during 2008-2010. However, to our knowledge, this program is focusing on collecting spring rate data, whereas both spring and summer data are needed. We have heard, but haven’t seen in writing, what may be approved by BEST and NPRB for summer. Thus we will identify those summer data needs that support our modeling, so as to make priority data needs clear. Those priority data needs are summer biomass and rate information. Summer primary production (e.g., as proposed by Lessard within the Hopcroft proposal) is extremely valuable, as ecosystem dynamics on the BS shelf are, in large part, determined by the post-spring bloom production. There is presently very little data available on summer production rates and biomass in this region. Summer plankton data (e.g., as proposed within the Hopcroft proposal) organized by the same guilds as represented in the NPZ model also are extremely valuable (small and large phytoplankton, small and large microzooplankton, small and large copepods and euphausiids). While the existing and expected data from the Bering Sea is sufficient to make headway in the NPZ modeling, the scarcity of summer biomass and rate data will limit our ability to tune the model properly, and hence will compromise forecasting reliability.

Upper Trophic Level BTU

The suite of 6 upper trophic level BTU models under consideration for NPRB funding are:

a. FEAST-Forage/euphausiid abundance in time and space (K. Aydin; A. Hermann; A. Hollowed; B. Fadely; M. Dalton)

b. Spatial economic models for pollock and cod (A. Haynie)

c. Behavioral foraging model (M. Mangel)

d. Correlative biomass dynamics model (G. Kruse; F. Mueter)

e. MSE I: Competitive existing models for blended forecasts, and management strategy evaluation (A. Punt; K. Aydin)

f. MSE II: Management resilience study (K. Criddle)

(a) The FEAST model of upper trophic level elements (capelin, eulachon, sandlance, myctophids, squid, shrimp, pollock, cod, arrowtooth flounder, salmon, kittiwakes, murres, fur seal, humpback whale, fin whale, transient killer whale) sounds like it will operate inside the common grid of the physical and NPZ model in a 2-way coupling.

(b) The spatial and economic model for pollock and cod sounds like it will operate inside the common grid of FEAST and the physical and NPZ model in a 2-way coupling.

So the climate-circulation-NPZ-FEAST-Spatial/economic models are intended to constitute a seamless main sequence for the BTU ecosystem modeling.

(c) The behavioral foraging model appears to provide an alternative set of foraging rules, based on optimization considerations. These conceivably might be substituted into FEAST, in place of the foraging rules that otherwise will be used in FEAST. But the proposal does not really state how the behavioral foraging model will be integrated into the program, or how different, in the end, its foraging rules will be from those already in FEAST.

***[Q15]** We need a discussion of how the behavioral foraging model will be integrated with the rest of the project, and what difference it will make.*

A15 (Mangel, Aydin): For upper trophic-level species, uncertainty in the scale of variation in predator/prey functional responses is a limiting factor in making successful predictions across regimes. Model error in specifying the functional response is of particular concern. When functional responses are calculated on a whole population/ecosystem scale (e.g., as in Ecosim), the "functional shape" (e.g., Type II, Type III) is assumed to be a direct response of a predator to local prey concentrations. In fact, the shape may be a phenomenological artifact of finer-scale mechanisms; for example, masking shifts in geography, species overlap, or age structure (see Aydin 2004 for an example of an Ecosim "functional response" disguising shifts in the age structure in a predator population).

Within the vertically-integrated model sequence, we will model predator/prey interactions on a finer spatial scale to separate the overall population "functional shape" into local-scale mechanisms. However, for central-place foragers (seabirds and fur seals), this approach may be insufficient. As described in the "species resolution" section of the revised proposal, inferences on bird and fur seal production will be made by comparing historical and BSIERP-collected data on their distribution with prey fields reconstructed by FEAST, to estimate changes in foraging range and success assuming continued concentration on the same prey sources. However as FEAST includes limited behavioral plasticity; prey switching across regimes cannot be predicted until it is either observed or predicted based on theoretical grounds; the behavioral foraging model provides the latter.

In hindcasting mode, the behavioral foraging model will be fit to observed population trajectories given hypothesized dietary switches, as described in Mangel and Wolf (2006), where a variation of this technique was used to examine the killer whale "serial depletion" hypothesis for Steller sea lion declines. In forecast mode, the behavioral foraging model will rely on the vertically integrated model for prediction of prey concentrations (NPZ -FEAST) but will use state-space modeling to construct an alternative set of predictions for the central-place foragers of local community interest (seabirds and fur seals). The predictions will necessarily be a plausible range of behavioral patterns and corresponding population trajectories (an ensemble of possibilities) rather than a single prediction with an error range. As the central-place foragers are not the subject of direct and explicit dynamic modeling in the vertically integrated set, but rather are modeled through inference, this alternate "competing" result (competing prey switching vs. inferences made from prey field predictions) is a valuable addition to the package.

(d) The correlative biomass model is essentially an extension of the ongoing retrospective statistical analysis. It might provide a statistical evaluation of the BTU "conceptual model," but the proposal does not explicitly state that the subproject will be structured that way. This subproject is not intended to integrate functionally with the other modeling, but its statistical evaluation of the "conceptual model" that is purported to underpin that modeling should be illuminating, since the data which will be examined in this statistical retrospective could be much the same data as will be used to calibrate the other models.

> ***[Q16]*** *We need an explanation of how the correlative biomass model effort, and the other retrospective statistical analyses (including the one ongoing), will relate to the BTU "conceptual model." To what extent will they be attempting to predict the same things as the main sequence model; to what extent will they be using the same data; how will temporal and spatial scales be handled to allow comparability?*

A16 (Mueter): The biomass dynamics approach is intended as an extension of the retrospective analysis to test hypotheses about how trophic interactions among gadids, flatfish, and crustaceans and their relative biomasses have changed in response to past climate variability (in retrospective mode) and how these interactions and biomass trajectories may change under different climate scenarios (predictive mode). As described in Collie and DeLong (1999), the scale of the correlations and predictions will be ecosystem-wide; e.g., using annual climate, biomass (from surveys and or stock assessment models), or other time series indices which are gleaned from the retrospective study (O.3.30). Therefore, the approach is both complementary to the vertically integrated modeling approach ("main sequence model"), as well as "competing" with that approach. It is complementary in that the variability in productivity and biomass "explained" by the correlative approach can confirm or question the underlying, finer-scale mechanistic parameterization of the vertically-integrated model; if the strongest correlations in this model run contrary to the explicit mechanisms of the vertically-integrated model, it would provide a significant challenge and an alternative model for consideration. The biomass dynamics approach "competes" with the vertically integrated modeling approach in that it provides alternative predictions of relative biomass trajectories of several of the focal species (Pacific cod, arrowtooth flounder), as well as for crab, at the ecosystem scale.

Important differences between the biomass dynamics model and the vertically integrated model are (1) that the latter includes an explicit representation of lower trophic dynamics and (2) that the latter is spatially explicit and uses a monthly time step, while the former is aggregated across space (at the scale of the southeast Bering Sea) and uses an annual time step. Both models predict some of the same quantities (such as biomass of key species), but at different spatial and temporal scales. By aggregating results from the spatially explicit model over time and space, predictions of the relative biomass trajectories of different species can be compared between models. This will allow an evaluation of what is gained (or lost) in terms of fitting observed abundance trends in the vertically integrated approach relative to the much simpler, but less realistic biomass dynamics approach. Moreover, the biomass dynamics approach can provide an independent check of some of the functional relationships (or "observed" relationships) in the vertically integrated model by testing how well models with different functional forms (competitive or predator-prey relationships) fit the available data.

The biomass dynamics modeling approach itself is more sophisticated than past correlative analyses performed in the Bering Sea by rigorously applying appropriate functional forms of species interference/interactions (e.g., logistic response variables). Its inclusion alongside the vertically-integrated model set specifically addresses (original) EMC question (e), in that it presents alternate parameterizations of relationships for formal model choice, especially for distinguishing competitive or "interference" effects between species that may not be evident from modeling using the process-oriented approach of the vertically-integrated model. For example, limitations due to habitat or other life-history bottlenecks of species that are also in the FEAST model may not be picked up by the vertically-integrated model but could be captured by this analysis, as it will include for example the inclusion of a "joint carrying capacity" variable in the regression which is independent of predator/prey concerns. Finally, this correlative modeling approach allows for the explicit consideration of energy pathways that were not included in the current BSIERP vertical-modeling due to funding limitations; in particular uncovering these relationships with benthic components (such as crab) will be a first step towards later, explicit examination of the benthic food web.

(e) The competitive existing models subproject will compare a number of alternative models of varying complexity. The alternatives cover an interesting spectrum. It is not clear from the proposal what will serve as the "ground truth" in the comparison. Perhaps the intention is to treat the main BTU model as the operating model, and evaluate the alternatives, or combinations of alternatives, as management strategies. If so, this would be interesting, as it does address the question whether alternative simpler models might give equivalent predictions to the BTU main model. But it does not get at the question of the predictive power of the BTU model itself.

> ***[Q17]*** *We need more specifics concerning the competitive existing models MSE, including a work plan (the staffing, with one post-doc, seems a little thin relative to the magnitude of the task).*

A17 (Aydin, Punt): We have increased the scope of the MSE component, first of all by adding Jim Ianelli (AFSC) as a co-PI for this project (in-kind contribution). We also have identified two research gaps in the work plan, in response to the EMC comment that "… MSE … seems a little thin …".

Two methods of decision analysis are proposed to address management responses to long-term projections and the usefulness of those projections. The first is a formal Management Strategy Evaluation (MSE) by a four-year postdoc, Andre Punt and Jim Ianelli. The evaluation consists of cross-testing of the whole model suite and its components using the **vertically-integrated models as an "operating model"** and currently developed methods (single-species stock assessments, multi-species modeling (MSVPA, MSM), and Ecosim) as "assessment" models. We will evaluate a set of models currently available for the Bering Sea: (a) Single species stock assessments w/ correlative recruitment indices (e.g., Ianelli et al. 2006; Wilderbuer et al. 2002); and (b) MSVPA and MSM (Jurado-Molina et al. 2005). Additionally, we will examine autocorrelative biomass dynamics/network models (Gaichas, 2006) and nonlinear correlative models (Hsieh et al. 2005) as "null" models for testing the value of more mechanistic approaches. Such a thorough analysis of competing models for the same ecosystem will provide value to

Bering Sea management efforts and future modeling advice for other ecosystems. This application of MSE will consider management strategies in a broader context than has been the case in the past and will specifically attempt to implement the guidelines of Marasco et al. (2007) as regards evaluating management strategies in an ecosystem context.

The metrics for evaluating success of the "simpler" models against the operating models in the MSE will be (1) success of hindcasts and predictions made by the simple models fit to data subsampled with observation error from the operating (FEAST) model, measured by likelihood/goodness of fit (i.e., improves or degrades goodness of fit). This will include individual model skill at determining past and current states (hindcast/nowcast) as well as success of models at predicting future states from current states; (2) for management models, relative success of determining fishing policies and reference points which keep fish populations and yields above a "best performance" reference point determined from the operating model; (3) for correlative models (including stock assessment models), we will perform an experiment in which the correlations are "ground-truthed" or "adjusted" periodically based on the state of the operating model. In other words, how often should an "intensive field and retrospective sampling season" be performed on the operating model to ensure that a correlative model continues to give a good answer. This final product is in reference to judging the potential success and necessity of future BSIERP-scale ecosystem projects.

The first research gap lies in the production of ensemble models. In the ensemble model component, blended model averages, including Bayesian integration and overall uncertainty, will be produced. We also will further evaluate model strengths, weaknesses, and uncertainties. This component will use the IPCC climate modeling reporting methods as an example for reporting model results, including tools for browsing outputs and more readily understood metrics such as "pollock droughts." The development will specifically include partnerships with outreach and LTK components to continually ensure that model results are relevant and understandable to stakeholders, including modeler and stakeholder workshops. We envision this component as a single, two-year postdoc **($150K research gap)** working with Andre Punt, Jim Ianelli, and Kerim Aydin.

The second research gap is a risk assessment component. The risk assessment is distinguished as separate from the MSE component above, despite similar techniques. The risk assessment will address the significant management challenge for fisheries management to avoid affecting marine mammal populations that are declining. The language of fisheries affecting marine mammals is the percent chance of becoming endangered or threatened, and will specifically encompass risks involved with behavioral changes identified by the behavioral foraging component, above. One critical piece of this risk assessment analysis is that it is a unique opportunity for the divergent management worlds (extractive vs. protective) to meet and form a common language outside of the crisis management mode (e.g., "interacting in crisis mode" means waiting until an ESA consultation is required). We envision this project as an additional two-year postdoc with Andre Punt **($150K research gap)**.

(f) The management resilience subproject is intended to examine how alternative incentive structures in the pollock (and possibly Tanner crab) fisheries would function in the contexts predicted by the main BTU model. Note that the main BTU model does not propose to model Tanner crab.

> ***[Q18]*** *How does the management resilience project link to the rest of the modeling, and how important is it that this linkage take place? Will the results of the management resilience model feed back into the BTU main model showing the effect of the fisheries on the ecosystem? In this respect, how will this submodel differ from the spatial fishing choice submodel (Haynie) which is coupled into the BTU main sequence?*

A18 (Aydin): The Criddle et al. proposal would provide management strategy information for valuable commercial species (crab) in the Bering Sea. However this project is less directly tied (and not as easily compared) to the vertically-integrated model. Thus the Criddle et al. proposal has been considered a lower priority and was dropped to stay within the NPRB modeling budget. This proposal should be revisited for other NPRB RFP, as it would provide useful information for management of IFQ and quasi-IFQ fisheries that is resilient to climate change and variability.

<u>Uncertainty Analysis in the Upper Trophic Level BTU Main Model</u>

The proposal is not very incisive about how predictive skill will be evaluated for the BTU main model sequence. Unless they have a better alternative, the BTU model team should consider extending to the upper trophic levels, the Bond proposal idea of testing whether a model fit to the pre-1977 data correctly hindcasts the mid 1970s regime shift, and also the suggestion provided in this review of testing whether a model fit to the rates observed in the late-ice-melt years correctly predicts observations in the early-ice-melt years, and vice versa. This will require adoption of a reasonably formal approach to fitting and "tuning."

[Q19] *We need formal specifications of how model fitting and "tuning" will be carried out.*

A19 (Aydin, Gibson, Hermann, Punt): Model fitting will be achieved using formal and semi-formal techniques. For lower trophic level components (i.e., NPZ), “model fitting” will necessarily involve limited tuning of parameter values guided by the ability to mimic monitoring data and formal fitting of submodels using maximum likelihood. In contrast, the models of upper trophic species will be fitted conditional on the predictions from the lower-trophic models using penalized maximum likelihood and / or Bayesian methods. The coupled modeling also is computationally intensive; further model testing will occur by iterating between 2-D and 3-D models (also see A11, 1-D testing).

As described for the NPZ component, the computational expense of a single run of the model will preclude a full, formal optimization of parameters to an objective function measuring goodness-of-fit. Fitting will be carried out in stages, especially in the 2-D transect and latitude/longitude grids, with sets of parameters fit in a step-wise manner; likelihood calculations for the data presented above will be used as an objective function. Some parameters will require prior probability distributions for model convergence; for parameters which the model is not able to fit and must be set, sensitivity analyses will be conducted. Several techniques for optimization (e.g., gradient descent, simulated annealing, neural networks, and genetic algorithms) are under consideration for the 2-dimensional version.

Examining the strength of two-way coupling between zooplankton and fish is an important computational issue in relation to fitting. We will only have sufficient computing power to produce about 30 “full runs” of the completely coupled 3-D runs through the supercomputing resources available; therefore the 2-dimensional runs will be used for much of the initial parameter exploration. However, the 2-D grid most appropriate for zooplankton is transect (depth and transect) while the most appropriate 2-D grid for fish is spatial (latitude/longitude). At the same time, one of the goals of this project is to produce coupled runs investigating the top-down control applied on zooplankton by fish.

To resolve this dilemma, we will proceed by (1) first fitting 2-D transect NPZ models without feedback from fish; (2) running the fit 3-D NPZ model without dynamic fish on the supercomputer to produce 3-D outputs; (3) using these outputs to produce depth-integrated spatial prey fields as 1-way coupling for fitting fish parameters; and finally (4) combining the resulting parameter sets in full 3-D runs. The comparison of results from (2) and (4) will indicate the amount of coupling needed between fish and zooplankton ultimately. Even if this process is repeated several times, most runs for fitting will be performed in the 2-D steps (both transect and spatial), which can run on standard workstations in a matter of several hours.

For the vertically-integrated models, model fitting will be assessed by the goodness of fit and likelihood functions for the upper trophic levels (above NPZ), and will be both temporal and spatial with respect to pollock and other forage species (FEAST Forage Species Component). The fitting will rely on data collected during standard NOAA surveys described in the BSIERP Study Plan (acoustic, bottom trawl, surface trawl, Appendix 2). The model will resolve space on a scale (3-km horizontal scale) finer than survey spacing for the bottom trawl (37-km between stations), acoustic (37-km between transects), and surface trawl (~56-km between stations) surveys; size-specific species CPUE from the surveys for the period 1982-2007 (and through 2010) will be a specific component of the likelihood. There will be a specific evaluation of autocorrelation and effective sample size to score goodness-of-fit (minimization of MSE error) appropriately.

The data for fitting or judging the goodness of fit of time series outputs (for both FEAST, behavioral foraging, and correlative biomass dynamics models) will include time series used in current single-species stock assessments (biomass indices, fisheries catch, age composition (fisheries and survey). For correlative biomass dynamics, time series from single-species outputs (e.g., recruitment, biomass) will be used while recognizing that these "data" are in fact model outputs from single-species assumptions for the most part. Therefore, correlation of "data" will be considered carefully for appropriate sample size calculations. Where appropriate (e.g., choosing between similar structured models with parameter differences), formal model selection through criteria such as the AIC will be employed.

There is a strong possibility that (as in stock assessment) different local minima ("best fits") will be found, each fitting to a different component of the data, and leading to different population projections. A formal criterion for selecting models (e.g., AIC) will be applied as statistically reasonable. However, sometimes the AIC will be unreasonable because determining effective sample sizes is an issue in fish data with spatial and temporal correlation. Therefore, parameter sets with different local minima will be represented as an "ensemble" for further evaluation, including the use of Delphi methods with stakeholders and/or outside scientists such as the NPFMC Scientific and Statistical Committee (see stakeholder design principles).

> ***[Q20]*** *We need a work plan and specifications for testing the ability of a main sequence upper trophic level BTU model fit to pre-1977 data to correctly hindcast the ecosystem changes observed in the mid 1970s. (Or a work plan and specifications for an alternative test that is at least equally informative).*

A20 (Aydin): On both the spatial and temporal scales, only limited pre-1977 survey data exists for formal tuning when compared to the 1982-present. Most pre-1977 abundance estimates result from stock assessment models with assumptions that may contradict more complicated BSIERP model analyses. Our BSIERP modeling approach will begin the coupled climate-NPZ-FEAST runs with climate/oceanography year 1950, allowing for biological "spin-up" of the model between 1950-1960, start fitting to available data from the 1960s (e.g., limited walleye pollock fisheries catch data and diet data) and then evaluating the ability of the model to accurately capture the post 1982-period. We will conduct fitting exercises where we put strong weighting on the limited pre-1977 data in order to ensure the model parameters can make a "regime transition" and specifically calculate fits for the pre- and post- regime period. Due to computational cost, much of this work will be performed in the 2-dimensional versions of the FEAST model.

We will also pay attention to shifts in 1989 and 1998 by performing fitting that withholds data from each regime and scores output by the fit to the withheld regime. Tables of parameters fit to each regime will be examined to see if parameters differ significantly; if so, it will be clear that the regime control rules are occurring at a scale finer than what the parameters capture or through a missing process; this outcome itself will be an important advance in determining the correct scale for both modeling and future field work (e.g., are shifts in the fall transition, or other unmeasured seasonal aspects, more critical than the current BSIERP hypotheses). We plan to provide these results prior to the final field year in order to provide feedback to field researchers on the BSIERP hypotheses. This work may produce data hypotheses (observations) which could determine whether new climate modes were operating in the North Pacific.

> ***[Q21]*** *We need a work plan and specifications for testing the ability of the main sequence upper trophic level BTU model to correctly hindcast the ecosystem state in early-ice-melt years when fit to late-ice-melt years, and vice versa. Since the upper trophic level organisms are primarily long lived, the by-year-type fitting and status evaluation will have to be primarily in terms of per capita or per unit biomass rates or changes rather than standing stocks. (Or a work plan and specifications for an alternative test that is at least equally informative).*

A21 (Aydin): Fish are longer-lived than zooplankton, and moreover their life-histories are adapted to multi-year variation. We will use two approaches to simulate these long life-histories with respect to ice melt. Our first approach will use the bioenergetics output of the FEAST model to predict pollock condition factor as an index of growth for single-year runs of the NPZ component in early and late ice-melt years, as well as to predict resulting per-capita rates, as suggested. Our second approach will use pollock and forage species standing stock as an indicator and will require multi-year runs. Using the 2-dimensional transects and 2-dimensional spatial implementations of FEAST, we will randomly choose 20-year intervals of ice conditions, and with varying degrees of autocorrelation, examine how the relative frequency of ice and non-ice year affects predictions of long-term predation. We also will examine the effect of treating 2001-2005 (low ice years) as a separate "regime" in the regime-scale fitting exercise. The approach of testing multi-year ice runs with differing frequencies of ice years addresses the Oscillating Control Hypothesis (OCH) and its extended cousin, the Multiple Life-history Control Hypothesis (MLCH; Aydin and Mueter in press), which are driven by alternation between early- and late-ice years as well as predation over a multi-year interval. We will determine what frequencies of oscillations between "ice" and "non-ice" conditions lead to long-term rises in pollock production, and extend this predictive theory (without explicit modeling) to hindcasts and forecasts of other life history strategists such as Pacific Cod, arrowtooth flounder and marine mammals. Overall, this approach will serve as a test of climate frequency analyses applied to biological systems data. In the long run, the ultimate fate of the Bering Sea ecosystem may depend on changes in variability as much as changes in long-term mean, and this extended frequency analysis and fitting will give a clearer view of a potential ice-free Bering Sea.

MANAGEMENT OF THE MAIN SEQUENCE OF THE INTEGRATED MODEL

The daisy-chained submodels in the BTU main sequence of coupled models (climate-ocean-NPZ-upperlevels-fishery) are proposed to be managed by Bond-Hermann-Gibson-Aydin-Haynie respectively. No management structure above the level of the respective subprojects has been described.

> ***[Q22]*** *The team needs to identify one lead person who is responsible for overall coordination and integration of the program, and for delivering "the prediction," whatever it is, and reporting "the validation" on that prediction, however they do it. That person should take responsibility for assembling the responses to this review.*

A22 (Aydin): Kerim Aydin (AFSC) will be the lead person responsible for coordination, integration and delivering results to NPRB. He led preparation of the responses to this review.

Table 1. Revised Modeling Budget and narrative

Project		Person	Original	Proposed
M.47	Forage euphausiid abundance in space and time	Aydin	$ 506,716	$ 506,716
M.48	Integrate economic-ecological models of pollock and cod	Dalton, Aydin	$ 360,710	$ 360,710
M.49	Spatially explicit integrated model of pollock and cod	Haynie	$ 298,349	$ 457,924
M.55	Management strategy evaluation	Punt, Ianelli	$ 478,626	$ 478,626
M.54	Competing fur seal-seabird-pollock model	Mangel	$ 375,699	$ 375,699
M.61	Correlative biomass dynamics model	Mueter, Kruse	$ 320,325	$ 320,325
M.12	Spatially explicit model development		$ 315,558	$ -
M.50	Management strategy resilience		$ 398,812	$ -
	TOTAL		**$ 3,054,795**	**$ 2,500,000**
Identified gaps				
M.55	Producing blended and/or ensemble model predictions	Punt, Ianelli		$ 150,000
M.55	Risk assessment of birds and mammal populations	Punt, Aydin		$ 150,000

Revised Budget (brief narrative)

The funding for most of the components remains as submitted in the original BSIERP proposal. The changes are as follows:

For project M.49 (Spatially explicit integrated model of pollock and cod) we have increased the original budget by $159,575 to perform the cost estimate analysis (cost engineering approach) proposed in the revised proposal under this subproject.

We have removed project M.12 (Spatially explicit pollock assessment model) as the necessary pollock tagging fieldwork is not currently funded by NPRB under BSIERP.

We have removed project M.50 (Management strategy resilience) as the focus on cod and crab does not match with the currently funded fieldwork. This is a worthwhile project to pick up in a future context for these species.

Finally, we have noted **two modeling gaps** as also noted by the EMC's response to the original proposal, suggesting that the management strategy evaluation was thinly funded (see EMC Q17). Each of the two gaps ($150,000 each) is described in detail both in the response to question 17 and in the revised proposal.

We will provide full budgeting (i.e. NPRB spreadsheets and detailed narratives) for the additions and proposed additions upon request.

Table 2. Modeling timeline, including timeline for integrating vertical modeling sequence.

	TASKS	Jan-08	Jun-08	Jan-09	Jun-09	Jan-10	Jun-10	Jan-11	Jun-11	Jan-12
Vertically-integrated sequence										
M.3	GCM (climate model) output		Initial conditions finished for inclusion							
M.4	ROMS Regional Oceanographic Modeling		Good ROMS hindcast produced		Major integration of vertically integrated model	*All fieldwork to be included in vertically integrated models must be available. Results to guide 2010 field season ready by late spring 2010*	Ensemble predictions made available			
M.5	NPZ		NPZ ready for initial runs							
M.47	Forage euphausiid abundance in space and time		2-d major model development		2-d uncoupled model run, calibrated/fit, begin integration		2 or 3 d driven by ensembles started			
M.48	Integrate economic-ecological models of pollock and cod		Development of model equations and integration with fish						All modeling complete / publication, prediction, and recommendations review	Final report/ presentation at Marine Science Symposium (presentation to NPFMC council, December 2011)
Parallel (complementary or competing) approaches										
M.49	Spatially explicit integrated model of pollock and cod									
M.55	Management strategy evaluation									
M.54	Competing fur seal-seabird-pollock model									
M.61	Correlative biomass dynamics model									
Dependent on funding:										
M.55	Producing blended and/or ensemble model predictions									
M.55	Risk assessment of birds and mammal populations									

Table 3. Components of the integrated forecast system and general category of predictions made by each model. We will compare results of proposed and existing models. "General category" of predictions is by trophic level, detailed explanation of prediction products is found in the EMC questions text.

General Category of Predictions	**Proposed Models**	**Existing Models to be compared in MSE component (M.55)**
Climate-oceanography	Climate downscaling (NSF) (M.3); ROMS (NSF) (M.4)	
Lower trophic level production	ROMS-NPZ (NSF) (M.5)	NEMURO 20km N. Pacific NPZ
Pollock recruitment	Biomass dynamics from retrospective correlations (M.61)	
Forage distribution, production	FEAST (M.47)	Ecospace
Fish, bird and mammal production	FEAST (M.47); Behavioral foraging (M.54); Biomass dynamics (M.61);	Single species assessments w/ correlative recruitment indices MSM/MSVPA; Ecosim w/bioenergetics; Ecospace
Economic/social consequences	FEAST w/fleet dynamics (M.47); Spatially explicit integrated economic (M.48, M.49)	Ecospace; Ecosim w/bioenergetics

Table 4. Models for predicting Bering Sea fish production in the "competing models" component of vertical integration (Fig. 1).

Models developed in this project, in order of increasing biological complexity/process detail	
Biomass dynamics (M.61)	Autocorrelative model including lags for organism life-history.
Behavioral foraging (M.54)	Top-down behavioral and energetics model, driven by predator behavior.
Forage/Euphausiid Abundance in Space and Time (FEAST; M.47)	Bottom-up process model, driven by prey supply and bioenergetics.
Existing Bering Sea models for comparative study (M.55), in order of increasing biological complexity	
Current AFSC stock assessments with environmental correlates	Single-species population models with recruitment correlates.
Multispecies Statistical Model (MSM; NPRB project #525)	Multispecies assessment/population model with recruitment correlates; feeding responses primarily "top-down" (fixed ration).
Ecopath/Ecosim	"Monolithic" ecosystem population model; feeding responses primarily "bottom-up" (strong response to lower trophic levels).
Ecosim with bioenergetics (Aydin et al. 2006)	"Monolithic" ecosystem population model, but allowing flexible parameterization from other models (e.g., behavioral foraging above).
Nemuro plankton model coupled to Ecospace ecosystem model	"Monolithic" spatial ecosystem model driven by detailed planktonic modeling.