Parameter sensitivity and identifiability for a biogeochemical model of hypoxia in the northern Gulf of Mexico

Marcus W. Beck[1], John C. Lehrter[1], Lisa L. Lowe[2]

[1]*USEPA National Health and Environmental Effects Research Laboratory*
Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL 32561
Phone: 850-934-2480, Fax: 850-934-2401
Emails: beck.marcus@epa.gov, lehrter.john@epa.gov

[2]*Lockheed Martin IS & GS - Civil, supporting the USEPA*
Resarch Triangle Park, NC 27709
Phone: 919-541-3985,
Email: lowe.lisa@epa.gov

Version Date: Sat Aug 13 13:27:22 2016 -0500

Abstract

Bio-geo-chemical models are useful tools in environmental sciences that can guide management and policy-making. Consequently, significant time and resources are spent developing these models in system-specific contexts. The optimization of model parameters to maximize precision, including transferability of these models to different systems, are fundamental concerns in the development and application of these tools. This study describes quantitative limitations of coupled hydrodynamic-ecological modelling by contrasting numeric and ecological certainty with a systematic framework for characterizing parameter sensitivity and identifability. We evaluate a simple bio-geo-chemical model that is the zero-dimensional (0-D) unit of a larger spatio-temporal model of hypoxia on the Louisiana continental shelf of Gulf of Mexico as an example. Results from analysis of the 0-D model are used to infer larger trends in dissolved oxygen dynamics over time, having implications for understanding factors that contribute to environmental conditions that are detrimental to aquatic resources. In particular, we focus on issues of parameter identifiability using local sensitivity analyses to provide quantitative descriptions of numerical constraints on model precision. We argue that quantitative and ecological certainty in model calibration are often at odds and the practitioner must explicitly choose model components to optimize given tradeoffs between the two. We further conclude that numerically optimal parameter sets for models of hypoxia are often small subsets of the complete parameter set because of redundancies in the unique effects of paramater perturbations on model output. As a result, we demonstrate that use of a model for inference into ecological mechanisms of observed or predicted changes in hypoxic condition can be potentially misguided in the absence of quantitative descriptions of identifiability. Although these concerns have been expressed in the literature, they are rarely explicitly addressed or included in model evaluations. In addition to immediate implications for regional models, we provide a framework for describing the effects of parameter uncertainty and identifiability that can be applied to similar models to better inform environmental management.

1 *Introduction*

Hypoxia formation in bottom waters of coastal oceans occurs primarily from excess nutrient inputs from land-based sources (Justíc et al. 1987, Diaz and Rosenberg 1995, Howarth et al. 1996). These events are detrimental to aquatic organisms and have significant negative

effects on economic resources derived from coastal ecosystems (Lipton and Hicks 2003, Diaz and Rosenberg 2011). An understanding of the biological, physical, and chemical processes that contribute to the growth of hypoxic areas is a critical concern for mitigating and preventing these negative impacts. Numerical ecosystem models have been important tools for describing current knowledge of ecosystem processes that contribute to hypoxia formation and for predicting the effects of proposed management activities or future scenarios (Scavia et al. 2004, Hagy and Murrell 2007, Pauer et al. 2016). Unlike statistical models that have more generic structures, simulation and process-based models include explicit descriptions of relevant processes that are constrained by empirical or observational data relevant to the system of interest (e.g. Omlin et al. 2001b, Eldridge and Roelke 2010). These models are often coupled with hydrodynamic grids to provide spatially-explicit representations of patterns in three dimensions (Warner et al. 2005, Zhao et al. 2010, Ganju et al. 2016). Combined hydrodynamic and bio-geo-chemical models have been developed specifically to describe hypoxic conditions on the Louisiana continental shelf (LCS) in the northern Gulf of Mexico (GOM) (Fennel et al. 2013, Obenour et al. 2015, Pauer et al. 2016, Lehrter et al. in review). This area drains a significant portion of the continental United States through the Mississippi-Atchafalaya River Basin (MARB) and is the second largest hypoxic area in the world (Rabalais et al. 2002). Understanding processes that contribute to the frequency and duration of hypoxic events remains a critical research goal for the region, including the application of process-based models to characterize the current knowledge domain.

The development and application of a model represents a tradeoff between characteristics expected from the output or provided by the structural components. An idealized model is sufficiently generalizable across systems, provides results that are precise given the inputs, and includes components that are realistic descriptions of actual processes (Levins 1966). Given that these characteristics cannot be simultaneously achieved, models are developed in partial dependence of reality and theoretical constructs, completely separate from both, or dependent on one or the other (Morrison and Morgan 1999, Ganju et al. 2016). These challenges are analagous to the well-known bias-variance tradeoff in statistical models that balances the competing objectives of over- and under-fitting to an observed dataset. Process-based models are more commonly imbalanced between reality and theory, such that most are over-parameterized in an attempt to completely describe reality (Denman 2003, Nossent and Bauwens 2012, Petrucci and

Bonhomme 2014). Such over-parameterization, including use of excessive structural equations, can have serious implications for practical applications. Quantitative limitations of over-parameterization are analagous to degrees of freedom in standard statistical models as free parameters cannot be numerically estimated when constrained to an observed dataset (Kirchner 2006). More importantly, over-parameterization can limit use across systems outside of the data domain and impose uncertainty in model predictions as realistic values for every variable may not be known or inaccurately applied from existing studies (Durand et al. 2002, Refsgaard et al. 2007, Wade et al. 2008). The application of process-based models to describe hypoxia dynamics has not been immune to these challenges and more comprehensive approaches are needed to develop models that more carefully balance theory with reality (e.g., Snowling and Kramer 2001).

Standard approaches for uncertainty analysis can be used to begin addressing model complexity issues. In the most general sense, uncertainty is evaluated relative to the effects of input conditions or the observed data used to calibrate a model, changes in parameter values, or variation in the structural components (Beck 1987). Evaluating parameter uncertainty is by far the most common and simplest means of evaluating model behavior. Although uncertainty analyses should be integrated throughout model development and application, parameters are more often evaluated post-hoc as a form of 'damage control' for further calibration. This approach is sometimes called inverse modelling where results from sensitivity analyses are used to guide calibration or fit of the developed model to observations (Soetaert and Petzoldt 2010, or confronting models with data, *sensu* Hilborn and Mangel 1997). Parameter sensitivity analysis combined with inverse modelling necessarily involves questions of parameter 'identifiability', where only a subset of parameters can be numerically constrained to the data as compared to the entire parameter set. Redundancies in parameter effects lead to unidentifiable models where optimal solutions may be empirically impossible (i.e., standard algorithms will not converge) or parameter values may be non-unique leading to the right answer for the wrong reason (Kirchner 2006). The concept of identifiable parameter subsets is not foreign to hypoxia or eutrophication models (Omlin et al. 2001a, Estrada and Diaz 2010, Mateus and Franz 2015), although there is a clear need for greater integration of these concepts in model development (Fasham et al. 2006). Moreover, the inclusion of sensitivity and identifiability analyses in model tuning will require the adoption of selection rules that are context-dependent given the subsets that are possible with

large parameter sets (e.g., Wagener et al. 2001a,b).

This study describes a parameter sensitivity analysis to evaluate identifiability for a bio-geo-chemical model of hypoxia for the northern GOM. We evaluate a simple zero-dimensional (0-D) unit of a larger spatial-temporal model to explore relationships between multiple parameter sets and hypoxia dynamics on the LCS. Specifically, we provide empirical results to support the assumption that models are generally over-parameterized and only a finite and smaller subset of the larger parameter set can be optimized for a given research question or dataset. We provide explicit guidance for choosing such subsets of the parameter space given constraints on identifiability as directly related to sensitivity analyses. The objectives are to 1) identify the parameters that have the greatest influence on dissolved oxygen (O_2) using local sensitivity analysis, 2) quantify the identifiability of subsets of the total parameter space based on sensitivity, 3) and provide a set of heuristics for choosing parameters based on sensitivity, identifiability, and parameter categories, including extension to other state variables provided by the model. A final analysis evaluates identifiability relative to structural uncertainty to provide an example of extending these methods to more complex uncertainty assessments. The optimum parameter space is defined as the chosen subset that represents the maximum number of identifiable parameters. Here, 'optimum' is both a qualitative description based on a research question or management goal and a quantitative objective based on numerical optimization criteria for fitting model output to a calibration dataset. These results can be used to refine existing models or guide application of models to novel contexts, such as downscaling or application to new environments. We conclude with a discussion of the implications for hypoxia formation in coastal regions, including management strategies for nutrient reduction and use of mechanistic models to inform decision-making.

2 *Methods*

2.1 Model description

Hypoxic events, defined as <2 mg L^{-1} of O_2 (< 64 mmol $m^{-}3$), occur seasonally in bottom waters in the northern GOM. The LCS receives high nutrient loads from the MARB that drains a significant portion of the continental United States. Nutrient-stimulated primary production in surface waters increases biological oxygen demand in bottom waters as sinking

organic matter is decomposed (Bierman et al. 1994, Murrell et al. 2013). The hypoxic area averages 15,540 km^2 annually (1993-2015) with minimum concentrations observed from late spring to early fall. Seasonal variation is strongly related to carbon and nutrient export from the MARB (Lohrenz et al. 2008, Bianchi et al. 2010), whereas hydrologic variation, currents, and wind patterns can affect vertical salinity gradients that contribute to the formation of hypoxia (Wiseman et al. 1997, Paerl et al. 1998, Obenour et al. 2015).

This study evaluated a recently developed hydrodynamic and ecological model that describes horizontal and vertical transport and mixing of state variables relevant for hypoxia in the northern GOM. The Coastal General Ecosystem Model (CGEM) includes elements from the Navy Coastal Ocean Model (Martin 2000) that describes hydrodynamics on the LCS and a biogeochemical model with multiple plankton groups, water-column metabolism, and sediment diagenesis (Eldridge and Roelke 2010). The hydrodynamic component of CGEM provides a spatially-explicit description of hypoxia using an orthogonal grid with an approximate horizontal resolution of 1.9 km^2 and twenty equally-spaced vertical sigma layers on the shelf (depth $\leq$ 100 m, with additional hybrid layers at deeper depths). The biogeochemical component includes equations for 36 state variables including six phytoplankton groups (with nitrogen and phosophorus quotas for each), two zooplankton groups, nitrate, ammonium, phosphate, dissolved inorganic carbon, oxygen, silica, and multiple variables for dissolved and particulate organic matter from different sources. Atmospheric and hydrological boundary conditions described in Hodur (1997) and Lehrter et al. (2013) are also included in CGEM.

The core unit of CGEM is FishTank, a 0-D model that implements the biogeochemical equations in Eldridge and Roelke (2010) and does not include any form of physical transport (i.e., advection, mixing, or surface flux). Although FishTank was developed for specific application in CGEM, it can easily be applied to other hydrodynamic grids. Accordingly, the sensitivity and identifiability analysis described below are informative for both the LCS gridded model as well as potential applications to different systems. The FishTank model provides estimates for the 36 state variables described above using a 0-D parcel that is uniformly mixed as a closed system. A set of initial conditions is provided as input to the model that was based on observations of relevant variables obtained from research cruises in April, June, and September 2006 (Table 1 in Murrell et al. 2014).

Results from FishTank are based on time-dependent differential equations that describe energy flow between phytoplankton and zooplankton groups as affected by nutrient uptake rates, organic matter inputs and losses, inherent optical properties, sediment diagenesis, and temperature (Penta et al. 2008, Eldridge and Roelke 2010, see appendix in Lehrter et al. in review). A total of 108 equations are estimated at each time step to return a value for each of the 36 state variables described by the model. In addition to the initial conditions, 251 parameter values for each of the equations are also supplied at model execution. These parameters define relationships among fixed effects in the equations and represent ecological properties described by the model that influence hypoxia formation. Values for each of the parameters were based on estimates from the literature, field or laboratory-based measurements, or expert knowledge in absence of the former. As such, a sensitivity analysis of parameter values is warranted given that, for example, literature or field-based estimates may not apply under all scenarios or expert knowledge is not completely certain (Refsgaard et al. 2007). The sensitivity of O_2 to perturbations of all relevant parameters for the 108 equations was estimated from January 1st to December 31st, 2006 by running FishTank at a timestep of five minutes. Initial screening removed 73 irrelevant parameters that were not relevant such that 178 were retained for the sensitivity analysis. Irrelevant parameters were removed for several reasons, e.g., a parameter was only relevant for the hydrodynamic model, was considered a physical constant, or had no effect given initial conditions. For simplicity, the remaining parameters were grouped into one of six categories based on their respective equations: optics ($n = 4$ parameters), organic matter (12), phytoplankton (132), temperature (8), and zooplankton (22). A full description is available as an appendix in (Lehrter et al. in review).

2.2 Local sensitivity analysis

The analysis focused on sensitivity of O_2 in the 0-D FishTank model to identify parameters that may affect spatial and temporal variation of hypoxia in the larger model. A local sensitivity analysis was performed for each of the 178 parameters using a simple perturbation approach to evaluate the change in O_2 from the original parameter values. The analyses relied exlusively on concepts used in the FME package developed for the R statistical programming language (Soetaert and Petzoldt 2010). Each parameter was perturbed by 50% of its original value and the model was executed to obtain an estimate of the effect on O_2. For each perturbation, a sensitivity value S was estimated for each time step i given a set value for parameter j as:

$$S_{ij} = \frac{\partial y_i}{\partial \Theta_j} \cdot \frac{w_{\Theta_j}}{w_{y_i}} \tag{1}$$

where the estimate depended on the change in the predicted value for response variable y divided by the change in the parameter Θ_j multiplied by the quotient of scaling factors w for each. The scaling factors, w_{Θ_j} for the parameter Θ_j and w_{y_i} for response variable y_i, were set as the default value of the unperturbed parameter and the predicted value of y_i after perturbation (Soetaert and Petzoldt 2010). The scaling ensures the estimates are unitless such that the relative magnitudes provide a comparison for model sensitivity to parameter changes that may vary in scale. Estimates for S_{ij} were summarized as $L1$ and $L2$ across the time series to obtain individual sensitivity values of O_2 in response to a change in parame ter j:

$$L1 = \sum |S_{ij}|/n \tag{2}$$

$$L2 = \sqrt{\sum \left(S_{ij}^2\right)|/n} \tag{3}$$

In general, positive sensitivity estimates suggested a parameter had a positive effect on O_2 for a given increase in the parameter, whereas the converse was true for negative sensitivity estimates. However, the effect of a parameter change may not be uniform over time such that S_{ij} can change in magnitude and sign depending on the temporal location. Time series of O_2 estimates before and after perturbation were also evaluated to identify patterns not captured by the summary statistics. All parameters for each of the six equation categories (optics, organic matter, phytoplankton, temperature, and zooplankton) that had non-zero $L1$ or $L2$ were retained for identifiability analysis.

2.3 Identifiability and selecting parameter subsets

Identifiability of parameter subsets was estimated from the minimum eigenvector of the cross-product of a selected sensitivity matrix (Brun et al. 2001, Omlin et al. 2001a):

$$\gamma = \frac{1}{\sqrt{\min\left(\mathrm{EV}[\hat{S}^\top \hat{S}]\right)}} \tag{4}$$

where γ ranges from one to infinity for perfectly identifiable (orthogonal) or unidentifiable

(perfectly collinear) results for parameters in a sensitivity matrix S. The sensitivity functions were supplied as a matrix $\hat{S}$ with rows i and columns j (eq. (1)) that described deviations of predicted O_2 from the default parameter values. The matrix $\hat{S}$ was first normalized by dividing by the square root of the summed residuals (Omlin et al. 2001a, Soetaert and Petzoldt 2010).

The collinearity index γ provides a measure of the linear dependence between sensitivity functions described above for subsets of parameters. Estimates of γ greater than 10-15 suggest parameter sets are poorly identifiable (Brun et al. 2001, Omlin et al. 2001a), meaning optimal values are inestimable given similar effects of the selected parameters on O_2. Greater sensitivity of a state variable to a subset of parameters does not always imply better identifiability if the individual effects are similar. An intuitive interpretation of γ is provided by Brun et al. (2001) such that a change in a state variable caused by a change in one parameter can be offset by the fraction $1 - 1/\gamma$ by the remaining parameters. That is, $\gamma = 10$ suggests the relative change in O_2 for a selected parameter can be compensated for by 90% with changes in the other parameters.

Initial analyses suggested that considerably limited subsets of parameters were identifiable of the 178 evaluated for the FishTank model. Given this limitation, parameter selection must consider the competing objectives of increased precision with parameter inclusion and reduced identifability as it relates to optimization. An additional challenge is the excessively high number of combinations of parameter sets, which complicates selection given sensitivity differences and desired ecological categories of each parameter. For example, Fig. 1 provides a simple graphic of the unique number of combinations that are possible for different subsets of 'complete' parameter sets of different sizes (i.e., based on n choose k combinations equal to $n!/\left(k!\left(n-k\right)!\right)$). The number of unique combinations increases with the total parameters in the set and is also maximized for moderate selections (e.g., selecting half the total). For example, over 10^{14} combinations are possible by selecting 25 parameters from a set of 50. Accordingly, parameter selection is complicated by differing sensitivity, identifiability, and the difficulty of choosing from many combinations.

A set of heuristics was developed to balance the tradeoff in model complexity and identifiability given the challenges described above (see also Wagener et al. 2001a). These rulesets were developed with the assumption that parameters will be selected with preference for those with high sensitivity and identifability based on $\gamma < 15$ as an acceptable threshold for

subsets (e.g., 93% accountability). Selection heurestics also recognized that parameter categories (i.e., optics, organic matter, phytoplankton, temperature, zooplankton) may have unequal preferences given questions of interest. In all selection scenarios, parameters were selected by decreasing sensitivity starting with the most sensitive until identifiability did not exceed $\gamma = 15$ where selections were 1) blocked within parameter category, 2) independent of parameter category, 3) or considering all categories equally . The selection rules produced seven subsets of parameters that could further be used to optimize model calibration for O_2.

The above analyses were repeated for additional state variables estimated by FishTank to provide further descriptions of ecological dynamics that are relevant for hypoxia. In addition to O_2, other state variables included chlorophyll *a* (chl-*a*), photosynthetically active radiation (PAR), nitrate, ammonium, particulate organic matter, dissolved organic matter, and phosphorus. Particulate and dissolved organic matter were estimated as the summation of the respective outputs for organic matter from phytoplankon (*OM1_A*, *OM2_A*), fecal pellets (*OM1_fp*, *OM2_fp*), river sources *OM1_rp*, *OM2_rp*), and boundary conditions (*OM1_bc*, *OM2_bc*, see Lehrter et al. in review).

2.4 Identifiability and structural uncertainty

As a final analysis, the effects of model structure on parameter identifiability were evaluated for O_2 predictions. The FishTank model includes several processes that can be included or excluded from the equations based on expected conditions or available data. These 'switches' are conceptually different from model parameters as they define the use of explicit equations or processes that affect the model structure. As such, a comparison of sensitivity and identifiability given different equations uses parameter uncertainty as a means to begin addressing componenents of structural uncertainty. Switches in FishTank include different structural equations for the vertical attenuation of light through the water column (inherent or apparent optical properties, Penta et al. 2009, Eldridge and Roelke 2010) and chlorophyll to carbon ratio models (fixed or dynamic given light and nutrients, Cloern et al. 1995). Several switches also affect phytoplankton growth including different models for specific growth and effects of temperature, light dependence, nutrient uptake, and internal cell quotas (Lehrter et al. in review, references therein). Parameter identifiability was evaluated using three scenarios of simple, intermediate, and complex combinations of model switches.

3 *Results*

3.1 Local sensitivity analysis

Local sensitivity analyses showed that O_2 was sensitive to perturbations in 105 of 178 parameters (59% of total) in FishTank (see Fig. S1 for other state variables). Within each parameter category, O_2 was sensitive to three parameters for optics (27% of all optic parameters, Table 1), eight for organic matter (25%, Table 2), 78 for phytoplankton (50%, Table 3), six for temperature (27%, Table 4), and 10 for zooplankton (33%, Table 5). Although O_2 had the greatest sensitivity to parameters in the zooplankton category (as percentage of total), the relative effects varied. Among all parameters, sensitivity values ranged from $L1 = 2.29 \times 10^{-5}$ for *respg*$_{p5}$ (phytoplankton) to 1.24 for *KG2* (organic matter), whereas average sensitivity among all parameters was $L1 = 0.12$. Within categories, sensitivity ranged from 6.78×10^{-4} (*astarOMZ*) to 0.43 (*astar490*) for optics, 0.01 (*KNO3*) to 1.24 (*KG2*) for organic matter, 2.29×10^{-5} (*respg*$_{p5}$) to 0.65 (*Qc*$_{p1}$) for phytopankton, 9.75×10^{-5} (*Tref(nospA+nospZ)*$_{p6}$) to 9.97×10^{-5} (*Tref(nospA+nospZ)*$_{p4}$) for temperature, and 5.75×10^{-5} (*ZQp*$_{z1}$) to 5.75×10^{-5} (*ZQp*$_{z2}$) for zooplankton (Fig. 2, bottom). Average sensitivity values in each category were $L1 = 0.17$ for optics, 0.34 for organic matter, 0.12 for phytoplankton, 9.89×10^{-5} for temperature, and 0 for zooplankton. Within the six phytoplankton groups, O_2 was sensitive to the same parameters within each group although the sensitivity magnitudes varied. Average sensitivity across parameters in each phytoplankton groups showed that O_2 was most sensitive to the first and third phytoplankton groups (average $L1 = 0.12, 0.12$), whereas sensitivity to parameters in the remaining groups was much lower (all with average $L1 < 1$). Sensitivity of O_2 did not vary considerably between parameters in the two zooplankton groups (average $L1 = 0, 0$ for groups one and two).

Response of O_2 to parameter perturbations was not uniform across the time series. Fig. 2 shows variation for the top parameters within each category. Because FishTank does not include a spatial component, the estimated O_2 trend describes a closed, heterotrophic system where respiration processes eventually remove all O_2 from the model space. The initial decrease in the time series reflects change from the initial conditions outside of the growing seasons (i.e., January, February), the spring/summer increase represents production associated with expected

seasonal growth, and the remaining time series from August to the end of year shows complete removal of O_2 as respiration processes dominate metabolic activity. Although this time series is not a realistic depiction of an actual system, the biogeochemical model behaves as expected in the absence of the hydrodynamic model. Accordingly, the interpretation of sensitivity results from the simple model has relevance in an ecological context. As expected, parameter perturbations had the largest effect during the summer months, although the effects varied. An inrease in 50% from the parameter default values generally caused a reduction in O_2 during the summer, with the exception of the zooplankton parameter, ZQp_{z2}, which caused an increase in O_2. The phytoplankton parameter, Qc_{p1}, had the largest effect such that the O_2 time series was similar to the default output in April/May, whereas a dramatic decrease was observed in the remaining months. The effects of perturbations early in the time series (January, February) showed similar patterns such that a reduction in O_2 was most common, particularly for the optics (*astar490*) and phytoplankton (Qc_{p1}) parameters.

3.2 Subset identifiability

The identifiability analyses suggested that most parameter subsets exceeded the thresholds of $\gamma = 10, 15$, providing further justification for using selection heuristics for parameter optimization. Parameter identifiability (as γ) increased at different rates depending on the parameter category or the number of top parameters that were selected (Fig. 3). By category, identifiability was lowest for parameter subsets in the phytoplankton (1.7% less than $\gamma = 15$, 1.2% less than $\gamma = 10$) and zooplankton categories (7.9%, less than $\gamma = 15$, 4.3%, less than $\gamma = 10$), whereas a majority of combinations for temperature were identifiable ($NaN \times 10^{NA}\%$ less than $\gamma = 15$, $NaN \times 10^{NA}\%$, less than $\gamma = 10$). All subset combinations for optics and organic matter parameters had $\gamma < 10$. All parameter subsets for choosing the top, top two, and top three parameters in each category were identifiable (Fig. 3), whereas a majority were identifiable for choosing the top four (2.7% less than $\gamma = 15$, 1.9% less than $\gamma = 10$) and top five (2.7% less than $\gamma = 15$, 1.9%, less than $\gamma = 10$) parameters.

A comparison of average and median identifiability by parameter category and top parameters in each category suggested that individual parameters had large effects on γ (Fig. 3). For example, a consistent increase in average γ from 2 to 7 parameters in a combination for temperature was observed, whereas median identiability remained low until 6 parameter

combinations were evaluated. Further evaluation showed that identifiability was greatly affected by the inclusion of one or two specific parameters in a subset combination. Fig. 4 shows the temperature collinearity (γ) ranges in detail for the parameter subsets in Fig. 3 before and after excluding individual parameters. Collinearity increases with more parameters included in a subset, although the increase varies depending on the specific parameter. Exclusion of the parameters *Tref(nospA+nospZ)*$_{p2}$ and *Tref(nospA+nospZ)*$_{p5}$ showed that γ remained well below the 10, 15 threshold for all parameter combinations. Morever, inclusion of *Tref(nospA+nospZ)*$_{p1}$, *Tref(nospA+nospZ)*$_{p4}$, *Tref(nospA+nospZ)*$_{z1}$, and *Tref(nospA+nospZ)*$_{z2}$ generally reduced collinearity relative to when the parameters were excluded. Similar analyses identified parameters in other categories that had disproportionate effects on identifiability if included in a subset (see supplementary information).

Comparison of identifiability between categories showed that phytoplankton and zooplankton had the least identifiable parameter subsets. As noted above, FishTank includes six phytoplankton and two zooplankton groups to characterize community structure and foodweb dynamics that likely have an important role in hypoxia development. However, structural equations for each group do not vary considerably such that variation in parameter values primarily control differences between the groups, e.g., large-bodied vs small-bodied plankton, slow-growing vs. fast-growing plankton. To obtain identifiability estimates of the plankton categories that were independent of groups, the identifiability analyses were re-evaluated using only one phytoplankton and one zooplankton group (Fig. 5). Compared to all groups, evaluating a single group improved identifiability such that a majority of parameter combinations were below the threshold (56.3% less than $\gamma = 15$, 47.1% less than $\gamma = 10$ for Phytoplankton; 83.4%, less than $\gamma = 15$, 72.6%, less than $\gamma = 10$ for zooplankton). Analysis of identifiability for top parameters in all categories did not show similar changes in identifiability. Although all combinations for the top one, two, and three parameters in each category were still well below the threshold, selecting the top four and five parameters showed an increase (84.2% less than $\gamma = 15$, 78.7% less than $\gamma = 10$) and a decrease (54.4% less than $\gamma = 15$, 51% less than $\gamma = 10$) in identifiability, respectively (Fig. 5), compared to results that included all phytoplankton and zooplankton groups (Fig. 3).

3.3 Parameter selection

Each of the three selection heuristics (blocked by parameter category, independent of category, all categories equally) for O_2 differed in the number of selected parameters and distribution of parameters within each category (if applicable, Fig. 6 and Table 6). For the first selection heuristic, all sensitive parameters from the optics ($n = 2$, $\gamma = 3.5$) and organic matter ($n = 3$, $\gamma = 10.5$) categories were selected, whereas one were selected for phytoplankton ($\gamma = 1$), one for temperature ($\gamma = 1$), and one for zooplankton ($\gamma = 1$). For the second selection heuristic, two parameters were selected ($\gamma = 8.5$, with a majority from the phytoplankton category, Table 6). For the third heuristic, four parameters were selected ($\gamma = 12.6$) with equal representation between categories. For the second and third selection heuristics, an individual parameter caused a disproportianate increase in γ that forced the selection to stop. Selection independent of category showed that including Qc_{p1} caused in increase in γ from 8.5 to 17.5 and equal selection within cateogies showed that including ZQc_{z1} caused in increase in γ from 12.6 to 72.2. Finally, parameter selection for other state variables (chl-*a*, PAR, nitrate, ammonium, particulate organic matter, dissolved organic matter, phosphorus) also showed that a limited number of parameters were identifiable (for brevity, only results using the second set of selection heuristics are shown, Fig. 7 and Table 7). Less than ten parameters were selected for each of the remaining state variables with most selected from the phytoplankton category. Interestingly, no parameters from the temperature category were selected.

4 Discussion

We showed only small subsets are identifiable, similar conclusions have been described by citations in (Wagener et al. 2001a), p. 14 for models that follow traditional calibration schemes (e.g., objective function minimization).

Emphasize that parameters that have the greatest effect on collinearity are not those that have the highest sensitivity (contrast the identifiability by category vs identifiability by top parameters) , also note that groups of parameters together can have large effects on collinearity, maybe some kind of bootstrap analysis could be done looking at doubletons, etc. The example in teh results highlights how redundant variables can be identified as a necessary part of the model calibration process.

Why did identifiability decrease for top five parameters in each category after removing redundant phyto/zoop groups?

(Denman 2003) describes over-parameterization for multiple phyto groups, use this to emphasize collinearity issues with the phyto groups despite differing sensitivity values between groups

Identifiability by category - varies with number of parameters in the category but some were more redundant than others (phytoplankton).

Questions specific to GOM - what initial conditions are important? How many phytoplankton groups do we need (e.g., related to structural uncertainty)?

How does the assimilation of additional parameters (e.g., other state variables) during calibration influence the conclusions? Wagener et al. (2001a) describes this as a potential approach to improving model performance by improving the availability of information for model calibration (p. 14).

How does uncertainty translate to what a model should provide (generality v precision)? The first step - find out what can be optimized but then do not overfit....

What about structural uncertainty - does sensitivity of a model to variation in a parameter imply parameter uncertainty and/or structural uncertainty?

A final point about optimization with identifiable parameter sets - optimization to fit the data still does not ensure a correct model. Failing in one way can be over-compensated by another feature, e.g., the parameter set that is optimized (see Flynn (2005), p. 1207, third paragraph), also (Durand et al. 2002, Arhonditsis et al. 2008), also note that over-parameterized models are not necessarily bad, see (Omlin et al. 2001a), models must be validated for the uses which they were intended where a comparison of observed and predicted is the simplest approach and other methods should be used depending on analysis needs (Jr. 1996)

Omlin et al. (2001a) state that the sensitivity, identifiability, estimation process is iterative (p. 113), need to rinse and repeat for proper calibration.

How to improve identifiability - get more/better observed data, include obs from other state variables in RSS minimization (eqn q in Omlin et al. (2001a))

Alternative methods for uncertainty analysis - bayesian, MCMC, nonlinear calibration-constrained optimization (Gallagher and Doherty 2007), (Arhonditsis et al. 2008)

Our heuristic products are partially analogous to the Rainfall-Runoff Modelling Toolbox (RRMT) presented by (Wagener et al. 2001b)2001 (cited on page 15, in Wagener et al. 2001a), although ours is simple in comparison

References

Arhonditsis GB, Perhar G, Zhang W, Massos E, Shi M, Das A. 2008. Addressing equifinality and uncertainty in eutrophication models. Water Resources Research, 44(1):W01420.

Beck MB. 1987. Water quality modeling: A review of the analysis of uncertainty. Water Resources Research, 23(8):1393–1442.

Bianchi TS, DiMarco SF, Jr JHC, Hetland RD, Chapman P, Day JW, Allison MA. 2010. The science of hypoxia in the Northern Gulf of Mexico: a review. Science of the Total Environment, 408(7):1471–1484.

Bierman VJ, Hinz SC, Zhu DW, Wiseman WJ, Rabalais NN, Turner RE. 1994. A preliminary mass-balance model of primary productivity and dissolved oxygen in the Mississippi River plume/ inner Gulf shelf region. Estuaries, 17(4):886–899.

Brun R, Reichert P, Künsch HR. 2001. Practical identifiability analysis of large environmental simulation models. Water Resources Research, 37(4):1015–1030.

Cloern JE, Grenz C, Vidergar-Lucas L. 1995. An empirical model of the phytoplankton:carbon ratio - the conversion factor between productivity and growth rate. Limnology and Oceanography, 40(7):1313–1321.

Denman KL. 2003. Modelling planktonic ecosystems: parameterizing complexity. Progress in Oceanography, 57(3-4):429–452.

Diaz RJ, Rosenberg R. 1995. Marine benthic hypoxia: A review of its ecological effects and the behavioural responses of benthic macrofauna. Oceanography and Marine Biology, 33:245–303.

Diaz RJ, Rosenberg R. 2011. Introduction to environmental and economic consequences of hypoxia. International Journal of Water Resources Development, 27(1):71–82.

Durand P, Gascuel-Odoux C, Cordier MO. 2002. Parameterisation of hydrological models: a review and lessons learned from studies of an agricultural catchment (Naisin, France. Agronomie, 22(2):217–228.

Eldridge PM, Roelke DL. 2010. Origins and scales of hypoxia on the Louisiana shelf: importance of seasonal plankton dynamics and river nutrients and discharge. Ecological Modelling, 221(7):1028–1042.

Estrada V, Diaz M. 2010. Global sensitivity analysis in the development of first principle-based eutrophication models. Environmental Modelling and Software, 25:1539–1551.

Fasham MJR, Flynn KJ, Pondaven P, Anderson TR, Boyd PW. 2006. Development of a robust marine ecosystem model to predict the role of iron in biogeochemical cycles: A comparison of results for iron-replete and iron-limited areas, and the SOIREE iron-enrichment experiment. Deep-Sea Research I, 53:333–366.

Fennel K, Hu J, Laurent A, Marta-Almeida M, Hetland R. 2013. Sensitivity of hypoxia predictions for the northern Gulf of Mexico to sediment oxygen consumption and model nesting. Journal of Geophysical Research: Oceans, 118(2):990–1002.

Flynn KJ. 2005. Castles built on sand: dysfunctionality in plankton models and the inadequacy of dialogue between biologists and modellers. Journal of Plankton Research, 27(12):1205–1210.

Gallagher M, Doherty J. 2007. Parameter estimation and uncertainty analysis for a watershed model. Environmental Modelling and Software, 22(7):1000–1020.

Ganju NK, Brush MJ, Rashleigh B, Aretxabaleta AL, del Barrio P, Grear JS, Harris LA, Lake SJ, McCardell G, O'Donnell J, Ralston DK, Signell RP, Testa JM, Vaudrey JMP. 2016. Progress and challenges in coupled hydrodynamic-ecological estuarine modeling. Estuaries and Coasts, 39(2):311–332.

Hagy JD, Murrell MC. 2007. Susceptibility of a northern Gulf of Mexico estuary to hypoxia: An analysis using box models. Estuarine Coastal and Shelf Science, 74:239–253.

Hilborn R, Mangel M. 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, New Jersey.

Hodur RM. 1997. The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). Monthly Weather Review, 125:1414–1430.

Howarth RW, Billen G, Swaney D, Townsend A, Jaworski N, Lajtha K, Downing JA, Elmgren R, Caraco N, Jordan T, Berendse F, Freney J, Kudeyarov V, Murdoch P, Zhao-Liang Z. 1996. Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: natural and human influences. Biogeochemistry, 35(1):75–139.

Jr. EJR. 1996. Testing ecological models: the meaning of validation. Ecological Modelling, 90(3):229–244.

Justíc D, Legovíc T, Rottini-Sandrini L. 1987. Trends in oxygen content 1911–1984 and occurrence of benthic mortality in the northern Adriatic Sea. Estuarine, Coastal and Shelf Science, 25(4):435–445.

Kirchner JW. 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3):W03S04.

Lehrter JC, Ko DS, Lowe L, Penta B. In review. Predicted effects of climate change on the severity of northern Gulf of Mexico hypoxia. In: Justic et al., editor, Modeling Coastal Hypoxia: Numerical Simulations of Patterns, Controls, and Effect of Dissolved Oxygen Dynamics. Springer, New York.

Lehrter JC, Ko DS, Murrell MC, III JDH, Schaeffer BA, Greene RM, Gould RW, Penta B. 2013. Nutrient distributions, transports, and budgets on the inner margin of a river-dominated continental shelf. Journal of Geophysical Research, 118(10):4822–4838.

Levins R. 1966. The strategy of model building in population biology. American Scientist, 54(4):421–431.

Lipton D, Hicks R. 2003. The cost of stress: low dissolved oxygen and eocnomic benefits of recreational striped bass (*Morone saxatilis* fishing in the Patuxent River. Estuaries, 26(2A):310–315.

Lohrenz SE, Redalje DG, Cai WJ, Acker J, Dagg M. 2008. A retrospective analysis of nutrients and phytoplankton productivity in the Mississippi River plume. Continental Shelf Research, 28(12):1466–1475.

Martin PJ. 2000. Description of the navy coastal ocean model version 1.0. Technical Report NRL/FR/7322-00-9962, Naval Research Lab, Stennis Space Center, Mississippi.

Mateus MD, Franz G. 2015. Sensitivity analysis in a complex marine ecological model. Water, 7:2060–2081.

Morrison M, Morgan MS. 1999. Models as mediating agents. In: Morgan MS, Morrison M, editors, Models as Mediators, page 401. Cambridge University Press, Cambridge.

Murrell MC, Beddick DL, Devereux R, Greene RM, III JDH, Jarvis BM, Kurtz JC, Lehrter JC, Yates DF. 2014. Gulf of mexico hypoxia research program data report: 2002-2007. Technical Report EPA/600/R-13/257, US Environmental Protection Agency, Washington, DC.

Murrell MC, Stanley RS, Lehrter JC, Hagy JD. 2013. Plankton community respiration, net ecosystem metabolism, and oxygen dynamics on the Louisiana continental shelf: Implications for hypoxia. Continental Shelf Research, 52:27–38.

Nossent J, Bauwens W. 2012. Multi-variable sensitivity and identifiability analysis for a complex environmental model in view of integrated water quantity and water quality modeling. Water Science & Technology, 65(3):539–549.

Obenour DR, Michalak AM, Scavia D. 2015. Assessing biophysical controls on Gulf of Mexico hypoxia through probabilistic modeling. Ecological Applications, 25(2):492–505.

Omlin M, Brun R, Reichert P. 2001a. Biogeochemical model of Lake Zürich: sensitivity, identifiability and uncertainty analysis. Ecological Modelling, 141(1-3):105–123.

Omlin M, Reichert P, Forster R. 2001b. Biogeochemical model of Lake Zürich: model equations and results. Ecological Modelling, 141(1-3):77–103.

Paerl HW, Pinckney JL, Fear JM, Peierls BL. 1998. Ecosystem responses to internal and watershed organic matter loading: consequences for hypoxia in the eutrophying Neuse River Estuary, North Carolina, USA. Marine Ecology Progress Series, 166:17–25.

Pauer JJ, Feist TJ, Anstead AM, DePetro PA, Melendez W, Lehrter JC, Murrell MC, Zhang X, Ko DS. 2016. A modeling study examining the impact of nutrient boundaries on primary production on the Louisiana continental shelf. Ecological Modelling, 328:136–147.

Penta B, Lee Z, Kudela RM, Palacios SL, Gray DJ, Jolliff JK, Shulman IG. 2008. An underwater light attenuation scheme for marine ecosystem models. Optical Express, 16(21):16581–16591.

Penta B, Lee Z, Kudela RM, Palacios SL, Gray DJ, Jolliff JK, Shulman IG. 2009. An underwater light attenuation scheme for marine ecosystem models: errata. Optical Express, 17(25):23351–23351.

Petrucci G, Bonhomme C. 2014. The dilemma of spatial representation for urban hydrology semi-distributed modelling: trade-offs among complexity, calibration and geographical data. Journal of Hydrology, 517:997–1007.

Rabalais NN, Turner RE, Scavia D. 2002. Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi river. BioScience, 52(2):129–142.

Refsgaard JC, van der Sluijs JP, Højberg AL, Vanrolleghem PA. 2007. Uncertainty in the environmental modelling process - a framework and guidance. Environmental Modelling & Software, 22(11):1543–1556.

Scavia D, Justic D, Bierman VJ. 2004. Reducing hypoxia in the Gulf of Mexico: Advice from three models. Estuaries, 27(3):419–425.

Snowling SD, Kramer JR. 2001. Evaluating modelling uncertainty for model selection. Ecological Modelling, 138:17–30.

Soetaert K, Petzoldt T. 2010. Inverse modelling, sensitivity, and Monte Carlo analysis in R using package FME. Journal of Statistical Software, 33(3):1–28.

Wade AJ, Jackson BM, Butterfield D. 2008. Over-parameterised, uncertain 'mathematical marionettes' - how can we best use catchment water quality models? An example of an 80-year catchment-scale nutrient balance. Science of the Total Environment, 400(1-3):52–74.

Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S. 2001a. A framework for development and application of hydrological models. Hydrology and Earth System Sciences, 5(1):13–26.

Wagener T, Lees MJ, Wheater HS. 2001b. A toolkit for the development and application of parsimonious hydrological models. In: Singh VP, Meyer DKFSP, editors, Mathematical Models of Small Watershed Hydrology and Applications - Volume 2, pages 91–140. Water Resources Publications LLC, USA.

Warner JC, Geyer WR, Lerczak JA. 2005. Numerical modeling of an estuary: a comprehensive skill assessment. Journal of Geophysical Research: Oceans, 110(C5):13.

Wiseman WJ, Rabalais NN, Turner RE, Dinnel SP, MacNaughton A. 1997. Seasonal and interannual variability within the Louisiana coastal current: stratification and hypoxia. Journal of Marine Systems, 12(1-4):237–248.

Zhao L, Chen C, Vallino J, Hopkinson C, Beardsley RC, Lin H, Lerczak J. 2010. Wetland-estuarine-shelf interactions on the Plum Island Sound and Merrimack River in the Massachusetts coast. Journal of Geophysical Research, 115(C10):13.

Table 1: Sensitivities of O_2 to perturbation of optics parameters. Sensitivities are based on a 50% increase from the default parameter value, where $L1$ and $L2$ summarize differences in model output from the default (see eqs. (2) and (3)). Parameters that did not affect O_2 are not shown.

Description	Parameter	Value	L1	L2
Chla specific absorption at 490 nm	*astar490*	0.04	0.43	1.95
OMA specific absorption at 490 nm	*astarOMA*	0.1	0.09	0.4
OMZ specific absorption at 490 nm	*astarOMZ*	0.1	6.78×10^{-4}	0

Table 2: Sensitivities of O_2 to perturbation of organic matter parameters. Sensitivities are based on a 50% increase from the default parameter value, where $L1$ and $L2$ summarize differences in model output from the default (see eqs. (2) and (3)). Parameters that did not affect O_2 are not shown.

Description	Parameter	Value	L1	L2
turnover rate for OM2A and OM2G	*KG2*	50	1.24	6.02
turnover rate for OM1A and OM1G	*KG1*	50	0.92	4.62
half-saturation concentration for O2 utilization	*KO2*	10	0.35	1.57
O2 concentration that inhibits denitrification	*KstarO2*	10	0.06	0.34
maximum rate of nitrification per day	*nitmax*	0.52	0.05	0.21
NH4 rate constant for nitrification	*KNH4*	1	0.04	0.17
decay rate of CDOM, 1/day	*KGcdom*	0.01	0.03	0.12
half-saturation concentration for NO3 used in denitrification	*KNO3*	10	0.01	0.05

Table 3: Sensitivities of O_2 to perturbation of phytoplankton parameters. Sensitivities are based on a 50% increase from the default parameter value, where $L1$ and $L2$ summarize differences in model output from the default (see eqs. (2) and (3)). Parameters that did not affect O_2 are not shown. Subscripts show the phytoplankton or zooplankton group that applies for the parameter. Parameters less than the 75th percentile (0.59) for L1 were removed for brevity.

Description, Parameter	Value	L1	L2
phytoplankton carbon/cell			
Qc_{p1}	1.35×10^{-6}	0.65	3.16
Qc_{p2}	1.35×10^{-6}	0.65	3.16
Qc_{p3}	1.35×10^{-6}	0.65	3.16
Qc_{p4}	1.35×10^{-6}	0.65	3.16
Qc_{p5}	1.35×10^{-6}	0.65	3.16
Qc_{p6}	1.35×10^{-6}	0.65	3.16

Table 4: Sensitivities of O_2 to perturbation of temperature parameters. Sensitivities are based on a 50% increase from the default parameter value, where $L1$ and $L2$ summarize differences in model output from the default (see eqs. (2) and (3)). Parameters that did not affect O_2 are not shown. Subscripts show the phytoplankton or zooplankton group that applies for the parameter.

Description, Parameter	Value	L1	L2
optimum temperature for growth(c)			
Tref(nospA+nospZ)$_{p4}$	22	9.97×10^{-5}	1.46×10^{-4}
Tref(nospA+nospZ)$_{p2}$	22	9.96×10^{-5}	1.46×10^{-4}
Tref(nospA+nospZ)$_{p1}$	22	9.96×10^{-5}	1.46×10^{-4}
Tref(nospA+nospZ)$_{p3}$	22	9.86×10^{-5}	1.45×10^{-4}
Tref(nospA+nospZ)$_{p5}$	22	9.86×10^{-5}	1.45×10^{-4}
Tref(nospA+nospZ)$_{p6}$	22	9.75×10^{-5}	1.43×10^{-4}

Table 5: Sensitivities of O_2 to perturbation of zooplankton parameters. Sensitivities are based on a 50% increase from the default parameter value, where $L1$ and $L2$ summarize differences in model output from the default (see eqs. (2) and (3)). Parameters that did not affect O_2 are not shown. Subscripts show the phytoplankton or zooplankton group that applies for the parameter.

Description, Parameter	Value	L1	L2
zooplankton biomass-dependent respiration factor			
Zrespb$_{z1}$	0.1	0	0.01
Zrespb$_{z2}$	0.1	0	0.01
zooplankton carbon/individual			
ZQc$_{z1}$	3.13×10^{-4}	0.01	0.04
ZQc$_{z2}$	3.13×10^{-4}	0.01	0.04
zooplankton mortality constant for quadratic mortality			
Zm$_{z1}$	7.2×10^{-4}	4.95×10^{-4}	0
Zm$_{z2}$	7.2×10^{-4}	4.95×10^{-4}	0
zooplankton nitrogen/individual			
ZQn$_{z1}$	6.95×10^{-5}	0	0.01
ZQn$_{z2}$	6.95×10^{-5}	0	0.01
zooplankton phosphorus/individual			
ZQp$_{z1}$	3.77×10^{-6}	5.75×10^{-5}	2.66×10^{-4}
ZQp$_{z2}$	3.77×10^{-6}	5.75×10^{-5}	2.66×10^{-4}

Table 6: Parameter identifiability for dissolved oxygen. Selection followed the second heuristic where parameters were selected by decreasing sensitivity up to maximum $\gamma < 15$. Rank describes the parameter sensitivity in each category (O: optics, OM: organic matter, P: phytoplankton, T: temperature, Z: zooplankton).

Parameter description	Parameter	Rank	L1	γ
turnover rate for OM2A and OM2G	*KG2*	1^{OM}	1.24	1
turnover rate for OM1A and OM1G	*KG1*	2^{OM}	0.92	8.48

Table 7: Parameter identifiability for ammonium, chlorophyll, irradiance, nitrate, OM1, OM2, and phosphate. Selection followed the second heuristic where parameters were selected by decreasing sensitivity up to maximum $\gamma < 15$. Rank describes the parameter sensitivity in each category (O: optics, OM: organic matter, P: phytoplankton, T: temperature, Z: zooplankton).

State variable, parameter description	Parameter	Rank	L1	γ
Ammonium				
NH4 rate constant for nitrification	*KNH4*	1^{OM}	0.81	1
Chlorophyll				
phytoplankton carbon/cell	Qc_{p5}	1^{P}	0.08	1
Irradiance				
OMA specific absorption at 490 nm	*astarOMA*	1^{O}	5.86×10^{14}	1
decay rate of CDOM, 1/day	*KGcdom*	1^{OM}	2.66×10^{14}	3.92
Nitrate				
turnover rate for OM2A and OM2G	*KG2*	1^{OM}	0.85	1
OM1				
phytoplankton carbon/cell	Qc_{p1}	3.5^{P}	43.06	1
OM2				
turnover rate for OM1A and OM1G	*KG1*	1^{OM}	35.62	1
phytoplankton carbon/cell	Qc_{p1}	3.5^{P}	33.53	10.36
Phosphate				
turnover rate for OM2A and OM2G	*KG2*	1^{OM}	0.21	1

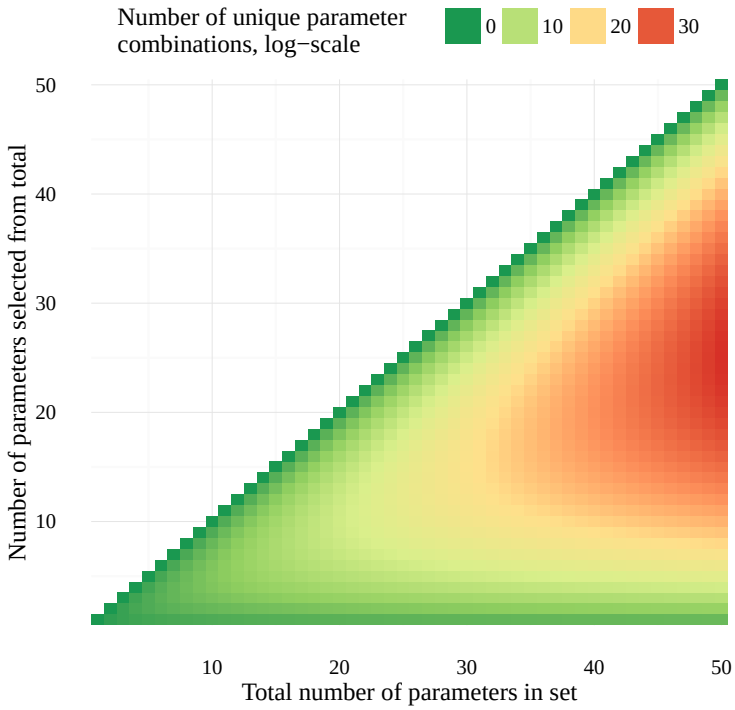

Fig. 1: Examples of unique parameter combinations from different parameter sets and number of selected parameters. The number of combinations are shown for increasing numbers of selected parameters from the total in the set, where 50 parameter sets are shown each with one through 50 total parameters. Note that the number of unique combinations is shown as the natural-log.

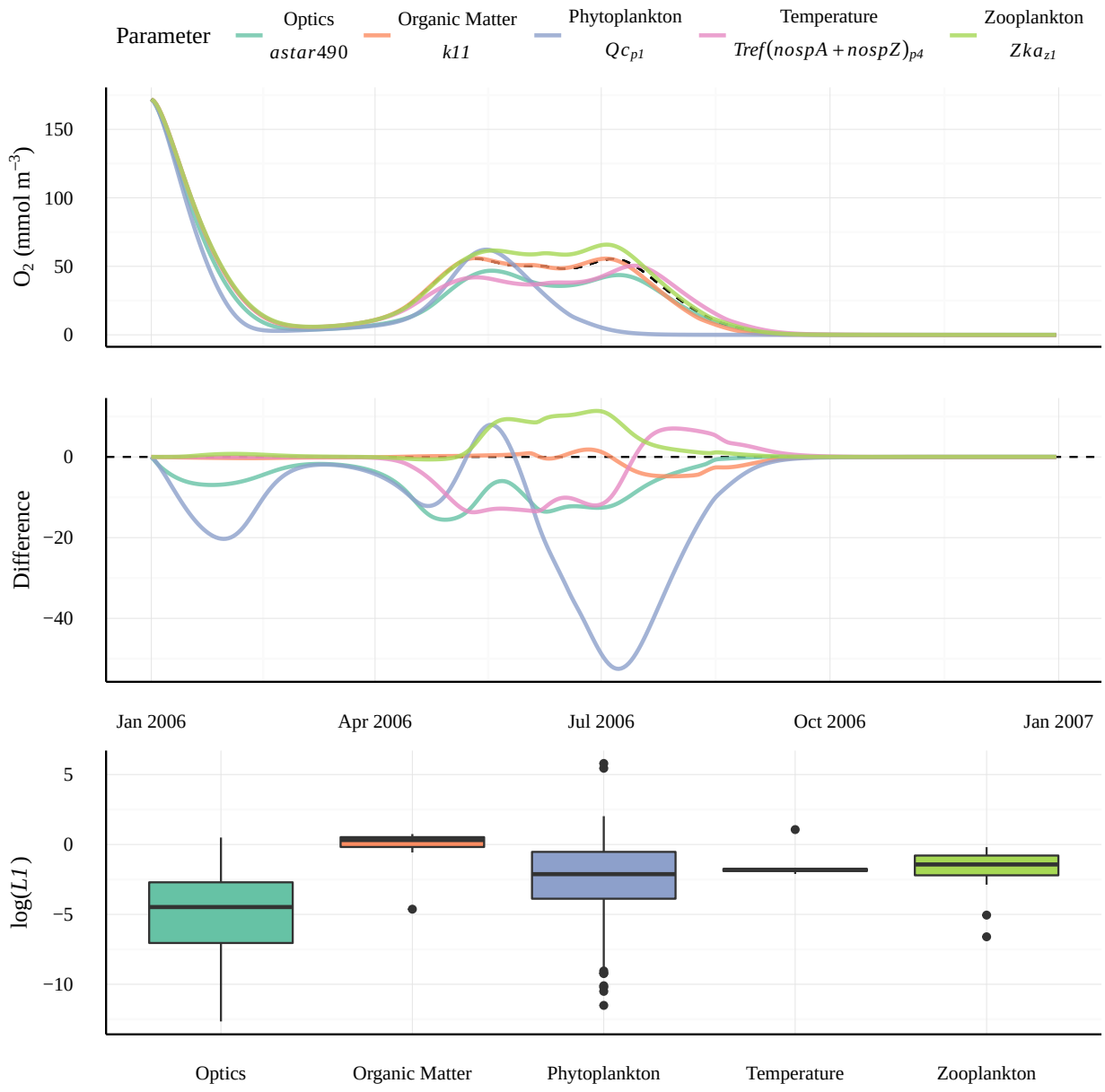

Fig. 2: Sensitivity of O_2 to parameter changes. The solid lines show the change in O_2 based on a 50% change from the default parameter values (dashed line) for each parameter. Individual parameters with the largest effect are shown for each category. The top plot shows the model output and the middle plot shows the estimated O_2 as a difference from the default. The bottom plot shows the distribution of error values (as $log\,(L1)$) for all parmaters in each category.

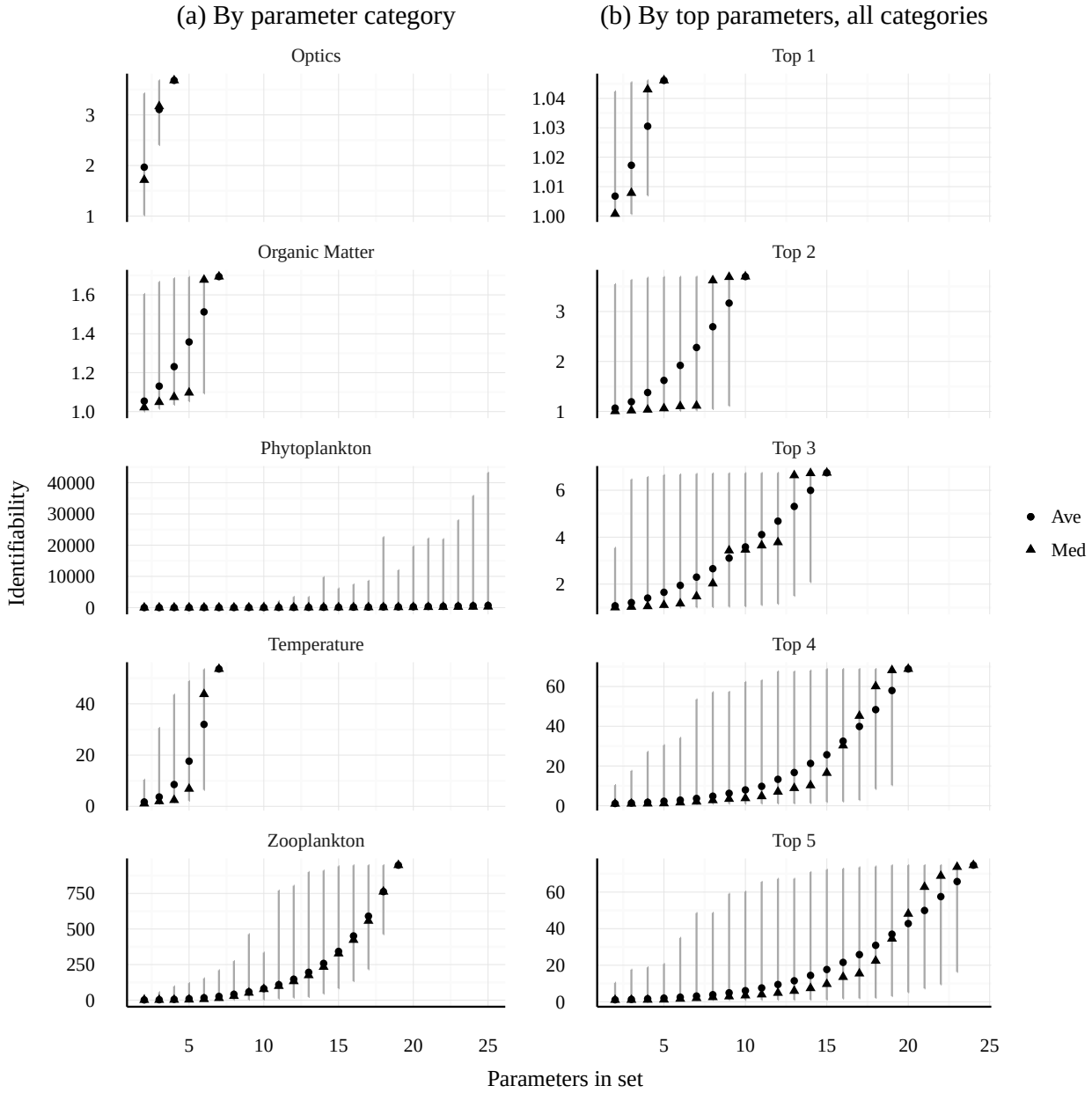

Fig. 3: Identifiability (as γ, eq. (4)) of parameter subsets for O_2. Plots in (a) show identifiability by parameter categories and (b) shows identifiability by selecting the top 1 through 5 parameters regardless of category. Lines represent identifiability ranges for the possible combinations given the number of parameters in the set. The phytoplankton category is limited to 25 total parameters.

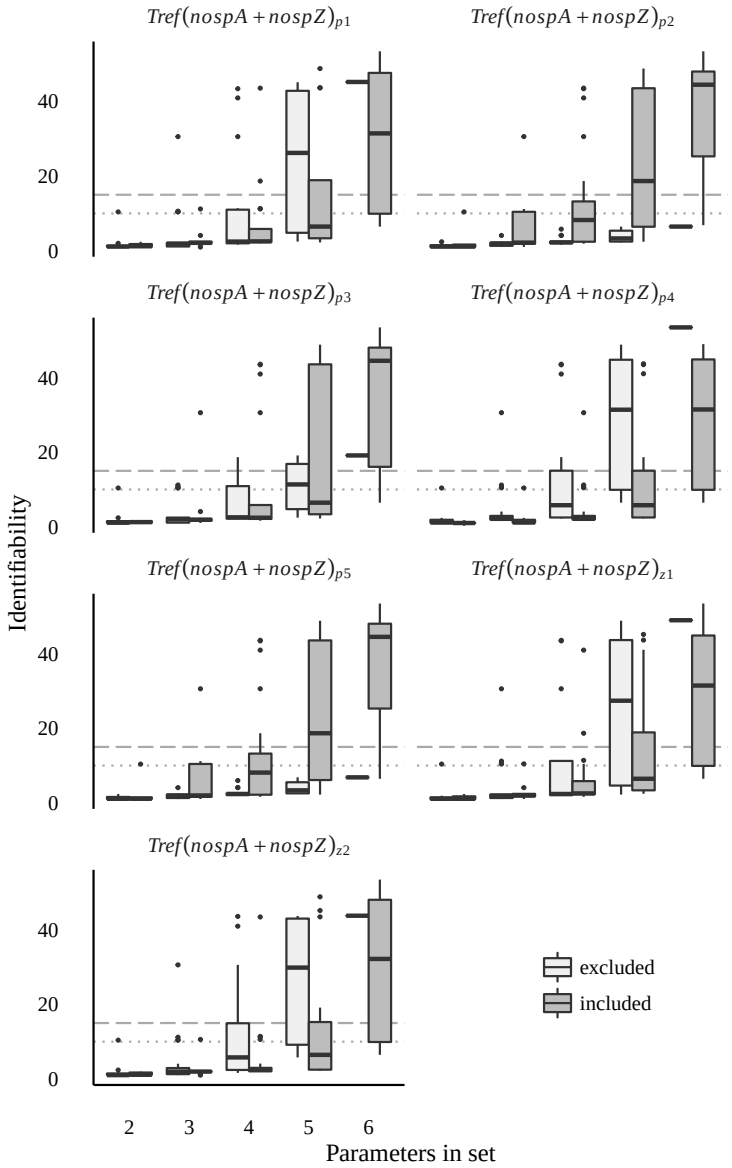

Fig. 4: Identifiability (as γ, eq. (4)) of temperature parameters for subset combinations in Fig. 3. Identifiability is evaluated for subsets that excluded and included the parameters at the top of each plot. Identifiability of including all seven parameters is in Fig. 3. Grey lines indicate potential thresholds at $\gamma = 10, 15$ for maximum acceptable identifiability.

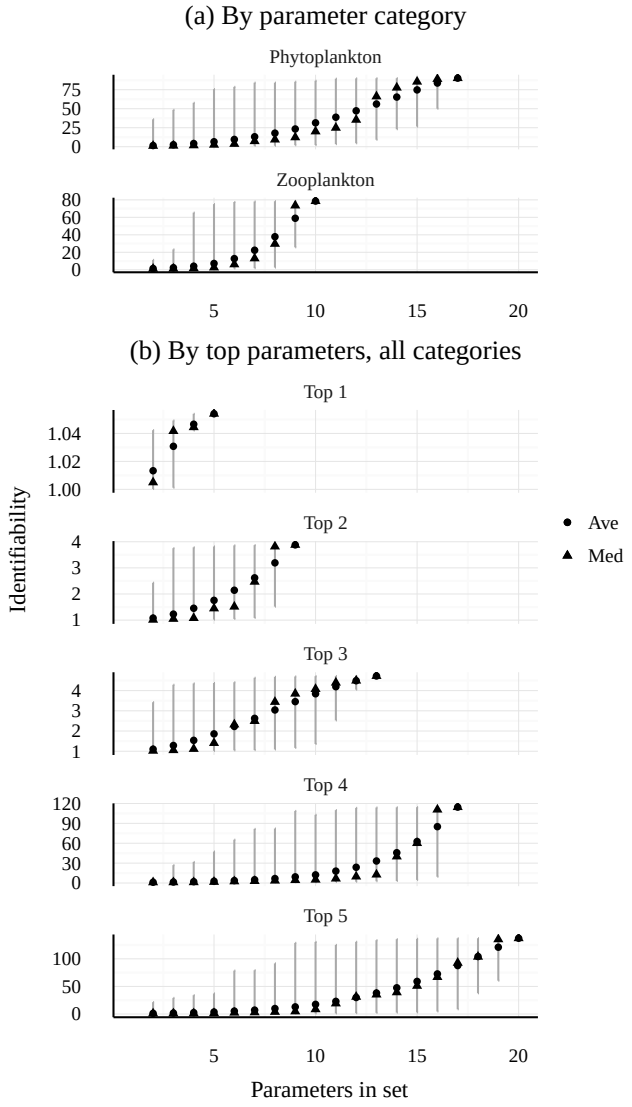

Fig. 5: Identifiability (as γ, eq. (4)) of parameter subsets for O_2 using only subsets from the first phytoplankon group and first zooplankton group. Plots in (a) show identifiability for only the phytoplankton and zooplankton categories and (b) shows identifiability by selecting the top 1 through 5 parameters regardless of category. Lines represent identifiability ranges for the possible combinations given the number of parameters in the set.

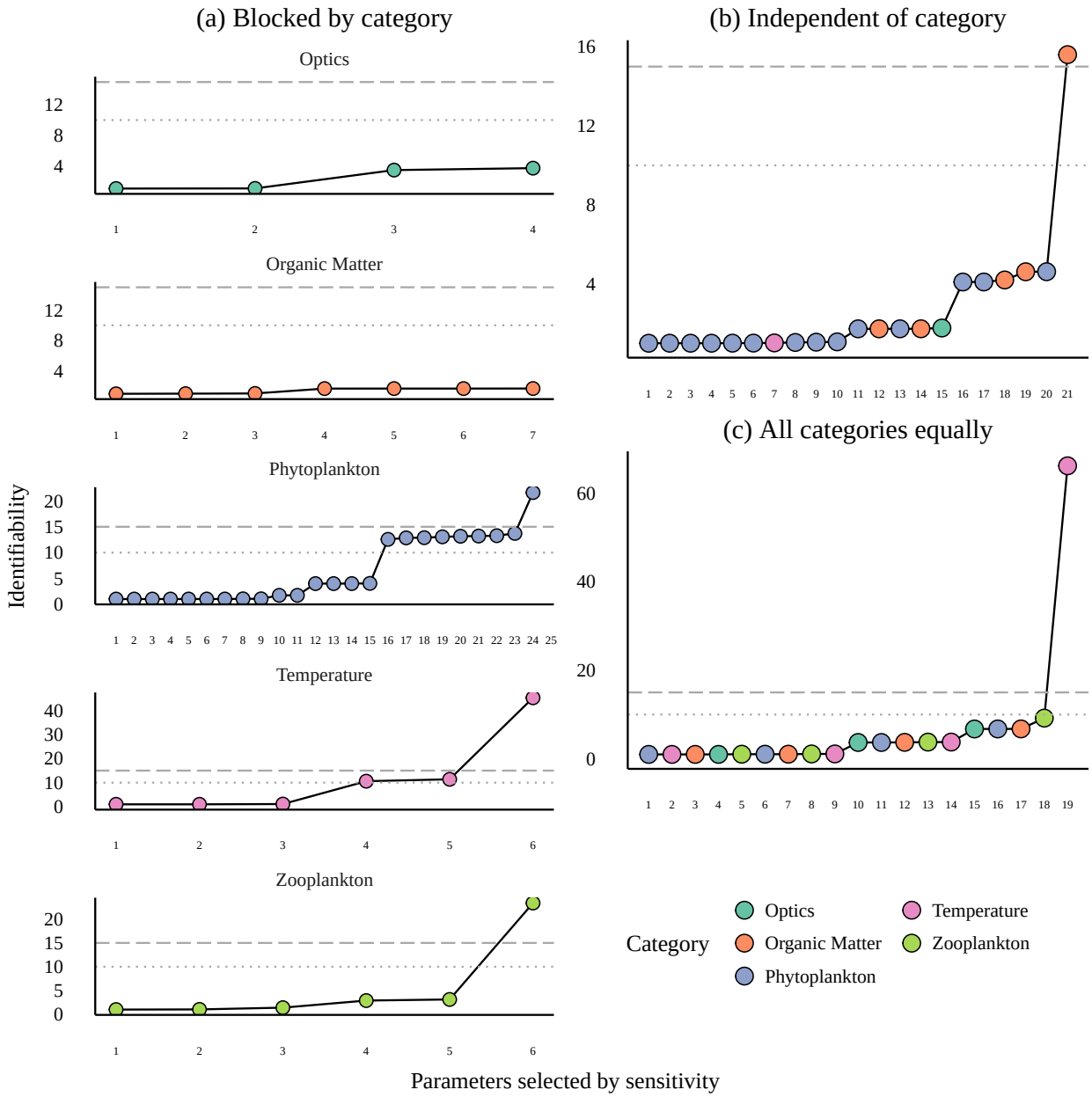

Fig. 6: Identifiability (as γ, eq. (4)) of selecting parameters with three different heuristics. Parameters are selected by decreasing sensitivity for all examples (Tables 1 to 5). The parameter selecions are blocked within each category (a), independent of category (b), or considering all categories equally (c). Grey lines indicate potential thresholds at $\gamma = 10, 15$ for maximum acceptable identifiability. Selection stops after $\gamma > 15$ or if the maximum number of possible parameters is selected. See Table 6 for values in (b).

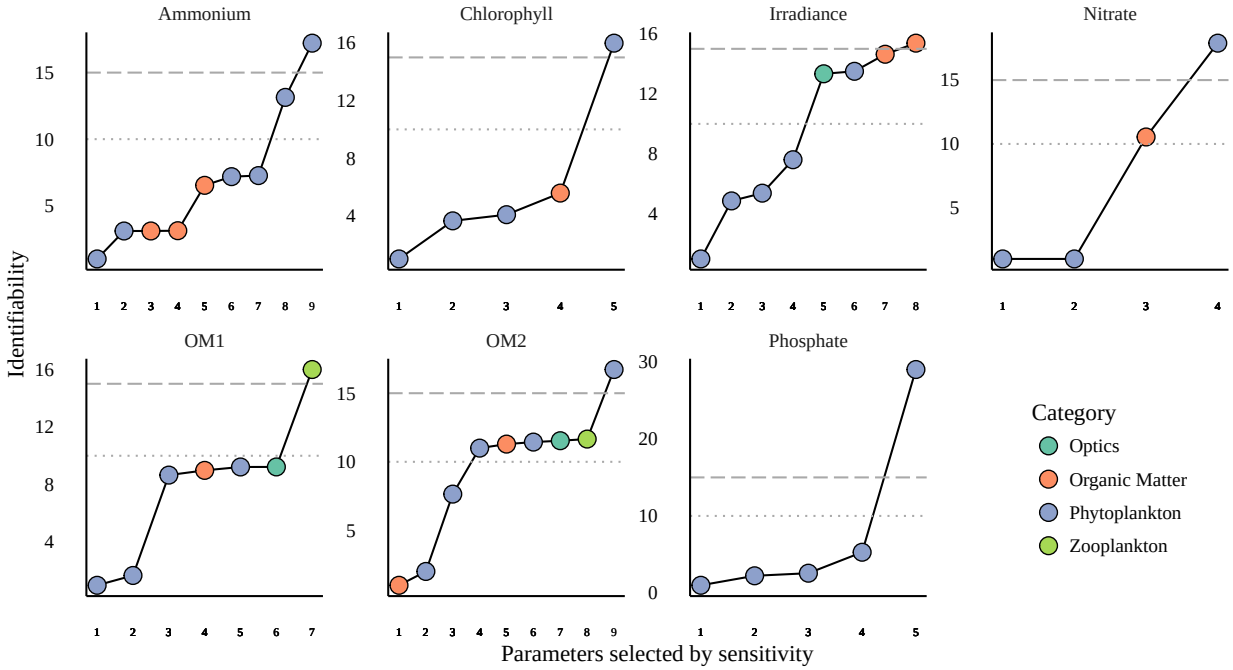

Fig. 7: Identifiability (as γ, eq. (4)) of selecting parameters for selected state variables. Parameters are selected by decreasing sensitivity indepent of parameter categories. Grey lines indicate potential thresholds at $\gamma = 10, 15$ for maximum acceptable identifiability. Selection stops after $\gamma > 15$. See Table 7 for values.

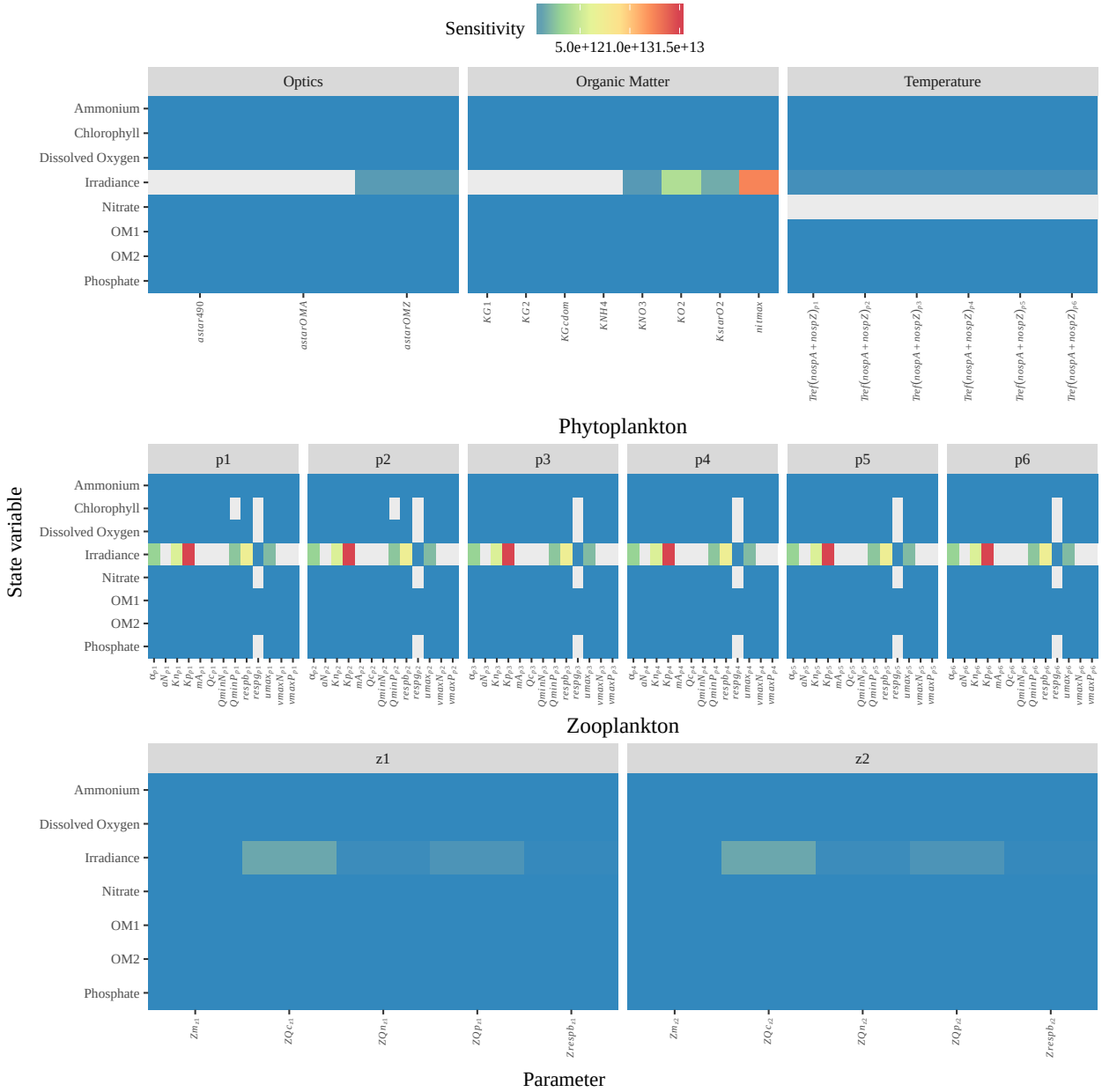

Fig. S1: Sensitivity values (L1, eq. (2)) for local analyses of all state variables. Values are truncated by the 5th and 95th sensitivity values between all parameters for visual appearance. Parameters are grouped by category (plot strips). Phytoplankton and zooplankton categories are further separated into one through six, and one and two groups (p1-p6, z1, z2), respectively.