Quantifying Uncertainty in Oxygen Predictions of an Ocean Numerical Model: A Case Study for Application to Anthropogenic Nutrient Effect Assessment

Martha Sutula

Fayçal Kessouri

Christina Frieder

Nicholas Lombardo

SOUTHERN CALIFORNIA COASTAL WATER RESEARCH PROJECT

Technical Report 1445

Quantifying Uncertainty in Oxygen Predictions of an Ocean Numerical Model: A Case Study for Application to Anthropogenic Nutrient Effect Assessment

Martha Sutula, Fayçal Kessouri, Christina Frieder, and Nicholas Lombardo

Southern California Coastal Water Research Project, Costa Mesa, CA

February 2026

Technical Report 1445

EXECUTIVE SUMMARY

Introduction and Goal

An ocean numerical model (Regional Ocean Modeling System with Biogeochemical Elemental Cycling; ROMS-BEC), used to predict ocean acidification, hypoxia and risk of toxic HABs in the Southern California Bight (SCB), has predicted that land-based nutrients are having a non-trivial effect on these environmental problems. Scientists are working to quantify uncertainty in modeled predictions, which becomes increasingly important when the water quality management decisions to be made are costly and irreversible. To date, the focus of efforts to quantify uncertainty of ROMS-BEC have yielded insights on overall model adequacy to investigate eutrophication. However, specific procedures are needed to quantify the "signal" or magnitude of the modeled effect relative to the "noise" of model uncertainty, procedures that ideally should be incorporated with every scenario in which ROMS-BEC is used to answer a management question. Thus, a key question is how to immediately incorporate uncertainty assessments that can generate error bars for routine applications, given that model applications are in an early stage and investments to improve on uncertainty information are just now being made.

To assess our readiness to routinely quantify uncertainty in ROMS-BEC management scenarios, we conducted a case study to (1) demonstrate which method(s) could be used now, using existing simulations in a retrospective assessment, and (2) identify what refinements are needed to improve uncertainty quantification and the investments required to make those refinements. The case study was the focus of yearlong discussions among the OAH Model Subcommittee of SCCWRP's Commission Technical Advisory Committee (CTAG).

Approach

In this case study, we identified a scientific question that the model was used to answer, then contextualized the answer to that question with estimates of uncertainty (see summary table below). We focused on oxygen (O_2) because: 1) hypoxia is of management concern, and 2) O_2 observational data are the most abundant. We calculated the difference between two scenarios, one with (ANTH) and one without land-based inputs (CTRL) and interpreted the change through two O_2 metrics: 1) the California Ocean Plan water quality objective (WQO), and 2) aerobic habitat thickness.

Summary of the science question and metrics used in the uncertainty case study

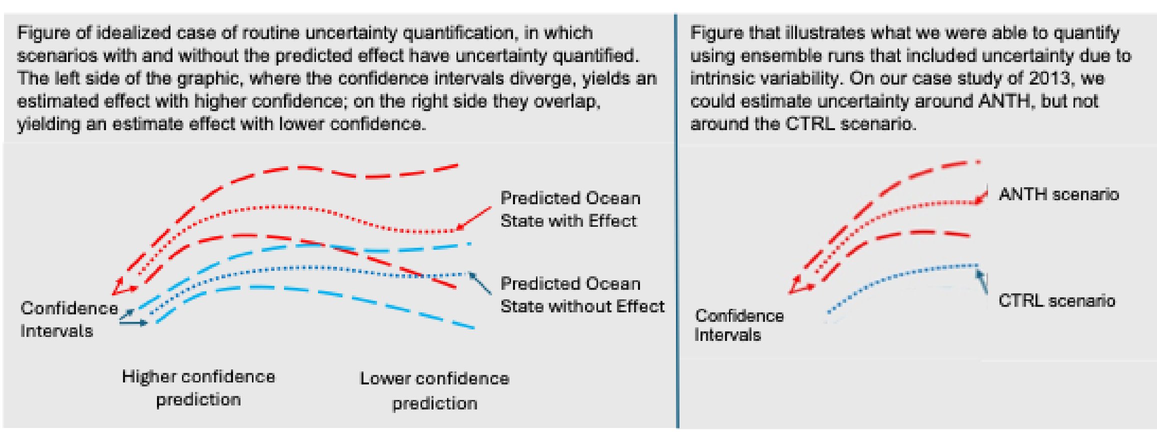
Decision category	Description
Question	"What is the predicted change in hypoxia due to land-based nutrients"?
Variable of interest	Oxygen
Endpoint of concern and threshold	Chemical endpoint: O_2 Water Quality Objective (WQO; -10% of natural background) Biological endpoint: Aerobic habitat thickness for Northern anchovy (total depth where $\Phi > \Phi_{CRIT}$)

Multiple sources of model uncertainty exist (intrinsic variability, forcing, parameter and structural/numerical). We investigated three that could encompass one or more of these sources of uncertainty: (1) data-model difference assessment, (2) ensemble runs of ROMS-BEC that incorporate one or more sources of uncertainty and (3) multiple model comparisons (i.e., ROMS-BEC versus other models). We conducted the change assessment of ANTH vs CTRL, then investigated which of three uncertainty quantification methods could be used to contextualize this estimate of change, with the understanding that this was the best estimate of uncertainty we could make at this time.

Findings

Our case study had three findings.

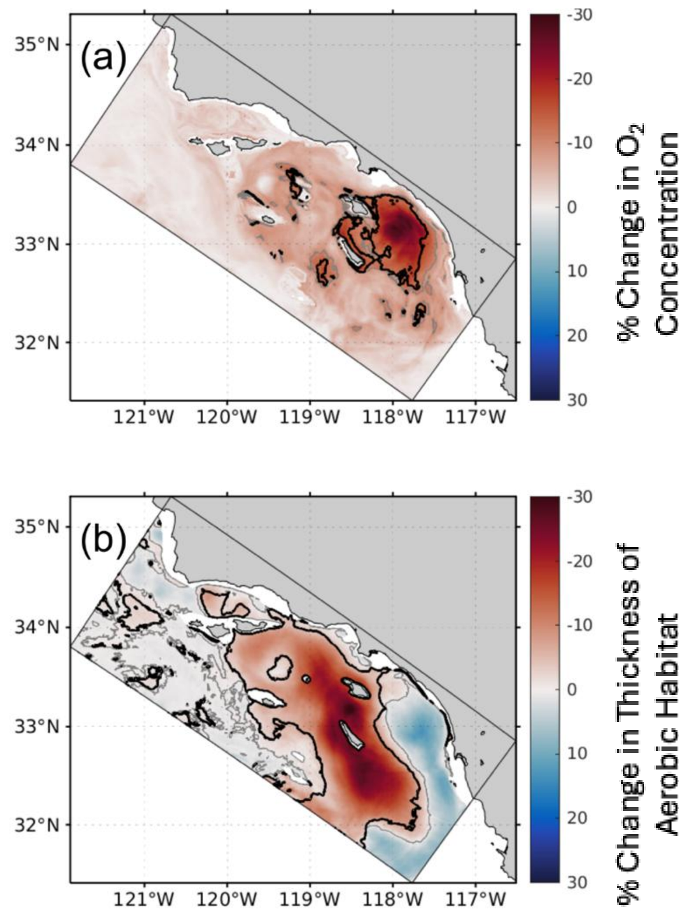
#1. Of three available methods to routinely quantify uncertainty, only ROMS-BEC ensemble runs that incorporate intrinsic variability (model noise due to stochasticity) can be used now, but only to estimate the uncertainty in ANTH during 2013. If interest exists in using intrinsic variability in the future, refinements to our current capacity are recommended, including expanded number of ensemble runs for both the management and comparator scenarios (i.e., ANTH or CTRL) and conducting a multiple year assessment to capture interannual variability.



#2. In 2013, exceedances of both chemical and biological criteria for ANTH minus CTRL were larger than intrinsic uncertainty estimates around the ANTH scenario, our best estimate of uncertainty for this case study at this time. The table below illustrates how the 2013 predicted change in O_2 WQO and aerobic habitat compression metrics due to land-based nutrients are contextualized with estimates of uncertainty in ANTH due to intrinsic variability. The change is expressed as: 1) the mean spatial extent where the metric threshold has been triggered (std. dev. among intrinsic variability runs) and b) mean vertical magnitude of change in which the threshold was triggered (std. dev. of combined error among intrinsic runs and across the spatial area).

Metric	O_2 WQO	Decrease in Aerobic Habitat Thickness
Spatial Area Exceeding Threshold: Mean area (± 1 SD) of effect due to land-based nutrients	11,718 (±399) km^2	44,456 (±7,862) km^2
Vertical Magnitude of Change: Mean vertical change (± 1 SD) where criterion is exceeded.	-15.1 (±5.2) %	-8.6 (±8.0) m

These 2013 estimates can be visualized in spatial maps (see inset graphic on left) for the maximum month of change for the O_2 WQO (Panel A) and the aerobic habitat assessment (Panel B). Heat maps show the magnitude of the vertical change. Contours show areas where the threshold was exceeded. In both maps, the grey contour is where the mean change exceeded the threshold. The black contour is where the mean change plus uncertainty exceeded the threshold, i.e., the spatially explicit area where confidence in the change is the greatest.



The magnitude of estimated change due to land-based inputs depends on season and year, as noted in Frieder et al. (2024), which first reported on predicted habitat compression in the SCB. O_2 and pH loss is most observed in the late summer and fall, shown in the inset panel as the maximum loss that occurred that year for the O_2 WQO (August) and the aerobic habitat compression (October). We only estimated uncertainty for 2013, but large interannual variability in compression occurs, depending on physics and its impact on biogeochemistry and primary productivity. Within the 20 years of ROMS-BEC predictions of ANTH, 2013 is a median productivity year.

#3. We identified the advantages, disadvantages and the scale of investments to utilize any of the three methods for management scenarios in the SCB (see table below). Investments could enhance our capacity to utilize ensemble ROMS-BEC model runs with uncertainty due to intrinsic variability, while use of data-model difference assessment and multiple model comparisons are not possible without investments.

Ultimately, which method or combination of methods to use in future model applications will depend on the specific questions to be answered, the endpoint of and the specific metric/criterion to be used, the types of scenarios, and modeling resources available to invest to improve uncertainty quantification, among other considerations.

We caveat this table summary and its assessment of resources needed requires much greater context and nuance than is provided, depending on the specific application intended. The CTAG OAH Model Subcommittee discussed in depth and agreed that providing this context is beyond the scope of this initial case study; the Subcommittee strongly recommended further discussion and elaboration of the magnitude of investments and associated costs, paired with discussion of the specific management scenarios, which is just beginning among the Management Scenarios Committee.

Summary of the routine uncertainty quantification approaches evaluated, including basic description, their advantages, whether the approach can be used now, and recommendations to build capacity to add these approaches to the ROMS-BEC uncertainty toolkit. Note that interpolation uncertainty is only applicable to data-model (DM) difference assessment; errors due to model resolution are combined with structural/numerical. For investments required, the number of “$” is intended to convey the order of magnitude of resource investment required, where 1$ is tens of thousands and each additional $ is an order of magnitude higher. Those investments are estimated for modeling services and for the in-kind services of different monitoring programs.

Type		**Data Model Difference Assessment**	**Ensemble Runs**	**Multiple Models**
Description		Comparison of observations with predictions of a realistic scenario	Supplemental runs of the same scenarios of ROMS-BEC, to which sources of uncertainty are added.	Simulations of other models compared to same scenarios of ROMS-BEC
Advantages	**Can be applied to any temporal or spatial scale**	No, limited to scale at which data can be justifiably interpolated	Yes	Yes
	Uncertainty types characterized	All but interpolation **error**	All but structural/ numerical	All but intrinsic and forcing uncertainty
	Applied to ANTH, CTRL, or other management scenario	No, Only ANTH	Yes	Yes
	Approach applied to eutrophication models with potential regulatory application?	No; skill assessment to evaluate model adequacy is frequently used, but not for this particular application	Yes	Yes
Can the approach be used right now?		No, observational uncertainty unquantified	Yes, but for intrinsic only	No, the two available CCS models have not reached “peer status”
Investments required include: **M = Modeling scientific services (e.g., SCCWRP or comparable)** **C = CalCOFI monitoring** **P = POTW monitoring** **SW = Stormwater monitoring**		Quantify observation uncertainty for CalCOFI data (5 years, C: $$$$, M: $$) Quantify observational data uncertainty for oxygen in CalCOFI and POTW data (5 years, C: $$$$; P: $$$$, M= $$) Improve POTW observational data quality to ensure suitability for data-model difference assessment (5 years, P: $$$$, M= $$)	Intrinsic variability: Develop a “bank” of ensemble runs (CTRL and ANTH) for a selected set of ocean base years that represent the “critical condition.” (1 year, M= $$) Forcing uncertainty: conduct sensitivity analyses; quantify uncertainty in most sensitive pathways; combine with intrinsic variability in ensemble scenario (5 years, M= $$, P=$$$, SW=$$$). Parameter uncertainty: Develop a data assimilation version of ROMS-BEC and identify parameter sets with an equivalent model skill as original model; combine parameter uncertainty of ROMS-BEC with forcing and/or intrinsic variability (5 years, M= $$$)	Onboard, set up, test, simulate, then validate model(s): (a) ROMS-MARBL: already set up for the SBC (3 years, M= $$) (b) ROMS-NEMURO: 4 years, M= $$$

Acknowledgements

This study was produced with funding from the Southern California Coastal Water Research Project, utilizing model scenarios that were funded through support from the Ocean Protection Council Grant C0831014. This work used the Expanse system at the San Diego Supercomputer Center through allocation TG-OCE170017 from the Advanced Cyber infrastructure Coordination Ecosystem: Services and Support (ACCESS) program, which is supported by National Science Foundation grants 2138259, 2138286, 2138307, 2137603, and 2138296. Additional computational support was provided by UCLA's Hoffman2 Shared Cluster, maintained by UCLA Office of Advanced Research Computing's Research Technology Group. The authors thank the SCCWRP Commission Technical Advisory Group (CTAG) Subcommittee on OAH Modeling for their thoughtful review and feedback on this uncertainty case study.

Citation

This report should be cited as: Sutula, M., F. Kessouri, C. Frieder, and N. Lombardo. 2026. Quantifying Uncertainty in Oxygen Predictions of an Ocean Numerical Model: A Case Study for Application to Anthropogenic Nutrient Effect Assessment. Technical Report 1445. Southern California Coastal Water Research Project. Costa Mesa, CA.

TABLE OF CONTENTS

List of Figures

List of Tables

LIST OF ACRONYMS AND ABBREVIATIONS

ANTH	Anthropogenic simulation
BMDL	Biogeochemical Model Development Lead
BTAL	Biological Tool Applications Lead
CalCOFI	California Cooperative Fisheries Investigation
CCS	California Current System
CM	Contract Manager
CTAG	Commission Technical Advisory Group
CTD	Conductivity, Temperature, and Depth
CTRL	Control simulation
LACSD	Los Angeles County Sanitation District
MAL	Modeling Applications Lead
MGD	Million Gallons per Day
MPA	Marine Protected Area
NWRI	National Water Research Institute
OAH	Ocean Acidification and Hypoxia
OC San	Orange County Sanitation District
OPC	Ocean Protection Council
PM	Project Manager
PNAS	Proceedings of the National Academy of Sciences
POTW	Publicly Owned Treatment Plants
QA	Quality Assurance
QAPrP	Quality Assurance Program Plan
QAPjP	Quality Assurance Project Plan
ROMS-BEC	Regional Ocean Modeling System with Biogeochemical Elemental Cycling
RSD	Ratio of Standard Divisions
SCB	Southern California Bight
SCCOOS	Southern California Coastal Ocean Observing System
SCCWRP	Southern California Coastal Water Research Project
SMBO	Santa Monica Bay Observatory
SOP	Standard Operating Procedure
SWBQAO	State Water Board Quality Assurance Officer
UCLA	University of California at Los Angeles, Dept. of Earth and Atmospheric Sciences
WQO	Water Quality Objectives
WRF	Weather Research and Forecast

INTRODUCTION

In the Southern California Bight (SCB), an ocean numerical model used to quantify effects of land-based nutrients on eutrophication is predicting that human sources of nutrients are having non-trivial effects on ocean acidification (OA), hypoxia and risk of toxic HABs (Kessouri et al. 2021a,b, 2024, Frieder et al. 2024). Use of this ocean numerical model, the Regional Ocean Modeling System with Biogeochemical Elemental Cycling (ROMS-BEC, Kessouri et al. 2021b) to support decisions on nutrient management could potentially result in billions of dollars of wastewater infrastructure upgrades. Members of the community have questioned whether the model adequacy for this application is commensurate with the potential economic consequences of decisions the model could support. An inherent part of all modeling exercises is understanding the degree of uncertainty in modeled predictions and in the observational data used to develop the models (see inset panel). For scientists, the public, managers, and policy makers to confidently use the model results, the uncertainty should be quantified. Multiple approaches exist to quantify model skill and assess sensitivity to model formulation and forcing decisions, producing a range of quantitative, semi-quantitative and qualitative information (Table 1). This basic information can yield key insights on model adequacy and inform managerial confidence in the model for its intended application, particularly when accompanied by the review of the model by an independent expert panel (NWRI 2025). However, these technical activities do not guarantee that managers can view the results of a given set of scenarios with confidence intervals representing model uncertainty (Fig. 2).

Figure 1. Sources of ocean numerical model uncertainty

Internal (or Intrinsic) Variability: Inherent, stochastic variations (e.g., mesoscale eddies) that are a source of uncertainty in predictions.

Input Data/Forcing Uncertainty: Inaccurate or limited resolution data used to drive model at its boundaries (e.g., oceanic, land inputs).

Parameter Uncertainty: Lack of knowledge about the true values of model parameters, which are often difficult to measure directly.

Model Structural and Numerical Uncertainty: Imperfect representation of processes in model equations from oversimplifications, numerical approximations, or lack of scientific understanding.

Additional tools and procedures are needed to quantify the “signal” or magnitude of modeled change in a given variable of interest relative to the “noise” of model uncertainty, particularly when investigating potential management options. Without the means to compare the magnitude of change to model noise, we risk overinterpreting the environmental significance of the predicted change (e.g., how many moles of oxygen change are actually meaningful). In addition, uncertainty quantification is application specific, i.e.,, dependent on the question being answered, and the temporal and spatial scales at which metrics and thresholds are being applied. Mature environmental management programs that have invested years and millions of dollars in modeling science and observational studies have developed a suite of tools to characterize model uncertainty for routine model applications (e.g., management scenarios). These tools typically move beyond model skill assessment (a.k.a. model validation) for the purpose of evaluating model adequacy. Common approaches incorporate the use of either (1) ensemble runs of the same model, run alongside the specific scenario, to which one or more sources of uncertainty are added (e.g., intrinsic, forcing, parameter uncertainty, etc.), (2) multiple model comparisons, or (3) combinations of these approaches. We investigated whether we could use an adaptation of the model skill assessment approach specifically to create confidence limits around the use of the scenarios to answer the questions, an approach we coined "*data-model difference assessment"* to be distinguished from the model evaluation or skill assessment conducted by Kessouri et al. (2021).

Table 1. Conceptual view of different model uncertainty quantification approaches and their intended uses. Both "intended use" categories contribute to managerial confidence in using the model for decision support.

<table>
<tr><th>Intended Use</th><th>Model Uncertainty Quantification</th></tr>
<tr><td>Evaluate model adequacy</td><td>Model skill assessment or evaluation

Sensitivity analyses
• Model forcing
• Parameter forcing
• Numerical estimation

Peer review (journal and/or comprehensive expert review)</td></tr>
<tr><td>To put error bars on managerially- relevant science questions</td><td>Data-model difference assessment

Ensemble runs that include error from one or more of the categories below, to compare with main simulation
• Internal or intrinsic variability
• Forcing uncertainty
• Parameter uncertainty

Multiple model comparisons</td></tr>
</table>

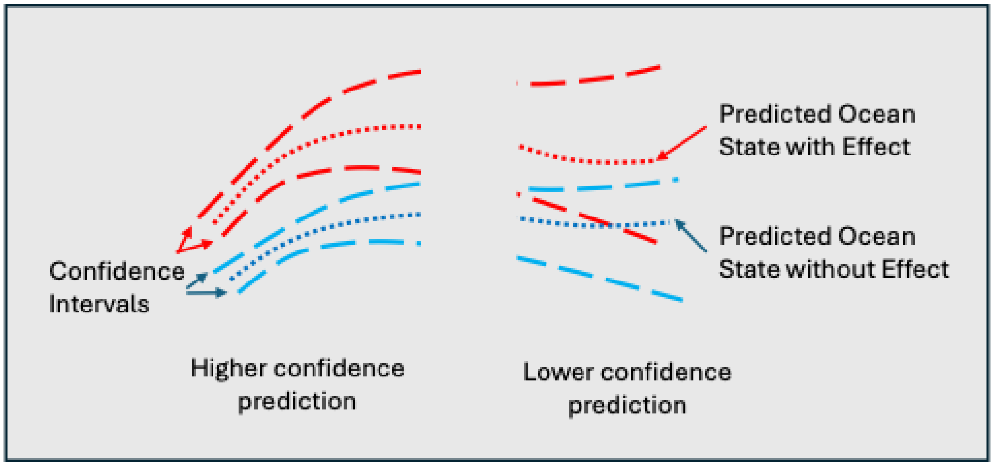


Figure 2. Idealized output of uncertainty quantification methods intended to show a confidence interval in the outcome of interest. The graph shows how the difference in predicted ocean state with versus without a given effect can yield a high confidence prediction when mean estimates and confidence intervals diverge (left side), versus a lower confidence prediction when the confidence intervals overlap (right side).

To date, the focus of ROMS-BEC uncertainty assessments has been on model validation and limited parameter sensitivity analyses. A key question is how to immediately incorporate uncertainty assessments that can generate error bars into routine management scenarios (hereto referred to as routine uncertainty assessments), given that model applications are in an early stage and investments to expand on uncertainty information are just now being made. The goal of this case study was to (1) investigate which methods can be used now to routinely quantify model uncertainty and demonstrate the method(s) in a case study and (2) identify what methods can be refined for uncertainty quantification in the future and discuss the investments to ready those methods. The case study was the focal point of discussions among the OAH Model Subcommittee of the SCCWRP Commission Technical Advisory Committee (CTAG).

Case Study Methods

Because a quality assurance project plan (QAPP) is a preferred way to describe how uncertainty will be captured and documented, we structured the methods as if they were presented in a model application QAPP. We identified a set of scientific questions, then contextualized the answer with estimates of uncertainty. A QAPP identifies an acceptable level of uncertainty for answering a question and then specifies the methods for uncertainty quantification, including which scenarios, endpoints and specific methodologies are used for their interpretation.

We focused on uncertainty quantification for dissolved oxygen (O_2) because: 1) hypoxia is of management concern, and 2) O_2 observational data are the most abundant and are more accurate and precise than acidification parameters. As such, uncertainty quantification is the most straightforward for this parameter.

Environmental and Management Context for Case Study

Rising atmospheric carbon dioxide (CO_2) concentrations are driving OA, global warming trends and oxygen (O_2) loss (a.k.a. hypoxia) in the ocean, with severe consequences for marine ecosystems. In coastal regions, these global drivers can combine with the effects of eutrophication to exacerbate hypoxia, particularly in shallow and enclosed bodies of water (Laurent et al. 2018, Brush et al. 2020). In eastern boundary current ecosystems, wind-driven upwelling drives high biological productivity, routinely exposing shelf waters to low-O_2 conditions (Chan et al. 2008, Fennel et al. 2019). The importance of upwelling, along with vigorous coastal circulation, would suggest a minor role for coastal eutrophication in exacerbating these stressors. However, in the SCB, in the southern range of the upwelling-California Current System, found that anthropogenically enhanced nutrient loads from a coastal population of 23 million people are predicted to amplify primary productivity and subsurface respiration within a 15 km band along the coast, exacerbating O_2 loss at a rate approaching that of climate change (Kessouri et al. 2021a,b).

Against this backdrop of rapid change, coastal managers are being challenged to assess the condition of marine resources and formulate place-based responses to climate change and local eutrophication stress caused by nutrient pollution, particularly with respect to sensitive coastal habitats and marine protected areas. Anthropogenic pollution controls are one of the few potential management actions that could potentially reduce effects of hypoxia, but the efficacy of these actions is unclear, while the cost of such actions is in the billions of dollars. Managers need increased certainty about how much anthropogenic nutrient inputs are resulting in changes in seawater chemistry and their biological effects, and how much of an alteration of nutrient loads could yield benefits (Sutula et al. 2021a).

The SCB is a marine open embayment, 94,000 km^2 in size, that spans from Point Conception (34.45◦N) to Baja California, Mexico (32.53◦N; Fig. 3). The U.S coastal waters of the SCB receives discharges from 19 U.S. permitted municipal wastewater facilities (point sources), non-point sources, and 75 rivers. Anthropogenic nutrient sources represent 98% of coastal nitrogen exports (Sutula et al. 2021a). The dominant contribution is from point sources discharged to ocean outfalls, representing 92% of total nitrogen discharged from land-based sources. The historical record captures the changes in effluent volume and constituent concentrations with progressive wastewater treatment upgrades (Sutula et al. 2021a). Among facilities permitted by the U.S. National Pollutant Discharge Elimination System (NPDES), publicly owned treatment works (POTWs) comprised the majority of discharges that occur via outfalls to the SCB. The four largest facilities each discharge over 100 MGD, and account for 86% of the total POTW effluent volume. These facilities are the Hyperion Treatment Plant (HTP) operated by the City of Los Angeles in the Santa Monica region, the A.K. Warren Water Resource Facility (Warren Facility, formerly Joint Water Pollution Control Plant) operated by the Los Angeles County Sanitation Districts in the San Pedro region, Orange County Sanitation District (OC San) Reclamation Plants in Orange County, and the City of San Diego's Point Loma Wastewater Treatment Plant (PLWTP) in South San Diego. Mexican transboundary inputs of primary treated wastewater contribute to riverine input via the Tijuana River watershed and to the International Wastewater Treatment Plant, both of which discharges to U.S. coastal waters. We note that inputs synthesized in Sutula et al. (2021a) do not reflect the recurrent sewage spills and loadings to the Tijuana River coastal confluence due to the massive wastewater infrastructure breakdown that occurred since 2017, nor the Mexico municipal wastewater discharges that enter coastal waters.

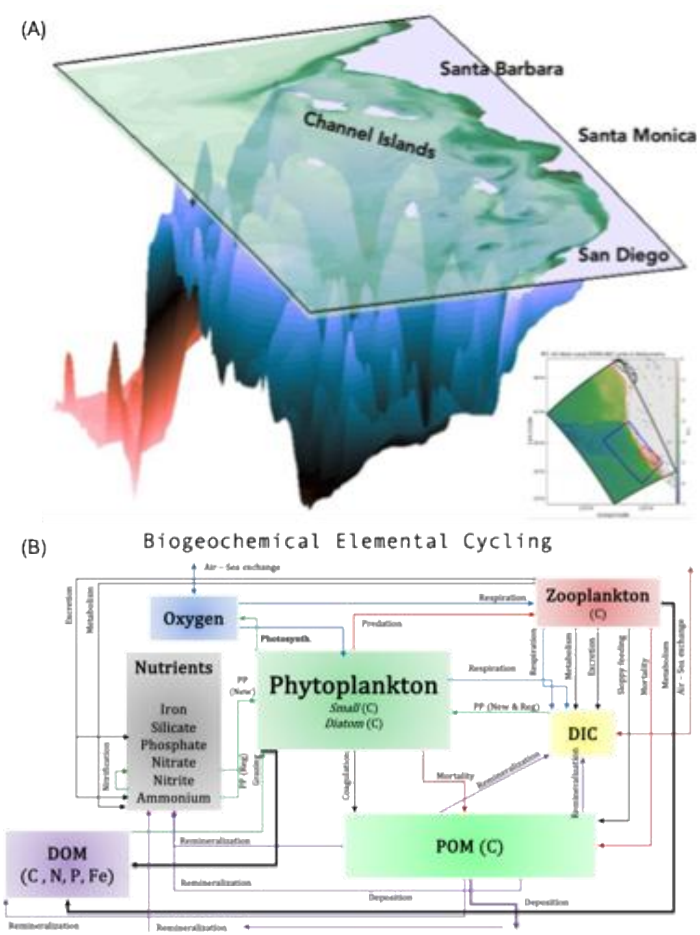


Figure 3. Panel A: Southern California Bight ROMS-BEC 300-m model domain, showing its location (inset figure, red rectangle) relative to the U.S. West coast 1-km model domain (blue boundary) and the 4-km resolution West coast wide domain. The land-ward edge of the southern boundary of the model terminates just north of Ensenada, Mexico and includes the region of Playas Rosario and the city of Tijuana, Mexico. The northern boundary is just north of Point Conception in Central Coast, California U.S.A. Panel B: Schematic of the Biogeochemical Elemental Cycling Model (Deutsch et al. 2021), showing O_2 as a state variable and the major processes that influence it.

Applied Science Questions

Two applied science questions are relevant to this case study:

- What is the effect of land-based nutrients on seawater O_2 and its management relevant endpoints?
- What is the quantifiable uncertainty in model predictions, specific to the spatial and temporal scales of the endpoints?

Assessment Approach

Our study quantified the magnitude of the effect of land-based nutrients on ocean water quality using published change assessment methods (Kessouri et al. 2021, 2024, Frieder et al. 2024), then compared that to the magnitude of uncertainty for two oxygen metrics: 1) a chemical endpoint, represented by the California Ocean Plan (COP) water quality objective (WQO) for O_2 (SWRCB 2019) and 2) a biological endpoint, namely aerobic habitat thickness. Aerobic habitat thickness is calculated from the metabolic index with emphasis on Northern anchovy as the sentinel species (Deutsch et al. 2015, Howard et al. 2020, Frieder et al. 2024). The metabolic index is the ratio of environmental oxygen supply to an aquatic organism's resting metabolic oxygen demand, used to predict marine species' habitat viability and geographic distributions.

We conduct a “change assessment” by differencing a control (CTRL) simulation of “ocean only” from one of the ocean plus the full suite of land-based inputs to U.S. coastal waters (ANTH). The magnitude in the change in O_2 metrics from land-based inputs is then contextualized with an estimate of uncertainty.

The scope of the case study was limited to existing simulations, observational data and corresponding documentation, with the intent to highlight what uncertainty methods are available now, versus what investments are needed to onboard new approaches. Table 2 summarizes the basic requirements for three routine uncertainty assessment methods considered in this case study. **Based on these requirements, only ensemble runs of intrinsic variability could be conducted now, for reasons further elaborated in the discussion.**

Table 2. Requirements for implementation of routine uncertainty assessment methods for this case study

Approach	Requirement(s)	Status Vis-à-vis Requirements
Data-model difference assessment	Observational uncertainty has been quantified. Data must be available to sufficiently characterize temporal and spatial scales of application.	O_2 observational uncertainty, including measurement, sampling and interpolation error, has not been comprehensively quantified for local CalCOFI or POTW ambient monitoring datasets.
Ensemble runs	Uncertainty from intrinsic variability, forcing, and/or parameter uncertainty must be quantified to incorporate into ensemble runs.	Intrinsic variability has been quantified via two ensemble ANTH runs for 2013 (Sutula et al. 2025) Forcing and parameter uncertainty has not yet been quantified.
Multiple model ensemble comparisons	Multiple models that are comparable to ROMS-BEC exist and have been validated for eutrophication applications within the SBC.	ROMS-MARBL and ROMS-NEMURO already exist and are comparable. However, they have not been set up and validated for eutrophication applications within the SCB.

Scenarios Employed to Quantify Effect and Uncertainty from Intrinsic Variability

We compared an existing set of ensemble ROMS-BEC model uncertainty simulations that represent internal or intrinsic variability. Intrinsic variability is defined as the variance in outcomes from multiple model simulations with the same or similar boundary forcing, due to the modeling of stochastic circulation processes. Stochastic (i.e.,, random) processes occur in ocean circulation. To induce this stochasticity, Kessouri et al. (2024) perturbated the initial conditions of oceanic boundary forcing for the ANTH scenario and repeated this exercise twice. The variance between these three simulations is used to estimate model noise from intrinsic variability from initial perturbation, a similar approach to that of Sérazin et al. (2015). Intrinsic variability is primarily developed at the mesoscale; ocean models capture this phenomenon even in large scale simulations (Waldman et al., 2018). Committee members inquired whether this intrinsic variability converges; the answer is no, despite the similarity of boundary conditions across all simulations. As an independent demonstration of this technique, Penduff et al. (2018) conducted a comparison of intrinsic variability in a global model simulation by perturbing only the initial conditions while maintaining the same boundary (atmospheric) conditions. After 50 years, the runs never converged.

The simulated ocean year of 2013 is the focus of the case study, as this is the year for which existing simulations can be used to calculate uncertainty due to intrinsic variability (Kessouri et al. 2024). This year represents a median year for ocean productivity among 1997-2017, the time period for which simulations have been run and validated (Kessouri et al. in review).

ROMS-BEC simulations used here comprise both the "main" and "ensemble uncertainty" runs. The main runs, the ANTH and the CTRL simulations, are conducted over a five-year period (08/1/2012 – 11/30/2017). The CTRL simulation does not include any terrestrial inputs and thus represents only natural ocean cycles of nutrient, carbon, and O_2, with the effects of global CO_2 superimposed. ANTH includes both natural oceanic cycles of nutrients that represent these same base conditions as the CTRL scenario, to which inputs from a spatially explicit data set of terrestrial sources of freshwater discharge, nitrate, ammonium, silicate, iron, inorganic carbon, alkalinity, and dissolved organic matter (N, C, and P) were added (Fig. 2). The model is "spun up" from initial conditions represented in the 1 km resolution model (without land-based inputs) that was initiated in 1996; thus, the initial conditions contain the physical suite of meso- and submesoscale eddies along with their biogeochemical variability. We assume the model equilibrates land-based inputs after one month, but our methods prescribe a 3 month spin up to avoid any artifacts. For this reason, we chose to only use simulations starting 1/1/2013 for the change assessment.

To assess intrinsic variability, two additional existing ANTH model runs were utilized, corresponding to the same 2013 period. The scenarios were set up identically to ANTH, but with perturbed initial conditions (ANTH2 and ANTH3) for which we add a random uniform variability to all state variables, including all the physical and biogeochemical state variables in the initial condition. The perturbation was created by averaging random 3 days between August 20 and September 10, 2012, from the original ANTH scenario. The additional submesoscale features develop in the 300 m model (ANTH) within days, with stochasticity driving differences in locale, and the model was stable after a few days.

Change Assessment and Use of Biological Thresholds to Interpret Effects

The California Ocean Plan (SWRCB 2019) has numeric objectives O_2 and a narrative biological objective but does not have prescribed policy guidance on how to interpret the objectives using model-based assessments. We used the basic framing of the objectives, but developed a scientific assessment approach to interpret them; therefore this should be considered a scientific assessment, not an impairment assessment and/or policy decision. The reader is referred to the primary literature for additional background.

Table 3. Summary of endpoints, metrics and thresholds employed to interpret model outputs for the project.

Endpoint	Metric	Threshold
Aerobic habitat	Change in thickness of aerobic pelagic habitat, defined by Metabolic Index (Deutsch et al. 2015)	Aerobic habitat thickness calculated as the thickness of the water column where $\Phi/\Phi_{CRIT} > 1$ (m) (Howard et al. 2020, Frieder et al. 2024). Compression is defined as any negative change in aerobic habitat thickness for Northern anchovy
O_2 concentration	Percent change in O_2 concentration (mmol m^{-3})	"The dissolved O_2 concentration shall not at any time be depressed more than 10 percent from that which occurs naturally, as the result of the discharge of O_2 demanding waste* materials." (California Ocean Plan, State Water Board 2019)

Protocol to Quantify Change and Uncertainty

For each endpoint, we assess its **vertical magnitude of change** between ANTH and CTRL as

$$\Delta H_{x,i,t} = \boldsymbol{H_{ANTHx,i,t}} - \boldsymbol{H_{CTRL,i,t}}$$

where *H* is the endpoint metric for grid cell *i* at month *t*, and $ANTH_x$ refers to the one of three ANTH scenarios. *H* is calculated for each endpoint metric according to Table 3. For aerobic habitat, ΔH is the change in thickness of the water column where $\Phi/\Phi_{CRIT} > 1$ (units in meters). For oxygen concentration, ΔH is the maximum difference in oxygen concentration (units in percent change) among density surfaces. We calculate a domain-wide severity metric for ΔH to identify the month of maximum change (*en sensu* Frieder et al. 2024), as

$$\boldsymbol{Severity_{x,t}} = \sum_{\boldsymbol{i=1}}^{\mathbf{max}\,(\boldsymbol{i})} \boldsymbol{\Delta H_{x,i,t}} \times \boldsymbol{Area_i}$$

Once the month with the maximum severity is identified for each endpoint, we evaluate the **spatial area (A) over which the endpoint exceeds its threshold** (Table 3, i.e., any aerobic habitat compression or a negative change greater than 10%in the O2 WQO)**.**

$$\boldsymbol{A_x} = \sum_{\boldsymbol{\Delta H_{x,i} > Threshold}} \boldsymbol{Area_i}$$

The mean and uncertainty in the spatial extent are reported as the mean and SD of A_x (n = 3). Within A_x we report the vertical magnitude of change as the mean of ΔH_x.

We then identify where there is agreement in the geographic area of effect among the change assessments. We first calculate the spatially explicit mean and standard deviation in the change among runs as

$$\overline{\Delta \boldsymbol{H}_{\iota}} = \boldsymbol{Mean}(\Delta \boldsymbol{H}_{x,i}); \ \boldsymbol{\sigma}_{\Delta \boldsymbol{H}\boldsymbol{i}} = \boldsymbol{SD}(\Delta \boldsymbol{H}_{x,i})$$

Those grid cells where $\overline{\Delta H_{\iota}}$ plus $\sigma_{\Delta Hi}$ exceeds the endpoint threshold are contoured as the geographic area of agreement among runs. We report the mean and standard deviation of $\overline{\Delta H}$ within this area. The standard deviation of $\overline{\Delta H}$ within the area of agreement is calculated as

$$\boldsymbol{\sigma}_{\Delta \boldsymbol{H}} = \sqrt{\boldsymbol{Var}(\overline{\boldsymbol{\Delta H}}) + \boldsymbol{Mean}(\boldsymbol{Var}(\boldsymbol{\Delta H}))}$$

For this assessment, the temporal window is one year; in the case of multiple years, the assessment would be repeated by year. For simplicity, we choose to express the magnitude of these metrics as the maximum for the year (Frieder et al. 2024). The temporal unit is monthly, and the spatial unit is a grid cell which corresponds with the horizontal resolution of the model, ~300 m x 300 m = 0.09 km^2. We only assess areas where the seafloor is greater than 100 m.

RESULTS

The time series of modeled ANTH versus CTRL (Fig. 4) highlights 1) the time period in which ANTH deviated from CTRL (July to November) and 2) the juxtaposition of the intrinsic variability of the mean ANTH (dark red) versus individual ANTH runs (shown in light red). August 2013 was the month with the maximum absolute severity in change for the O2 WQO metric, while October 2013 was the most severe month for the aerobic habitat metric (Figure 4).

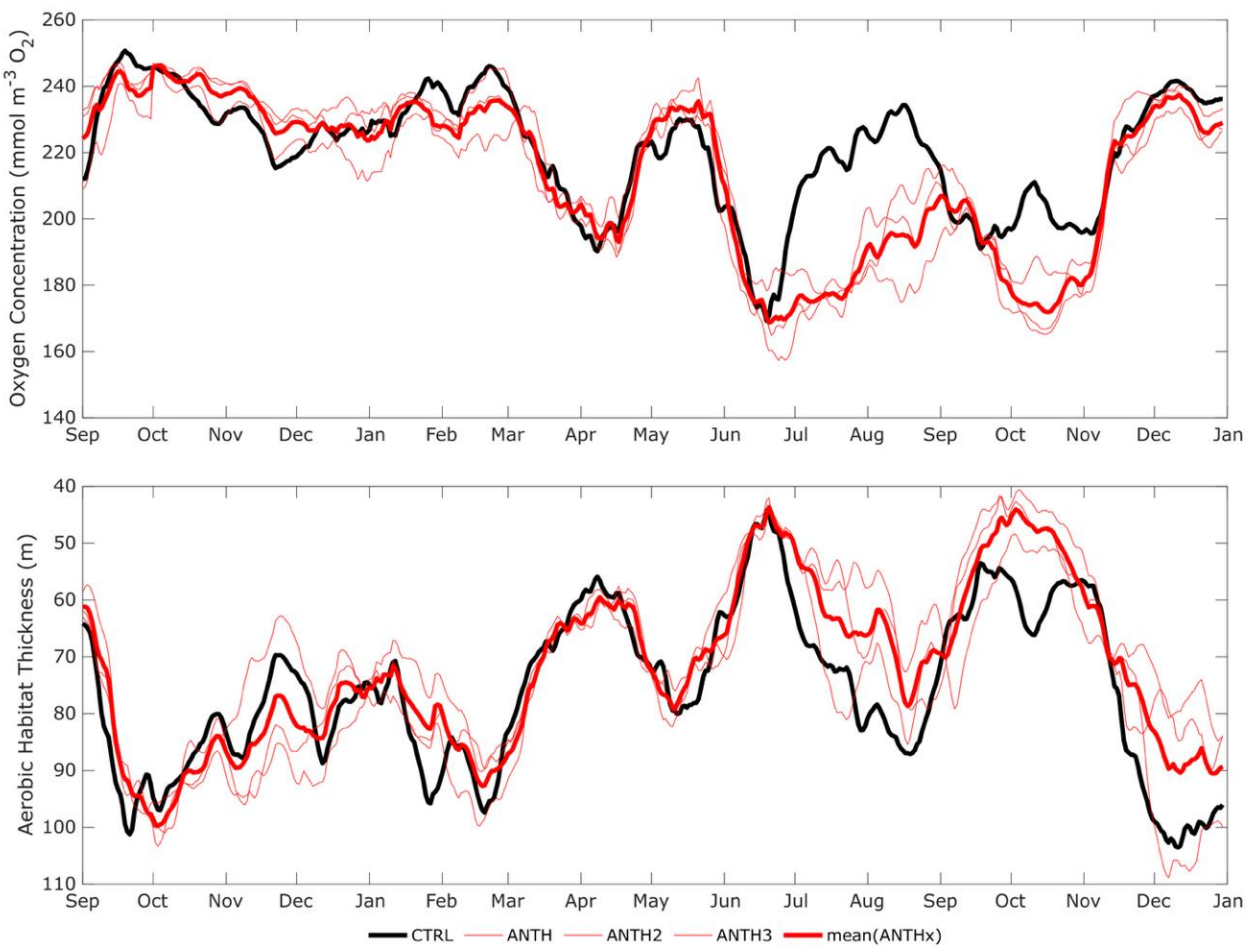


Figure 4. Time series showing the mean (dark red line) of ANTH, ANTH2 and ANTH3 and the intrinsic variability (change between thin red lines) versus the CTRL scenario (black line) for O_2 concentration at isopycnal surfaces (top panel) and aerobic habitat change over 200 m (bottom panel). The figure illustrates that the signal emerges from noise of intrinsic variability in the summer and fall of 2013, where the black and red lines decouple.

Spatial Area Exceeding Threshold

The spatial extent over which the percent change in O_2 exceeded its threshold ranged from 11,282 to 12,065 km^2 among the change assessments (Figure 5a; Table 4). The mean and uncertainty of the change assessments were 11,718 ± 399 km^2 (1 SD; Table 4). The spatial extent over which aerobic habitat thickness exceeded its threshold ranged from 35,790 to 51,133 km^2 among the change assessments (Table 4; Figure 5b). The mean and uncertainty of the change assessments were 44,456 ± 7,862 km^2 (Table 4).

The geographic area where there is agreement among the change assessments is shown in the black contour in Figure 6. This is the area where the mean change in the endpoint not only exceeds the threshold but also exceeds the uncertainty due to intrinsic variability. For the O_2 WQO metric, this area is 5,918 km^2 while for aerobic habitat loss, the area is 28,552 km^2.

Table 4. Mean and intrinsic variability uncertainty in the spatial extent of the affected area for the O_2 WQO and aerobic habitat thickness. The affected areas reported for each change scenario correspond with the contours shown in Fig. 5.

Change Scenario	Average Affected Area (km^2)	
	O_2 WQO	Aerobic Habitat
$ANTH_1$-CTRL	12,065	51,133
$ANTH_2$-CTRL	11,805	35,790
$ANTH_3$-CTRL	11,282	46,445
Mean (St. Dev)	**11,718 (399)**	**44,456 (7,862)**

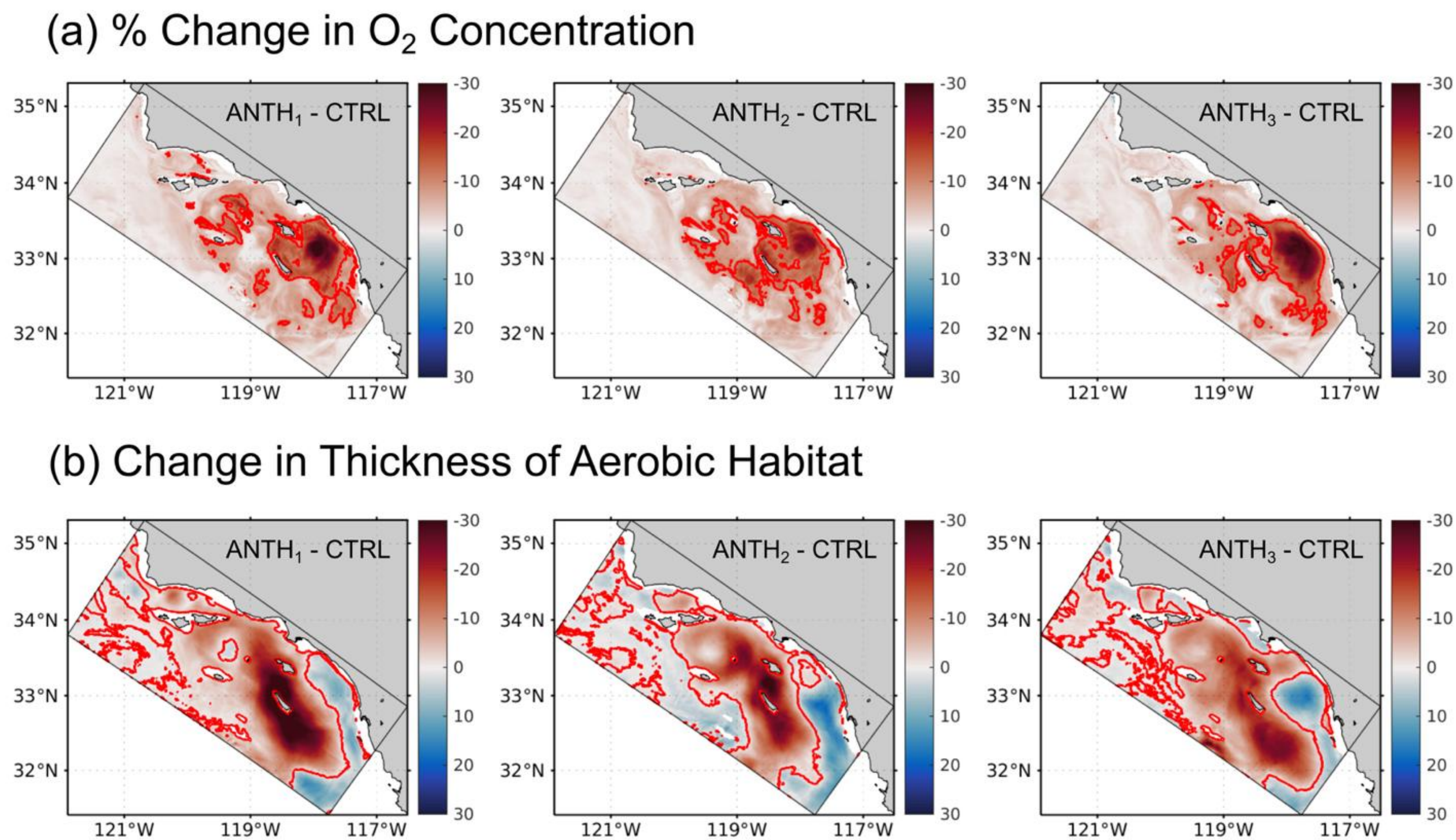


Figure 5. Spatial extent (red contour) over which the (a) percent change in oxygen concentration WQO metrics > -10% and (b) aerobic habitat compression occurs.

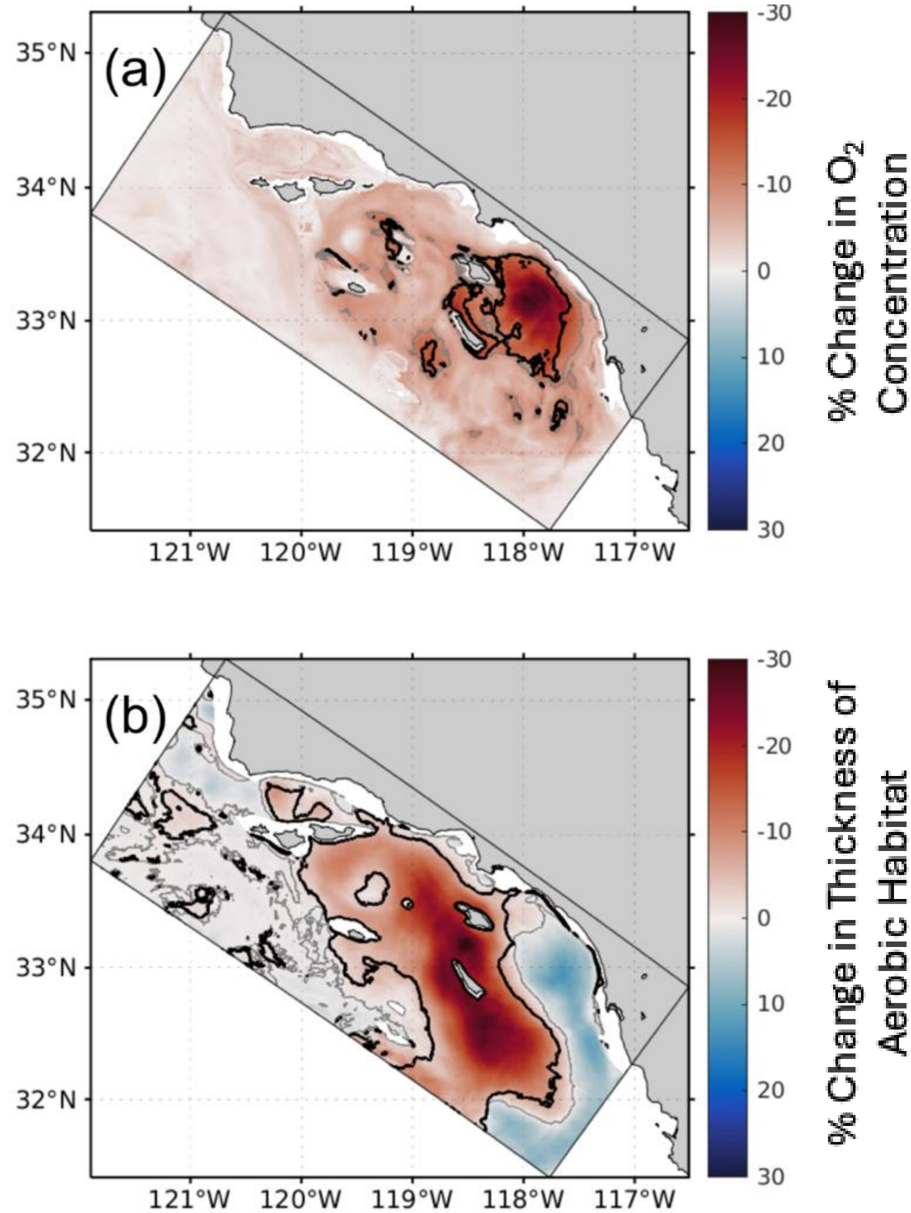


Figure 6.1 Spatial representation of the O_2 WQO change assessment (Panel A, August 2013) and the aerobic habitat assessment (Panel B, October 2013). Heat map shows the magnitude of the vertical change. Contours show areas where the threshold was exceeded. In both maps, the grey contour is where the mean change [mean(ANTHx-CTRL)] exceeded the threshold, The black contour is where the mean change plus uncertainty (mean(ANTHx-CTRL) + SD(ANTHx-CTRL)) exceeded the threshold, i.e., the spatially explicit area where confidence in the change is the greatest.

Vertical Magnitude of Change

Among the change assessments, the average magnitude of change for O_2 within the area that exceeded its threshold ranged from -13% to -16% (Table 5; average change within red contours in Figure 5); the mean and uncertainty were -15 ± 5% (1 SD; Table 5). The range in the average magnitude of change in aerobic habitat thickness was less than a meter among the change assessments (Table 5; average change within grey contours in Figure 5); the mean and std. dev. were -8.6 ± 8.0 m (Table 5).

Table 5. Mean and uncertainty in the vertical magnitude of change for areas where the endpoint exceeded its threshold. Note that error due to intrinsic variability is substantially less than the variance in the metric within the polygon where the metric was triggered.

Change Scenario	**Avg. Magnitude of Change**	
	O_2 (%)	**Aerobic Habitat (m)**
$ANTH_1$-CTRL (std. deviation across contour)	-15.1 (5.1)	-8.8 (9.0)
$ANTH_2$-CTRL (std. deviation across contour)	-13.7 (3.6)	-8.8 (8.0)
$ANTH_3$-CTRL (std. deviation across contour)	-16.4 (6.0)	-8.3 (6.8)
Mean (std. dev. in the mean change due to intrinsic variability)	-15.1 (1.4)	-8.6 (0.3)
Total St. Dev., including intrinsic variability and variability across the contour	**5.2**	**8.0**

Within the geographic area of agreement among change assessments, the mean change in O_2 was -17.3 ± 5.8 % and the mean change in aerobic habitat thickness was -11.6 ± 8.3 m (± 1 SD; both calculated within black contours shown in Figure 6).

Summary of Change Assessments and Uncertainty

For both endpoints, the quantified exceedance of the criteria was larger than the uncertainty estimates (Table 6). Thus, for the purposes of this case study, the uncertainty estimate contextualizes the magnitude of the change in oxygen metrics attributed to land-based inputs. In this particular case study, the choice in O_2 metric can make the assessment more or less stringent, as illustrated in this exercise.

Table 6. Summary of 2013 predicted change in O_2 WQO and aerobic habitat compression metrics due to land-based nutrients; estimates of uncertainty due to intrinsic variability are included in parentheses. The change in O_2 metrics is expressed as: 1) the mean spatial extent and std. dev. among intrinsic variability runs in which the metric threshold has been triggered and b) mean vertical magnitude of change (combined error among intrinsic runs and across the spatial area in which the threshold was triggered). The spatially explicit area in which all reruns agree is the black contoured polygon in Fig. 6.

Metric	O_2 WQO	Aerobic Habitat Thickness
Spatial Area Exceeding Threshold		
(a) Mean area (± 1 SD) among ANTHx – CTRL	11,718 (399) km^2	44,456 (7,862) km^2
(b) Spatially explicit area of agreement among runs (emerges from uncertainty of intrinsic variability)	5,918 km^2	28,552 km^2
Vertical Magnitude of Change		
(a) Mean change (± 1 SD) in area that exceeds threshold among ANTHx – CTRL	-15.1 (5.2) %	-8.6 (8.0) m
(b) Mean change (± 1 SD) within spatially explicit area of agreement	-17.3 (5.8) %	-11.6 (8.3) m

Discussion

Over the past five years, significant investments have been made to document ROMS-BEC model performance (Kessouri et al. 2021b, 2024). This information can yield key insights on model adequacy and inform managerial confidence in the model, particularly when accompanied by in-depth review of the model by an independent expert panel (NWRI 2025). This case study expands on this work via generation of confidence estimates generated for a science question, at the same temporal and spatial scale of the metric. Options exist for refining the toolkit for uncertainty quantification and we specify those investments needed to make those refinements.

Currently Only Ensemble Runs with Intrinsic Variability Can Be Used to Quantify Uncertainty

Among the three currently available methods of quantifying model uncertainty that we assessed, we found that ensemble runs with uncertainty due to intrinsic variability was the only applicable method for this case study. Even for this method, only two intrinsic runs for ANTH were available and only for 2013, limiting the scope of the uncertainty assessment. Figure 7 (left panel) demonstrates the idealized case of routine uncertainty quantification, in which scenarios with and without the predicted effect have uncertainty quantified. The right panel of Figure 7 illustrates what we were able to quantify using ensemble runs that included uncertainty due to intrinsic variability. On our case study of 2013, we could estimate uncertainty around ANTH, but not around the CTRL scenario. If managers are interested in using intrinsic variability in the future, refinements to our current capacity are recommended, which include an expanded number of ensemble runs for both the management scenario of interest and the comparator scenario (be it ANTH or CTRL) for multiple years.

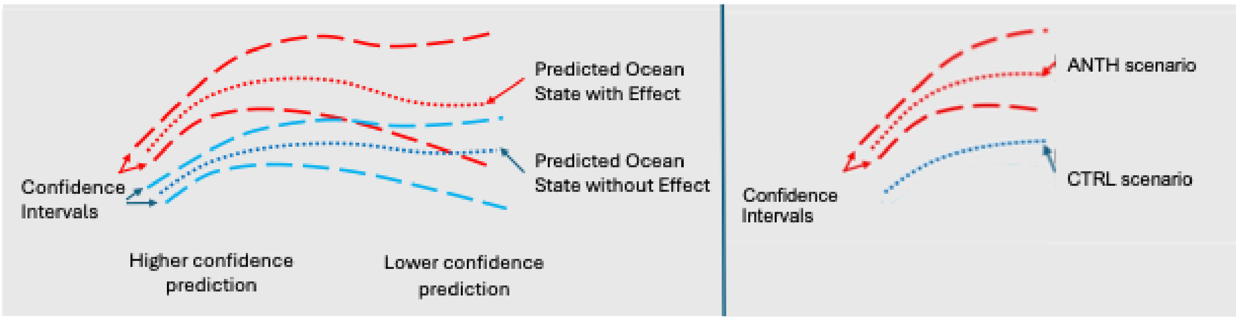


Figure 7 Depiction of an idealized routine uncertainty quantification, in which scenarios with and without the predicted effect have uncertainty quantified. The right panel illustrates what we were able to quantify using ensemble runs that included uncertainty due to intrinsic variability. On our case study of 2013, we could estimate uncertainty around ANTH, but not around the CTRL scenario.

Case Study Demonstrates How Error Estimates Can Be Used to Contextualize the Predicted Effect

Using the published change assessment methods of Kessouri et al. (2021, 2024) and Frieder et al. (2024), the predicted exceedances of the chemical and biological metrics of O_2 due to land-based nutrients were larger in 2013 than the intrinsic uncertainty estimates around the ANTH scenario, our best estimate of uncertainty for this case study at this time. We caveat that these estimates are not comprehensive of all sources of uncertainty. We note that the magnitude of estimated change due to land-based inputs depends on season and year, as noted in Frieder et al. (2024), which first reported on predicted habitat compression in the Bight. Oxygen and pH loss is most frequently predicted in the late summer and fall; in 2013, the maximum month differed by metrics (O_2 WQO in August and the aerobic habitat compression in October). Both Kessouri et al. (2024) and Frieder et al. (2024) noted that O_2 loss and habitat compression have large interannual variability, depending on SCB physics and its impact on biogeochemistry and primary productivity (Kessouri et al. 2024). Within the 20 years of ROMS-BEC predictions of ANTH, 2013 is a median productivity year.

Managers may use both performance criteria for model adequacy, defined in the ROMS-BEC quality assurance procedures (Sutula et al. 2025). The combination of O_2 metric and how it is expressed (general versus spatially explicit) can make the assessment more or less stringent, as illustrated in this exercise.

Expanding the Uncertainty Quantification Toolkit: Options and Required Investments

Table 7 summarizes the three methods, their relative advantages, and the scale of investments to utilize any of the three methods for management scenarios in the SCB. Investments could enhance our capacity to utilize uncertainty due to intrinsic variability for management scenarios, while use of data-model difference assessment and multiple model comparisons are not possible without additional investments.

Ultimately, which methods to use in future model applications will depend on a number of different factors (e.g., the questions to be answered, the variable of interest and the specific criterion to be used, the amount and quality of observational data available, the types of scenarios, and modeling resources available to invest to improve uncertainty quantification, among other considerations). We provide detailed discussion of these potential investments to expand the ROMS-BEC uncertainty toolkit; however, we caveat that the assessment of resources needed requires much greater context and nuance, depending on the specific application intended. The CTAG OAH Model Subcommittee discussed in depth and agreed that providing this context is beyond the scope of this initial case study; **the Subcommittee strongly recommended further discussion and elaboration of the magnitude of investments and associated costs, paired with discussion of the specific management scenarios, which is just beginning among the Management Scenarios Committee.**

Table 7. Summary of routine uncertainty quantification approaches, including basic description, their advantages, whether the approach can be used now, and recommendations to build capacity to add these approaches to the ROMS-BEC uncertainty toolkit. Note that interpolation uncertainty is only applicable to data-model difference assessment; errors due to model resolution are combined with structural/numerical. For investments required, the number of "$" is intended to convey the order of magnitude of resource investment required, where 1$ is tens of thousands and each additional $ is an order of magnitude higher. Those investments are estimated for modeling services and for the in-kind services of different monitoring programs.

Comparators		**Data-Model Difference Assessment**	**Ensemble Runs**	**Multiple Models**
Description		Comparison of observations with predictions of a realistic scenario	Supplemental runs of the same scenarios of ROMS-BEC, to which sources of uncertainty are added.	Simulations of other models compared to same scenarios of ROMS-BEC
Advantages	**Can be applied to any temporal or spatial scale**	No, limited to scale at which data can be justifiably interpolated	Yes	Yes
	Uncertainty types characterized	All but interpolation **error**	All but structural/ numerical	All but intrinsic and forcing uncertainty
	Can be applied to comparator or management scenario	No, Only realistic comparator only (e.g., ANTH)	Yes	Yes
	Used with eutrophication models with potential regulatory application?	No	Yes	Yes
Can the approach be used right now?		No, observational uncertainty unquantifie**d,**	Yes, but for intrinsic only	No, the two available CCS models have not reached "peer status"
Investments required include: **M = Modeling scientific services (e.g., SCCWRP or comparable)** **C = CalCOFI monitoring** **P = POTW monitoring** **SW = Stormwater monitoring**		Quantify observation uncertainty for CalCOFI data (5 years, C: $$ to $$$$, M: $$) Quantify observational data uncertainty for oxygen in CalCOFI and POTW data (5 years, C: $$$$; P: $$$$, M= $$) Improve POTW observational data quality to ensure suitability for data-model difference assessment (5 years, P: $$$$, M= $$)	Intrinsic variability: Develop a "bank" of ensemble runs (CTRL and ANTH) for a selected set of ocean base years that represent the "critical condition." (1 year, M= $$) Forcing uncertainty: conduct sensitivity analyses; quantify uncertainty in most sensitive pathways; combine with intrinsic variability in ensemble scenario (5 years, M= $$, P=$$$, SW=$$$). Parameter uncertainty: Develop a data assimilation version of ROMS-BEC and identify parameter sets with an equivalent model skill as original model; combine parameter uncertainty of ROMS-BEC with forcing and/or intrinsic variability (5 years, M= $$$)	Onboard, set up, test, simulate, then validate model(s): (a) ROMS-MARBL: already set up for the SBC (3 years, M= $$) (b) ROMS-NEMURO: 4 years, M= $$$

Data-Model Difference Assessment

Skill assessment remains one of the most routine methods to quantify model uncertainty to assess model adequacy, the most easily communicated, and straightforward methods to apply (Stow et al. 2009). One of the core advantages is that it integrates across multiple sources of model uncertainty, thus providing a simple approach that can communicate a basic uncertainty measure. It can be used to quantify uncertainty or communicate model confidence, as part of a comprehensive model evaluation (e.g., Kessouri et al. 2021b, 2024).

We attempted to utilize the principles of skill assessment to estimate the error in a model prediction, an application that we specifically refer to a "data-model difference assessment." We determined that insufficient published documentation of oxygen observational uncertainty exists for both CalCOFI and POTW monitoring data sets for this data-model difference assessment to be included in the uncertainty case study, for the reasons summarized in Table 2 and detailed below. Therefore, this method was not used in the case study of routine uncertainty quantification, although we note that skill assessment itself is still the primary mechanism to judge model adequacy (Sutula et al. 2025).

Why Observational Uncertainty Matters. Figure 8 provides a conceptual view of how observational uncertainty matters for quantification of model uncertainty. Scenario A shows the ideal case, in which observational uncertainty is smaller than both ANTH minus CTRL and data-model difference (O minus P). In this case, the uncertainty band can be calculated around the prediction and determined to be less or more than the change in ANTH versus CTRL. Less useful is Scenario B, in which observational uncertainty is too large to quantify prediction uncertainty but may still be smaller than ANTH minus CTRL. In that case, the confidence in the change in ANTH minus CTRL is driven by observational uncertainty rather than model uncertainty. Finally, Scenario C shows the case where observational uncertainty is so large that it cannot be useful to quantify model uncertainty because it is both larger than data-model difference AND larger than ANTH minus CTRL. In each case, the prediction error (data-model difference) remained the same, but the magnitude of observational uncertainty determines the interpretability and ultimate utility of using a data-model difference assessment in this manner.

How Is Observational Uncertainty Quantified? Observational uncertainty is dependent on the technique or specific instrument. The amount of seawater O_2 can be measured from discrete bottles using the Carpenter modification of the Winkler method (hereafter referred to as 'bottle oxygen'; Carpenter 1973), which is considered the gold standard. Vertical oxygen profiles measured via in situ sensors on CTDs are the industry standard, but to be reliable for uncertainty assessment must be calibrated with bottle samples, or a value from another calibrated O2 sensor in a side-by-side validation or controlled bath. Measurement error, sampling error, and (mapping (interpolation) error contribute to observational uncertainty.

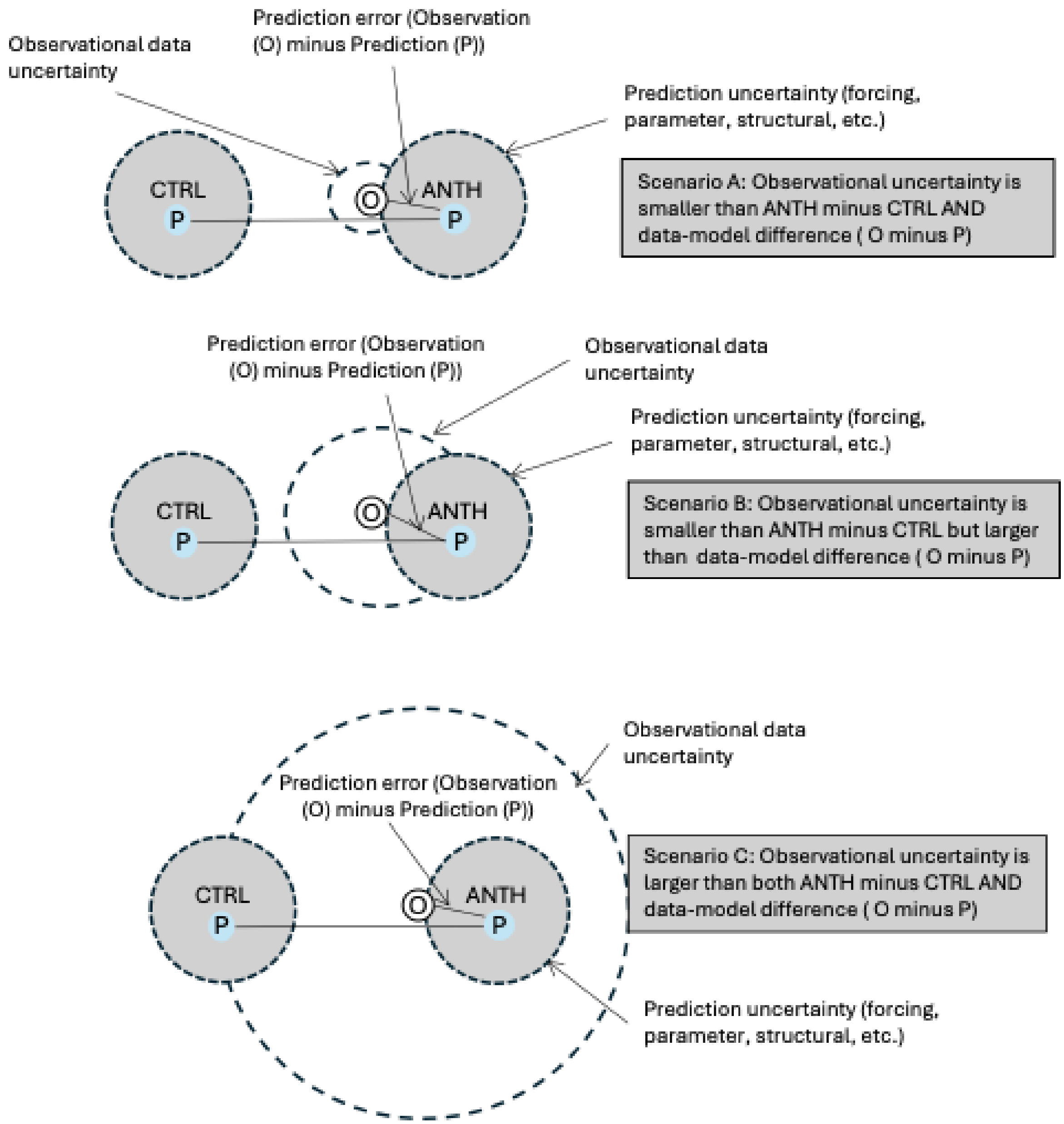


Figure 8. "How magnitude of observational uncertainty matters," as context for why it must be quantified to use it for assessment of uncertainty due to data-model differences. Schematic diagram of hypothetical relationships between model prediction (P), observations (O). Both O and P of ANTH and CTRL are assumed to have a halo of uncertainty, which is unknown. In this particular hypothetical example, scenario A shows the ideal case, in which observational uncertainty is smaller than both ANTH minus CTRL and data-model difference (O minus P). Less useful is Scenario B, in which observational uncertainty is too large to quantify prediction uncertainty but may still be smaller than ANTH minus CTRL. Scenario C shows the case where observational uncertainty is clearly so large that it cannot be useful to quantify model uncertainty because it is both larger than data-model difference AND larger than ANTH minus CTRL.

1. *Measurement error* for oxygen is technique or instrument specific. CalCOFI reports a bottle oxygen precision of 0.005-0.1 ml/L and a standard deviation typically under 0.010 ml/L (an equivalence of <0.5 µmol/kg).
2. *Sampling error* refers to errors associated with sampling because sampling is not always a perfect representation of the environment. For example, events can happen during sampling that cause error in the result, like contamination (e.g., oxygen intrusion). Global oxygen datasets assess this source of error by binning multiple oxygen profiles across a designated area and time period (e.g., within a month). The global mean for sampling error is 3.8 μmol/kg. However, this value is typically higher along eastern boundary upwelling systems where high spatial and horizontal oxygen gradients are present. We could evaluate sampling error specific to the CalCOFI oxygen dataset by evaluating whether there are replicate profiles or replicate bottle data available within the dataset.
3. *Interpolation error* represents error resulting from statistical methods used to interpolate datasets across space and time. This is typically the largest source of error. There is a parallel concept for data-model comparisons where model and data are integrated across space and time scales for comparison, but that relies on having data to sample the variability that represents that interpolation (e.g., from a single CTD drop to a 0.09 km^2 grid cell, from a single day to a monthly average, and from multiple stations needed to capture variability in O_2 state, particularly in areas with and without O_2 loss). Interpolation error in the CalCOFI dataset has not been documented for very specific ROMS-BEC application and it is highly likely that new data collection would be required.

Ito et al. 2024 can be used to illustrate the three types of error in global observations of oxygen (Table 8). The combined uncertainty in oxygen observations can be calculated as the square root of the sum of the square of each of the three individual sources. The amount of replication and sampling effort impacts confidence intervals across all three categories. Among the three types, interpolation error is the dominant sources of error; care in characterizing these error sources is key to utilizing data-model difference as an uncertainty assessment method.

Table 8. Example of uncertainty characterized in bottle oxygen datasets to generate global mean values (µmol/kg) and to conduct status and trends, from Ito et al. (2024).

Error Type	Global Mean, Bottle
Measured	1
Sampling	3.8
Interpolation	13-18
TOTAL	~13.5

What Investments are Needed? In order to use a data-model difference assessment to quantify model uncertainty, observational uncertainty must be quantified. Oxygen observational uncertainty from the regional CalCOFI and POTW monitoring datasets have not yet been comprehensively quantified, for the purposes of application at the temporal and spatial scales of O2 metrics used in this case stud. This gap is most easily addressed for CalCOFI O_2, carbonate chemistry and chlorophyll-a quarterly time series data, because they pair bottle data with CTD data, which would allow us to at least quantify measurement (and likely sampling) error over a 60-year time series. In addition, CalCOFI oxygen and pH data are collocated within the offshore regions where the greatest impacts are predicted to occur. We recommend consultation with both CalCOFI and modeling experts to determine the most appropriate means to invest in additional data collection to determine interpolation uncertainty, should such an option be considered.

A more significant effort is required to characterize observational uncertainty of POTW monitoring data. These data have been collected to comply with a regulatory program, so the quality standards are different from oceanographic research. Quality assurance practices considerably vary by agency and have evolved over time. If observational uncertainty is larger than both data-model differences and the differences between scenarios of interest (e.g., ANTH minus CTRL), then even with investments, the observational data uncertainty might not be of use (see Scenario C, Fig. 7). For some POTW agencies, the CTD monitoring data that appropriate calibration are likely to have much larger or unquantifiable uncertainty and cannot be used to quantify uncertainty. This is a critical data gap for metrics that rely on chlorophyll-a (i.e.,, risk of toxic HABs, Kessouri et al. in review), because the anthropogenic effect is nearshore and POTW monitoring data from agencies might not be able to be recalibrated to a gold standard. Addressing that gap would require investments in improved field sampling and laboratory analytical protocols over a 5+ year timeframe to generate enough data to be useful for model uncertainty characterization.

Finally, an even greater investment would be required to expand the density of data, particularly offshore and near the channel islands where the zones of OAH habitat compression are predicted to be the most severe.

Ensemble ROMS-BEC Simulations

The general idea behind ensemble runs of the same model is that different components of uncertainty (forcing, parameter, etc.) can be incrementally added the runs to provide a specific error estimate. Intrinsic variability, used in this case study, represents irreducible model noise due to ocean stochasticity. It is a type of forcing uncertainty but does not represent a comprehensive assessment of ROMS-BEC forcing, parameter and structural/numerical uncertainty and thus likely underestimates uncertainty. Forcing and parameter uncertainty cannot currently be incorporated into a ROMS-BEC ensemble run, because those types of uncertainty have not yet been quantified.

Investments Needed for Intrinsic Variability. Short-term, practical improvements could enhance how this approach is more routinely applied to support decisions. For example, we utilized three existing ANTH scenarios of 15 months in duration, repurposed from another project, to illustrate the concept. Extending the simulations to encompass multiple climate states and increasing the number of uncertainty runs would improve estimates of intrinsic variability but come with a cost. A similar approach could be taken to simulate intrinsic variability in the CTRL scenario, so that the difference in model noise could be accounted across both ANTH and CTRL runs.

We note that the application of these intrinsic variability assessments must be carefully tailored to the application. In the ideal case, you would have a confidence interval around both of the management scenario and the comparator scenario (ANTH, CTRL or other). Our example quantified uncertainty only in ANTH but not in the CTRL; this approach was used to quantify the source to be added to ANTH (Sutula et al. 2026), to determine whether that source has a quantifiable effect over the baseline ANTH. For other scenarios such as reductions in nutrient inputs (e.g., Ho et al. 2023), it might be more relevant to compare a management scenario to multiple intrinsic CTRL simulations, in order to detect the anthropogenic signal that emerges from the noise of natural background. Thus, multiple ANTH and CTRL intrinsic variability simulations, of sufficient length to capture interannual variability, would be needed; this could only be cost effective if a bank of intrinsic variability runs were produced that could be used for an established set of ocean years across multiple sets of projects and their respective scenarios. These improvements could be implemented within a year, forming a bank of runs used for routine management scenarios.

Investments Needed to Quantify Forcing Uncertainty. Quantifying forcing uncertainty requires effort but is relatively straightforward. For example, model and observational data are used to force ROMS-BEC at the 4 km scale (Deutsch et al. 2021). Error could be propagated from that scale, in addition to perturbated mesoscale intrinsic variability, in order to simulate a suite of ocean boundary forcing uncertainty simulations. Similarly, uncertainty in terrestrial forcing data exists, but has not been quantified. Observational error in the outfall volume and dissolved inorganic nitrogen constituents could be used to perturbate a suite of scenarios to bracket error in that source category. Assessments can be made of riverine forcing uncertainty; although they represent ~2% Bight wide (Sutula et al. 2021a), they can be important on local scales, especially for locations such as the Tijuana, San Gabriel, Los Angeles Rivers, and Calleguas Creek watersheds. In this case, costs could be reduced if the choice were to focus on the top 5 to 10 watersheds that contribute 75-90% of the loads, rather than characterize uncertainty in all riverine inputs.

Investments Needed to Quantify Parameter Uncertainty. Quantifying parameter uncertainty, while important, is more complicated. Two general approaches could be taken to quantify and incorporate this uncertainty into ensemble runs. The first is comprised of sensitivity analyses to quantify which parameters have disproportionate impact on management endpoints (OAH and chlorophyll-a), then a trial-and-error approach to determine the combination of these parameters that preserves a reasonable model skill, compared to the original ROMS-BEC (NWRI 2025). The second option is to create a data assimilation version of ROMS-BEC, then generate multiple sets of possible parameter combinations, bound by literature ranges, and choose the version that optimizes model skill, as was done for ROMS-NEMURO (Jacox et al. 2023). Ultimately the latter may represent a more reasonable path, because the outcome is both an ensemble ROMS-BEC model that can be used in uncertainty quantification, as well as a form of ROMS-BEC that can be used for nowcasts and short-term forecasts.

Thus, there is an inherent investment required to embrace and utilize this methodology for routine uncertainty characterization, with expanded investments needed to be more inclusive of multiple types of model uncertainty.

Multiple Model Comparisons

Multiple models are one of the most widely used methods for quantifying uncertainty in ocean prediction systems and are the favored method for predicting climate change. Instead of producing a single deterministic prediction via ROMS-BEC, an ensemble comparison generates many simulations—each with slightly different initial conditions, parameters, or model formulations—to explore the range of plausible ocean states. This approach addresses the inherent complexity and chaotic nature of ocean dynamics. This is the single most efficient method to capture the effect of different model structural and parameter uncertainty. Separate steps would still be required to incorporate forcing and intrinsic variability, increasing cost, but we note that programs like Chesapeake Bay have combined multiple approaches into ensemble runs in order to achieve a more comprehensive assessment of uncertainty (Sutula et al. 2022, NWRI 2025).

Two intermediate complexity coupled ROMS and biogeochemistry models have been developed for the California Current System and are available for multiple model comparisons: 1) ROMS coupled with the Marine Biogeochemistry library (MARBL); Damien et al. (in prep) and 2) ROMS coupled with the North Pacific Ecosystem Model for Understanding Regional Oceanography (NEMURO). ROMS-MARBL is structurally very similar to BEC, while NEMURO is a simpler biogeochemical model (11 state variables instead of 60) that has been frequently used with data assimilation in Central and Northern California for seasonal forecasting and fisheries applications (Jacox et al. 2023).

While both models have the potential be useful ensemble comparators with ROMS-BEC, the main issue is use of either model or both models at this time is not a “pairing of equals;” ROMS-BEC has been extensively validated for the SCB (Kessouri et al. 2021b). Model set up with equivalent forcing and multiannual year validation would be required in order for these models to be determined to be an appropriate comparison with ROMS-BEC. ROMS-MARBL has been set up for the SCB 300 m configuration with rivers and outfalls implemented, so for this model, an ensemble comparison would require running a 5-year simulation and validation, while for ROMS-NEMURO, more work would be involved to set up the model, run the model and validate it.

Summary and Next Steps

The case study demonstrated that we could produce a modeled change due to land-based nutrients, contextualized with an estimate with uncertainty due to intrinsic variability. For this example, the quantified exceedance of the criteria was larger than the intrinsic uncertainty estimates. Ensemble runs with intrinsic variability can be implemented at this time. However, for this method and for the two other remaining methods, investments should be considered to ready these approaches for ROMS-BEC applications in the future.

CTAG should consider prioritizing those investments, with an understanding that the optimal choice of which methods to use in future model applications will depend on the questions to be answered, the endpoint of and the specific metric/criterion to be used, the types of scenarios, and modeling resources available to invest to improve uncertainty quantification, among other considerations. As noted in the discussion, the Subcommittee strongly recommended further discussion and elaboration of the magnitude of investments and associated costs to provide the needed nuance and realistic cost estimates, paired with discussion of the specific management scenarios, which is just beginning among the Management Scenarios Committee.

LITERATURE CITED

Allen, J., Jimmieson, N. L., Bordia, P., & Irmer, B. E. (2007). Uncertainty during Organizational Change: Managing Perceptions through Communication. Journal of Change Management, 7(2), 187–210. https://doi.org/10.1080/14697010701563379

Brush, M.J., Giani, M., Totti, C., Testa, J.M., Faganeli, J., Ogrinc, N., Kemp, W.M. and Umani, S.F. (2020). Eutrophication, Harmful Algae, Oxygen Depletion, and Acidification. In Coastal Ecosystems in Transition (eds T.C. Malone, A. Malej and J. Faganeli). https://doi.org/10.1002/9781119543626.ch5

Chan, F., Barth, J. A., Lubchenco, J., Kirincich, A., Weeks, H., Peterson, W. T., & Menge, B. A. (2008). Emergence of Anoxia in the California Current Large Marine Ecosystem. Science, 319(5865), 920–920. https://doi.org/10.1126/science.1149016

Deutsch, C., Frenzel, H., McWilliams, J. C., Renault, L., Kessouri, F., Howard, E., Liang, J.-H., Bianchi, D., & Yang, S. (2021). Biogeochemical variability in the California Current System. Progress in Oceanography, 196, 102565. https://doi.org/10.1016/j.pocean.2021.102565

Deutsch, C., Penn, J. L., & Seibel, B. (2020). Metabolic trait diversity shapes marine biogeography. Nature, 585(7826), 557–562. https://doi.org/10.1038/s41586-020-2721-y

Howard, E. M., Penn, J. L., Frenzel, H., Seibel, B. A., Bianchi, D., Renault, L., Kessouri, F., Sutula, M. A., McWilliams, J. C., & Deutsch, C. (2020). Climate-driven aerobic habitat loss in the California Current System. Science Advances, 6(20), eaay3188. https://doi.org/10.1126/sciadv.aay3188

Jacox MG, Buil MP, Brodie S, Alexander MA, Amaya DJ, Bograd SJ, et al. (2023) Downscaled seasonal forecasts for the California Current System: Skill assessment and prospects for living marine resource applications. PLOS Clim 2(10): e0000245. https://doi.org/10.1371/journal.pclm.0000245

Kessouri, F., McLaughlin, K., Sutula, M., Bianchi, D., Ho, M., McWilliams, J. C., Renault, L., Molemaker, J., Deutsch, C., & Leinweber, A. (2021a). Configuration and Validation of an Oceanic Physical and Biogeochemical Model to Investigate Coastal Eutrophication in the Southern California Bight. Journal of Advances in Modeling Earth Systems, 13(12). https://doi.org/10.1029/2020ms002296

Kessouri, F., McWilliams, J. C., Bianchi, D., Sutula, M., Renault, L., Deutsch, C., Feely, R. A., McLaughlin, K., Ho, M., Howard, E. M., Bednaršek, N., Damien, P., Molemaker, J., & Weisberg, S. B. (2021b). Coastal eutrophication drives acidification, oxygen loss, and ecosystem change in a major oceanic upwelling system. Proceedings of the National Academy of Sciences, 118(21). https://doi.org/10.1073/pnas.2018856118

Kessouri, F., Sutula, M. A., Bianchi, D., Ho, M., Damien, P., McWilliams, J. C., Frieder, C. A., Renault, L., Frenzel, H., McLaughlin, K., & Deutsch, C. (2024). Cross-shore transport and eddies promote large scale response to urban eutrophication. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-57626-6

McClatchie, S., Jacox, M., Leising, A., Schneider, S., Thompson, A., Golightly, R., & Gomez-Valdes, J. (2016). State of the California Current 2015–16: Comparisons with the 1997–98 El Niño. California Cooperative Oceanic Fisheries Investigations Reports, 57, 1–57.

National Water Research Institute. (2025). Consensus Findings and Recommendations for the ROMS-BEC Model (24-398-CASA-13). National Water Research Institute.

Schiff, K., & Bay, S. (2003). Impacts of stormwater discharges on the nearshore benthic environment of Santa Monica Bay. Marine Environmental Research, 56(1–2), 225–243. https://doi.org/10.1016/s0141-1136(02)00332-x

Stow, C.A., J. Jolliff, D.J. McGillicuddy, S.C. Doney, J. I. Allen, M.A.M. Friedrichs, K.A. Rose, P. Wallhead. (2009). Skill assessment for coupled biological/physical models of marine systems, Journal of Marine Systems, Volume 76, Issues 1–2, Pages 4-15, https://doi.org/10.1016/j.jmarsys.2008.03.011.

Sutula, M., Howard, M., Crowder, L., McAfee, S. (2014). Modeling in support of the management of coastal hypoxia and acidification in the California Current Ecosystem. December 10-11, 2013, Workshop Proceedings, published March 2014. SCCWRP Technical Report 829.

Sutula, M., Ho, M., Sengupta, A., Kessouri, F., McLaughlin, K., McCune, K., & Bianchi, D. (2021a). A baseline of terrestrial freshwater and nitrogen fluxes to the Southern California Bight, USA. Marine Pollution Bulletin, 170, 112669. https://doi.org/10.1016/j.marpolbul.2021.112669

Sutula, M., Senn, D., Fono, L., Ho, M., Karimpour, F., Kessouri, F., King, A., Latker, A., and Markle, P. (2021b). Approaches to Quantifying Uncertainty in Coastal Eutrophication Numerical Modeling: A Workshop Summary. Joint Report of the Southern California Coastal Water