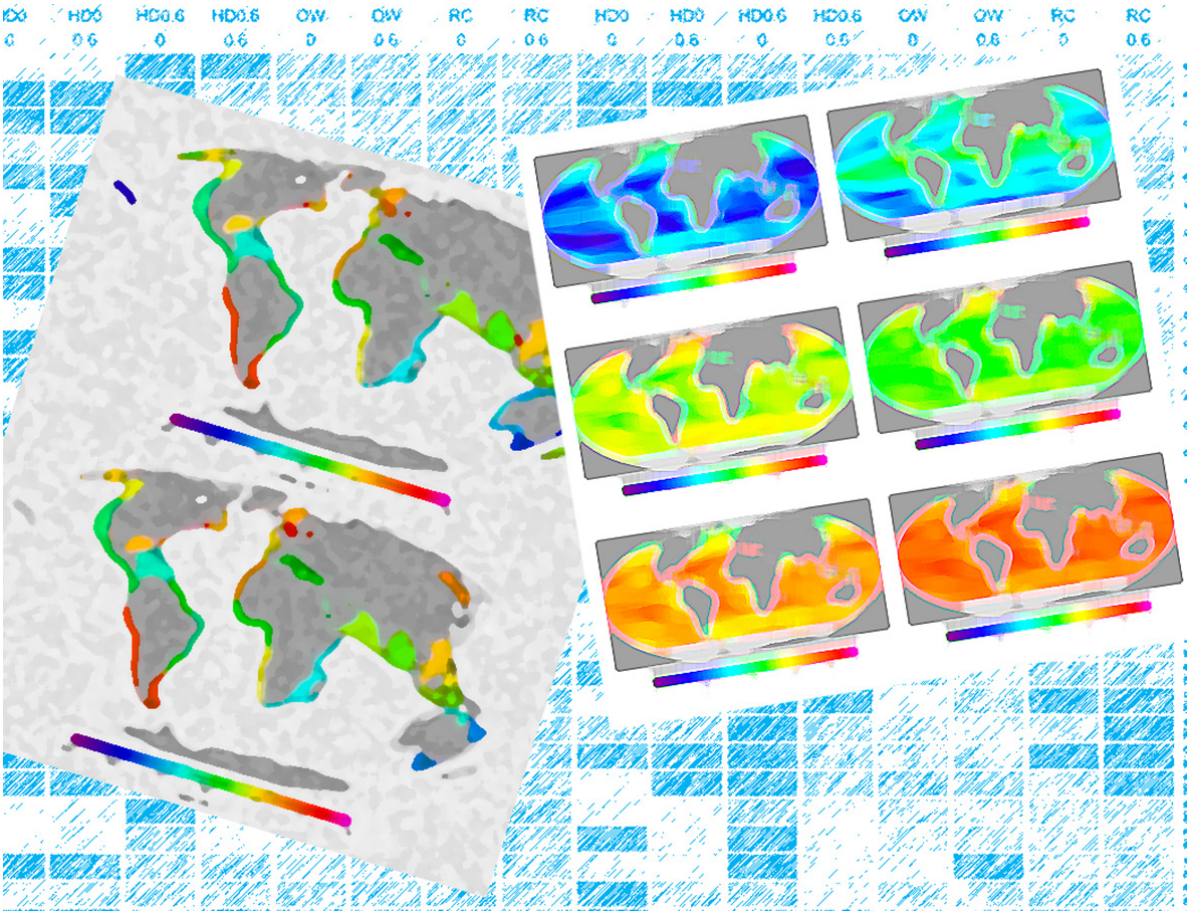
FAO Fisheries and Aquaculture Circular No. 1086 **FIRF/C1086 (En)**

ISSN 2070-6065

DEVELOPING NEW APPROACHES TO GLOBAL STOCK STATUS ASSESSMENT AND FISHERY PRODUCTION POTENTIAL OF THE SEAS



Copies of FAO publications can be requested from:

Sales and Marketing Group
FAO, Viale delle Terme di Caracalla
00153 Rome, Italy
E-mail: publications-sales@fao.org
Fax: +39 06 57053360
Website: www.fao.org/icatalog/inter-e.htm

DEVELOPING NEW APPROACHES TO GLOBAL STOCK STATUS ASSESSMENT AND FISHERY PRODUCTION POTENTIAL OF THE SEAS

Andrew A. Rosenberg
UCS, United States of America

Michael J. Fogarty
NFSC, NMFS, NOAA, United States of America

Andrew B. Cooper
SFU, Canada

Mark Dickey-Collas
ICES, Denmark

Elizabeth A. Fulton
CSIRO, Australia

Nicolás L. Gutiérrez
MSC, United Kingdom

Kimberly J.W. Hyde
NMFS, NOAA, United States of America

Kristin M. Kleisner
Sea Around Us Project, UBC, Canada

Trond Kristiansen
IMR, Norway

Catherine Longo
NCEAS, United States of America

Carolina V. Minte-Vera
UEMa Brazil, and IATTC, United States of America

Cóilín Minto
GMIT, Ireland

Iago Mosqueira
EC JRC, IPSC, MAU, Italy

Giacomo Chato Osio
EC JRC, IPSC, MAU, Italy

Daniel Ovando
SFG, UC, United States of America

Elizabeth R. Selig
Gordon and Betty Moore Center for Science and Oceans, CI, United States of America

James T. Thorson
FRMD, NFSCm NMFS, NOAA, United States of America

Yimin Ye
FAO, Italy

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 2014

ISBN 978-92-5-107992-8 (print)
E-ISBN 978-92-5-107993-5 (PDF)

PREPARATION OF THIS DOCUMENT

FAO has been monitoring the state of the world's marine fish stocks since 1974, and it periodically produces the most authoritative report on the subject – *The State of World Fisheries and Aquaculture*. Information on the state of fishery sustainability is not only important for policy formulation, but also crucial to guide the fishing industry and its managers to develop effective harvest strategies. Moreover, sustainable fisheries require healthy ecosystems. To monitor ecosystem health, it is necessary to conduct ecosystem-level assessments that take into consideration both targeted and non-targeted species, interspecies interactions, and other factors that cannot be determined by looking at each stock in isolation. With these objectives, the Fisheries and Aquaculture Department of the FAO commissioned a study – Developing New Approaches to Global Stock Status Assessment and Fishery Production Potential of the Seas.

This circular presents the results of the study. It consists of two parts. Part 1 focuses on determining single-stock status and summarizes the results of simulation testing of four methods that can be applied to data-poor fisheries. Part 2 reports the results on the estimation of ecosystem-level production potentials based on satellite-based estimates of primary productivity.

Eighteen scientists around the world participated in this study: Andrew A. Rosenberg, Union of Concerned Scientists, United States of America; Michael J. Fogarty, Northeast Fisheries Science Center, National Marine Fisheries Services, National Oceanic and Atmospheric Administration, United States of America; Andrew B. Cooper, Simon Fraser University, Canada; Mark Dickey-Collas, International Council for the Exploration of the Sea, Denmark; Elizabeth A. Fulton, CSIRO, Australia; Nicolás L. Gutiérrez, Marine Stewardship Council, United Kingdom; Kimberly J.W. Hyde, National Marine Fisheries Services, National Oceanic and Atmospheric Administration, United States of America; Kristin M. Kleisner, Sea Around Us Project, University of British Columbia, Canada; Trond Kristiansen, Institute of Marine Research, Norway; Catherine Longo, National Center for Ecological Analysis and Synthesis, United States of America; Carolina V. Minte-Vera, Universidade Estadual de Maringá, Brazil, and Inter-American Tropical Tuna Commission, United States of America; Cóilín Minto, Galway-Mayo Institute of Technology, Ireland; Iago Mosqueira, European Commission Joint Research Centre, Institute for Protection and Security of the Citizen, Maritime Affairs Unit, Italy; Giacomo Chato Osio, European Commission Joint Research Centre, Institute for Protection and Security of the Citizen, Maritime Affairs Unit, Italy; Daniel Ovando, Sustainable Fisheries Group, University of California, Santa Barbara, United States of America; Elizabeth R. Selig, Betty and Gordon Moore Center for Science and Oceans, Conservation International, United States of America; James T. Thorson, Fisheries Resource and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, United States of America; Yimin Ye, Food and Agriculture Organization of the United Nations, Italy.

Rosenberg, A.A., Fogarty, M.J., Cooper, A.B., Dickey-Collas, M., Fulton, E.A., Gutiérrez, N.L., Hyde, K.J.W., Kleisner, K.M., Kristiansen, T., Longo, C., Minte-Vera, C., Minto, C., Mosqueira, I., Chato Osio, G., Ovando, D., Selig, E.R., Thorson, J.T. & Ye, Y. 2014. *Developing new approaches to global stock status assessment and fishery production potential of the seas.* FAO Fisheries and Aquaculture Circular No. 1086. Rome, FAO. 175 pp.

ABSTRACT

Stock status is a key parameter for evaluating the sustainability of fishery resources and developing corresponding management plans. However, the majority of stocks are not assessed, often as a result of insufficient data and a lack of resources needed to execute formal stock assessments. The working group involved in this publication focused on two approaches to estimating fisheries status: one based on single-stock status, and the other based on ecosystem production.

For the single-stock status work, a fully factorial simulation testing framework was developed to assess four potential data-limited models. The results suggest that Catch-MSY, a catch-based method, was the best performer, although the different models performed similarly in many cases. Catch-MSY was more effective in estimating status over short time scales and could be particularly applicable for use in developing countries where data time series are often shorter. Harvest dynamics was the most important explanatory variable in determining performance, which emphasizes the importance of having accurate information on fishing effort and total removals.

For the ecosystem-level production analysis, the working group used satellite-based estimates of primary productivity by size classes and a more complete food web, which included more complete microbial pathways than earlier approaches. The working group also assembled estimates of ecological transfer efficiencies from a large number of energy flow network models to characterize uncertainty. The first-order estimates of fishery production potential indicated a potential yield of up to 180 million tonnes of fish, which could vary depending on the capacity to sustainably diversify the suite of species that are currently exploited. Planktivorous species provide the largest scope for growth. However, consideration of factors such as the ecological impact on other food web components, profitability of harvest operations, and marketability for these species must first be resolved. The realized production potential for planktivores may be much lower than their potential levels depending on the outcome of these considerations. The working group estimated that up to 50 million tonnes of benthic production could be potentially harvested, although this estimate is subject to similar constraints as those for planktivores. The greatest scope for growth in the benthic component may be found in the mariculture sector, subject to suitable environmental safeguards.

Ecosystem exploitation rates should not exceed 20–25 percent of available production, considering basic energetic constraints in marine ecosystems. Current harvest levels for benthivorous and piscivorous species (principally fish) exceeded these levels in higher-latitude ecosystems (subarctic-boreal and temperate) and were near or slightly below them in lower latitudes and upwelling systems. The estimates of the ratio of current catches to available production for planktivorous species are substantially lower, reflecting the production potential of currently underutilized species. However, targeted harvesting of selected planktivorous species does lead to relatively high exploitation rates for some species. Together, these results provide globally applicable methods for estimating fish stock status and fishery production potential.

CONTENTS

TABLES

FIGURES

ACKNOWLEDGEMENTS

The working group members would like to thank their respective organizations for supporting their work on the analyses and preparation of this report. These organizations include the Commonwealth Scientific and Industrial Research Organisation, EC Joint Research Center, Galway-Mayo Institute of Technology, Institute of Marine Research – Bergen, Norway, Inter-American Tropical Tuna Commission, International Council for the Exploration of the Sea, Marine Stewardship Council, National Center for Ecological Analysis and Synthesis, National Oceanic and Atmospheric Administration (NOAA) Northwest and Northeast Fisheries Science Centers and National Marine Fisheries Service, Oceanographic Institute at the University of São Paulo, Simon Fraser University, Universidade Estadual de Maringá, Sea Around Us Project at University of British Columbia, University of California – Santa Barbara, Union of Concerned Scientists, and Conservation International.

Special thanks go to Trevor Branch (University of Washington), Ray Hilborn (University of Washington), Hiroyuki Kurota (Seikai National Fisheries Research Institute, Fisheries Research Agency), Meredith Lopuch (Gordon and Betty Moore Foundation) and James Rising (Columbia University) for attending working group meetings and providing valuable feedback.

Gratitude also goes to the Norwegian Metacenter for Computational Science for its technical and logistical support during the working group's use of the supercomputer Hexagon for conducting the analyses.

Reg Watson (IMAS, UTAS) and Tilla Roy (LOCEAN-IPSL, Université Pierre et Marie Curie) are acknowledged for their help in calculating transfer efficiency statistics, as are Villy Christensen, Joe Buszkowski and Jeroen Steenbeek of the Fisheries Center, University of British Columbia, Canada, for their assistance in extracting Ecopath statistics.

Robert Gamble (NOAA Northeast Fisheries Science Center) assisted with the ecosystem production potential simulations, and Mariano Koen-Alonso (Northwest Atlantic Fisheries Centre, Fisheries and Oceans Canada) provided the WinBugs code adapted for those analyses.

The working group would also like to thank the University of Washington, the EC Joint Research Center, the New England Aquarium and the Marine Biological Laboratory for hosting working group meetings.

The working group was generously funded by FAO and by the Gordon and Betty Moore Foundation.

ABBREVIATIONS AND ACRONYMS

APE	absolute proportional error
AR	autoregression
BMSY	biomass at maximum sustainable yield
CMSY	catch-MSY
COM	catch only model
COM-SIR	catch only model – sampling importance resampling
DB-SRA	depletion-based stock reduction analysis
DCAC	depletion-corrected average catch
EwE	Ecopath with Ecosim
HD	harvest dynamics
ICES	International Council for the Exploration of the Sea
ID	initial depletion
LH	life history
LME	large marine ecosystem
MAPE	mean absolute proportional error
MPE	mean proportional error
mPRM	modified panel regression model
MSD	maximum single density
MSY	maximum sustainable yield
MTL	mean trophic level
NES LME	Northeast US Continental Shelf LME
OW	one-way trip (a type of harvest dynamic)
PE	proportional error
PPR	primary production required
PRM	panel regression model
RC	roller coaster (a type of harvest dynamic)
SIR	sampling importance resampling
SSCOM	state–space catch only model
TL	trophic level
TS	time series length
VGPM	Vertically Generalized Productivity Model

OVERALL INTRODUCTION

Wild-capture fisheries provide a critical source of nutritional and economic benefits to people worldwide. In 2010, fisheries generated livelihoods and income for almost 38.5 million people (FAO, 2012) and currently fish provide approximately 3 billion people with almost 20 percent of their intake of animal protein. In the last half century, marine fisheries have been rapidly expanding and developing (Swartz *et al.*, 2010). Fishing fleets have also been increasing, both in number and extent, since the 1970s (Anticamara *et al.*, 2011; Watson *et al.*, 2013), although this growth has stabilized in the last decade (FAO, 2010). Concurrently, total landings increased from 16.8 million tonnes in 1950 to a peak of 86.4 million tonnes in 1996, but subsequently declined to 77.4 million tonnes in 2010 (FAO, 2012). With coastal populations projected to grow by 35 percent in the next 20 years, the demand for fisheries resources is likely to continue to increase.[1] The combined intensification in both pressures on and demand for fisheries resources necessitates a broad understanding of the state of global fisheries to support policy formulation and the development of effective marine management.

In spite of their importance, it remains a major challenge to determine the status and potential production of wild-capture fish stocks. Managers and policy-makers need information on individual fish stocks to evaluate their status so that effective management strategies can be developed. At the same time, it is also necessary to undertake ecosystem-scale assessments that account for the interactions between stocks, the impact of fishing on non-target fish, and other factors that cannot be determined by looking at each stock in isolation.

Costello *et al.* (2012) estimated that more than 80 percent of the global catch comes from stocks that have not been formally assessed. Formal stock assessments require substantial data and resources to complete. Therefore, data-limited approaches are needed to assess the status of global fish stocks and to develop benchmarks for the fishery production potential of the oceans. The working group addressed these challenges using two approaches to estimate fisheries status: one based on single-stock status, and the other based on ecosystem production. The single-species work stream focused on evaluating the operational performance of different methods for estimating stock status within a simulation framework to evaluate their performance robustly. This simulation framework can also be used to examine the performance of other data-limited and data-rich approaches. The ecosystem production work stream was tasked with developing estimates of fishery production for each large marine ecosystem (LME) and FAO statistical area based on overall primary production in each area. This information allows for the extracted production to be compared with the estimated total production in an LME or FAO area, which is useful for developing food security policies, for effective marine stewardship, and for understanding the potential gains in fishery production from enhanced ocean management. Results from both work streams can be used to compare current exploitation rates with estimated fishery production potential.

There is always a trade-off between risk and exploitation, and this study provides a suite of methods for evaluating fish stocks at greatest risk so that they can be prioritized for management and increased data collection. Estimating stock status and identifying regions that may be at risk for overexploitation are key components of moving towards ensuring sustainable exploitation. The work described in this report is an important step in investigating the performance of methods that can be used to estimate stock status. The results are not intended to provide direct advice to motivate management measures on specific fisheries, but to give an indication of the health of fish stocks and their production potential.

The approaches from the two work streams provide a more quantitative and consistent basis for evaluating global fish stock status than has previously been available. These estimates are vital for efforts to assess the health of marine ecosystems globally under data-limited situations.

[1] www.earth.columbia.edu/news/2006/story07-11-06.php

PART 1

DETERMINING SINGLE-STOCK STATUS

1. INTRODUCTION

Managers and policy-makers need information on the status of individual fish stocks in order to manage marine fisheries resources sustainably, implement rebuilding plans for overfished species and increase production where possible. Formal stock assessments, often considered to be the gold standard in fisheries science, are available for a relatively small proportion of global stocks. Assessed stocks account for about 16 percent of harvested fish taxa (Ricard *et al.*, 2012), although the proportion of stocks assessed is likely to be lower for developing countries (Mora *et al.*, 2009). These assessments use all available data (e.g. catches, size and age distributions, surveys and tagging information) to quantify the rate of exploitation (F) in relation to that which is considered sustainable (F_{MSY}) and the relationship between historical and current stock biomass and the biomass that can produce maximum sustainable yield (MSY) (Branch *et al.*, 2011). This biomass ratio is commonly referred to as B/B_{MSY}.

In order to assess the status of fish stocks at the global level, FAO uses a combination of quantitative (formal) and qualitative stock assessments, using available information such as catch, abundance indices, spawning potential and age and size composition (FAO, 2012). In some cases, numerous types of data of varying quality are used for these assessments, but sometimes the only information that may be available is catch data (Branch *et al.*, 2011). These FAO assessments, which have been applied to 445 fish stocks since 1974, revealed that 30 percent of marine capture fisheries were overexploited and 57 percent of stocks were fully exploited in 2009 (FAO, 2012). Other research has estimated that 63 percent of assessed stocks require rebuilding to B_{MSY}; therefore, greater efforts are needed to improve the health of fisheries (Worm *et al.*, 2009). In general, these global assessments provide an important overall picture of the health of fish stocks, but they are based only on a limited number of stocks. In some cases, these assessments do not provide target or limit reference points that can be used for management. However, both the formal assessment methods and the FAO assessments still omit many small stocks, many of which are vital for food security, especially in developing countries and small island nations.

The majority of commercially exploited species have never been assessed and no reference points have been established for them. Most methods for calculating stock status in data-limited fisheries rely solely on catch data. There has been considerable controversy over the use of catch data to estimate stock status for unassessed fisheries (Branch *et al.*, 2011; Pauly, Hilborn and Branch, 2013). Nonetheless, some studies show that small, unassessed stocks may be in poorer condition than suggested by global estimates of fisheries status, based largely on assessed stocks (Costello *et al.*, 2012; Froese *et al.*, 2012). Although formal stock assessments remain the standard for determining stock status and exploitation rates that can be used to inform management action, they will continue to be unfeasible for many of the world's fisheries because of the data and technical capacity required.

Determining stock status typically requires time-series information on historical removals (e.g. catch and discards), information on trends in abundance (e.g. catch per unit effort) and assumptions about the underlying processes that regulate or affect fish stocks (e.g. a production function such as a Schaefer production model, recruitment and/or assumptions about the economic drivers of fisheries). Only landings data exist for many data-limited stocks, which require additional assumptions, information and methods in order to estimate stock status.

There are both mechanistic and non-mechanistic methods that use only catch data to obtain a picture of stock status. Non-mechanistic approaches to assessing stock status include stock status plots, which use catch time series to assign development stages to individual stocks based on catch levels in relation to the maximum or peak catch of the time series (e.g. Froese and Kesner-Reyes, 2002; Pauly, 2007; Kleisner *et al.*, 2013). However, these methods have been criticized for their lack of mechanistic underpinnings (Branch *et al.*, 2011). In the United States of America, Congress tasked the National Marine Fisheries Service with the setting of annual catch limits and accountability measures for each managed fishery by fishing year 2010 for all stocks experiencing overfishing and by fishing year 2011 for all other stocks in the fishery (Berkson *et al.*, 2011). This mandate affected both data-rich stocks for which traditional stock assessments could be conducted as well as data-limited stocks.

As a result, methods with a more mechanistic underpinning were developed to be applied to data-limited stocks. These methods included depletion-corrected average catch (DCAC; MacCall, 2009) and depletion-based stock reduction analysis (DB-SRA; Dick and MacCall, 2011). The DCAC method requires: (i) average catch over some period; (ii) an estimate of natural mortality; (iii) an estimate of the ratio of fishing mortality at MSY (F_{MSY}); and (iv) an estimate of the depletion over the catch time series. The DB-SRA approach requires similar inputs but needs a complete time series of catch and an estimate of the age of maturity. While these methods are considered to be data-limited, some of the inputs required for these methods would be too difficult to obtain for the majority of catch-only stocks on a global basis. Therefore, the working group focused on the use and development of other methods with fewer data requirements.

The working group considered numerous models and tested four. These four models can be categorized into two groups: (i) empirical models; and (ii) catch-based analyses. The working group felt that these models represented the range of data-limited approaches. Empirical models aim to predict the status of unassessed stocks by transferring knowledge derived from assessed stocks and covariates (Costello *et al.*, 2012; Thorson *et al.*, 2012). Catch-based models consider a Schaefer-like biomass dynamics model together with assumptions such as resilience (Catch-MSY; Martell and Froese, 2013) or harvest dynamics (COM, Catch Only Model; Vasconcellos and Cochrane, 2005, initially tested by Minte-Vera *et al*., unpublished data) to extract information from the catch time-series. Thorson *et al.* (forthcoming) developed a new state–space catch-based model. The working group modified these models where necessary to allow them to be broadly applied to most global fisheries. Then, a simulation-testing framework was created that allowed consistent comparison of performance across models for a set of simulated stocks. This framework is a critical element of the approach adopted here and it can be used when assessing the performance of any assessment methodology. The overall goals of this work were to: (i) test how the models performed in estimating stock status across a range of simulated characteristics of fish stocks and fisheries; and (ii) consider the applicability of each model under different scenarios.

2. METHODS: EVALUATION OF THE PERFORMANCE OF DIFFERENT MODELS FOR ESTIMATING THE STATUS OF UNASSESSED FISHERIES

DESCRIPTION OF CANDIDATE MODELS

The performance of four models (Table 1) was investigated in a full factorial simulation framework. These four models were:

- modified panel regression model (mPRM);
- catch-MSY model (CMSY);
- catch only model – sampling importance resampling model (COM-SIR);
- state–space catch only model (SSCOM).

The descriptions below include adaptations made by the working group to the original models to make them comparable across the simulation framework. The code to run the models is available at: ftp://ftp.fao.org/FI/STAT/Rfiles/C1086.zip

TABLE 1
Overall model comparison

	Empirical	**Mechanistic**		
Model	**Modified panel regression (mPRM)**	**Modified catch-MSY (CMSY)**	**COM-SIR**	**SSCOM**
Method	Log-linear regression model	Schaefer model	Schaefer model with harvest dynamics Bayesian model – sampling importance resampling (SIR) algorithm	Schaefer model with harvest dynamics State–space Bayesian model
Input	Catch; life history and fishing history parameters	Catch; priors for *r* and *K*; initial and final depletion rates	Catch; priors for *r* *K*; parameter *a* and *x* of harvest dynamics model	Catch; priors for *r*, *K*; parameter *a* and *x*, of harvest dynamics model; process error variability
Reference	Costello *et al*. (2012)	Martell and Froese (2013)	Vasconcellos and Cochrane (2005)	Thorson *et al.* (forthcoming)

Modified panel regression model (mPRM)

Costello *et al.* (2012) developed a series of six tiered panel regression models (PRMs) to estimate B/B_{MSY} using only basic life-history information along with a time series of catch. A regression-based approach allows the observed relationships between B/B_{MSY} and the explanatory variables to inform the model, rather than specifying a specific mechanistic form. This approach was not developed as a replacement for traditional stock assessments, but it has been demonstrated to be a robust means for estimating B/B_{MSY} using minimal data (Costello *et al.*, 2012). The PRM models were applied to a set of 1 793 previously unassessed fisheries.

While the PRM methods presented in Costello *et al.* (2012) use limited life-history data, they do require some life-history information that is not often available for many fisheries. To facilitate implementation for highly data limited stocks, the working group developed an mPRM that requires only a catch time series and broad life-history information that can be easily applied to almost any

fishery. The mPRM uses the same catch-related variables as those published for the PRM. However, it omits all life-history information except fixed effects for the three life-history categories available in this analysis: demersal, small pelagic, and large pelagic (fixed effects are relative to large pelagics, see Table 2). These three species categories were the only life-history data used in the mPRM, as they were the only life-history data provided by the simulated stocks for this analysis. The simulated stocks were made extremely data-limited in order to make the results of this report as applicable to as many fisheries as possible. The PRM models developed by Costello *et al*. (2012) could be applied to stocks with more available life-history information, which would most probably result in more accurate estimates of stock status.

The mPRM only uses the "developed" period of a fishery, defined as beginning once catch exceeds 15 percent of the maximum catch recorded for that fishery. Within the developed period, where a fishery was missing catch data in less than 10 percent of its years, the missing catches were filled in by interpolation. Fisheries missing more than 10 percent of their catch history were omitted from the analysis. The final processed catch history had to be greater than or equal to seven years. For a given fishery, all catch variables except for maximum catch are scaled relative to the maximum catch recorded for that fishery.

The mPRM was trained on a subset of fisheries from the RAM Legacy database (Ricard *et al.*, 2012) (N = 166) using a linear regression predicting log(B/B_{MSY}) in the form:

$$\log\left(\frac{\mathrm{B}}{\mathrm{B_{MSY}}}\right)_{\mathrm{ijt}} = \alpha + \beta \mathrm{X_{ijt}} + \gamma_{\mathrm{j}} + \varepsilon_{\mathrm{ijt}} \log\left(\frac{B}{B_{MSY}}\right)_{ijt} = \alpha + \beta X_{ijt} + \gamma_j + \varepsilon_{ijt} \quad (1)$$

where i denotes fishery, j species type, t time, α is a constant, β are the regression effects on variables X, γ is the species-type fixed effect, and ε is an error term (Table 2).

Because mPRM predicts log (B/B_{MSY}), in order to present summary data on the median B/B_{MSY} values for groups of fisheries, it was necessary to correct for a retransformation bias. Median or mean values cannot be calculated by simply using the raw individual predictions of log (B/B_{MSY}). The methodology used to correct the retransformation bias is detailed in the supplementary material for Costello et al. (2012), although the simulation modelling approach did not aggregate estimates from multiple stocks and, therefore, generally did not require use of the retransformation bias correction method.

The application of mPRM and the retransformation bias correction to catch histories of fish stocks provides an estimate of B/B_{MSY} together with 95 percent confidence intervals across the time series of individual or aggregated groups of fisheries. The mPRM produced a significant fit to the training dataset ($P < 0.001$), although the high significance of the individual model variables may be a result of the large dataset used in the model fitting (Table 2). The mPRM did not provide as robust a fit as any of the full PRM models (mPRM $R^2 = 0.14$). The R^2 values for the PRM models ranged from 0.21 to 0.45 (from most-data-poor to most-data-rich models, respectively). However, while these results may suggest that the mPRM does not perform as well as the PRM models, this cannot be inferred from the R^2 values alone, as the mPRM and PRM were trained on different datasets and so are not directly comparable.

TABLE 2
Summary of variables and model fits for the modified panel regression model (mPRM)

Variable	Description	Coefficient
Catch4	Scaled catch 4 years prior to present year	4.830e-1 ***
Catch3	Scaled catch 3 years prior to present year	4.842e-1***
Catch2	Scaled catch 2 years prior to present year	5.319e-1***
Catch1	Scaled catch 1 years prior to present year	7.300e-1***
CatchNow	Scaled catch in the present year	-2.061***
TToMax	Number of years until maximum catch occurs from the developed period of the fishery	-1.040e-2***
InitSlope	The slope of the catch over the initial 6 years of the fishery	9.815e-2***
CMax	The raw maximum catch recorded	-1.677e-7***
MeanCatch	The average scaled catch from the developed period of the fishery	-1.414***
RunningRatio	Ratio of catch in the present year to largest catch prior to the present year	1.726***
Small Pelagic	Fixed effect for small pelagic species type	-2.294e-1**
Demersal	Fixed effect for demersal species type	-3.584e-1***
Intercept		-1.120e-1
Adjusted R-squared		0.143

*** = $p < 0.001$; ** $p < 0.01$.

Catch-MSY model (CMSY)

The most basic model-based approaches for estimating *MSY* are production models such as Schaefer's (1954) surplus-production model:

$$\hat{B}_{t+1} = B_t + rB_t\left(1 - \frac{B_t}{B_0}\right) - C_t \qquad (2)$$

In this model, B_t is the biomass in the beginning of year t, r is the intrinsic rate of growth, K is the carrying capacity, and C_t is the catch in year t. These models require time series data of removals and depletion levels to estimate two model parameters: the carrying capacity, K, and the maximum rate of population increase, r, for a given stock in a given ecosystem.

Martell and Froese (2013) modified existing stock-reduction analysis methods (Kimura and Tagart, 1982; Kimura, Balsiger and Ito, 1984) for estimating *MSY* from a time series of catch data of a specific area (Table 3; item 1), normally defined as a unit stock where the population is closed to immigration and emigration. Their model uses resilience estimates from FishBase (Musick, 1999) to create prior distributions of the r parameter for each species, and estimations about depletion (i.e. relative stock abundances at the beginning and the end of the time series that are derived from the

relationship between current catch and maximum catch). These depletion levels are denoted by λ_{01} and λ_{02} (Table 3; item 2) for the initial stock size and by λ_1 and λ_2 (Table 3; item 3) for the final lower and upper limits, respectively. Following Martell and Froese (2013), the default values for λ_{01} and λ_{02} are 0.5 and 0.9 when first-year catch was less than 0.5 of the maximum catch and 0.3 and 0.6 otherwise. The default values for λ_1 and λ_2 are 0.3 and 0.7 when last-year catch was higher than 0.5 of the maximum catch and 0.01 and 0.4 otherwise. The Bayesian priors for *r* and *K* parameters are uniformly distributed in log space, as opposed to COMSIR and SSCOM, which have priors distributed in normal space (Table 3; items 4, 5). The bounds for the priors on *r* are determined by the "resilience" categories (Table 3) and the bounds for the prior on *K* are given by the maximum catch and 100 times maximum catch. The working group also ran the CMSY model with uniform unlogged priors, and found that the results were very similar to logged uniform priors (< 2 percent difference in best performance summaries) with no changes in performance inference. The process errors are assumed to be lognormal, independent, and identically distributed (Table 3; item 6). In all instances in this report, $\sigma\nu = 0$ (i.e. there is zero process error variance) such that the model has deterministic population dynamics. The model parameters of interest are the carrying capacity, *K*, and the maximum intrinsic rate of population growth, *r* (Table 3; item 7). Starting with an assumed relative biomass of $B1 = \lambda_0$ times *K* in the first year (Table 3; item 8), biomass in subsequent years is calculated based on an annual Schaefer difference equation (Table 3; item 9), where the observed catch is subtracted from the biomass at the start of the year. This model assumes the catch is measured without error.

TABLE 3
A simple Schaefer production model and corresponding management parameters relevant to the CMSY approach

Data		**Item**
c_t observed catch from $t = 1$ to $t =$ n years		1
λ_{01}, λ_{02} lower and upper bounds for depletion level in year 1		2
λ_1, λ_2 lower and upper bounds for depletion level in the final year of the time-series		3
Prior densities		
$p(\log(K)) \sim \text{uniform}(\log(l\ K), \log(u\ K))$		4
$p(\log(r)) \sim \text{uniform}(\log(l\ r), \log(u\ r))$		5
$p(v\ t) \sim \text{normal}(0, \sigma\nu)$		6
Parameters		
$\Theta = K, r$		7
Initial states $t = 1$		
$B_1 = \lambda_0\ K \exp(v\ t)$		8
Dynamic states $t > 1$		
$B_{t+1} = [B_t + r\ B_t(1 - B_t/K) - c_t] \exp(v\ t)$		9
Likelihood		
$l(\Theta\|c_t) = 1$	$\lambda_1 \leq B_{n+1}/K \leq \lambda_2$	10
$= 0$	$\lambda_1 > B_{n+1}/K > \lambda_2$	
Management quantities		
$MSY = ¼\ rK$		11
$B_{MSY} = ½\ K$		
$F_{MSY} = ½\ r$		

Source: Adapted from Table A1 in Martell and Froese (2013).

The joint distribution of model parameters (in this case, *r* and *K* of the Schaefer production model) that lead to current depletion levels between λ_1 and λ_2 was used to identify cases where combinations of (*r*, *K*) lead to the population going extinct or overshooting *K* before the end of the time series. In these cases, a 0 is assigned for that parameter combination. A value of 1 is assigned for combinations of (*r*, *K*) that result in final stock sizes between λ_1 and λ_2 (Table 3; item 10). Estimates of *MSY* can be calculated from the population parameters for each parameter combination that results in a viable population at the end of the time series (Table 3; item 11). Using this approach, the model was also extended to produce biomass (and B/B_{MSY}) time series. To do this, the working group ran the Schaefer model for each stock using the catch time series and each viable *r-K* pairs and then computed the arithmetic mean biomass ratio in each year and upper and lower quartiles.

Catch only model – sampling importance resampling model (COM-SIR)

The COM is a coupled harvest–biomass dynamics model proposed by Vasconcellos and Cochrane (2005). The model predicts catches based on a combination of a biomass dynamics model and a harvest rate dynamics model. Two patterns of how harvest rate may change over time were proposed by Vasconcellos and Cochrane (2005): linear and logistic. The logistic model mimics the harvest rate behaviour of a developing fishery that has no management as described in (Caddy and Gulland, 1983). The biomass dynamics follows a Schaefer model, where the maximum surplus production is half of the virgin biomass or carrying capacity. Given that preliminary simulation-testing of the COMs showed that the logistic model outperformed the linear model (Minte-Vera, unpublished report; Medley *et al.*, 2009), the linear model was not considered.

Catches are predicted by the following equation:

$$\hat{C}_{t+1} = P_{t+1}\left[B_t + rB_t\left(1-\frac{B_t}{K}\right) - \hat{C}_t\right] \tag{3}$$

where $\hat{C}_{t+1}$ is the predicted catch in year *t+1*, P_{t+1} is the harvest rate or proportion of the biomass caught in year *t+1*, *K* is the carrying capacity or biomass at which the growth of the population is zero, and *r* is the intrinsic rate of population biomass change.

The harvest rate evolves over time according to a logistic model:

$$P_{t+1} = P\left[1 + x\left(\frac{B_t}{aK} - 1\right)\right] \tag{4}$$

where P_t is the proportion of biomass caught at time t, and *a* (where $0 < a < 1$) is the bioeconomic equilibrium as a proportion of *K*, assuming no subsidies or constant subsidies in the fishery. The working group defined *x* as a multiplier that expresses the increase in the harvest rate over time. In cases where catchability is assumed to be constant over time, *x* is the intrinsic rate of effort change. The four parameters to be estimated are *r*, *K*, *x* and *a*. The initial exploitation rate is given by:

$$P_0 = \frac{C_0}{B_0} \tag{5}$$

The initial catch C_0 is assumed to be equal to the first observed catch in the time series and $B_0 = K$. In addition, the working group assumed that no harvest control regulations were in place, or that if there were any existing regulations, they had only negligible effects. Thus, the harvest rate dynamics were designed to respond only to economic/market stimulus. If regulations had been implemented and they had a significant impact on the evolution of harvest, the harvest model would need to be either modified to take this into account or fitted only to the period without regulations. Different eras would need to be fitted twice (Thorson *et al.*, forthcoming). Biomass at maximum sustainable yield (B_{MSY}) is obtained by:

$$B_{MSY} = \frac{K}{2} \tag{6}$$

The parameters were estimated using the Bayesian algorithm, sampling importance resampling (SIR; McAllister *et al.*, 1994; Gelman *et al.*, 2004), implemented in R. The observed catches were assumed to follow a lognormal likelihood function (Casella and Berger, 2002) with expected values equal to the catches predicted by the models:

$$L(\phi|w) = \prod_{t=1}^{n} \frac{1}{\sigma C_t \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} (\ln C_t - \mu)^2 \right] \tag{7}$$

where $\mu = \ln E(C_t) - \frac{\sigma^2}{2}$, C_t is the observed catch in year t, $E(C_t)$ is the expected catch for year t given by equations (1), σ^2 is the variability parameter assumed known and equal to 0.4 (Vasconcellos and Cochrane, 2005).

The prior probability distributions were *a~U(0,1)*, *x~(0.000001,1)*. For the biomass dynamics parameter *K*, the same priors were used as for the CMSY (Martell and Froese, 2013). *K* was assumed to have uniform distribution in log space *so ln(K) ~ U (ln(maximum Catch), ln(100*maximum Catch))*. The prior for *r* was set using the "Resilience" category from FishBase (Table 4) (Martell and Froese, 2013). The working group assumed a uniform prior according to different resilience categories assigned to the stocks, based on the rules proposed in ICES (2012a; Table 3.2.3.1). As only the L_∞ was provided, the *k* parameter of the von Bertalanffy growth function was obtained using the following equations: $k_{min} = exp(0.696-0.602*log(L_\infty))$ and $k_{max} = exp(1.474-0.602*log(L_\infty))$ (Froese *et al.*, unpublished data).

TABLE 4
Resilience category by stock parameters (*k* parameter of the von Bertalanffy growth function, age at maturity T_{mat}, maximum age T_{max}) and corresponding prior distribution for *r*

Stock parameters and rules	**Resilience category**	**Prior probability distribution for *r***
k_{max} < 0.05 or k_{min} < 0.05 or T_{mat} >10 or T_{max} > 30	Very low	U ~ (0.015, 0.1)
(k_{max} < 0.15 and k_{max} >= 0.05) or (k_{min} < 0.15 and k_{min} >= 0.05) or (T_{mat} >5 and T_{mat} <=10) or (T_{max} > 11 and T_{max} <=30)	Low	U ~ (0.05, 0.5)
(k_{max} <0.3 and k_{max} >= 0.16) or (k_{min} < 0.3 and k_{min} >= 0.16) or (T_{mat} >2 and T_{mat} <=4) or (T_{max} > 4 and T_{max} <=10)	Medium	U ~ (0.1, 1)
(max.k.growth >= 0.3 or min.k.growth >= 0.3 or T_{mat} <=1 or T_{max} <=3)	High	U ~ (0.6, 1.5)

The importance function was equal to the joint prior function, and thus the importance ratio was equal to the likelihood. Approximately 1–5 million parameter vectors were randomly sampled from the joint prior distribution. Of those, 5 000 samples were taken with replacement and probability proportional to the importance ratio. Punt and Hilborn (1997) found that the resampling needs to be done until no vector is assigned more than one percent of the posterior probability (MSD – maximum single density). The working group ensured that this condition was met. Other diagnostics for convergence were also used including the coefficient of variation in the average importance weight (McAllister and

Kirchnet, 2002), the maximum importance ratio (McAllister and Pikitch, 1997), and the effective sample size (Raftery and Bao, 2010).

State–space catch only model (SSCOM)

Models for using catch data in data-limited estimates of stock status are broadly classified as mechanistic or empirical. Mechanistic models such as DB-SRA (Dick and MacCall, 2011) and CMSY (Martell and Froese, 2013) use well-established population dynamics models, and a value or prior on final status, to reconstruct stock productivity (Kimura and Tagart, 1982). They are transparent, but require a strong assumption of current status. Empirical models such as Thorson *et al.* (2012) and Costello *et al.* (2012) use assessed stocks to train a model, which then can be used to predict status from catch data. Results are less interpretable, but can easily use meta-analytic information from assessed stocks.

The working group sought to combine these strengths in a single model (Thorson *et al.*, forthcoming), to which the working group refers as a state–space catch only model (SSCOM). This model combines explicit population dynamics equations with a proposed model for changes in fishing mortality over time. It integrates across unknown states for biomass and fishing effort, while solving for biological productivity and fishing mortality parameters.

The working group first assumed that exploitation rates were semi-predictable but unobserved. Specifically, the working group used the following harvest dynamics model:

$$\hat{E}_{t+1} = E_t \left(\frac{B_t}{a \cdot B_0 / 2} \right)^x \tag{8}$$

Where a and x are parameters governing the evolution of fishing mortality over time. This harvest dynamics model was selected so that the logarithm of changes in exploitation rates over time is a linear model of parameters a and x, thus facilitating future estimation of priors for these parameters, e.g. using data in the RAM Legacy Stock Assessment Database (Ricard *et al.*, 2012). Given the assumption that harvest dynamics are not perfectly explained by this model, E_t must be treated as a random variable:

$$\Pr\left(E_{t+1} \middle| \hat{E}_{t+1} \right) = \frac{1}{\sqrt{2\pi\sigma_\tau^2}} e^{-\left(\frac{\left(\ln(E_{t+1}) - \ln(\hat{E}_{t+1}) \right)^2}{\sigma_\tau^2} \right)} \tag{9}$$

where σ_τ is the magnitude of variability in harvest dynamics in year t that is unexplained by the harvest dynamics model (a "process error").

The working group next assumed that catch followed the typical equation for exploitation rates:

$$\hat{C}_t = E_t B_t \tag{10}$$

where C_t, B_t and E_t are estimated catch, true biomass and exploitation rates respectively in year t. However, the working group again assumed that catch may not be perfectly explained by this model, i.e. due to variability in the scaling between fishing effort and exploitation rate ("catchability"). This results in the following likelihood:

$$L\left(C_{t+1} \middle| \hat{C}_{t+1} \right) = \frac{1}{\sqrt{2\pi\sigma_\omega^2}} e^{-\left(\frac{\left(\ln(C_{t+1}) - \ln(\hat{C}_{t+1}) \right)^2}{\sigma_\omega^2} \right)} \tag{11}$$

where σ_ω is the magnitude of variability in catchability. If σ_ω is fixed at *0*, this is equivalent to the assumption of constant catchability over time.

Last, the working group specified that population dynamics follow a conventional Schaefer surplus production model (Equation 2). The working group assumed, however, that biomass dynamics might not be perfectly explained by this surplus production model, and estimated true biomass as a random effect:

$$\Pr\left(B_t \middle| \hat{B}_t\right) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\left(\frac{\left(\ln(B_{t+1})-\ln(\hat{B}_{t+1})\right)^2}{\sigma_\varepsilon^2}\right)} \tag{12}$$

where σ_ε is the magnitude of variability in the population dynamics that is unexplained by the Schaefer population–dynamics model. The working group estimated parameters after re-scaling catch to have a maximum of 1.0 so that scale parameters were defined relative to maximum catch. Initial effort was given a broad uniform prior in log space, *x* was uniformly distributed from 0.01 to 0.50, *a* was uniformly distributed from 0.1 to 2.0, $ln\ (B_0)$ was given a uniform prior from –4.6 to 4.6, and *r* was uniformly distributed between minimum and maximum values derived from life history (Table 5). Code for running this model either with or without variability in catchability can be found in Thorson *et al.* (forthcoming). These equations were fitted simultaneously to catch data and priors for all parameters using Markov chain Monte Carlo implemented using JAGS (Plummer *et al.*, 2009), using an annealing algorithm, in which a first run was conducted using 100 short chains, and the posterior from these runs was used to define starting locations for parameters in subsequent runs. Subsequent runs used three chains with three million iterations per chain, the first million of which were discarded for burn-in and the remainder thinned at an interval of 1 000.

SIMULATION FRAMEWORK AND IMPLEMENTATION

To test the operational performance of the four models in estimating true stock status, the working group developed a set of simulated stocks of known properties and fishery histories. In contrast to typical approaches for simulation testing of stock assessment methodologies (National Research Council, 1998; ICES, 2012b), the approach implemented here is a full factorial experimental simulation design, in which the main and higher-level interaction effects of independent variables on model performance can be investigated. Previous investigations of the performance of data-limited assessment models have largely focused on either performance compared with full stock assessments (e.g. Dick and MacCall, 2011) or performance in a limited number of pertinent simulation scenarios, typically based on select real stocks (Wetzel and Punt, 2011; ICES, 2012a). Although useful for testing the performance of different models under given scenarios, inferences from these approaches are necessarily restricted to the cases examined, as opposed to a general appraisal of performance under a wide range of scenarios. These scenarios may interact and confound, as is likely the case with real stock assessments (National Research Council, 1998).

There are several benefits to adopting a factorial testing design. First, a wider range of scenarios can be investigated, including those that may be overlooked when simulations are based on real assessments. Second, the main and higher-level interaction effects can be examined, and confounding effects can be isolated. Last, a factorial design allows a wide array of analytical methods to be used for performance evaluation (e.g. ANOVA, and regression trees). From the factorial design, the working group was able to develop clear evaluation criteria and performance indicators. However, this methodology depends on the ability of the simulations to mimic the most relevant dynamics that drive real stocks, and all results are conditioned on this assumption.

Data generation framework using the Fisheries Library for R tools

The working group developed a two-stage simulation framework with a deterministic set of simulations (without any random variation) and a stochastic set of simulations (where process and measurement errors were incorporated). Simulated stocks were generated using the Fisheries Library

for *R* (*FLR*[2]) (Kell *et al.*, 2007). The FLR Project developed two new functions within the *FLBRP* package[3] that allow for the simulation of a fish stock based on life-history parameters and fishery characteristics. An overview of the process of stock simulation is shown in Figure 1. The core simulation function *gislasim* takes the following arguments:

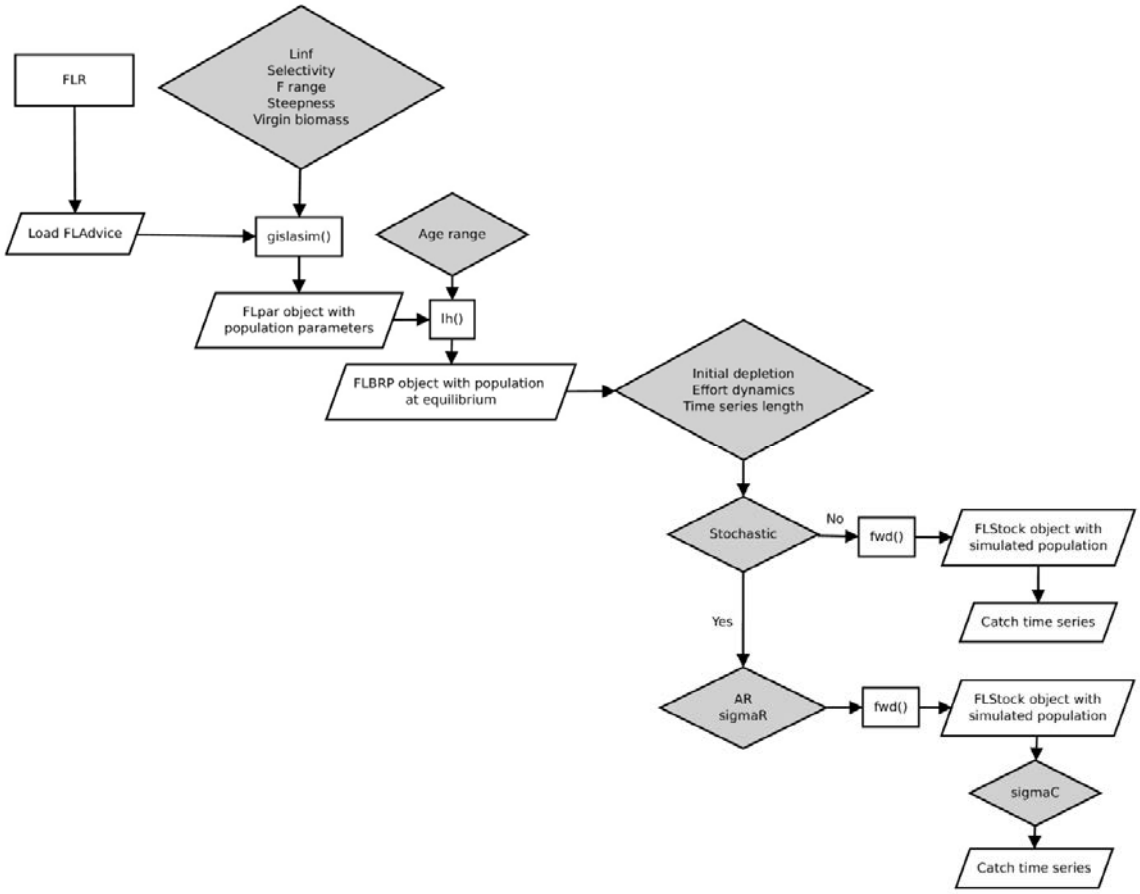
- mean asymptotic length (L_∞);
- selectivity parameters;
- fishing mortality age range;
- steepness and virgin biomass parameters of the stock recruitment curve.

Based on these values, *gislasim* generates a complete set of parameters for the: corresponding growth model (von Bertalanffy), natural mortality, maturity ogive, selectivity and stock recruitment function. These are all based on derived life history relationships (Gislason *et al.*, 2008).

The returned object, of class *FLPar*, is then passed to the *lh* function (Figure 1), which generates the corresponding reference points and population structure at equilibrium. The resulting object, of class *FLBRP*, contains a set of biological reference points and a matrix of abundances at age at equilibrium. These outputs can then be used to generate a simulated population trajectory by projecting the stock forwards under a given scenario of fishing mortality, the stock recruitment relationship, including variability therein and natural mortality. The population projection uses the *fwd()* FLR function (Figure 1).The catch series extracted from the projection provides the basis for the simulation testing of the four data-limited models.

[2] http://flr-project.org
[3] http://github.com/flr/FLBRP

FIGURE 1
Flow chart of stock simulation



Note: Processing steps are in rectangles with functions denoted by empty parentheses; decisions are grey diamonds; input/output are parallelograms.

Deterministic independent variable design

In designing factorial simulations, trade-offs exist between increasingly realistic scenarios and an increasing number of combinations, which require increased computational resource requirements. The factorial design chosen for the deterministic runs tried to balance these constraints by including several factors, while maintaining a streamlined and computationally manageable design. With these considerations in mind, the working group generated 72 stocks for the full factorial interactions between life history, initial depletion, harvest dynamics and time series length. The levels of each of these factors are also described below.

Life history (LH)

The working group generated life-history types based on considerations of: size, age range and productivity. Size relates to a large number of metabolic processes and fish life-history evolution (Charnov and Gillooly, 2004). The working group therefore chose three mean asymptotic sizes to cover generic small, medium and large commercial species (Table 5). Age ranges were based on those typically reported for the life histories generated based on the experience of the group. Stock-recruitment productivity was parameterized based on the steepness meta-analysis of Myers, Bowen and Barrowman (1999). The working group chose these life-history dynamics with a bias towards commercial stocks with existing stock assessments. However, the working group considered the variety of traits to be different enough to illustrate any probable impact of life-history type on the evaluation of the performance of the models. For future work, sexual dimorphism, sequential

hermaphroditism and longer-lived stocks should be considered and developed using this simulation framework.

TABLE 5
Life-history specifications implemented in the simulations

Species type	Generic name	Factor level	L_∞ (cm)	Age range (years)	$\bar{F}$ ages (years)	Steepness
Large pelagic	Scombrid	LP	150	1:20	4:20	0.8
Small pelagic	Clupeoid	SP	30	1:8	2:8	0.7
Demersal	Gadoid	DE	70	1:20	4:20	0.8

Note: $\bar{F}$ ages refer to the ages over which fishing mortality is spread; steepness refers to how steep the Beverton-Holt stock recruitment relationship is in the descending portion towards the origin (i.e. the proportion of the maximum recruitment that is obtained on average when the spawning biomass is at 20 percent of the unfished biomass).

Initial depletion (ID)

Initial depletion determines how depleted the biomass of a given stock is at the start of the catch series. Many models either make assumptions about initial depletion or are sensitive to the assumptions about depletion. Therefore, the working group felt that initial depletion scenarios had to be included in the factorial design. All generated stocks started from a similar virgin biomass of 1 000 tonnes and three levels of initial depletion were implemented (0 percent, denoted as ID:1; 30 percent, denoted as ID:0.7; 60 percent, denoted as ID:0.4). As the level of stock depletion is often an unknown but critical parameter for many stock assessment models, the criterion was to test cases where the stock was at virgin biomass (ID:1), at moderate depletion (ID:0.7), and very depleted (ID:0.4) in the first year of the landings records.

Harvest dynamics (HD)

Harvest dynamics govern removals of biomass by a fishery and are therefore central to the simulation design. The working group constructed harvest dynamic scenarios that reflected commonly perceived scenarios in fishing effort (Hilborn and Walters, 1991). The working group examined four harvest dynamic scenarios (Table 6):

- Constant harvest dynamics (HD:0) where the harvest rate (proportional to fishing effort, given fixed catchability and instantaneous fishing) remains constant irrespective of biomass, e.g. bycatch species harvest.

- Bioeconomic coupling where harvest ratio has a dynamic relationship with biomass, described by the following difference equation:

$$E_{t+1} = E_t \left(\frac{B_t}{aB_{MSY}}\right)^x \quad (13)$$

  where E_t is the harvest rate at time *t*, B_t is total stock biomass, *a* is a portion of B_{MSY} at which bioeconomic equilibrium occurs, and *x* is an exponent governing the rate of response of effort to changes in biomass (factor level denoted HD:0.6; see Thorson *et al.* [forthcoming] for further details of the coupled system), e.g. open-access single-species harvest. Having an explicit bioeconomic coupling may appear to bias this simulation level in favour of SSCOM and COM-SIR, but the realized dynamics of the coupled system are highly flexible (e.g. stable, unstable, oscillatory dynamics) depending on the parameter values for *a* and *x*, which are not passed to the models. A coupled dynamic is a necessary scenario to test but, as such, the working group does not expect the structure to unduly bias results in favour of these models.

- One-way trip, where the harvest rate increases 5 percent per year to 80 percent of the harvest level at which the stock crashes, F_{crash} (factor level denoted as HD:OW), e.g. a stock where harvest rate has continually increased, such as a high value large pelagic stock.

- Roller-coaster or dome-shaped, where the harvest rate increases 25 percent per year to 80 percent of the harvest rate at which the stock crashes, F_{crash}, stays at this level for 5 years, and then decreases to FMSY levels by 30 percent per year (factor level denoted as HD:RC), e.g. a stock where management began following extensive depletion, such as with some small pelagic species.

Time series length (TS)

The length of the catch time series can affect the estimates of biomass, especially if the catch time series does not capture all of the phases of the fishery. For the analyses, the working group used two time series lengths for the catch series: 20 years (denoted TS:20) and 60 years (denoted TS:60). These two values were chosen to reflect two contrasting levels of data availability: 20 years as an intermediate time series; and 60 years to simulate a fishery with a long time series. The TS:20 was simulated by projecting the dynamics for 60 years and providing only the last 20 years of catch data for the estimation methods. The full deterministic experimental design (Table 6) resulted in 72 simulated stocks with time series of stock biomass, fishing mortality and catch (Figure 2).

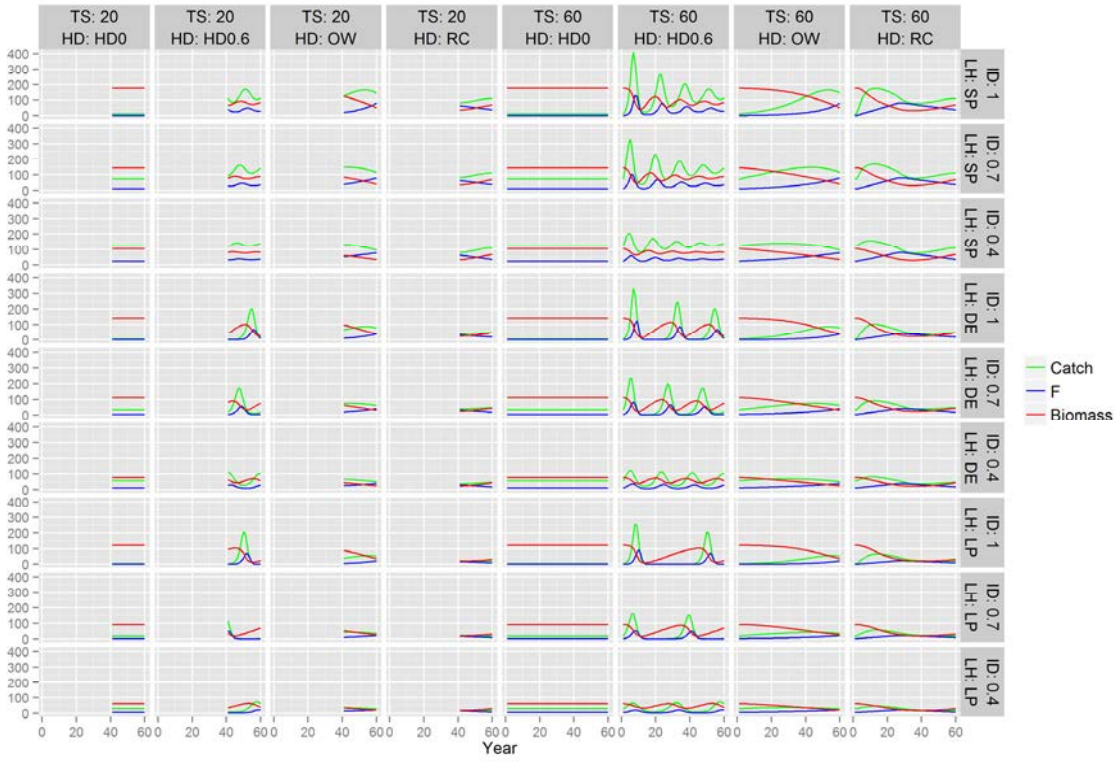
TABLE 6
Deterministic experimental design

Factor	**Level 1**	**Level 2**	**Level 3**	**Level 4**
Initial depletion (ID: percent)	1	0.7	0.4	
Harvest dynamics (HD)	Constant harvest rate (**HD0**)	HD model with a = 1 (corresponding to bioeconomic equilibrium at B_{MSY}) and x = 0.6 (**HD0.6**)	One-way trip (**OW**)	Roller-coaster (**RC**)
Time series length (TS:years)	20	60		
Life history (LH)	Clupeoid Small pelagic	Gadoid Demersal	Scombrid Large pelagic	

Notes: Factor levels differ by column for a given factor (row). Life-history parameters are based on the designations from Gislason *et al.* (2008).

FIGURE 2

Deterministic simulation trajectories of rescaled stock biomass B/B_{MSY}, fishing mortality (F) and resulting catch for each combination of the design given in Table 6



Notes: Harvest dynamics levels (HD) and length of time series (TS) structure the columns and initial depletion (ID) and life history (LH) the rows. The simulation is driven by the F levels and in the case where HD = 0, effort is decoupled from biomass and presents no variation in time. When HD=0.6, there is a tight coupling between biomass and F.

Stochastic independent variable design

Variability in a simulated fishery is encapsulated in the process error on the biomass and harvest dynamics and measurement error on observed variables. Reflecting the central role of recruitment on population dynamics (Myers, 2001), the working group included stochasticity on recruitment both in the form of random and autocorrelated variability to reflect environmental effects (Pyper and Peterman, 1998; Vert-Pre *et al.*, 2013). From a measurement perspective, the dominant source of information provided to the methods is catch, which is likely uncertain. The working group therefore implemented random measurement errors on the catches.

Recruitment variability

Lognormal recruitment variability was incorporated as:

$$R_t = f\left(S_{t-\tau}\right)e^{\varepsilon_t} \quad (14)$$

where $f(S_{t-\tau})$ is the Beverton-Holt stock recruitment model using spawner biomass lagged to the age of recruitment τ; with independent residuals $\varepsilon_t \sim N(0,\sigma_R^2)$, where σ_R^2 controls the level of recruitment variability. Two levels were implemented: $\sigma_R = 0.2$ (sigmaR:0.2) and $\sigma_R = 0.6$ (sigmaR:0.6). The sigmaR:0.6 level was chosen based on Pyper and Peterman (1998).

Autocorrelation on recruitment residuals

Persistent environmental variability likely results in non-independent recruitment residuals. This was incorporated by assuming a first-order autoregressive form for the recruitment residuals:

$$\varepsilon_t \approx N\left(\varphi\varepsilon_{t-1}, \sigma_\eta^2\right) \quad (15)$$

where $\varphi$is the autoregressive coefficient with factor levels $\varphi = 0$ (AR:0) and $\varphi = 0.6$ (AR:0.6); σ_η^2 was chosen to maintain the same marginal variance with and without autocorrelation by setting:

$$\sigma_\eta^2 = \sigma_R^2\left(1-\varphi^2\right) \quad (16)$$

Measurement error on catch

Measurement error on catch was implemented by sampling observed catches $c_t$from a lognormal distribution with a mean given by the true catch C_t with lognormal variability:

$$c_t = C_t e^{\xi_t}, \quad (17)$$

where $\xi_t \sim N(0,\sigma_C^2)$, and where σ_ξ^2 controls the level of catch variability. Two levels were implemented: $\sigma_\xi = 0$ (sigmaC:0) and $\sigma_\xi = 0.2$ (sigmaC:0.2), corresponding to a coefficient of variation of approximately 20 percent on the catch data.

Iterations

Stochastic simulations include variation across realizations as characterized by the probability distributions outlined above. Datasets were generated with a total of 250 iterations, although the number of replicates could be easily augmented if needed in the available source code. Running all methods on such a relatively large number of replicates, for all methods and on the full factorial design was too computationally demanding, so the working group settled on using ten iterations per stock scenario, resulting in a total of 5 760 stock trajectories. The relatively low number of iterations per stock scenario was a compromise between iteration number and computational expense. The full stochastic experimental design table (Table 7) resulted in 576 combinations, with ten replicas each, which produced 5 760 catch time-series (stocks) to be used by each of the four estimations methods.

TABLE 7
Stochastic simulation experimental design factor levels

Factor	Level 1	Level 2	Level 3	Level 4
Initial depletion	ID:1	ID:0.7	ID:0.4	–
Harvest dynamics	HD:HD0	HD:HD0.6	HD:OW	HD:RC
Time-series length	TS:20	TS:60	–	-
Life history	LH:SP	LH:D	LH:LP	-
Autoregressive process error on recruitment variability (AR(1))	0	0.6 (Pyper & Peterman 1998)	-	-
Recruitment variability (sigmaR)	0.2	0.6	-	-
Catch error (sigmaC)	0	0.2	-	-

Notes: Stochastic simulations included the same four factors as the deterministic design and three additional factors presented below. The "–" denotes no given level for that factor.

The simulation functions are freely accessible and can be easily modified in the levels of the factorial design, number of iterations, level of recruitment variability, the autoregressive coefficient of the autocorrelation on recruitment residuals and the level of measurement error on catch. More in depth modifications of the stock simulation structure will require ad hoc code development. All source code is available at the FLR repository at https://github.com/flr/StockSims

Estimation platform

A total of 5 760 stocks were generated for the stochastic set. To test the four data-limited assessment methods therefore required 23 040 assessment runs. Trial runs of the methods indicated that some would take a considerable amount of CPU and processing time (up to ten hours per assessment for a given level of convergence) to run on a local machine. The simulation team therefore used parallel computation on the Hexagon cluster at the University of Bergen, Norway (http://docs.notur.no/uib). Given that simulation runs are independent, the problem is therefore "embarrassingly parallel" (Moler, 1986) meaning that individual assessments can be run in isolation from others with no communication needed during a given run. Batches of up to 640 stocks were submitted at a time. Parallel implementation was conducted using message passing interface linked to R through the R package Rmpi (Yu, 2002). Example scripts for the implementation of SSCOM on Hexagon are available at the repository.[4] Parallel implementations were conducted for the methods: SSCOM, COM-SIR and CMSY. The mPRM was instead implemented and run locally as it takes less computing time to predict based on the previously fitted model.

Data provided to method developers

Researchers in charge of building the simulation infrastructure were not involved in method development or coding, and method developers were only provided with the necessary inputs to run each method, which included catch time series and minimal life history information that should be known in any case to a fisheries biologist working on a given stock, such as length at infinity, maximum age and age at maturity.

[4] http://github.com/flr/StockSims

EVALUATION OF METHOD PERFORMANCE

Visualizations

To create visualizations of the performance of the methods, the working group plotted time series of the estimated 95 percent confidence/credible interval of B/B_{MSY} for the deterministic results for each scenario. The working group used the R package *ggplot2* for the time series plotting (Wickham, 2009), The high dimensionality of the stochastic runs precluded such visualizations, although the working group did examine individual means for selected scenarios (Figures A1.1–A1.7).

Performance metrics

Generally, the performance of the methods was judged on the difference between the true and estimated B_t/B_{MSY} where t denotes a given year. Various measures of bias and precision of estimates are available, but the working group chose to focus primarily on proportional and absolute proportional error because they are more intuitive. The proportional error is a dimensionless value, which gives the proportional difference between the estimate and the true value. The absolute value or square of the proportional errors provides a combined measure of bias and precision similar to mean squared error. The working group used the absolute value because it is influenced less by outliers compared with taking the square.

Proportional error (PE)

The proportional error (PE; also called relative error) of an estimate is defined (Abramowitz and Stegun, 1972) as:

$$\sigma_{\theta_t} = \frac{\hat{\theta}_t - \theta_t}{\theta_t} \tag{18}$$

where $\widehat{\theta_t}$ and θ_t is the estimated and true B_t/B_{MSY}, respectively. A value greater than zero indicates that the estimated status is greater than the true status (i.e. the stock is predicted to be healthier than it actually is). Conversely, a proportional error less than zero indicates an estimated status lower than the true value. For example, a proportional error value of 0.25 denotes a 25 percent overestimate of B_t/B_{MSY}. When summarized over a set of values (e.g. the mean proportional error [MPE]) provides the average bias of the estimate:

$$\overline{\delta\theta} = \frac{\sum_{t=t_{\min}}^{t_{\max}} \delta\theta_t}{t_{\max} - t_{\min} + 1} \tag{19}$$

where t_{min} is the first year and t_{max} the last year.

Absolute proportional error (APE)

Similarly, the absolute proportional error (APE) is calculated as:

$$\left|\delta_{\theta_t}\right| = \left|\frac{\hat{\theta}_t - \theta_t}{\theta_t}\right| \tag{20}$$

and the mean absolute proportional error (MAPE) is defined as:

$$\left|\overline{\delta\theta}\right| = \frac{\sum_{t=t_{\min}}^{t_{\max}} \left|\delta\theta_t\right|}{t_{\max} - t_{\min} + 1} \tag{21}$$

Note that MPE indicates whether there is a positive or negative bias, but may mask the presence of a large absolute error when it is evenly distributed across negative and positive values. This error can be detected using MAPE.

Time period (last five years versus entire time series)

Both PE and APE (and MPE and MAPE) are calculated over the entire time series $[t_{\min} = 1, t_{\max} = T]$ and for the last five years $[t_{\min} = T - 4, t_{\max} = T]$. It is important to conduct both analyses, as calculations over the entire time series include initial years, reflecting assumptions of the model at the start of the time series (e.g. assumptions regarding initial depletion) and periodicity and thus provide an overall summary of performance. The calculations based only on the final five years provide a snapshot of recent performance distal to the initial conditions. The most recent five years are also more reflective of management time scales as they tend to be informative at a time scale that is conducive to management advice. Where the same metric showed different results when calculated over the two time periods, results were further inspected visually to determine the reasons for the effect of time period on method performance.

Indices

To maintain interpretability of the results, the working group presented a synopsis of the indices to reflect: time series length, stock ID, iteration number, time period and analytical method (Table 8).

TABLE 8
Indices for results

Time	$t = \{41, \dots, 60\}$ (20 year series)	$t = \{1, \dots, 60\}$ (60 year series)
Stock	$i = \{1, \dots, 72\}$ (deterministic)	$i = \{1, \dots, 576\}$ (stochastic)
Iteration	$j = \{1, \dots, 10\}$ (stochastic)	
Time period	$k = \{1 = \text{full}, 2 = \text{last 5 years}\}$	
Method	$l = \{1, \dots, 4\}$	

Note: For example, $\overline{|\delta\theta|}_{i=33,j=1,k=2,l=2}$ is the MAPE value for method 2 on stock 33, iteration one over the last five years.

Regression trees

Explanatory variables

To investigate how the independent variables of the simulation design affect the performance metrics, the working group modelled each performance metric (proportional error: PE and absolute proportional error: APE) as a function of the independent variables using non-parametric regression trees (De'ath and Fabricius, 2000). Because the data were time series data with various levels of nesting, the working group implemented random effect regression trees with an AR(1) structure on the residuals. Regression trees were implemented in the R package *REEMtree* (Sela and Simonoff, 2012). The working group created two sets of regression trees, both with performance as the dependent variable. In the first set, performance was investigated separately by method, which resulted in 16 regression trees (four methods over two time periods – all years and the last five years for two performance diagnostics: PE and APE). The second set of regression trees compared performance across methods. Therefore, method was included as an explanatory variable, resulting in four regression trees (all years and the last five years for two performance diagnostics: PE and APE).

Emergent properties

To investigate whether emergent properties of the simulated data explain variability in performance, regression trees were performed on a set of variables for the set of across-method regression trees only. The selected variables that reflect emergent properties of the time series (Table 9) include measures of the mean, range and variance of biomass (stock), catch (catch) and effort (harvest) (Figures A1.10–A1.13).

TABLE 9
Emergent variables estimated for biomass, catch and effort series

Variable	Short name	Description
Coefficient of variance (CV) in year to year change in catch	YtYcv	The CV of the year-to-year changes in the time series
CV of the time series statistic	Cv	The overall CV of the time series
Mean	Mean	The mean of the time series
Minimum	Min	The minimum value of the time series
Maximum	Max	The maximum value of the time series
Minimum relative to maximum	MintoMax	The ratio of the minimum value of the time series to the maximum value of the time series
Standard Deviation	STD	The standard deviation of the time series

For a given stock and iteration, the working group defined the best performing method as the method with the lowest mean statistic $\dot{y}_{i,j,k}$. (e.g. MPE or MAPE) per stock, iteration and time period. To provide a map of best performance, the working group used classification trees to explain variability in $\dot{y}_{i,j,k}$. The working group identified the best performing model per scenario as a function of the design variables in a classification tree. This provides the ability to map the best performing method relative to the simulation design variables. The classification trees were trimmed to have at least 500 observations in the terminal nodes. The classification trees were computed using the R package *rpart* (Therneau, Atkinson and Ripley, 2012).

Best performance visualization (tile plots)

In order to synthesize the results from both the deterministic and the stochastic runs, the working group used "best performance" and "relative performance" tile plots (Wickham and Hofmann, 2011).

In both plots, each tile corresponds to an individual stock, and in the “best performance” plots, the colour of the tiles corresponds to the model that performs the best. For the statistics MPE and MAPE, lower values correspond to best performance, so the working group selected the top performer by taking the minimum of the absolute value of each model score. The relative rank for the performance plots was computed as:

$$\text{Relativerank} = 1 - \left(\frac{|statistic| - \min|statistic|}{\max|statistic|}\right) \quad (22)$$

where *statistic* is either MPE or MAPE.

3. RESULTS

The working group assessed method performance with four primary metrics. The proportional error (PE) and the mean proportional error (MPE) across stocks estimate the bias of the method. The absolute proportional error (APE) and the mean absolute proportion error (MAPE) across stocks estimate the bias and precision of the method. These statistics were calculated for all years and for the last five years of the catch time series. Using PE or MPE, the working group determined how different the status is from the true simulated stock status. The working group also used APE or MAPE because a large absolute error may not be detected using only PE or MPE. This section focuses first on overall method performance. It provides simple frequencies, tile plot visualizations and performance maps for both metrics, classifying best performance according to the simulation design variables. It then discusses in detail the performance of each method within each scenario. For each method, the working group also explored how the combination of the choice of method (i.e. mPRM, CMSY, COM-SIR and SSCOM) and particular simulation variables (i.e. depletion, harvest dynamics, time series length and life history) could explain variability in the performance metrics.

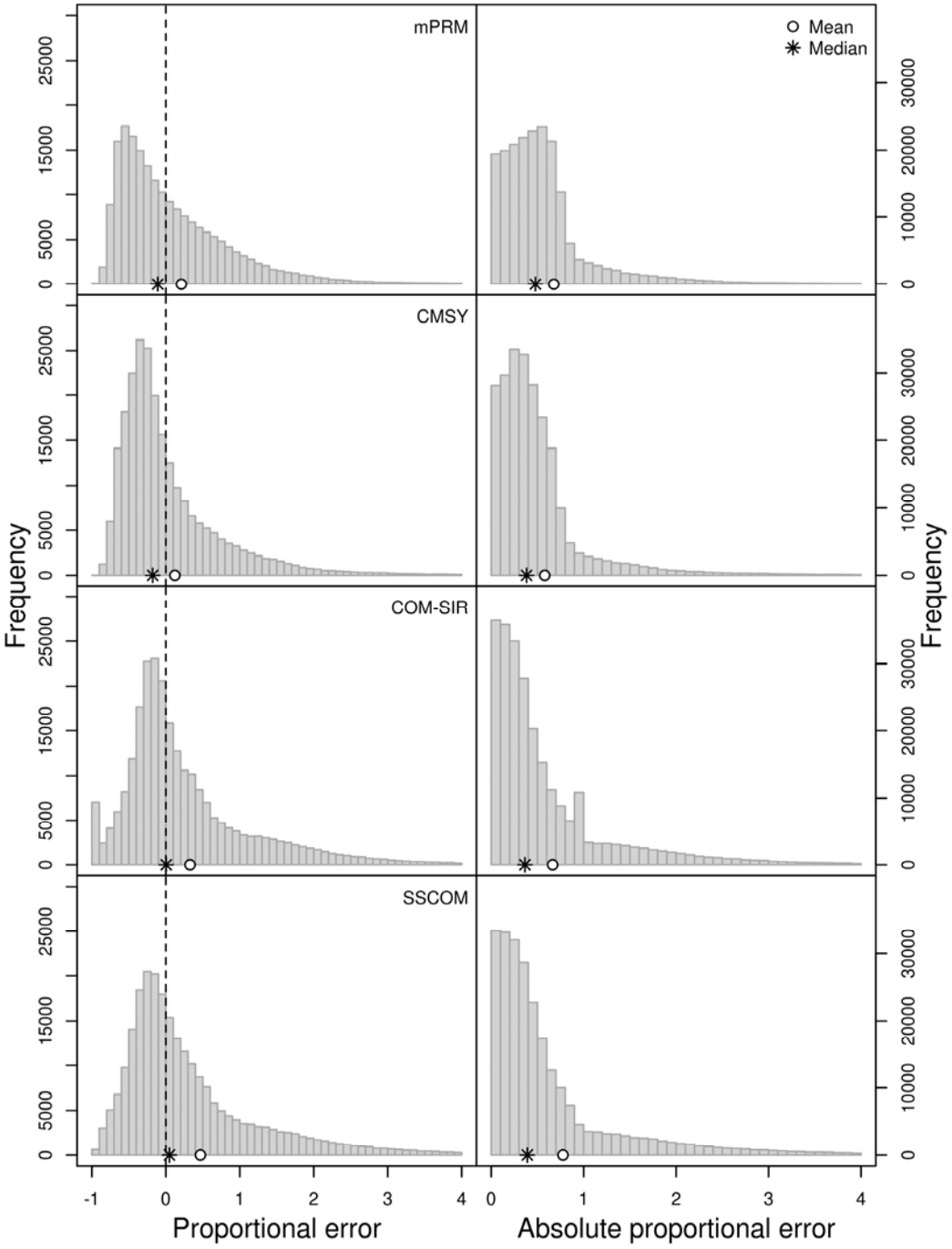
OVERALL PERFORMANCE

The distribution of PEs across all simulated fishery types and for all methods were positively skewed distributions with a lower bound of –1 and long tails to the right (Figure 3). The methods all had positive bias, as judged by the MPE. The CMSY method had the smallest bias (lowest proportional error of 0.122, interpreted as an average overestimate of B/B_{MSY} of 12.2 percent across all simulations), followed by mPRM (21 percent overestimate), COM-SIR (32.7 percent overestimate) and SSCOM (46.8 percent overestimate). The long tails of the distributions have a strong influence on the MPEs (Figure 3), whereas the median proportional errors are closer to unbiased or negatively biased (median PE: mPRM 10.9 percent underestimate of status; CMSY 17.8 percent underestimate; COM-SIR 0.5 percent overestimate; and SSCOM 4.9 percent overestimate). It is important to note that the lower bound arises because the minimum values for both the true and estimated B/B_{MSY} values are zero. Therefore, the most that a method could underestimate is by 100 percent.

The APE encompasses both bias and precision with lower values being favoured in the same manner as a mean squared error. Again, CMSY had the lowest MAPE (0.579, implying that the average over or underestimate of CMSY was 57.9 percent), mPRM and COM-SIR had similar MAPEs (0.677 and 0.664, respectively), and SSCOM had the largest MAPE (0.776). In contrast, the median absolute proportional errors were very similar among CMSY (0.382), COM-SIR (0.365) and SSCOM (0.390), while mPRM had a larger (0.478) median absolute proportional error.

As judged by the mean of PE or APE, CMSY performed best. Conversely, COM-SIR had the lowest bias when a robust estimate of central tendency (median) was used. The decision of whether the mean or median is a better performance measure overall depends on whether one is concerned with the length of the tails. Because values in the tails were large overestimates of status (i.e. the estimate of B/B_{MSY} is large), a measure of central tendency that is responsive to these values (i.e. the mean) may be preferable from a precautionary management point of view. Under this assumption, the working group found that CMSY performed best. On the other hand, it is necessary to urge caution in the application of large values of estimated B/B_{MSY} because of the skewed pattern of these distributions, particularly in the absence of additional corroborating information.

FIGURE 3
Histograms of mean proportional error (MPE: left column) and mean absolute proportional error (MAPE: right column) across all stochastic simulations



Notes: The mean (white circle) and median (star) are plotted at the base of each histogram. Unbiased (PE=0) dotted line is shown for reference in the left column. Plots are truncated at an upper bound of 4; 1.1 percent of values exceeded 4. Means and medians are calculated on untruncated data.

BEST PERFORMANCE

Frequency of best performance

When MPE was used to assess performance, CMSY had the highest frequency of best performance (i.e. it was the best performer in 43 percent of scenarios; Table 10) followed by COM-SIR and mPRM with similar levels of best performance (21.7 percent and 22.1 percent, respectively). The SSCOM method had the lowest percentage of best performance (12.3 percent, Table 10). When judging performance by MAPE, CMSY remained the top-performing method (35 percent) and SSCOM increased to 18.4 percent (Table 10). When using MAPE or MPE over the last five years of the time series, CMSY was the best performer with mPRM a close second.

TABLE 10

Number of times a given method performed best as judged by mean absolute proportional error (MAPE) or mean proportional error (MPE) over all years

	All years		Last 5 years	
Method	**MPE**	**MAPE**	**MPE**	**MAPE**
CMSY	253 (43.9%)	202 (35.1%)	206 (35.7%)	204 (35.4%)
COM-SIR	125 (21.7%)	138 (24.0%)	129 (22.4%)	120 (20.8%)
mPRM	127 (22.1%)	130 (22.6%)	178 (30.9%)	185 (32.1%)
SSCOM	71 (12.3%)	106 (18.4%)	63 (10.9%)	67 (11.6%)

Notes: Total number of scenarios was 576. Percentage of best performance is given in parentheses.

Tile plots

All years

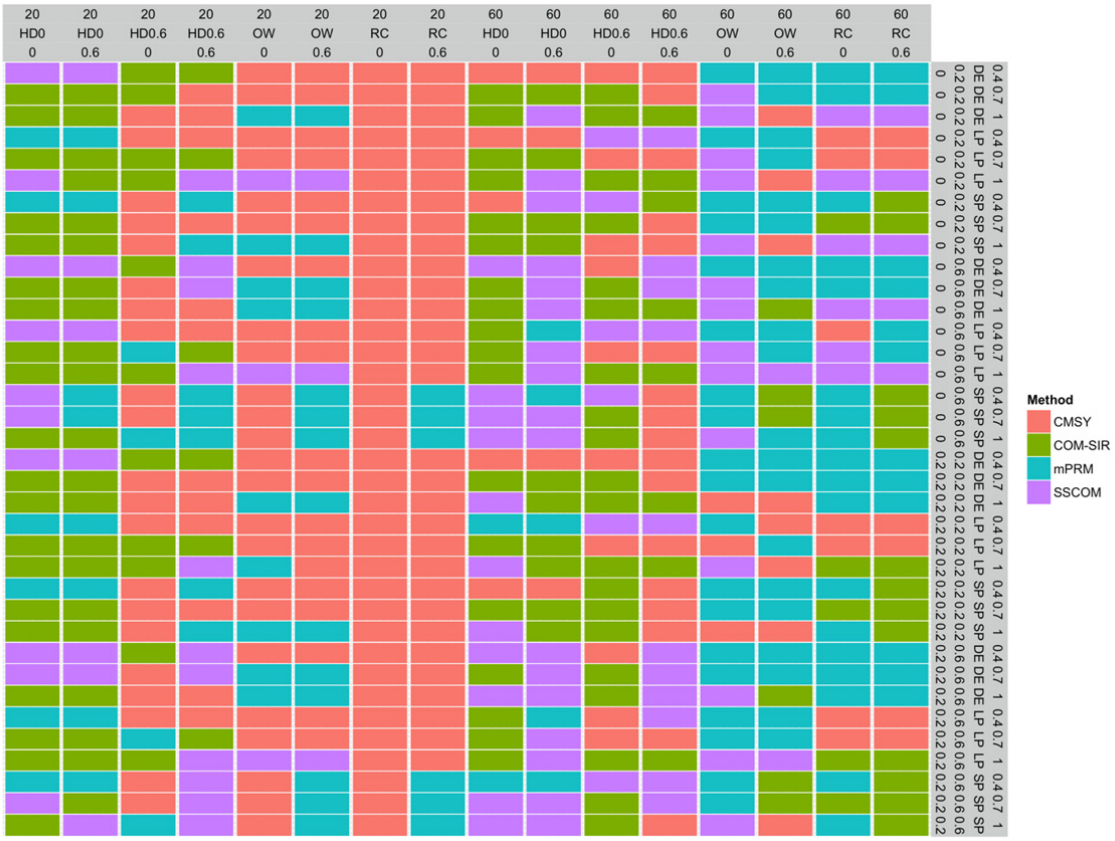
The best performance plots for MAPE and MPE illustrated very similar patterns (see Figure 4 for MAPE). Harvest dynamics and time series length played a dominant role in differentiating between scenarios. Overall, the CMSY method performed best for the short time series, with the exception of the flat harvest dynamics scenarios (HD:0). The COM-SIR method was the best performer under the short time series with flat harvest dynamics (with or without autocorrelation on recruitment variability). The CMSY model also performed best for scenarios with long time series, one-way trip harvest dynamics and no autocorrelation. With the addition of autocorrelation, mPRM performed better in more scenarios. The mPRM also performed best for long time series and roller-coaster harvest dynamics (with or without autocorrelation on recruitment variability). However, the relative performance of the models within each stochastic scenario (e.g. MAPE all years, Figure 5), illustrated the model space to be quite diverse. In other words, there are no models that dominate as the top performer within the majority of scenarios. Other metrics and time periods had a similar diversity.

Last five years

For MAPE over the past five years, the patterns were similar to the performance over the whole time series. However, the mPRM performed best more often for scenarios with long time series, one-way trip or roller-coaster harvest dynamics, and with or without autocorrelation.

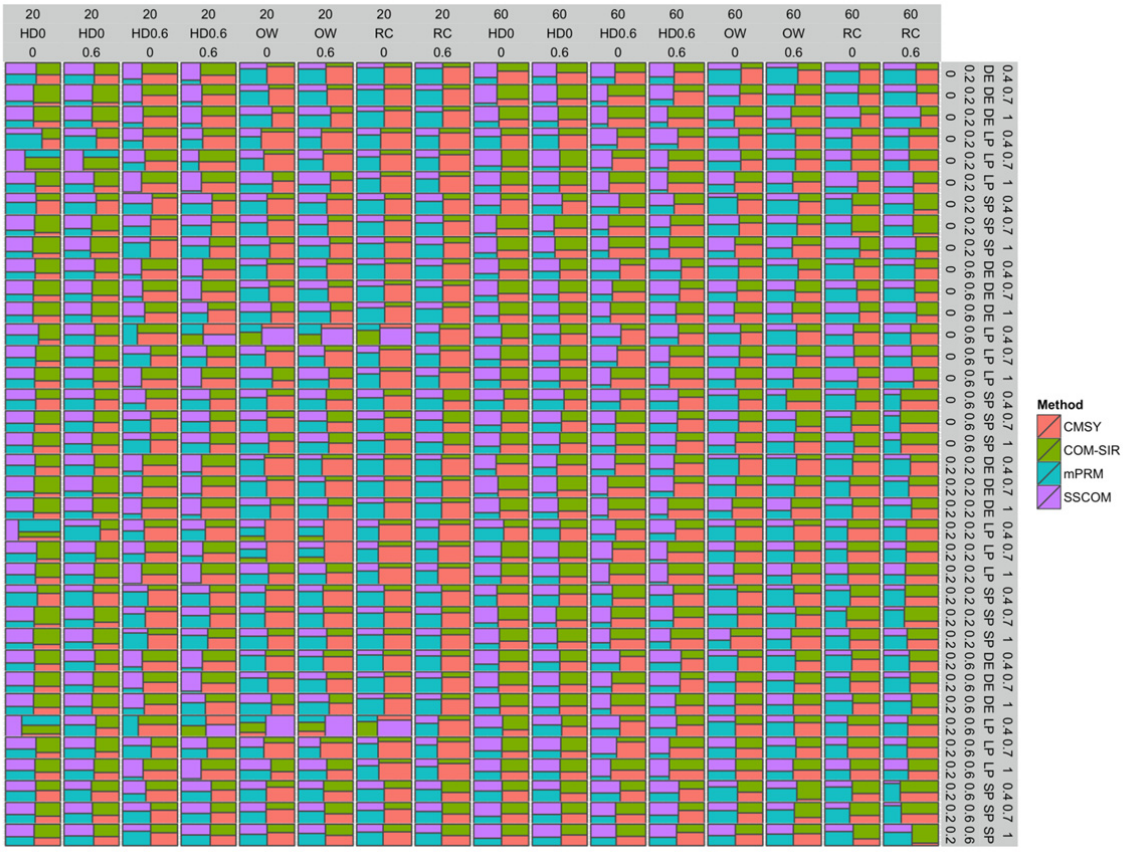
FIGURE 4

Best performance in each stochastic scenario according to the mean absolute proportional error (MAPE) statistic over the entire time series



Notes: The columns are time series length, harvest dynamics and autoregressive process error on recruitment variability. The rows are combinations of initial depletion, life history strategies, recruitment variability and measurement error on catch.

FIGURE 5
Relative performance in each stochastic scenario according to the mean absolute proportional error (MPE) statistic over the entire time series



Notes: Proportional area of the square occupied by a given method denotes how well that method performed relative to the others. The columns are time series length, harvest dynamics and autoregressive process error on recruitment variability. The rows are combinations of initial depletion, life history strategies, recruitment variability and measurement error on catch.

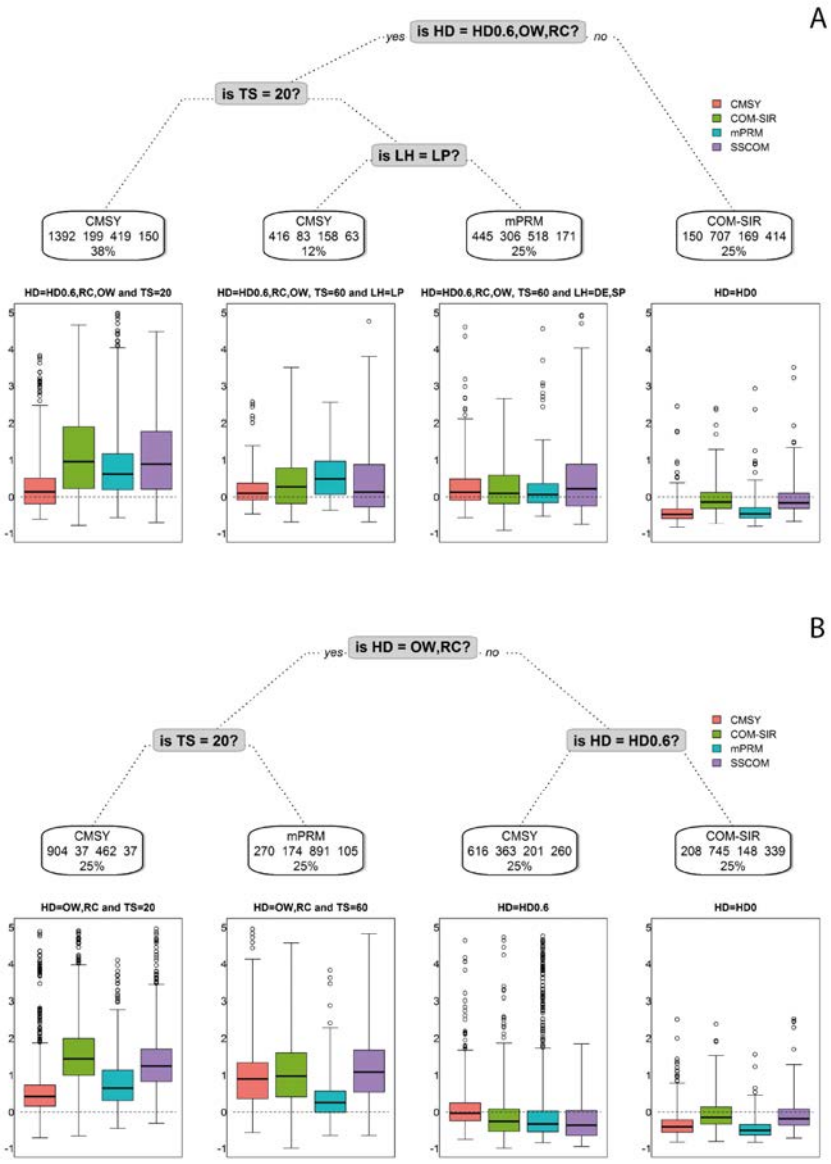
Performance maps

When looking at performance maps for both the full time series and the last five years, CMSY was the most frequent best performer. For example, in the case HD=OW or RC and short time series (TS=20), the median MPE (Figures 6 and 7) for CMSY was 10 percent whereas the next best performer, mPRM, was 60 percent. However, no model was clearly superior in all scenarios (Figures 6 and 7). Based on the boxplots of the final nodes, the "best performer" was not consistently the best in any meaningful way. Therefore, managers must ask whether the choice of model will improve accuracy or bias in a way that will actually affect the outcome. Thus, the best performer will produce less biased results than the other methods in some situations, while in other situations the methods will perform similarly. For example, the variability of MPE shown in the boxplots overlapped for all methods under HD0.6, OW or RC, TS=20 and LH=DE or SP scenarios. When looking only at the last five years, the analyses for MAPE and MPE produced identical trees (Figures 6 and 7, bottom panels).

For MAPE, the trees for the last five years and all years were quite similar (Figure 7), but not identical. However, the five-year and all years MPE trees (Figure 6) indicated that life history differentiates best performance for long time series with harvest dynamics other than HD0. Best performance differed between MPE across all years (Figure 6, top panel) and MAPE (Figure 7, top panel). A major difference between these statistics was the importance of initial depletion in

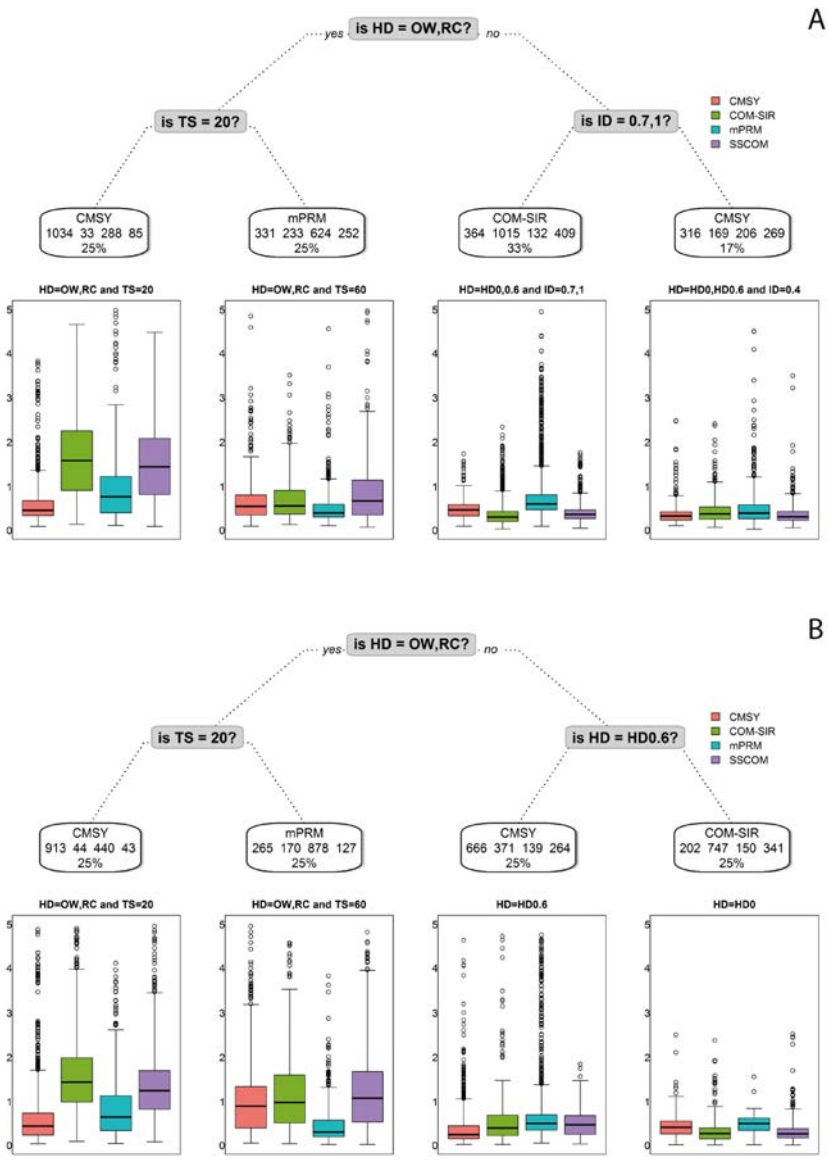
differentiating best performance for MAPE for HD0 and HD0.6. Initial depletion (ID) will affect the precision and bias of the estimate. However, this factor was likely to have less of an influence on the bias because the high, medium and low ID levels generally balanced out. The CMSY method performed best in cases of high ID, which is a natural consequence of the method having ID included in contrast to COM-SIR and SSCOM, which assumed no ID. In practice, therefore, CMSY performed better by relaxing the assumption that the biomass starts at carrying capacity.

FIGURE 6
Performance maps by mean proportional error (MPE)



Notes: Maps for the mean over the whole time-series of the PE (top) and mean over the last five years of the time-series (bottom) are shown. The method listed within each terminal node or box is the best performing model (i.e. the highest number of observations of MPE values close to zero). The numbers within each terminal node are the frequency that CMSY, COM-SIR, mPRM and SSCOM are the best performers, respectively. The percentages are based on the total number of observations that fall within each terminal node. The boxplots below each terminal node show the median (line), 1st and 3rd quartiles (box), two times the interquartile range (whiskers) and extreme values of MPE by method. Values of MPE higher than five are not shown (0.5 percent of all observations for top panel and 0.8 percent for bottom panel). HD = harvest dynamics (HD0 = harvest dynamics uncoupled with biomass, HD0.6 = harvest dynamics with high coupling with biomass, OW = one-way trip, RC = roller coaster), TS = time-series length (years), LH = life history (LP = large pelagic, DE = demersal, SP = small pelagic).

FIGURE 7
Performance maps by mean absolute proportional error (MAPE)



Notes: The mean over the whole time series of the APE (top) and mean over the last five years of the time series (bottom) are shown. Values of MAPE greater than five are not shown (0.5 percent of all observations for top panel and 0.8 percent for bottom panel). HD = harvest dynamics (HD0 = harvest dynamics uncoupled with biomass, HD0.6 = harvest dynamics with high coupling with biomass, OW = one-way trip, RC = roller coaster), TS = time-series length (years), LH = life history (LP = large pelagic, DE = demersal, SP = small pelagic), ID – initial depletion).

PERFORMANCE ACROSS MODELS

Harvest dynamics was the most important factor related to performance among the set of variables that were used in designing the simulations (Table 7). However, the clustering of the harvest dynamics differed depending on whether performance was measured across all years of the catch time series or only the last five years (the period more likely to be used for management advice). When performance was measured across all years (Figure 8), HD0, HD0.6 and OW clustered together and

separated from RC. However, the harvest dynamic OW did separate from HD0 and HD0.6 at the next break.

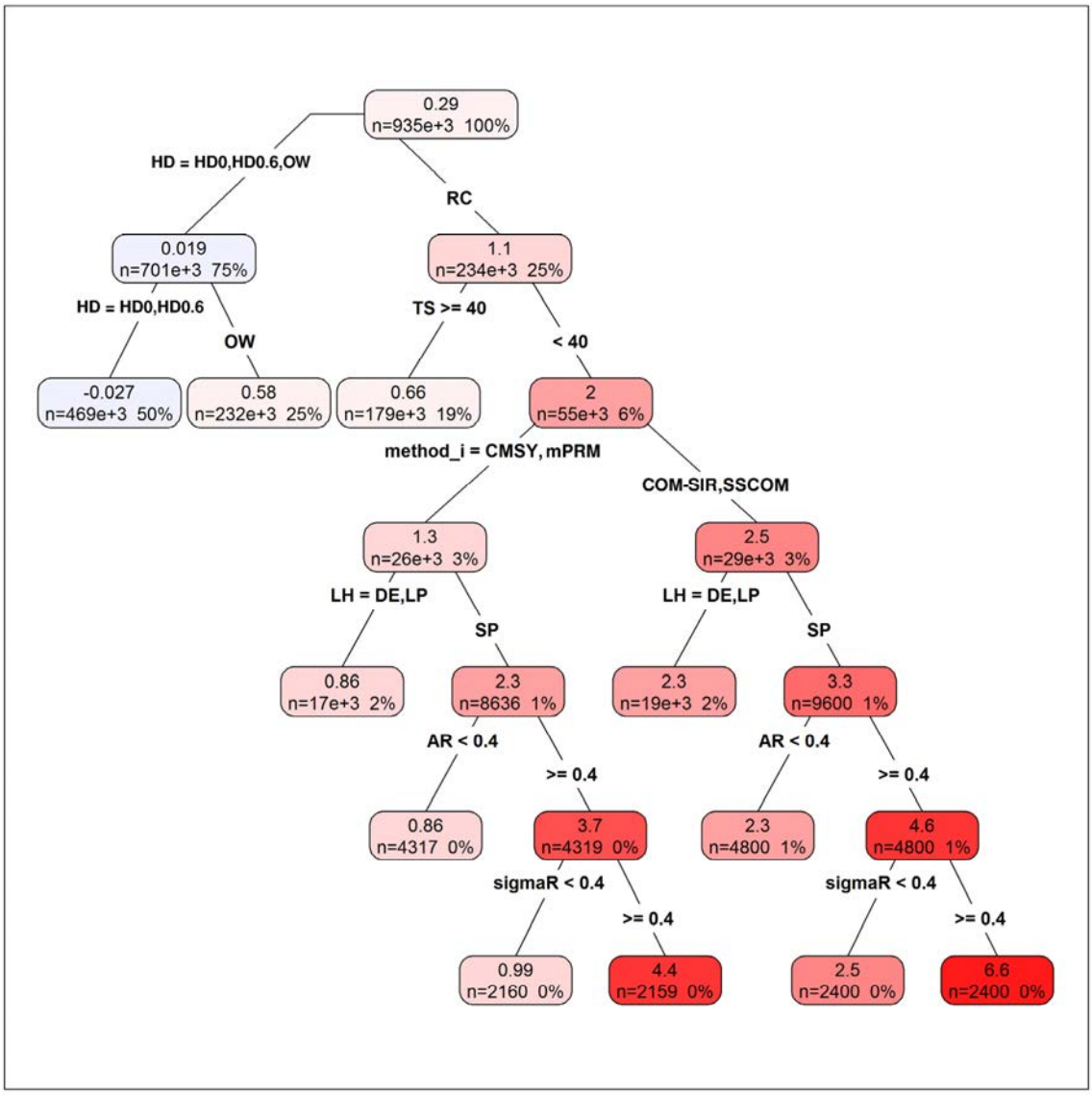
In contrast, when measuring performance over only the last five years (Figure 9), HD0 and HD0.6 clustered together and separated from OW and RC, which clustered together and did not separate. HD0.6 had a strong coupling between biomass and harvest dynamics and was therefore expected to allow for improved estimates of biomass dynamics based on catch data. The fact that HD0.6 and HD0 harvest dynamics explained such a large portion of the variability in all regression trees (i.e. regardless of statistic or whether one looks at all years or the last five years) can likely be attributed to several factors. First, there was a strong simulated coupling between biomass and harvest dynamics in the HD0.6 scenarios, which was an inherent assumption of the COM-SIR and SSCOM models. This is probably why COM-SIR performed well in the HD0.6 scenarios (see the COM-SIR section in "Determinants of performance for each of the four assessment models" below). On the other hand, in the flat harvest dynamics (HD0) scenarios, the true B/B_{MSY} was sometimes relatively flat and rarely dropped below a value of one (resulting in an approximately constant time series of catch). Because there is very little information in a flat catch series (except for strong environmental signals, such as small pelagics with strong autocorrelation) for many of the models, the B/B_{MSY} estimates were typically flat and varied around a "default" level, ranging from zero to carrying capacity. Therefore, with B/B_{MSY} estimates greater than one, the proportional error was generally bounded between –1 and 1 for the HD0 case, which may be an artefact of the selection of the a parameter in the harvest dynamics model (Equation 8). If a higher constant value for harvest dynamics had been used, it would be expected that true B/B_{MSY} could range much lower than one, resulting in a much larger PE.

Time series length was the second branch in the regression trees across all years (Figures 8 and 10). In the last five years (Figures 9 and 11), life history was the next most important variable after harvest dynamics with large pelagic and demersal life histories always clustering together and separately from small pelagics. When looking either at PE or APE, for the last five years, recruitment variability branched on the third tiers, which explained a proportion of the variability for the small pelagic life-history scenarios.

The working group observed differences between the regression trees for PE and APE when the working group considered the entire time series. On the third tier of the PE regression tree for all years (Figure 8), the choice of assessment model explains some of the variability tree, which only applies to a small percentage of the scenarios. Unlike the performance maps that will always select a best model regardless of the degree of relative performance, model explained the variability in only 6 percent of the scenarios. Both CMSY and mPRM (both of which currently rely on *a priori* information) clustered together and have a lower MPE than COM-SIR and SSCOM (1.3 vs. 2.5), which also clustered together and do not currently rely on any *a priori* information. After breaking on model, the remainder of the tree was identical in structure, regardless of the model cluster. In contrast, for APE for all years (Figure 10), life history explained the variability in 6 percent of the scenarios, and choice of model was not a factor.

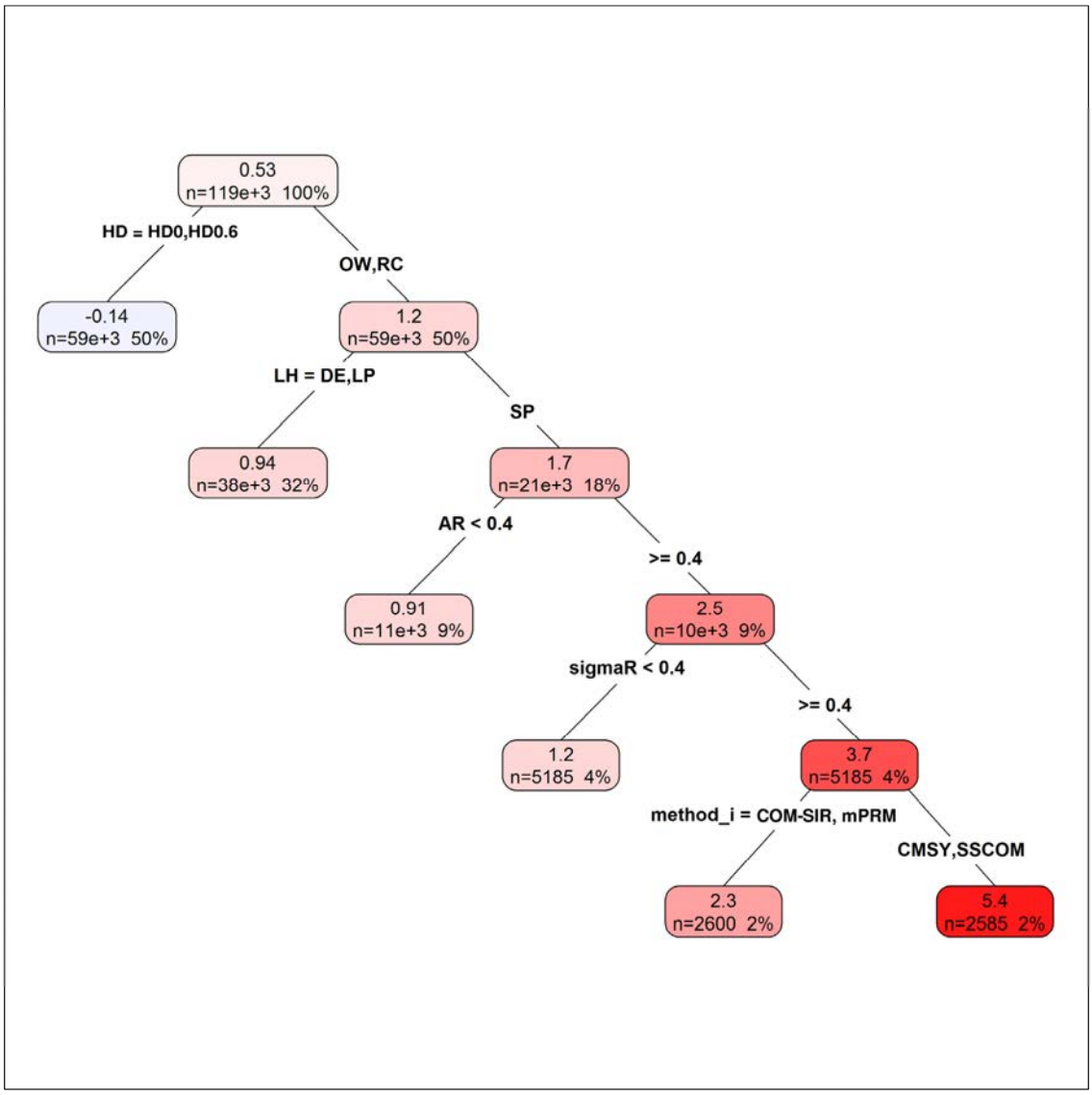
The fact that which model explains less variability (or does not show up as a factor) in the regression trees makes sense when the histograms are considered (Figure 3). In general, the regressions are based on the weight of these distributions and the mean values for each model, all of which are quite similar (i.e. the tails do not have a strong influence). Therefore, while choice of model may be important, the regression trees pick up on the factor that is contributing the most to the variability in the scenarios (i.e. harvest dynamics).

FIGURE 8
Regression tree of proportional error (PE) across all methods for all years



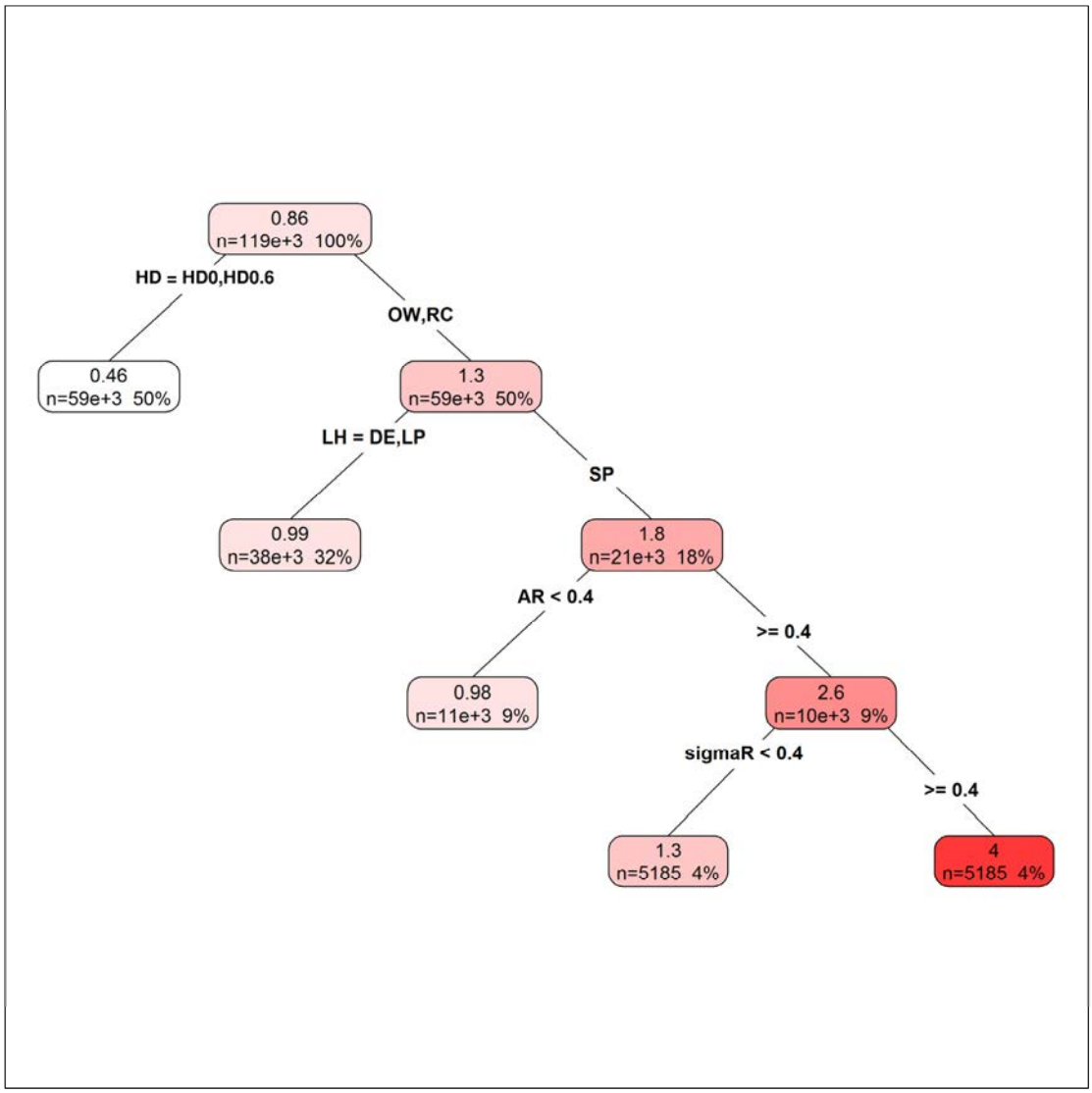
Notes: The top number in each box is the average PE for a set of simulation scenarios (i.e. the averaged PE across all methods and simulations was 0.29). The numbers in the second row of the boxes list the number of data points and percentage of simulation scenarios in that set (i.e. the top box has 9 350 scenarios representing 100 percent of the scenarios), and each box either has no boxes below it (i.e. it is a terminal node), or has two boxes below it (i.e. it has additional branching). The percentages in each box of a single tier sum to 100 percent.

FIGURE 9
Regression tree of proportional error (PE) across all methods for the last five years



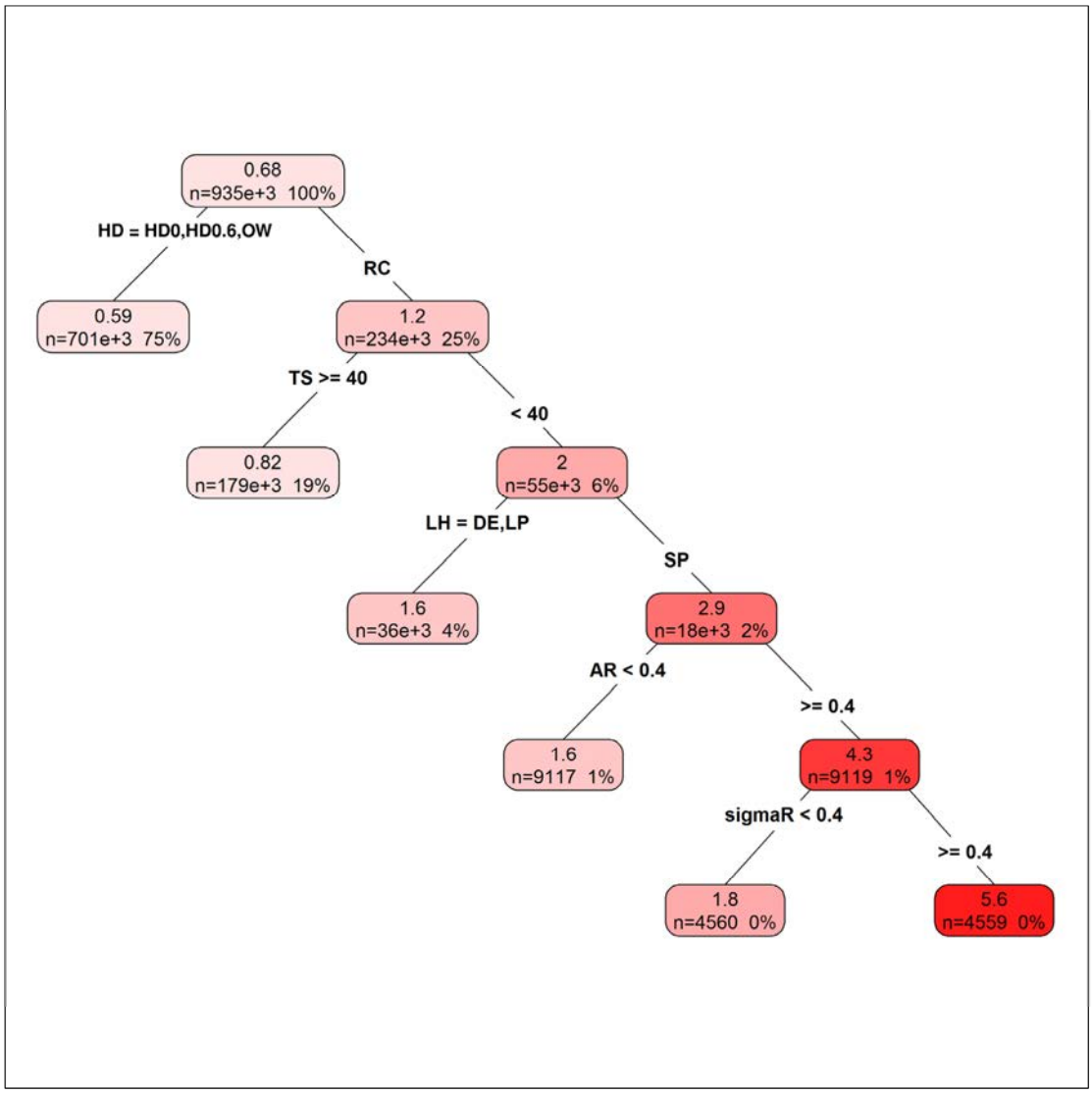
Note: See Figure 8 caption for tree interpretation.

FIGURE 10
Regression tree of absolute proportional error (APE) across all methods for all years



Note: See Figure 8 caption for tree interpretation.

FIGURE 11
Regression tree of absolute proportional error (APE) across all methods for the last five years



Note: See Figure 8 caption for tree interpretation.

DETERMINANTS OF PERFORMANCE FOR EACH OF THE FOUR ASSESSMENT METHODS

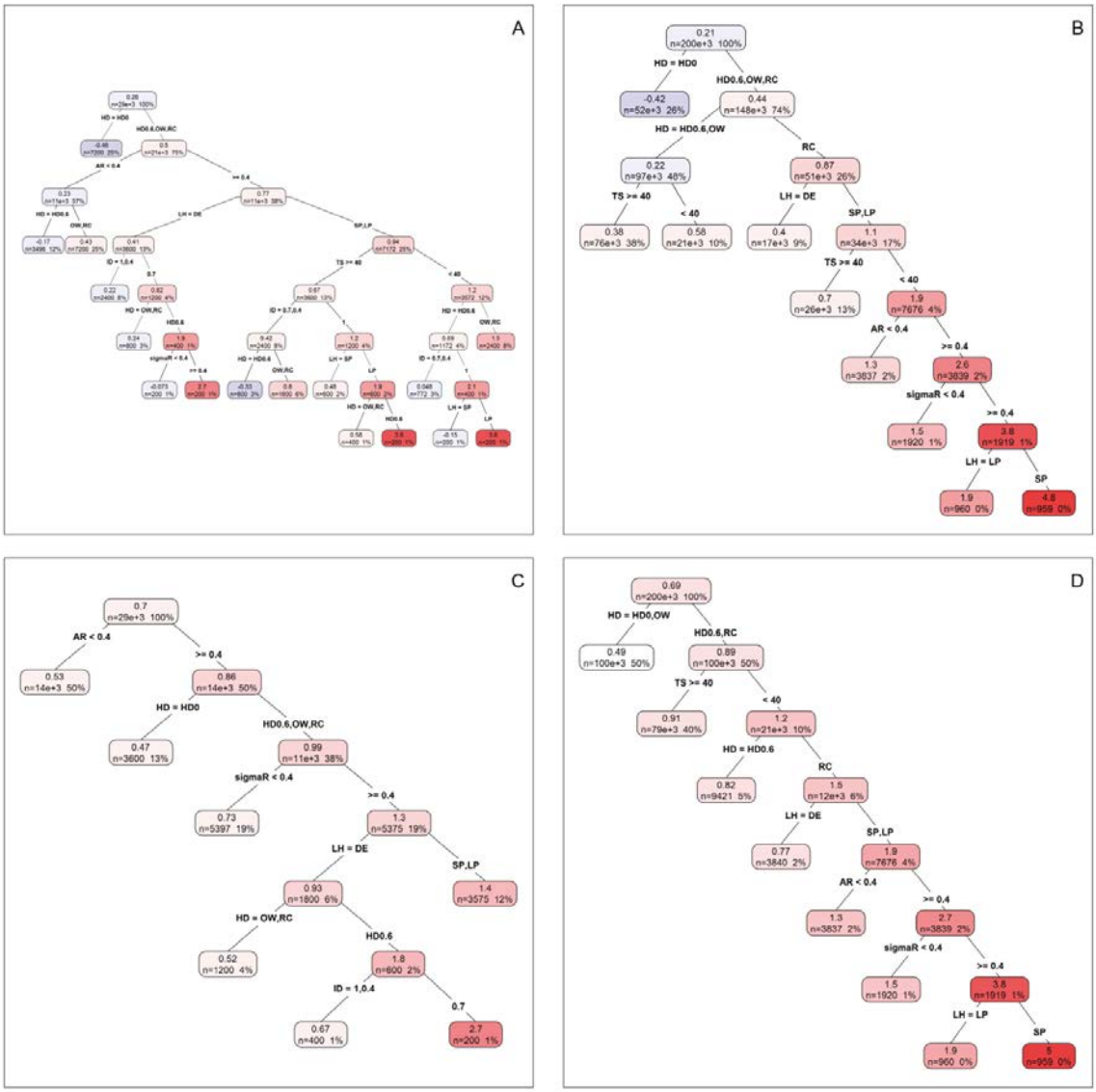
Modified panel regression model (mPRM)

Harvest dynamics was the most important explanatory variable affecting the performance of the mPRM method for proportional error and absolute proportional error for the last five years and for the whole catch time-series (Figure 12). The harvest dynamic grouping was different compared with other methods. For example, for PE in the last five years and in all years (Figure 12 A and B), the constant harvest rate (HD0; e.g. decoupled harvest dynamics as in a bycatch fishery) split from the other cases: HD0.6 (tightly coupled harvest dynamics), OW (one-way trip), and RC (roller coaster). For APE, the strength of autocorrelation in recruitment was the most important variable explaining performance over the last five years (Figure 12 B and 12 D). Autocorrelation was of secondary importance over the last five years for PE as well (Figure 12 A), although performance was better for low autocorrelation.

Several variables included in the mPRM could be affected by strong autocorrelation. One possibility is that the initial slope of the catches (an explanatory variable in mPRM) could be biased by strong autocorrelation, resulting in the model over- or under-estimating status. The performance of the mPRM also relies on a relationship between recent historical catches and current biomass. It is possible that strong autocorrelation results in current biomass being more influenced by recent strings of especially strong or weak recruitment events than by historic catches, thereby reducing the predictive power of the mPRM.

Figure 12

Modified panel regression method (mPRM) regression trees for: (A) proportional error (PE) last five years; (B) PE all years of the catch time series; (C) absolute proportional error (APE) last five years; (D) APE all years in the catch time series

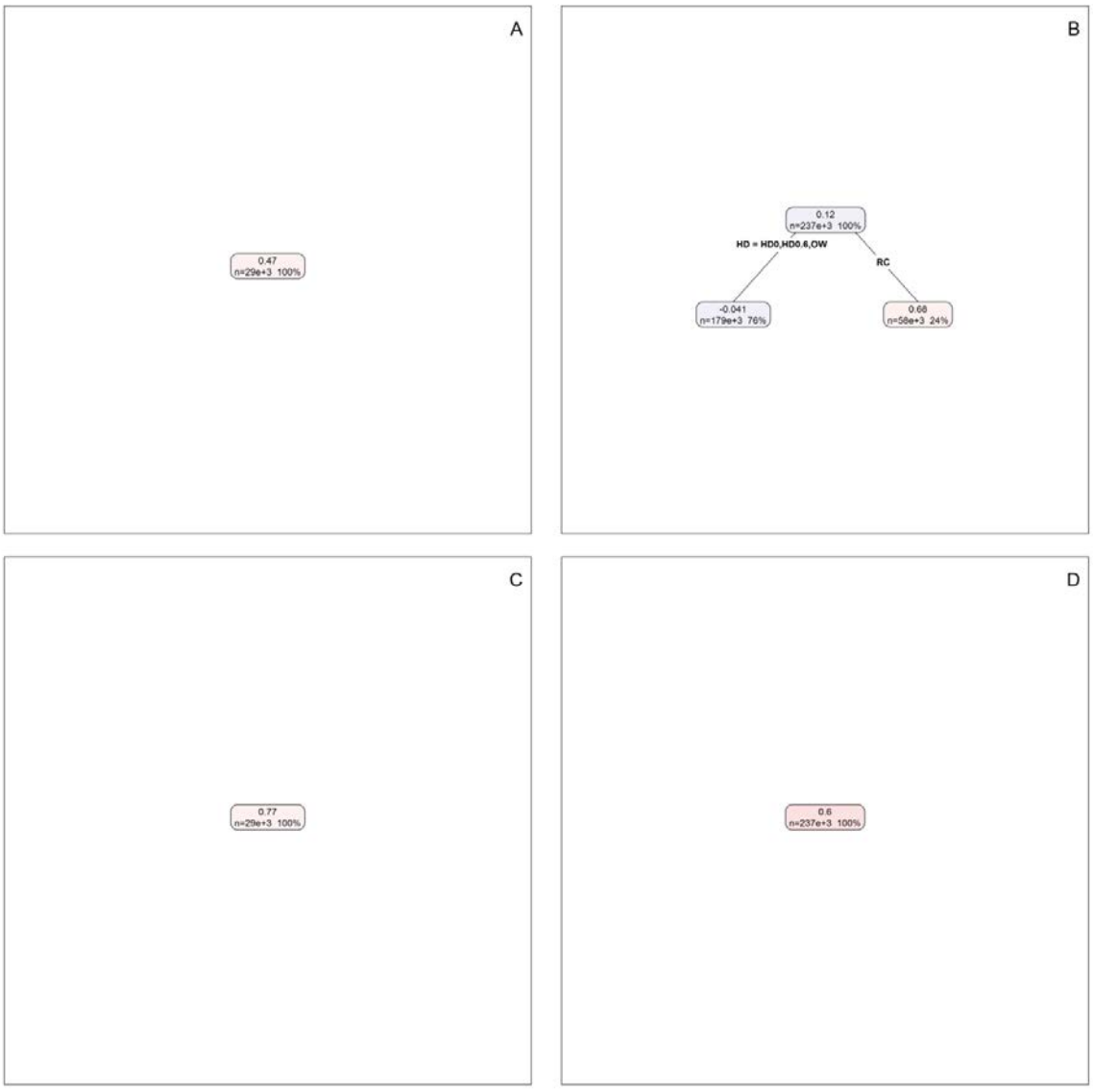


Notes: Higher branches denote greater explanatory power. Mean level per group is shown in each node with the number of years and percentage of years.

Catch-MSY (CMSY)

No splits occurred in the regression trees of APE for either all years in the catch time series or the last five years (Figure 13 A and B) or PE for the last five years for CMSY (Figure 13 C). This indicated that no variables appeared to affect performance consistently in these cases. Harvest dynamics was the only variable affecting performance in PE for all years (Figure 13 B) with HD0, HD0.6, and OW harvest dynamics grouping and splitting from RC harvest dynamics. The performance for RC was considerably poorer (Figure 13 B). In the RC scenarios, the assessment methods overestimated the relative biomass (B/B_{MSY}) by an average of 68 percent. CMSY behaved similarly to SSCOM (below) for this scenario in that recovery of status is difficult when RC harvest dynamics were present. It appears that, in this case, the models confused low catches at high fishing effort on a low-biomass stock with low fishing effort and a high-biomass stock.

FIGURE 13
Catch-MSY (CMSY) regression trees for: (A) proportional error (PE) last five years; (B) PE all years of the catch time series; (C) absolute proportional error (APE) last five years; (D) APE all years in the catch time series



Notes: Higher branches denote greater explanatory power. Tree interpretation is the same as that for Figure 12.

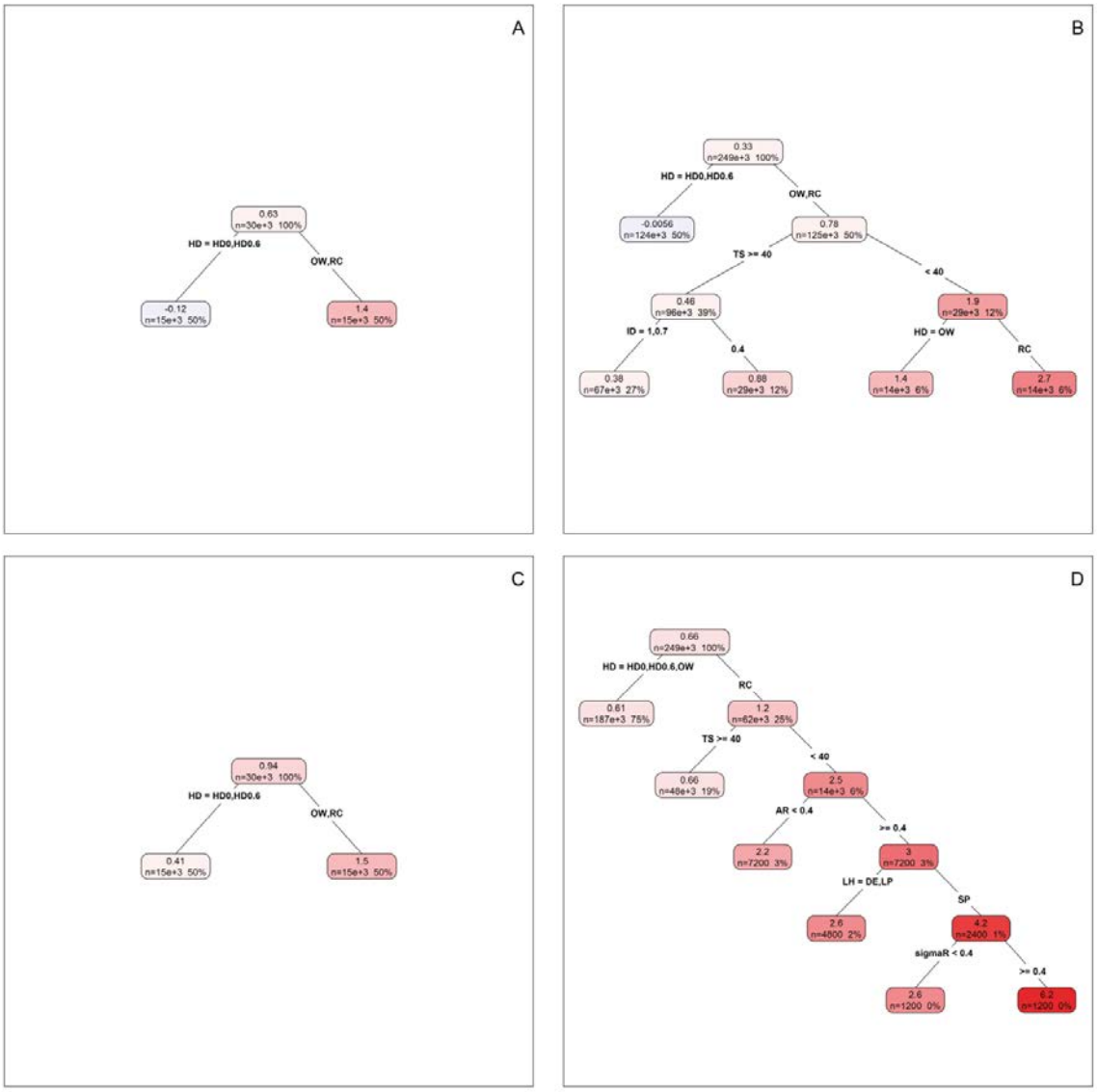
Catch-only model (COM-SIR)

Harvest dynamics was the main variable that affected the performance of the COM-SIR (Figure 14). For PE in the last five years (Figure 14 A) and all years of the catch time series (Figure 14 B) and APE in the last five years (Figure 14 C), the main split was between HD0/HD0.6 and RC/OW, with HD0/HD0.6 having better performance. COM-SIR had lower APE for OW harvest dynamics over 60 years (Figure 14 D) than over the last five years, hence the split with HD0 and HD0.6 (Figure 14 C). Time series length (20 versus 60 years) was the second-most important variable as judged by PE and APE across all years only. Neither time series length nor any other variables other than harvest dynamics affected performance over the last five years in COM-SIR.

The grouping of HD0/HD0.6 requires explanation as to why those scenarios resulted in better performance for COM-SIR. In the HD0.6 scenarios, there was a strong simulated coupling between biomass and harvest dynamics, which was an inherent assumption of COM-SIR. Therefore, it is not surprising that the model performed well in that situation. For flat harvest dynamics (HD0), the true B/B_{MSY} was relatively flat and rarely dropped below a value of one, resulting in an approximately constant time series of catch. The catch time series for HD0 was often flat. As there was very little information in a flat catch series (except for strong environmental signals, such as small pelagic species with strong autocorrelation), COM-SIR estimates were also typically flat varying around a default level for the model. The default level ranges from zero to carrying capacity, so with B/B_{MSY} typically greater than one, the PE was mostly bounded between –1 and 1 for the HD0 case. If a higher constant value for harvest dynamics had been used, one would have expected that true B/B_{MSY} could range much lower than one, resulting in a much larger PE. For COM-SIR, the conclusion is that the grouping of HD0/HD0.6 into superior performance was largely an artefact of the value chosen for constant harvest dynamics. Note that the true biomass ratio can dip well below one for small pelagic species with strong autocorrelation in residuals; however, this pertains to a relatively small number of scenarios.

Time series length affected performance (as judged by PE and APE) for COM-SIR in the full time series analysis in that the methods performed considerably better over a longer time series for OW and RC (Figure 14 B and D) because the model had a longer period of time to estimate the dynamics in these cases.

FIGURE 14
Catch-only model (COM-SIR) regression trees for: (A) proportional error (PE) last five years; (B) PE all years of the catch time series; (C) absolute proportional error (APE) last five years; (D) APE all years in the catch time series



Notes: Higher branches denote greater explanatory power. Tree interpretation is the same as that for Figure 12.

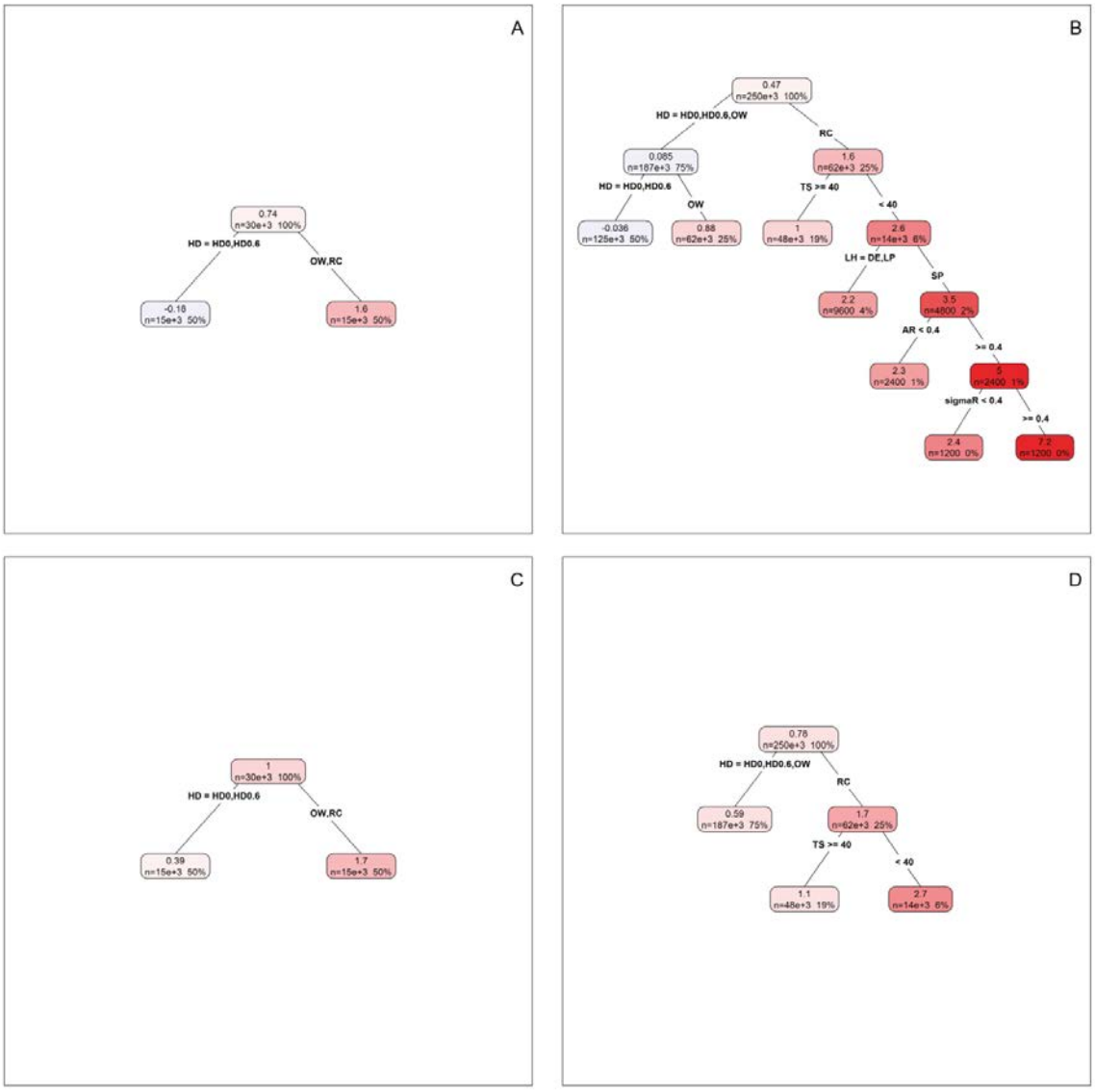
State–space catch-only model (SSCOM)

Harvest dynamics was again the main variable influencing the performance of SSCOM (Figure 15) (i.e. how the fishery responds to changes in biomass). The main grouping was the same for PE and APE for the last five years (Figure 15 A and C) and APE and PE for all years (Figure 15 B and D). The grouping differed within performance metric, e.g. PE had different grouping depending on whether the last five years or all years were analysed (Figure 15 A and B). For all years for both PE and APE, HD0, HD0.6, OW grouped separately from RC, with RC having the lower performance (Figure 15 B and D). For the last five years, HD0 and HD0.6 harvest dynamics grouped separately from OW and RC. The explanation for better performance of HD0 and HD0.6 is the same as that for COM-SIR presented above. The grouping of HD0, HD0.6, and OW for all years (Figure 15 B and D) appears to be driven by the very poor performance of the method for RC harvest dynamics, which formed the main split. For PE for all years, the HD0, HD0.6, and OW group further split into an

HD0/HD0.6 group separate from OW at the second branching. For the RC harvest dynamics branch, longer time series resulted in considerably better performance (Figure 15 B and D) because the model had a longer period of time to estimate the dynamics in these cases.

FIGURE 15

State–space catch-only model (SSCOM) regression trees for: (A) proportional error (PE) last five years; (B) PE all years of the catch time series; (C) absolute proportional error (APE) last five years; (D) APE all years in the catch time series



Notes: Higher branches denote greater explanatory power. Tree interpretation is the same as that for Figure 12.

4. DISCUSSION

This analysis has shown that combinations of catch data and ancillary information (e.g. on fishery dynamics or the life history of the target species) can provide preliminary estimates of relative biomass for exploited fish stocks globally, including many stocks that are currently unassessed. Although there can be significant error in the estimates and there is no one-size-fits-all approach, the working group has identified the Catch-MSY (CMSY) method as an overall best performer in estimating exploitation status for fish stocks globally, within a region or national jurisdiction. The analysis provided here for four data-limited methods can guide application for estimating the exploitation status of fisheries.

Testing of different plausible data scenarios in a full factorial simulation framework was a critical part of the working group's efforts, and the working group strongly advocates that any proposed methods for determining fisheries status undergo such rigorous testing before application. To that end, the simulation framework is available online at ftp://ftp.fao.org/FI/STAT/Rfiles/C1086.zip to the broader fishery science community so that it can be used to assess the performance of other data-limited and potentially data-rich methods that were not tested here.

As illustrated with the methods presented here, any extra information on the dynamics of a fishery or the productivity of a fish stock may improve estimation if more informative priors for r and K can be determined. Thus, it is useful to note that the methods tested are flexible and can readily incorporate additional information such as priors for parameters like r and K or direct indices on harvest dynamics. Moreover, they can be extended to more sophisticated models as more information becomes available. It is also clear from the results that efforts to improve catch data, as well as better information on historical depletion and life-history parameters, will result in better estimates and should be actively pursued by national and international fishery science institutions, in cooperation with fishers and the fishing industry where possible.

Although most of the models performed similarly (Figure 3), the CMSY model performed better as measured by the two metrics used in this study (i.e. MPE and MAPE; Table 10). However, if the medians of these statistics are considered, COM-SIR performed very well, and could be considered a good alternative choice. Using the more detailed analyses of performance (e.g. regression trees) can help users consider which specific factors in the factorial design are most helpful in guiding application for a particular set of stocks or a particular region. Model choice was particularly important when considering bias in short time series where management had affected harvest rates. CMSY was more effective in estimating status over short time scales (five years), except for relatively flat harvest dynamic scenarios, presumably because it used more-informative prior distributions than the other models, and thus had more information given short time series. Therefore, CMSY may be more suitable for fisheries in developing countries where data collection programmes have recently been implemented and only short time series of catches are available. It is important to note that the working group did not formally test these models for rebuilding scenarios, and it is possible that these types of situations (e.g. where catches remain low due to management interventions while biomass increases) would reduce performance. Across all models, harvest dynamics was the most important explanatory variable determining performance, which emphasizes the importance of having accurate information on fishing effort and total removals.

Generally, the models perform relatively well as a group in determining B/B_{MSY}. Histograms illustrate that all of the models were positively biased in estimating B/B_{MSY} (Figure 3). Although there are long tails on the distributions, the mean and in particular the median (PE and APE were very close to unbiased. The long tails indicate large under- or over-estimates of B/B_{MSY}. Estimates in the tails of the distributions should be viewed with caution when judging performance based on the mean or median PE or APE. However, it is difficult to determine whether a single B/B_{MSY} value for a particular stock is itself a "tail". Overall, the working group used the mean of the performance metrics rather than the median in judging performance here because the mean may be more responsive to values in the tails and, therefore, preferable from a precautionary management point of view.

Although the simulated fisheries used here provide a clear and robust method for testing assessment performance, applying data-limited assessment models to real data is a logical next step in verifying their abilities and limitations. In addition, the simulated data used to assess the data-limited models originated from single, non-interacting populations. As such, the definition of a stock is a single population unit. The application of these models to catch data derived from aggregations or subsets of populations (as might occur in some global databases such as the FAO catch database) has not been tested.

With one exception, the models tested were based on Schaefer surplus production curves, but others might be more suitable, e.g. Pella-Tomlinson (Thorson, Branch and Jensen, 2012). From the outset, these models were not being tested as management tools, but as tools for determining stock status. As such, the working group did not test these models within a management strategy evaluation framework to determine under what conditions these data-limited approaches might adequately perform as management tools for individual fisheries. This is an important follow-up activity in the ongoing development of these models.

By making all code freely available at ftp://ftp.fao.org/FI/STAT/Rfiles/C1086.zip, the methods can be considered transparent. In addition, the working group attempted to use best practices in simulation testing. Therefore, the simulation developers remained independent from the working group members who developed the models in order to ensure that the simulations were not influenced by the models and vice versa. Correspondence between these groups was limited to checks that the models were being applied and tested accurately.

Although the characteristics of the simulated fish populations could be regarded as skewed towards "temperate" species, the code can be easily adapted to apply to other types of stocks. Future work can build on the results, including testing the performance of these models on a regional basis to see whether some models perform better under certain geographic or environmental conditions. In addition, it will be important to test performance on specific stock types, such as tropical reef fish, deep-sea fish, invertebrates or highly migratory transboundary stocks in further applications. As indicated above, comparing the status estimates from these models with FAO stocks or with the stocks in the RAM Legacy database will also provide key information on model performance. Future work may also use meta-analytic techniques to develop priors for harvest dynamics, initial depletion, and final depletion, either using assessed stocks, local and traditional knowledge, or survey data where available.

Even with these limitations, the results suggest that the evaluated models are capable of capturing broad patterns in relative biomass using very basic data. These models are not meant to replace data-rich stock assessments, but can, instead, provide guidance in regions or for stocks where it is not yet possible to implement a full stock assessment. Future research could seek to develop control rules that can use these models in real-world management scenarios (Wiedenmann, Wilberg and Miller, 2013). Some of the models applied here can contribute to improving the quality of all assessment models via the transfer of ideas such as explicit harvest dynamics and state–space analyses of harvest dynamics. In addition, full factorial simulation testing is rare even within data-rich scenarios, and the approach adopted can provide a model for future efforts. The models tested are not a perfect solution or a replacement for robust data collection or stock assessment methods. More knowledge on fishery and fish stock dynamics is crucial if fishers want to maximize yield. Because these approaches are not precise management tools, if fishers choose a riskier approach, it must be offset with greater investment in data collection and knowledge creation. On their own, these models are not designed to be used to set specific management tactics (e.g. quotas) for individual fisheries. Currently, these models are best used to assess broad patterns and highlight key regions or fisheries that need more detailed assessment and management.

PART 2

FISHERY PRODUCTION POTENTIAL

5. INTRODUCTION

Many experts have attempted to define the fishery production potential of marine systems based on energetic considerations (Kesteven and Holt, 1955; Graham and Edwards, 1962; Schaefer, 1965; Moiseev, 1969, 1994; Ricker, 1969; Ryther, 1969; Gulland, 1970, 1971). Bottom-up control of fish production has now been demonstrated in many regions of the world ocean (Ware, 2000), supporting the general approach of tracing pathways involved in the translation from primary production to fishery yields. The ability to estimate primary production was revolutionized by Steeman-Nielsen's (1951) development and application of the ^{14}C method, paving the way for elaboration of simple models of energy flow from the base of the food web to fish production.

Earlier estimates of annual fishery production potential based on energetic principles utilized estimates of primary production over all phytoplankton size classes, inferred ecological transfer efficiencies from laboratory experiments and other observations, and observed or assumed levels of the mean trophic level (MTL) of the catch. Kesteven and Holt (1955) laid out this general strategy. Using simple models of energy flow, Graham and Edwards (1962) provided an estimate of potential global fish yield of 115 million tonnes for bony (teleost) fishes for conventional fisheries. In contrast, their estimate of potential yield based on extrapolations of catch histories in space and time was less than half this value (55 million tonnes). Schaefer (1965) applied somewhat higher estimates of transfer efficiencies and estimated the potential yield to be about 200 million tonnes. Ricker (1969) followed with a projection of about 150 million tonnes. In a widely cited evaluation, Ryther (1969) estimated the world fish production potential to be about 100 million tonnes. This study was the first to apply a partitioning of fishery production potential among different oceanic domains including coastal, offshore, upwelling and open ocean systems. Ryther (1969) further applied different estimates of food chain length in these different system types to reflect fundamental differences in ecosystem structure and patterns of energy flow. At the time of Ryther's projection, the global marine fish catch was approximately 60 million tonnes (Ryther, 1969). The landings from marine capture fisheries are now approximately 85 million tonnes (FAO, 2012), approaching Ryther's estimate after accounting for discarded catch. The differences in Ryther's estimates from those of Graham and Edwards (1962), Schaefer (1965) and (Ricker, 1969) reflect, in part, different characterizations of the harvestable component of marine ecosystems owing to varying assumptions about economic efficiency and availability to harvest.

An extensive FAO initiative to estimate global marine fishery production potential (Gulland, 1970) based on an extrapolation of catch trends, an extrapolation from moderately or heavily fished to lightly fished regions, and food web considerations resulted in an estimate of 100 million tonnes for conventionally harvested species and up to 260–350 million tonnes if species currently lightly exploited or unexploited (including krill, mesopelagic fish species, squid and others) were included. An overall reliance on key elements of the analysis such as transfer efficiencies and MTL of the catch, which were characterized by high levels of uncertainty in food-web-based analyses, led Pauly (1996) to infer that the concordance of Ryther's estimates (1969) with current observations may largely reflect countervailing errors (i.e. the answers may be right for the wrong reasons).

One objective of the present analysis is to more fully characterize uncertainty in the elements of production and to assess the overall uncertainty in global estimates of fishery production potential and exploitation rates at the ecosystem level. The working group therefore estimated the fishery productivity for LMEs for which satellite-derived estimates of primary production could be derived. The working group also updated earlier estimates using satellite-derived values for primary production partitioned into two major phytoplankton size classes: microplankton (> 20 μm) and nano-picoplankton (< 20 μm). The former includes diatoms and larger dinoflagellates while the latter includes smaller flagellates, autotrophic bacteria, etc. Relative to earlier estimates, this approach permitted fuller consideration of the microbial food web and related energetic pathways. The working group also assembled estimates of ecological transfer efficiencies from a large number of energy flow network models and used these estimates to characterize uncertainty. In all cases, the working group stratified the primary production estimates by designated LMEs for continental shelf regions. Within LME boundaries, the working group further stratified by depth (> 300 m and < 300 m) to reflect

overall differences in production and in fishery characteristics. The working group focused on these regions for detailed analysis because of the dominance of the LME regions in world fisheries production and the availability of more extensive ecological information for the continental shelves and adjacent areas of the world ocean (Christensen *et al.*, 2009). Approximately 80-90 percent of the global fish catch is derived from the LME regions of the world ocean (Christensen *et al.*, 2008).

It is essential to assess production potential at different trophic levels in order to evaluate exploitation status at an ecosystem level. This work stream is intended to complement the analysis of the species- and stock-based approach in the first part of this report. Here, the focus is not on the status of individual species or stocks, but on the system as a whole. Ultimately, production at the population level is connected to, and controlled by, broader considerations of production at the ecosystem level. Energetic constraints place clear limits on the overall fishery production potential of the global ocean.

METHODS AND DATA SOURCES

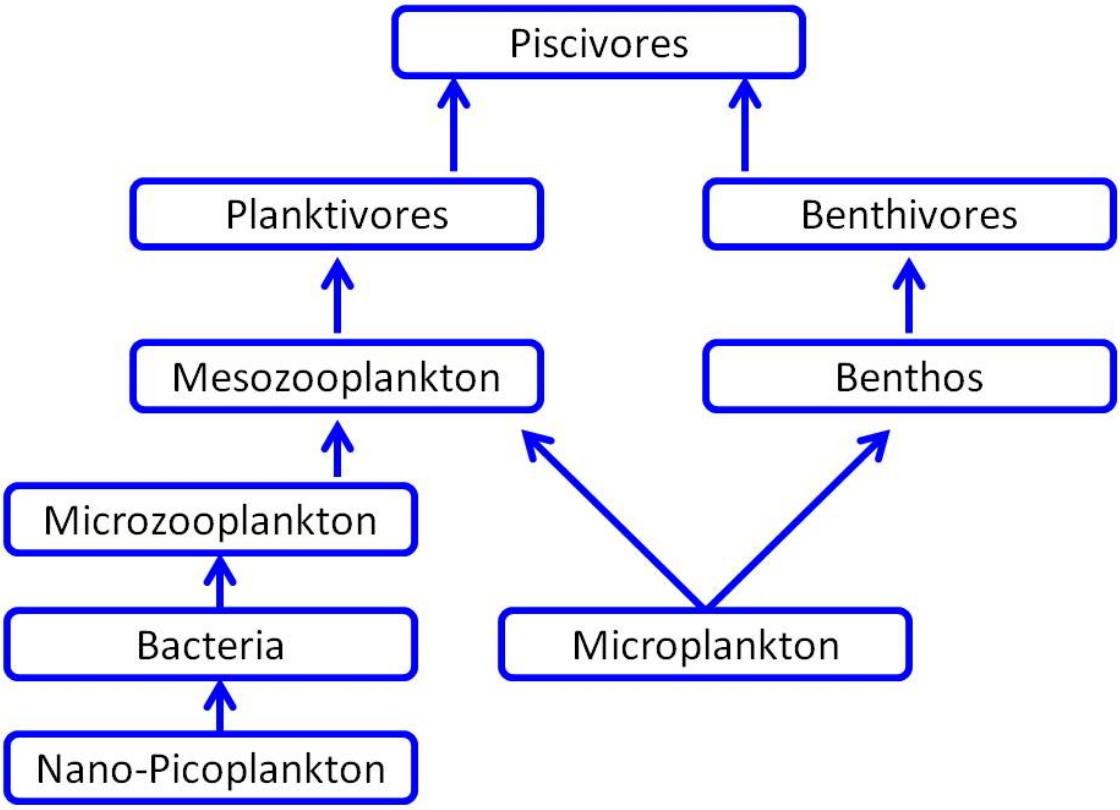
Ecosystem network models have now been applied for all the LMEs considered in this report using the well-known Ecopath with Ecosim (EwE; Christensen *et al.*, 2008; Christensen *et al.*, 2009) formulation based on the original developments by Polovina (1984) and Christensen and Pauly (1992). Here, the attempt is made to complement these analyses by using a simple and broadly applicable characterization of fishery production systems. The approach entails projections of available production at different trophic levels given information on estimates of primary production. This method is therefore consistent with the earlier analyses noted above (Graham and Edwards, 1962; Schaefer, 1965; Ricker, 1969; Ryther, 1969). The working group expanded the implicit food chain approach in these analyses to a very simple but broadly applicable food web model (see Figure 16). The working group also specified removals from discrete ecosystem components (including benthos, planktivores, benthivores and piscivores) to more fully characterize fishery dynamics directed at different functional groups, often by different fleet sectors. However, the working group ignored potential production coming from detrital or demersal primary production, as it was not possible to obtain global estimates for them. Nor did the working group explicitly account for recycling in the estimates of production. The estimates will be conservative in systems where these elements collectively are a significant proportion of the primary basal resources. The production at node *i* is a function of the transfer efficiency from other nodes (*j*) to node *i*, the inputs from other locations and losses from the i^{th} node:

$$P_i = TP_j + A_j - LP_i \qquad (23)$$

where P_i is a vector of production values over all nodes; T is a matrix of ecological transfer efficiencies from node j to node i; A_i represents the addition of production to node i from other sources; and L represents a fractional loss term from node i (e.g. advective loss, removals due to harvest).

In the analysis, the working group recognizes two pathways for transfer of primary production in the system: the classical grazing food web tracing the fate of production of microplankton (phytoplankton cells greater than 20 μm, principally diatoms and large dinoflagellates); and production involving transfer through the microbial food web originating with combined nano-picoplankton (2–20 μm) and picoplankton (less than 20 μm) production (i.e. nano-picoplankton, see Figure 16). The first pathway involves grazing by mesozooplankton and filtering of diatom production by benthic invertebrates, particularly bivalves. The second entails consumption of nano-picoplankton by heterotrophic bacteria and feeding of microzooplankton on bacteria. In this representation, carnivorous zooplankton (mesozooplankton) prey on microzooplankton. The microbial pathway therefore involves two or more trophic transfer steps before reaching mesozooplankton as a bridge to higher trophic levels. The working group notes that the functional groups represented in the upper food web depicted in Figure 16 do not strictly correspond to taxonomic groups. Individual taxa may feed at multiple trophic levels, reflecting both ontogenetic shifts in diet and generalist feeding strategies with life stages.

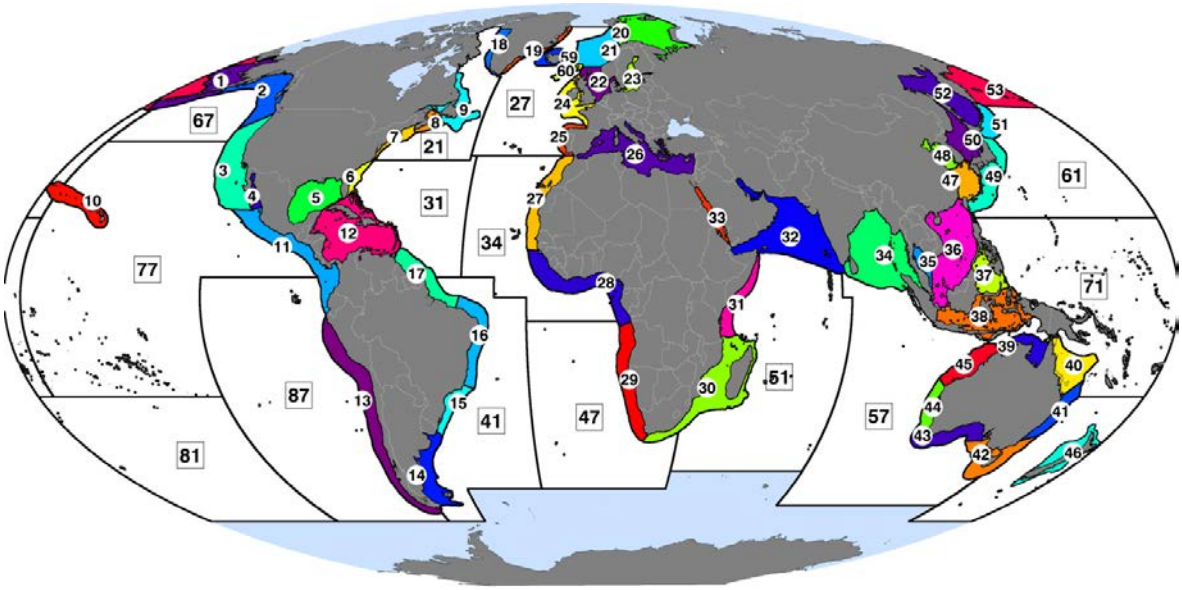
FIGURE 16
Food web structure employed in this analysis



Notes: Nano-picoplankton, bacteria, and microzooplankton comprise the microbial food web in this representation. The classical grazing food web is fuelled by microplankton production. Species characterized by ontogenetic shifts in diet and/or mixed feeding strategies can occupy multiple compartments in this representation.

For this analysis, the working group used designated LMEs as strata (Figure 17). LMEs are differentiated by similar physical and ecological features, such as hydrography, productivity and tropically dependent populations (Sherman and Alexander, 1986; Sherman, 1991), and account for about 80–90 percent of the global fisheries catch (Christensen *et al.*, 2008). To account for some of the nearshore versus offshore variability in production within some regions, each LME was subdivided using the 300 m isobath. Subareas less than 300 m depth included the characteristically more-productive continental shelf areas and the nearshore areas of the upwelling regions. In general, subareas greater than 300 m depth were characterized by lower overall levels of microplankton production. Inland seas and high-latitude regions, including Hudson Bay, Black Sea, Arctic Ocean, Kara Sea, Laptev Sea, East Siberian Sea, Beaufort Sea, Chukchi Sea and Antarctica, were not included in this analysis owing to the seasonal effects of cloud cover and high solar zenith angles on estimates derived from satellite coverage in these regions.

FIGURE 17
Strata used in estimating global fishery production potential based on large marine ecosystem boundaries



Notes: Individual large marine ecosystems (LMEs) designated by colour; LME numbers in circles; FAO Major Statistical Areas number in squares. For LME names see Table 11.

Subsequently, the working group further assigned LMEs to major ocean ecotypes: "subarctic and boreal", "temperate", "subtropical", "tropical" and "upwelling" (see Table 11). Some LMEs straddled traditional dividing lines between ecotypes; the working group made assignments based on relative areas within each of the ecotypes in these cases. For some of the upwelling systems considered here, the relative importance of upwelling is not uniform throughout the LME. In others, the seasonal importance of upwelling can be high (e.g. during the monsoon season in the Arabian Sea), but the working group did not classify these as upwelling systems. However, the seasonal importance of upwelling in phytoplankton production dynamics is still represented in the estimates.

TABLE 11
Large marine ecosystems (LMEs) and designated ecotypes used in determining transfer probability estimates

LME	LME NAME	ECOTYPE
1	East Bering Sea	Subarctic-boreal
2	Gulf of Alaska	Subarctic-boreal
3	California Current	Upwelling
4	Gulf of California	Subtropical
5	Gulf of Mexico	Subtropical
6	Southeast US Continental	Temperate
7	Northeast US Continental	Temperate
8	Scotian Shelf	Temperate
9	Newfoundland-Labrador Shelf	Subarctic-boreal
10	Insular Pacific-Hawaiian	Subtropical
11	Pacific Central-American Coastal	Upwelling
12	Caribbean Sea	Tropical

LME	LME NAME	ECOTYPE
13	Humboldt Current	Upwelling
14	Patagonian Shelf	Temperate
15	South Brazil Shelf	Temperate
16	East Brazil Shelf	Subtropical
17	North Brazil Shelf	Subtropical
18	West Greenland Shelf	Subarctic-boreal
19	East Greenland Shelf	Subarctic-boreal
20	Barents Sea	Subarctic-boreal
21	Norwegian Sea	Subarctic-boreal
22	North Sea	Temperate
23	Baltic Sea	Subarctic-boreal
24	Celtic-Biscay Shelf	Temperate
25	Iberian Coastal	Temperate
26	Mediterranean Sea	Temperate
27	Canary Current	Upwelling
28	Guinea Current	Upwelling
29	Benguela Current	Upwelling
30	Agulhas Current	Temperate
31	Somali Coastal Current	Upwelling
32	Arabian Sea	Tropical
33	Red Sea	Tropical
34	Bay of Bengal	Tropical
35	Gulf of Thailand	Tropical
36	South China Sea	Tropical
37	Sulu-Celebes Sea	Tropical
38	Indonesian Sea	Tropical
39	North Australian Shelf	Tropical
40	Northeast Australian Shelf	Tropical
41	East Central Australian	Subtropical
42	Southeast Australian Shelf	Temperate
43	Southwest Australian Shelf	Temperate
44	West Central Australian	Subtropical
45	Northwest Australian Shelf	Tropical
46	New Zealand Shelf	Temperate
47	East China Sea	Subtropical
48	Yellow Sea	Temperate
49	Kuroshio Current	Temperate
50	Sea of Japan	Temperate
51	Oyashio Current	Temperate
52	Sea of Okhotsk	Subarctic-boreal
53	West Bering Sea	Subarctic-boreal
59	Iceland Shelf	Subarctic-boreal
60	Faroe Plateau	Subarctic-boreal

ESTIMATING PRIMARY PRODUCTION

Ocean colour remote sensors provide an unprecedented view of the global ocean and are the only means of obtaining basin-scale, synoptic high-frequency measurements of global primary production. The working group calculated annual estimates of primary production using data from the Sea-viewing Wide Field-of-View Sensor (SeaWiFS, NASA) and a modified version of the Vertically Generalized Productivity Model (VGPM; Behrenfeld and Falkowski, 1997). This modified VGPM model replaces the original temperature-dependent description of photosynthetic efficiencies with the exponential Eppley function (Eppley, 1972), which was modified by Morel (1991):

$$P^{b}_{max}(T) = 4.6 * 1.065^{T-20^{\circ}} \tag{24}$$

where P^{B}_{max} is the maximum carbon fixation rate and T is sea surface temperature. The VGPM calculates the daily amount of carbon fixed based on: the maximum rate of chlorophyll-specific carbon fixation in the water column; sea surface daily photosynthetically available radiation; the euphotic depth (the depth where light is 1 percent of that at the surface); chlorophyll *a* concentration; and the number of daylight hours:

$$PP_{eu} = 0.66125 * P^{b}_{max} * \frac{I_0}{I_0 + 4.1} * Z_{eu} * Chl * DL \tag{25}$$

where, PP is the daily amount of carbon fixed integrated from the surface to the euphotic depth (mgC m^{-2} day^{-1}), P^{B}_{max} is the maximum carbon fixation rate within the water column (mgC $mgChl^{-1}$ hr^{-1}), I_0 is the daily integrated molar photon flux of sea surface PAR (mol quanta m^{-2} day^{-1}), Z_{eu} is the euphotic depth (m), Chl is the chlorophyll concentration (mg m^{-3}), and DL is the photoperiod (hours) calculated for the day of the year and latitude, according to Kirk (1994). The light dependent function ($\frac{I_0}{I_0+4.1}$) describes the relative change in the light saturation fraction of the euphotic zone as a function of surface PAR (I_0).

To estimate the proportion of primary production attributed to the microplankton (> 20 µm) component, the working group first estimated the microplankton total chlorophyll *a* (i.e. biomass) fraction and then used an empirical relationship to calculate the percentage of microplankton production. Recent advances in ocean colour remote sensing have led to the development of several phytoplankton size class and phytoplankton functional type models. However, most of these models are not recommended for use in continental shelf waters (i.e. less than 200 m depth). Owing to the importance of the continental shelf regions to global fisheries, the working group used a regional model by Pan *et al.* (2011), who developed a chemotaxonomic method to measure phytoplankton functional types on the Northeast US Continental Shelf LME (NES LME) using satellite-derived phytoplankton pigment measurements (Pan *et al.*, 2010). The working group combined diatom and dinoflagellate biomasses to represent the microplankton fraction and combined the remaining functional groups in the nano-picoplankton (< 20 µm) group (Vidussi *et al.*, 2001). The working group then estimated the fraction of the total primary production associated with the microplankton size class using a relationship based on more than 600 measurements of size-fractionated chlorophyll and primary production in the NES LME (O'Reilly, Evans-Zetlin and Busch, 1987). These regional relationships may not precisely represent size-fractionated production in all LME subareas. However, *in situ* size-fractionated production data are not readily available on a global scale to develop these relationships for each LME.

The working group computed monthly chlorophyll concentration and primary production by size class on a pixel-by-pixel basis for each month. The working group then summed over all pixels within an LME and integrated over months to generate annual estimates of the total for each subregion. This approach, therefore, differed from many previous analyses of fishery production potential in that the working group treated the satellite-derived data as a census rather than a sample in which mean

productivity was applied over selected coastal and ocean areas. The approach preserves the spatial variability in phytoplankton production dynamics within an LME.

TRANSFER EFFICIENCIES

Energy is transferred from lower trophic levels to higher trophic levels through consumption. In this way, lower trophic levels support the higher trophic levels in a process commonly thought of as an energy pyramid. Energy that is transferred through each level up the pyramid also approximates the transfer of biomass through the ecosystem. In EwE, the transfer efficiencies between successive groups can be calculated as "the ratio between the sum of exports from a given trophic level, plus the flow that is transferred from one trophic level to the next, and the throughput on the trophic level" (Christensen, Walters and Pauly, 2005).

Early laboratory studies by Slobodkin (1961) indicated that the expected transfer efficiency was of the order of 10 percent. Clear thermodynamic constraints place limits on the transfer efficiency between successive levels in the food chains. Pauly and Christensen (1995) supported the canonical value of 10 percent as an ecological transfer efficiency; they estimated that the transfer efficiency of biomass between trophic levels in aquatic ecosystems, although variable, had a mean of 10 percent. To objectively assess trophic transfer efficiencies throughout the generic food web, the working group evaluated estimates of transfer efficiencies derived from 234 published EwE models collected by the Sea Around Us Project of the University of British Columbia for 209 models.[5] These models were spatially explicit and represented 30 of the LMEs considered here. In addition, the working group included results from another 25 published and validated EwE models that have been used in previous global fisheries analyses (Worm *et al.*, 2009; Garcia *et al.*, 2012). Rather than assume or assign trophic transfer efficiencies at different steps in the food web for the models for each LME, the working group used these model estimates to define probability distributions characterizing transfer probabilities at different steps in the food web. The characterization of transfer efficiencies between discrete trophic levels based on these EwE models followed the approach in Ulanowicz (1993).

BENTHIC-MESOZOOPLANKTON PATHWAY

For all models, the working group calculated transfer efficiencies from primary producers and from detritus. To determine transfer efficiencies from the microplankton, the working group examined energetic pathways in the 234 EwE models and assigned a proportion to the microplankton group and determined the production flowing to mesozooplankton and benthos (see Figure 16). Given that there are three main food chains, the working group needed the proportion of the primary production flowing to zooplankton versus the proportion flowing to benthic invertebrates in addition to the transfer efficiencies. These proportions were estimated from the consumption tables taken from the EwE models following five steps:

1. Classify producers in the EwE model as nano-picoplankton, microplankton (principally diatoms and large dinoflagellates), large phytoplankton, or benthic primary producers. Most EwE models specify a single phytoplankton component, in which case the working group assumed they are microplankton in the assessment of flows to the benthos and to mesozooplankton.

2. Identify EwE groups that are zooplankton.

3. Identify EwE groups that are benthos.

4. Calculate the proportion of consumption of microplankton that is due to zooplankton (K_Z). The consumption by benthos (K_B) will then be one minus this value, i.e.:

[5] www.ecopath.org/biomasspnas

$$K_Z = \frac{C_{P\_by\_Z}}{C_P}$$
$$K_B = 1 - K_Z \quad (26)$$

where C_P is the total consumption of microplankton, and $C_{P\_by\_Z}$ is the consumption of microplankton by zooplankton.

5. For many EwE models, there will be no primary production from phytoplankton flowing to benthos. As the assumed generic food web above does not allow for that, the working group also needed to calculate the proportion of all primary production flowing to zooplankton (K_{ZT}), and the proportion flowing to benthos (K_{BT}), i.e.:

$$K_Z = \frac{C_{T\_by\_Z}}{C_T}$$
$$K_{BT} = 1 - K_{ZT} \quad (27)$$

where C_T is the sum of consumption of all primary producers, and $C_{P\_by\_Z}$ is the total consumption across all primary producers by zooplankton.

LANDINGS DATA

To assess exploitation status within the LMEs and adjacent open ocean areas covered in this analysis, the working group assembled landings statistics compiled by the Sea Around Us Project that are based on the FAO global landings. The LMEs are differentiated by similar physical and ecological features, such as hydrography, productivity and tropically dependent populations (Sherman and Alexander, 1986; Sherman, 1991).

Watson *et al.* (2004) described the process used by the Sea Around Us Project and collaborators to allocate landings statistics from the FAO, which are available at the FAO statistical area scale and reported by country fishing and taxa. They used a rule-based approach that relied on maps of the global distribution of commercial taxa and a database of fishing access agreements to estimate values for global 0.5-degree grid cells. This allocation process produced spatial time-series of landings data from 1950 through 2006 that could be aggregated to the exclusive economic zone, LME, or other scales and which discerned between landings by foreign and domestic fleets. In addition, each grid cell contains the minimum, maximum and mean depth based on ETOPO2v2 2-minute gridded global relief data.[6] The working group was therefore able to aggregate landings for each subarea investigated in this study using the landings at the 0.5-degree grid cells. Within each LME, the working group created subset cells based on whether the mean depth was greater or less than 300 m depth. The working group did this in order to match the satellite-derived estimates of chlorophyll and primary production.

The working group assigned trophic levels (TLs) to taxonomic groups in the catch as a simple means of expressing where fish and other organisms tend to operate in their respective food webs. The estimates of TLs for fish or invertebrates, therefore, considered both their diet composition and the TL of their food items. The TL of a given group of animals (individuals, population, species) was estimated by:

TL = 1 + mean TL of the food items (28)

where the mean was weighted by the contribution of the different food items.

[6] Data can be found at: www.ngdc.noaa.gov/mgg/fliers/06mgg01.html

Following a convention established in the 1960s by the International Biological Program, primary producers and detritus (including associated bacteria) were defined as having a TL of one. FishBase TL estimates for finfish are as follows:[7]

1. For species with one or more sets of diet composition data, the TL taken from FishBase was either the only estimate, or the median of values pertaining to the juveniles/adults, or adult stages. In a few cases, pertaining to very low (2.0) or very high (4.5) estimates, the TL values were adjusted upwards or downwards, respectively, if closely related species had less extreme values. For species with only food item data, the TL estimate was taken as is only if it fell within the TL range of other, closely related species. If not, the estimated TL was adjusted as above. The working group used the FishBase estimates rather than estimates from the EwE models employed for transfer efficiency analyses because of the variable taxonomic resolution that the models employ.

2. For fish genera, the working group derived TL estimates from the mean TL of the component species with estimates; the working group gave more weight to species with estimates of TL based on diet composition data.

3. For fish families, the working group derived TL estimates by averaging the mean TL of the component genera or taxonomic order; for invertebrates, where less-direct diet composition data were available, the working group estimated TL based on the "ISCCAAP Table" of FishBase 2000 (Froese and Pauly, 2000), itself based largely on estimates from EwE models. These estimates were then complemented by data from more recent models, documented in www.ecopath.org, and in Sea Around Us Project reports (e.g. www.seaaroundus.org/report/impactmodels.htm).

ASSIGNMENT OF LANDINGS DATA TO TAXONOMIC GROUPS

Within the model, production flows through several pathways (Figure 16); beginning from the microbial loop, energy flows to nano-picoplankton to bacteria and microzooplankton and up to mesozooplankton. This energy is transferred to the upper trophic levels, including planktivores and pelagic fish. To determine removals from the systems, the working group assigned each taxon in the landings data to a category of "benthos" (i.e. taxa that feed on phytoplankton and detritus), "benthivore" (i.e. taxa that feed on benthos), "planktivore" or "upper trophic levels" (i.e. taxa that feed on finfish and benthivores). To make these assignments, the working group first examined the general feeding strategies of the taxa in each of the 29 functional group designations used by the Sea Around Us Project (Table 12). In some cases (e.g. "large pelagics"), the assignment to diet boxes was straightforward and all taxa within the functional group were assigned to a single diet box. For other groups (e.g. "medium reef-associated fish"), there were taxa in the group that belonged to more than one group (e.g. benthivores and upper trophic levels).

Assigning individual species into the functional categories was challenging, especially where species exhibit ontogenetic shifts in diet and mixed feeding strategies associated with generalist predators. As described above, the functional groups were not intended to map to individual taxa, but rather to ascertain the trophic position of the assemblages feeding in whole or in part at different levels of the food web.

[7] www.fishbase.org

TABLE 12.
Designation of 29 functional groups from the Sea Around Us Project

Small pelagics (< 30 cm)	Small to medium rays (< 90 cm)	Small to medium sharks (< 90 cm)
Medium pelagics (30–89 cm)	Large rays (≥ 90 cm)	Large sharks (≥ 90 cm)
Large pelagics (≥ 90 cm)	Small demersals (< 30 cm)	Small reef associated fish (< 30 cm)
Small benthopelagics (< 30 cm)	Medium demersals (30–89 cm)	Medium reef associated fish (30–89 cm)
Medium benthopelagics (30–89 cm)	Large demersals (≥ 90 cm)	Large reef associated fish (≥ 90 cm)
Large benthopelagics (≥ 90 cm)	Small bathydemersals (< 30 cm)	Lobsters, crabs
Small bathypelagics (< 30 cm)	Medium bathydemersals (30–89 cm)	Shrimps
Medium bathypelagics (30–89 cm)	Large bathydemersals (≥ 90 cm)	Krill
Large bathypelagics (≥ 90 cm)	Small to medium flatfishes (< 90 cm)	Other demersal invertebrates
Cephalopods	Large flatfishes (≥ 90 cm)	

DISCARD DATA

FAO-funded research on estimating global discards (Alverson *et al.*, 1994; Kelleher, 2005) can be used to consider this critical element of the catch. The working group focused on recent catch and discard information linked to the available satellite-derived primary production estimates; accordingly, the working group concentrated on the estimates provided by Kelleher (2005), who provides discard data for many but not all LMEs. The available data were not disaggregated to the functional group level, and the working group made no attempt to use taxon-specific discard information.

MEAN TROPHIC LEVEL AND SPECIES DOMINANCE OF LANDINGS

The trophic level at which catch was extracted holds important implications both for the amount of available production that can be sustainably taken and for its effect on other ecosystem components (including marine mammals, reptiles and seabirds) that consume prey at this level. To examine temporal patterns in the overall composition of the species comprising the landings, the working group computed the MTL (partitioned according to all species) and including the upper trophic level (TL > 3.5). The objective was to characterize empirically the species mix in the landings and not to draw inferences concerning the underlying abundance or biomass levels of species making up the landings.

The overall pattern of diversity in the catch is also critically important given recent concerns over highly selective fishing patterns that place high fishing pressure on restricted elements of the food web (Zhou *et al.*, 2010; Garcia *et al.*, 2012). These highly species-selective harvesting patterns can result in large-scale disruption of ecosystem structure and function. Moreover, any consideration of increasing sustainable yield in capture fisheries will require both a reduction of fishing pressure on currently overexploited species and, generally, a diversification of the species mix included in the assemblage of exploited species. The working group therefore examined a measure of species dominance in the landings based on an adaptation of the Berger-Parker Index (Berger and Parker, 1970). The dominance index in year *t* is:

$$D_t = \frac{L_{\max,t}}{L_t} \tag{29}$$

where $L_{max,t}$ is the maximum landings of any species in the catch in year t, and L_t is the total landings level in year t. May (1975) indicated that this simple measure of dominance is highly robust. The inverse of the dominance index ($1/D_t$) gives a measure of diversity in the landings. Lower levels of dominance reflect higher levels of overall diversification in the species mix landed in the fishery. The working group looked for evidence of changing levels of specialization in fishing or marketing practices as reflected in the landed composition of the catch. To test for changes in the level of taxonomic resolution in the landings over time, the working group first examined decadal-scale measures of species resolution in reported catch statistics for each LME (Figure A2.1).

CATCH–PRODUCTION RATIOS

The working group examined two measures of catch in relation to the amount of available production. The first examines the ratio of the catch within the LME to the amount of total primary production. This simple measure is a catch efficiency ratio:

$$CE_{TPP} = \frac{C_T}{TPP} \tag{30}$$

where CE_{TPP} is the catch efficiency relative to total primary production, C_T is the catch for all species harvested and TPP is total primary production. The working group also computed:

$$CE_{MPP} = \frac{C_T}{MPP} \tag{31}$$

where MPP is microplankton primary production.

The amount of production actually appropriated for catch at each trophic level affords a more proximal view of the potential impact of harvesting on the system. To assess this factor more directly, the working group rewrote Equation 23 to explicitly consider removals due to harvest as:

$$P_i = TP_j + A_i - L'P_i - C_i \tag{32}$$

where L' now represents all losses from node i other than removals due to fishing, C_i is the total catch from node i (including discarded and landed components), and all of other terms are defined as before. For the purposes of the analysis, the working group assumed that inputs and losses from sources other than fishing were in balance at each node. The landings extracted from node i can be expressed as the product of a fractional exploitation rate E_i and the production at node i:

$$C_i = E_i P_i \tag{33}$$

The working group was interested in devising a sustainable exploitation strategy for extraction rates at different nodes. Given estimates of production at each node and the observed removals from each exploited node, the realized ecosystem exploitation is:

$$\tilde{E}_i = \frac{\tilde{C}_i}{P_i} \tag{34}$$

TREATING UNCERTAINTY

To represent uncertainty in key input parameters to the production model, the working group specified empirically derived probability distributions for primary production, transfer efficiencies, and the split between transfer of energy from microplankton to benthos and mesozooplankton. The working group used normal probability distributions to represent interannual variability in microplankton and nano-

picoplankton production for the period between 1998 and 2007. The working group computed the mean and variance of interannual phytoplankton production for both phytoplankton components to specify the parameters of the normal distributions.

The working group used Beta distributions at each level based on the compilation of EwE to obtain transfer efficiencies between microplankton and higher components of the food web. Transfer estimates are constrained between 0 and 1 and are appropriate for application of the Beta distribution. To obtain sufficiently large sample sizes that characterize these probability distributions, the working group pooled model estimates over the five major ocean ecotypes described above. The working group then used the resulting distributions to represent the transfer probabilities in each LME and adjacent open ocean region in the analysis.

Energetic pathways involving the benthos differed substantially between the food web models the working group examined. In recognition of the limitations of these models used to characterize uncertainty in energetic pathways involving the benthos, the analysis used uniform probability distributions bounded by the upper and lower quartiles of the range of observed splits between the benthos and mesozooplankton.

Many of the EwE models the working group examined did not partition phytoplankton production by size class and therefore did not allow treatment of the microbial food web as specified in the model (Figure 16). In those cases, the working group used literature values for ecotrophic efficiencies (proportion of production consumed within the microbial food web) and the gross growth efficiency of bacteria and microzooplankton (e.g. Straile, 1997; Ware, 2000). It was not possible to define these elements according to ecotype, nor was it possible to fully represent the uncertainty in these estimates.

ECOSYSTEM-BASED EXPLOITATION REFERENCE LEVELS

As noted above, the estimates of fishery production potential described above typically assumed that between 50 and 70 percent of production at a defined MTL could be extracted as catch (Graham and Edwards, 1962; Schaefer, 1965; Ricker, 1969; Ryther, 1969; Moiseev, 1994). These proposed extraction rates were predicated on prevailing single-species recommendations based on the implicit assumption that fishing mortality rates could equal natural mortality for the stock (Pauly and Christensen, 1995). It is now recognized that these earlier target levels for single-species management were too high and led to risk-prone decisions (Pauly and Christensen, 1995). Standard reference points have not been fully established to guide overall extraction policies for marine ecosystems. Iverson (1990) proposed that exploitation rates should not exceed the *f*-ratio (the ratio of new primary production to total primary production) in marine systems. This suggestion was based on the underlying recognition that new production (primarily by larger phytoplankton species) is more readily available to fuel production at the higher trophic levels of principal economic interest. Although direct estimates of the *f*-ratio are not broadly available for large marine ecosystems throughout the world ocean, the working group used the ratio of microplankton production to total primary production as a first-order approximation.

6. RESULTS

PRIMARY PRODUCTION

Chlorophyll concentration and primary production were highest in coastal locations characterized by important inputs of nutrients from land and strong mixing processes driven by winds and tides (Figure 18). High chlorophyll and production levels were concentrated in upwelling regions. Overall primary production was dominated by nano-picoplankton production, especially in the deeper coastal locations and the ocean basins. Within the 300 m isobath, microplankton production accounted for 25.1 percent of the total production on average. For deeper-water components (> 300 m) within individual LMEs, microplankton production accounted for 20.1 percent of the total production. As expected, the microplankton contribution to production was smallest (14.2 percent) in the open ocean regions outside LME boundaries. Further results by LME are provided in Figures A2.2–A2.56 and by FAO region in Figures A2.57–A2.70. The working group also estimated the ratio of microplankton to total production primary production by phytoplankton size class along with coefficients of variation for nano-picoplankton production and microplankton production for each LME (Table 13). The working group used these estimates to characterize the uncertainty in phytoplankton production on interannual time scales and to assess the relative importance of microplankton production. These estimates of the proportion of microplankton production suggested mean reference levels of about 20–25 percent.

FIGURE 18

Distribution patterns for total chlorophyll *a*, primary production, percentage microplankton chlorophyll *a*, percentage microplankton primary production, percentage nano-picoplankton chlorophyll *a*, and percentage nano-picoplankton primary production

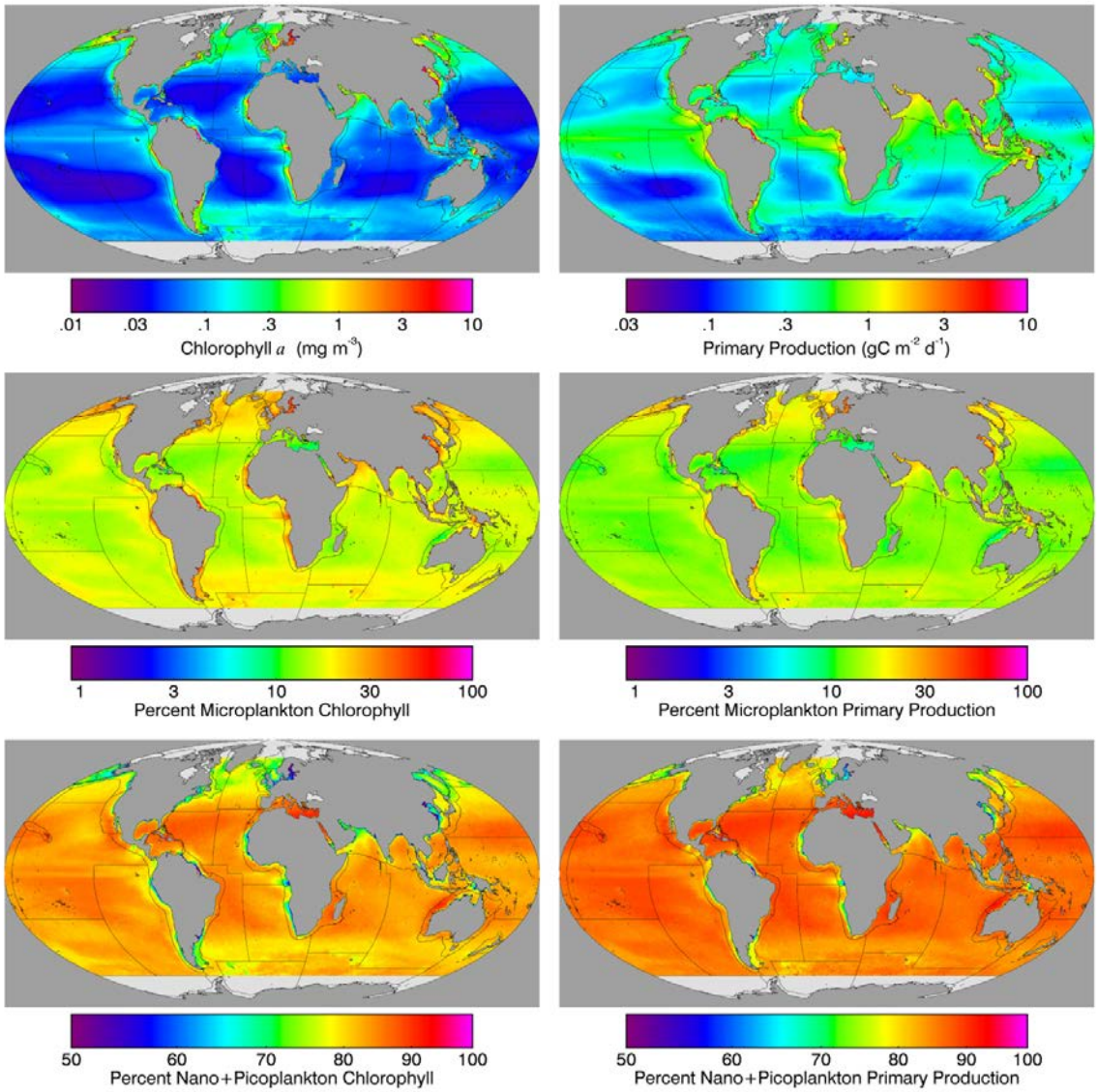


TABLE 13
Ratio of microplankton to total production, nano-picoplankton coefficient of variation (CV) and microplankton CV by LME

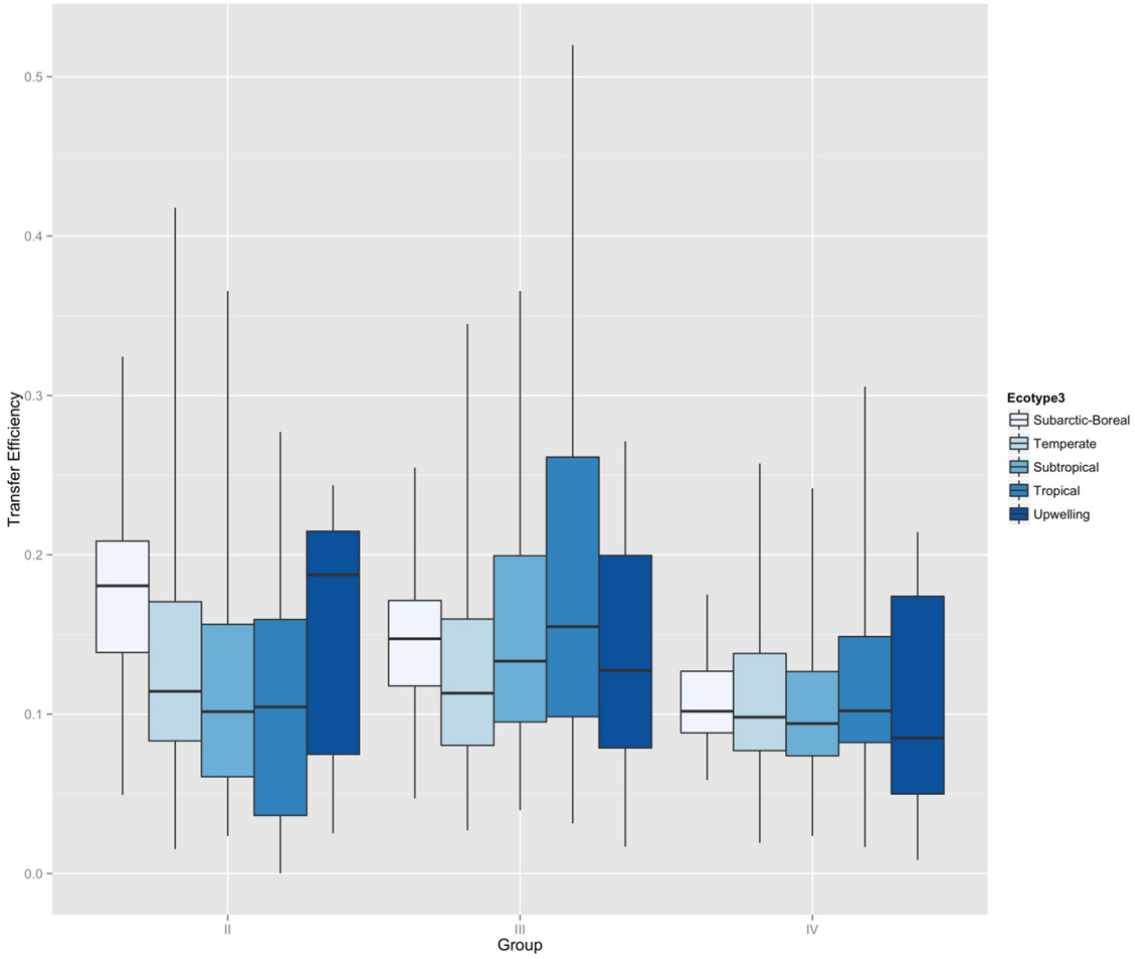
LME name	Ratio of microplankton to total production	Nano-picoplankton production CV	Microplankton production CV
East Bering Sea	0.29	0.4	0.31
Gulf of Alaska	0.24	0.32	0.17
Gulf of California	0.22	0.28	0.15
Gulf of Mexico	0.2	0.24	0.09
Southeast US Continental	0.19	0.23	0.21
Northeast US Continental	0.27	0.38	0.12
Scotian Shelf	0.25	0.33	0.25
Newfoundland-Labrador Shelf	0.22	0.28	0.27
Insular Pacific-Hawaiian	0.11	0.12	0.06
Pacific Central-American Coastal	0.18	0.22	0.24
Caribbean Sea	0.16	0.19	0.04
Humboldt Current	0.23	0.3	0.14
Patagonian Shelf	0.27	0.37	0.1
South Brazil Shelf	0.21	0.26	0.12
East Brazil Shelf	0.13	0.15	0.06
North Brazil Shelf	0.28	0.38	0.13
West Greenland Shelf	0.23	0.3	0.49
East Greenland Shelf	0.2	0.25	0.41
Barents Sea	0.28	0.38	0.01
Norwegian Sea	0.24	0.32	0.27
North Sea	0.28	0.38	0.17
Baltic Sea	0.38	0.61	0.34
Celtic-Biscay Shelf	0.24	0.32	0.13
Iberian Coastal	0.17	0.21	0.09
Mediterranean Sea	0.12	0.14	0.07
Canary Current	0.25	0.34	0.09
Guinea Current	0.24	0.31	0.08
Benguela Current	0.25	0.33	0.07
Agulhas Current	0.14	0.17	0.06
Somali Coastal Current	0.16	0.19	0.15
Arabian Sea	0.22	0.28	0.11
Red Sea	0.18	0.22	0.07
Bay of Bengal	0.19	0.23	0.09
Gulf of Thailand	0.19	0.24	0.11
South China Sea	0.18	0.22	0.08
Sulu-Celebes Sea	0.16	0.19	0.13
Indonesian Sea	0.18	0.22	0.12
North Australian Shelf	0.2	0.26	0.11

LME name	Ratio of microplankton to total production	Nano-picoplankton production CV	Microplankton production CV
Northeast Australian Shelf	0.15	0.17	0.06
East Central Australian	0.13	0.15	0.07
Southeast Australian Shelf	0.16	0.19	0.07
Southwest Australian Shelf	0.15	0.17	0.07
West Central Australian	0.13	0.15	0.09
Northwest Australian Shelf	0.14	0.17	0.07
New Zealand Shelf	0.17	0.2	0.1
East China Sea	0.25	0.33	0.13
Yellow Sea	0.33	0.49	0.12
Kuroshio Current	0.16	0.18	0.07
Sea of Japan	0.19	0.24	0.17
Oyashio Current	0.23	0.3	0.27
Sea of Okhotsk	0.26	0.35	0.48
West Bering Sea	0.25	0.33	0.37
Iceland Shelf	0.25	0.34	0.19
Faroe Plateau	0.24	0.31	0.34

TRANSFER EFFICIENCIES

The working group estimated ecological transfer efficiencies by ecotype and trophic level based on EwE models (Figure 19). Median transfer efficiencies generally declined with increasing trophic level in each ecotype. By TL IV, the median transfer efficiencies were about 10 percent, which is the canonical level used in many of the earlier production potential calculations mentioned above. Ryther (1969) applied constant transfer efficiencies over all trophic levels within his ocean domains but specified different estimates for different domains: 15 percent in coastal areas, 10 percent in oceanic regions, and 20 percent in upwelling areas.

FIGURE 19
Box plots of ecological transfer efficiencies



Notes: Based on EwE models compiled by the Sea Around Us Project by ecotype for microplankton and secondary producers (T II), mesozooplankton and tertiary producers (TE III) and between tertiary producers and piscivores. Vertical lines indicate range; coloured boxes show lower and upper quartile range, and horizontal line gives median value.

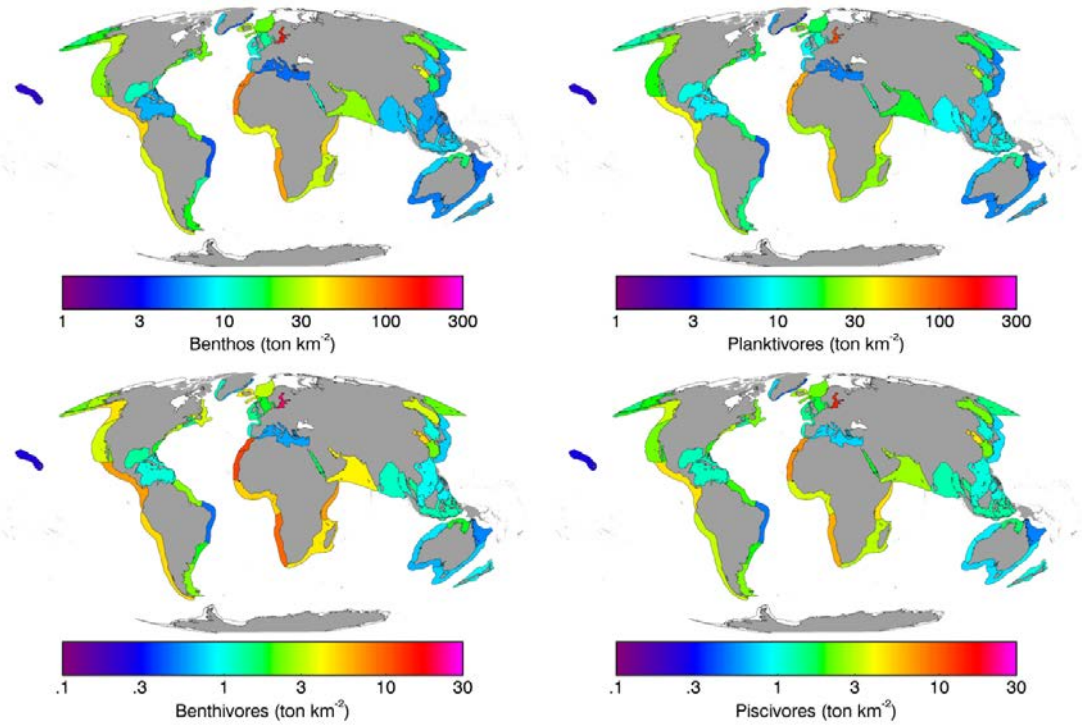
Estimates of the parameters for the Beta distributions for 3 of the 15 estimates indicated a significant departure from the expected values for the Beta distribution ($p < 0.05$). These departures reflected higher or lower observed frequencies for one or more transfer efficiency bins, but not a general deviation from the expected overall form of the Beta distribution. Accordingly, the working group used the resulting parameter estimated in the analyses rather than using a non-informative prior distribution.

PRODUCTION POTENTIAL

The working group calculated production estimates for the major functional groups of potential or realized importance to harvesting for each LME (Figure 20). Characteristically high production levels were found in the dominant upwelling regions of the world ocean and in regions where at least seasonal upwelling patterns are important (e.g. the Arabian Sea.). Western boundary current regions exhibited moderately high production levels (e.g. the Oyashio and Kuroshio Current systems, the Northwest Atlantic LMEs and the Agulhas Current region). Intermittent and localized upwelling

patterns in these regions coupled with high nutrient concentrations in several of these systems contribute to relatively high production levels.

FIGURE 20
Estimated production levels in the absence of exploitation by functional group for LMEs represented in this study



Note: Change to logarithmic scale for the benthivore and piscivore functional groups.

The working group estimated a mean total production within the LMEs included in this study of 2.43 gigatonnes year^{-1} for benthos, 0.35 gigatonnes year^{-1} for benthivores, 1.22 gigatonnes year^{-1} of planktivores, and 0.18 gigatonnes year^{-1} of piscivores. A critical consideration in determining fishery production potential is the amount of production comprising harvestable ecosystem components. For the benthos in particular, meiobenthic and macrobenthic organisms accounted for a dominant fraction of the benthic production, but are not suitable for exploitation. Available biomass estimates of meio-, macro- and mega-benthic (> 1 cm length) were compiled by Peters-Mason (2007) and used to infer biomass by LME and major ocean areas based on predictive models for factors affecting benthic production including depth, sea surface temperature, and chlorophyll *a* concentration. The working group converted these estimates to production using production to biomass ratios based on compilations from Christensen *et al.* (2009). In general, megabenthic organisms comprised less than 20 percent of the production in these systems. Moreover, the species of potential or realized commercial interest (principally molluscs and crustaceans) comprised an even smaller fraction of the remaining megabenthic fauna.

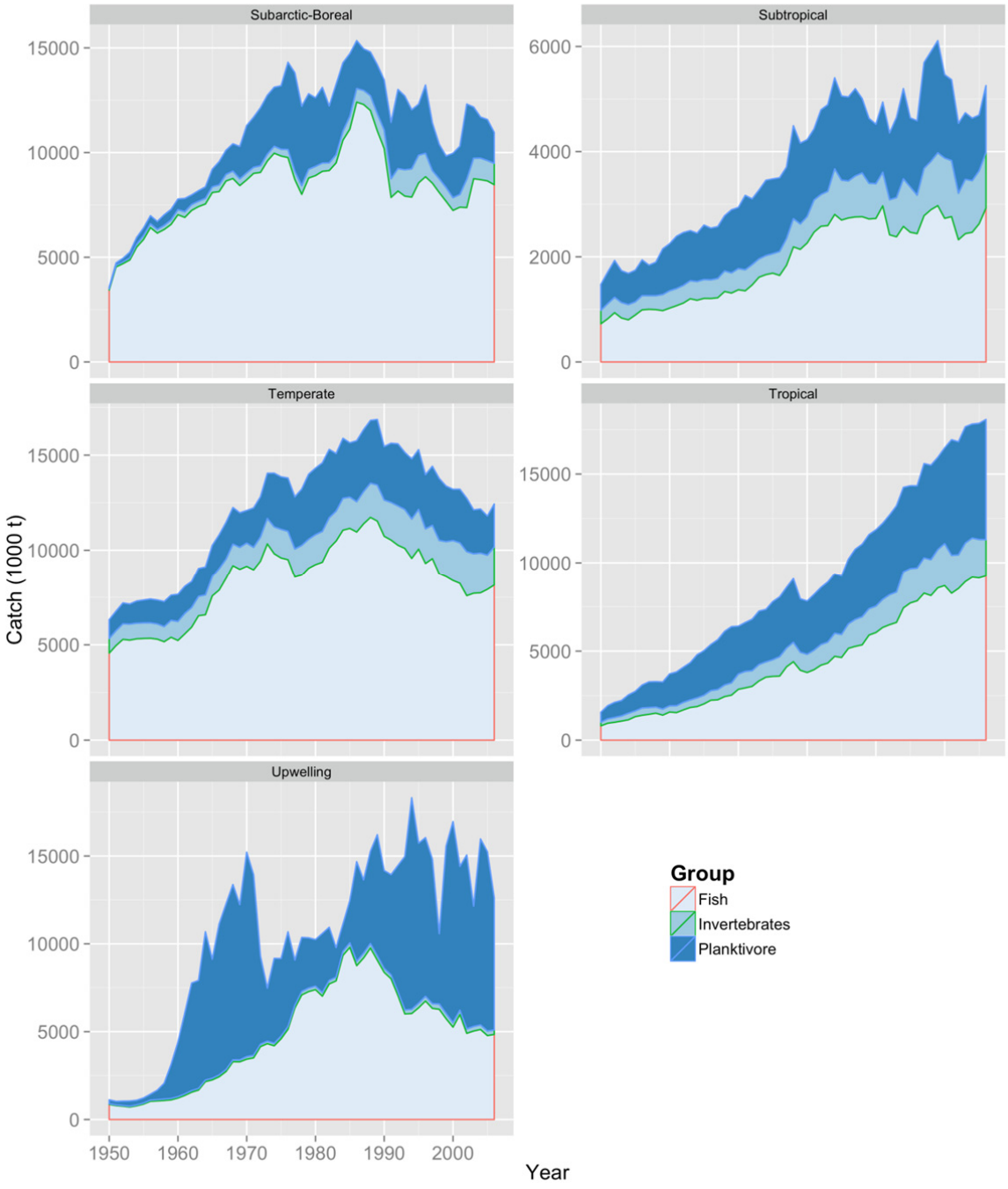
Planktivore production was large, but a substantial part of this production was also not available for harvest. For example, the larvae of many exploited fish populations are planktivorous and derive a significant part of their overall production in the plankton during this life-history stage. Similarly, planktivorous invertebrates, including chaetognaths, gelatinous zooplankton and other species account for significant levels of overall production within this functional group. A number of planktivorous

fish species remained unexploited or are exploited regionally, but not consistently throughout all LMEs. The estimated biomass of planktivorous mesopelagic fish on a global scale (including deeper water areas outside the LMEs) is extremely high, of the order of 1 billion tonnes (Gjøsæter and Kawaguchi, 1980; Moiseev, 1994), but are not currently exploited to any significant degree. Within the LMEs considered in this report, mesopelagic biomass levels in the Arabian Sea were high and clearly of potential significance to fishery production (Gjøsæter and Kawaguchi, 1980). The working group return to these considerations later in this report in the evaluation of fishery production potential.

LANDINGS AND CATCH

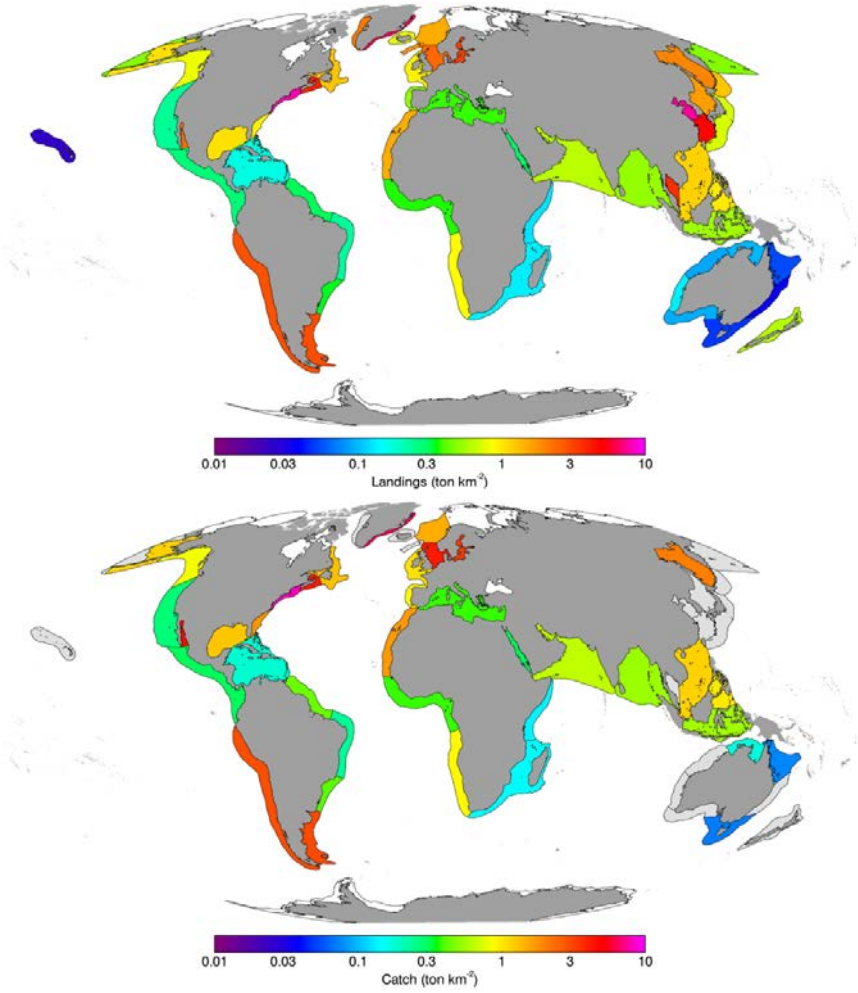
The working group calculated trends in landings since 1950 by ecotype for the following categories: (i) invertebrates; (ii) planktivorous fish; and (iii) benthivorous and piscivorous fish (Figure 21). Overall, landings have declined from a peak in the late 1980s to early 1990s for the subarctic-boreal, temperate and subtropical ecotypes. Landings have increased steadily in the tropical ecotype and have been highly variable, but have stabilized in upwelling systems. Global trends in landings have shown an overall decrease since the peak about two decades ago. Overall, the planktivores and the combined benthivore and piscivore fish categories have dominated landings. Planktivores have been landed typically in high-volume and low-value fisheries, while other fish have represented higher-value products. A very significant fraction of the planktivore landings has contributed to fishmeal, fish oil, and other products not used for human consumption. Although invertebrate landings have been relatively low volume, they have accounted for a disproportionate share of the landed value (Christensen *et al.*, 2008; Christensen *et al.*, 2009).

FIGURE 21
Trends in landings by ecotype for invertebrates, planktivores, and other fish (benthivores and piscivores) by ecotype



Geographical patterns in total landings (tonnes km^{-2}) within the LMEs examined in this report are shown in Figure 22 (upper panel). Total landings generally reflect production levels for planktivores, benthivores and piscivores, with the highest in upwelling zones, certain western boundary current regions and areas of high nutrient concentration. The working group also calculated information on the total catch where discard information (Kelleher, 2005) was available by LME (Figure 22).

FIGURE 22
Average landings by LME for the period 1998–2007(upper) and discard estimates (lower) from Kelleher (2005). Areas shaded in grey do not have available discard estimates.



Note: Areas shaded in grey do not have available discard estimates.

MEAN TROPHIC LEVEL AND DOMINANCE OF LANDINGS

Global fishery landings have remained relatively constant or have slightly declined in recent decades. Effective utilization of the production potential of the harvestable components of marine ecosystems will require reduction in overexploitation of some ecosystem components and a diversification of the suite of species harvested for others. FAO (2012) estimated that about 30 percent of the stocks for which evaluations were possible were overexploited or depleted. Consideration of options for diversification of the species harvested will require careful consideration of overall ecosystem impacts of harvesting on other components of the system. Potential utilization of forage fish species has garnered particular interest for this reason (e.g. Cury *et al.*, 2011; Pikitch *et al.*, 2012).

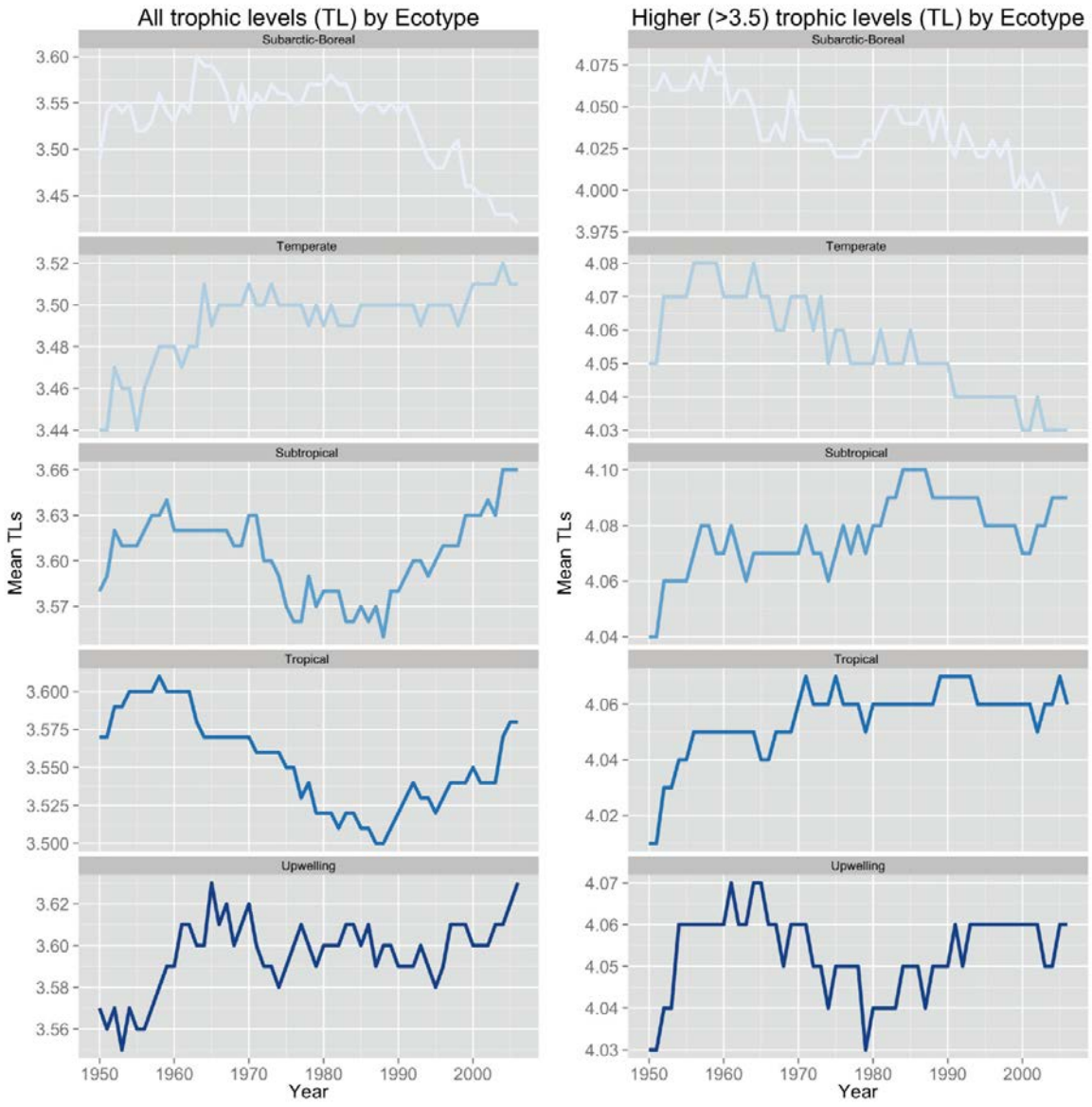
Declines in catch in three of the five ecotypes reflect overexploitation, particularly in the planktivore and other fish categories in these systems. There has been an accompanying reduction in the MTL in the landings for the subarctic-boreal system in both the total and upper (i.e. > 3.5) trophic level components (Figure 23). The working group also documented a decline in the MTL of the upper food web for the temperate system, but not over all trophic levels. In both the subtropical and tropical ecotypes, the working group observed general declines in the MTL over all trophic levels from the 1960s through the late 1980s to the early 1990s, followed by a significant increase. For these

ecotypes, the estimated MTL increased during the 1950s, then stabilized or increased for the upper trophic level components after the mid-1960s. Finally, the MTL for all trophic levels in upwelling systems increased through the 1950s and then fluctuated around a relatively stable level. For this ecotype, the upper trophic level component exhibited a more complex pattern with a pronounced decline and recovery in the period between 1960 and 1990.

The patterns observed here were generally consistent with those reported by Pauly *et al.* (1998) in their earlier characterization of 'fishing down marine food webs'. Essington, Beaudreau and Wiedenmann (2006) noted that declines in MTL can occur through the development of new fisheries on lower trophic levels and not simply through overfishing of high trophic level species. Essington, Beaudreau and Wiedenmann (2006) referred to this phenomenon as "fishing through marine food webs". This issue is related to the diversification of fisheries, a key consideration in any potential increase in yield from global fisheries.

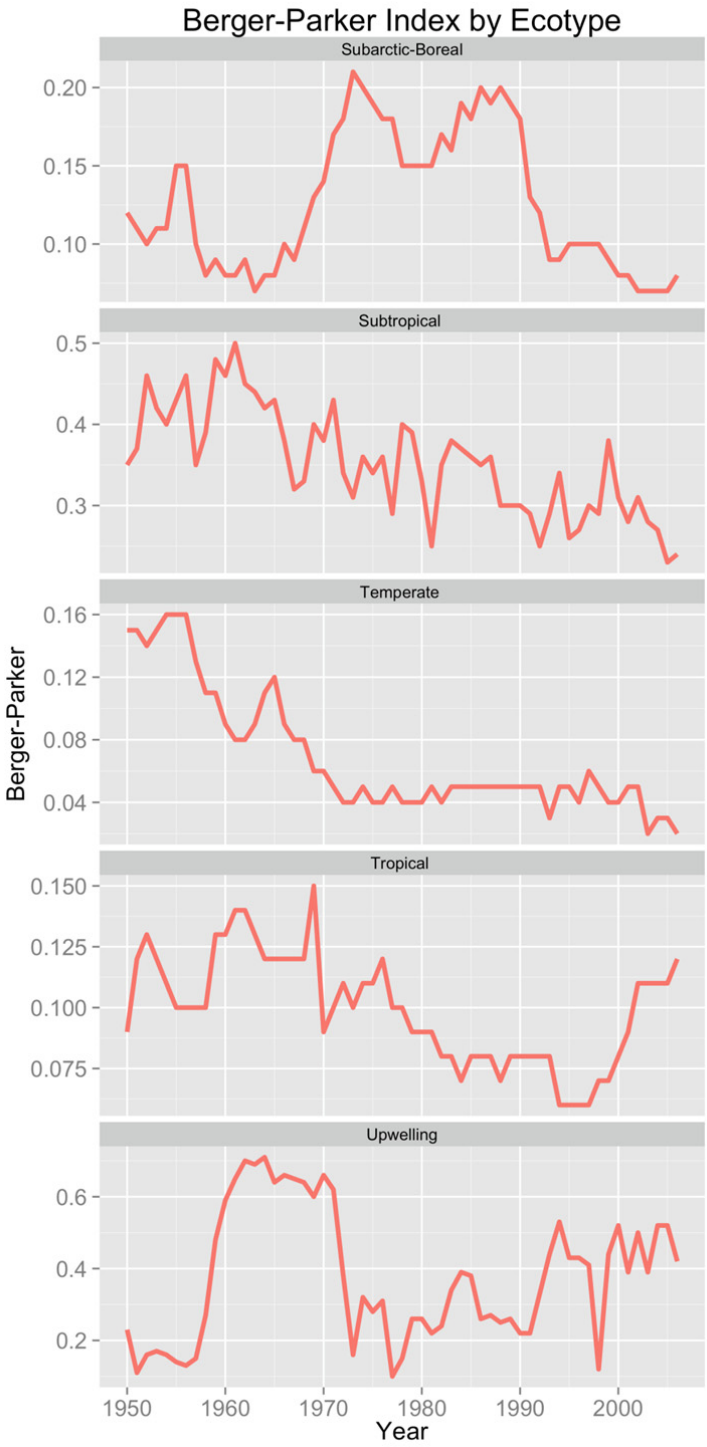
FIGURE 23

Trends in the mean trophic level of the catch (left panel) for all species and upper trophic level species (right panel) by ecotype



Changes in the diversity of landings can directly affect the utilization of the available production. A more diverse pattern in the landings can indicate a beneficial broadening of harvesting levels throughout the system if it also entails lowering exploitation rates on a more select group of species. Although the working group presented results from the Berger-Parker index for the entire time series of landings for each of the ecotypes (Figure 24), the period since 1990 is considered to be more reliable. Potential concerns related to the level of species identification in the landings prior to this time indicate that caution is necessary in interpreting the taxonomic information in the earlier period for at least some of the LMEs considered (Figure A2.1). For the most recent period, the working group observed a general diversification of the landings in the subarctic-boreal and subtropical ecotypes, and in the temperate ecotype since 2000 (the inverse of the Berger-Parker index is a measure of diversification). For the tropical systems, there has been an increase in species dominance in landings since the mid-1990s, and for the upwelling ecotype, species dominance has fluctuated without apparent trend.

FIGURE 24
Trends in species dominance of the catch as measured by a modification of the Berger-Parker diversity index

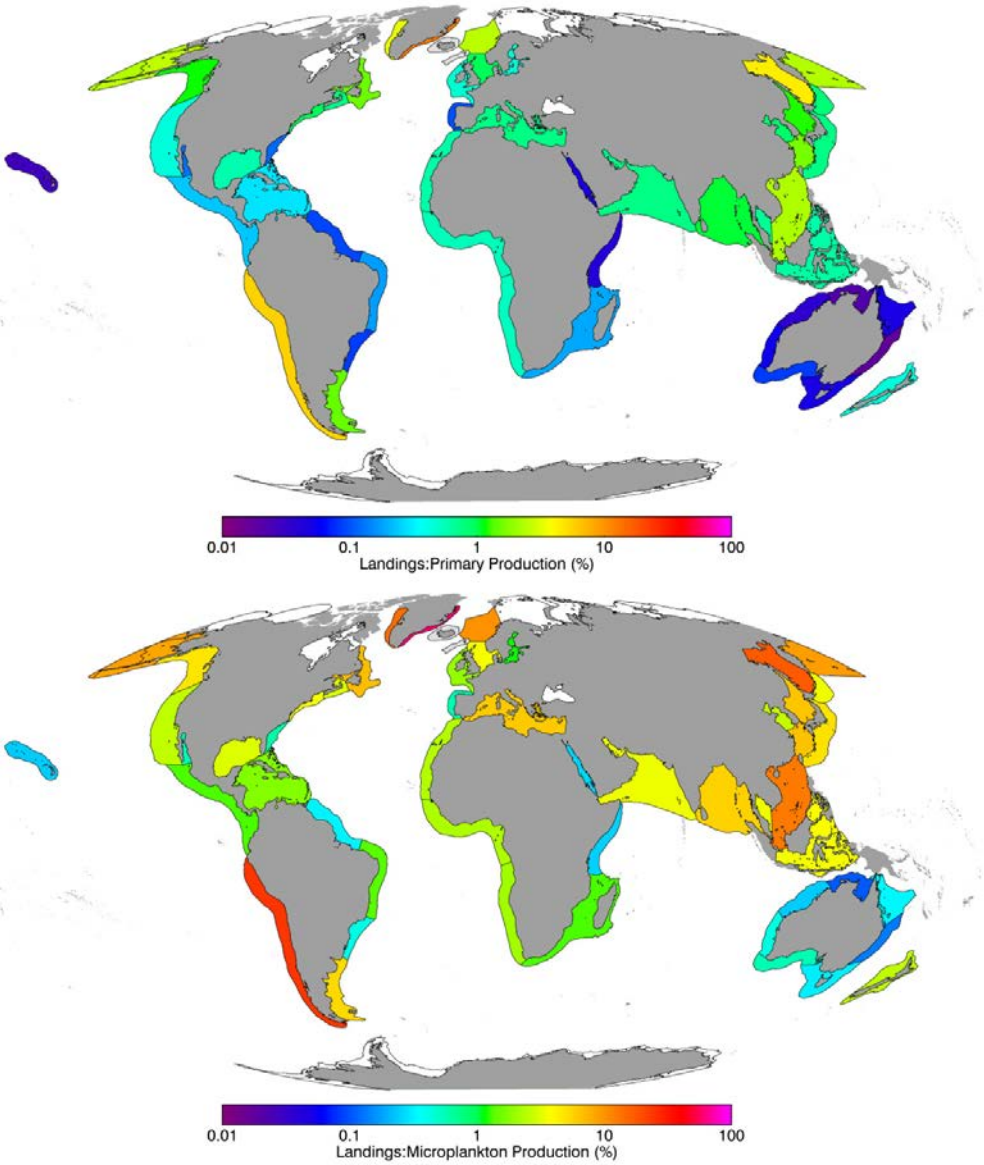


Note: Lower levels of the index reflect higher levels of overall diversification in the species mix landed in the fishery.

YIELD EFFICIENCY INDEX

The ratio of landings or catch to the level of primary production in a system has been used as a simple index of utilization patterns in marine ecosystems (e.g. Ware, 2000). This index differs from the concept of primary production required (PPR; Pauly and Christensen, 1995) to support observed catch levels. The PPR concept associates yields at a specified trophic level to the production at that trophic level. The yield efficiency index calculated here was relatively low, reflecting the dissipation of energy in successive steps in the food web. However, the working group also considered that the perception of the yield efficiency differs depending on whether total primary production (from both nano-picoplankton and microplankton) or microplankton production alone (Figure 25). The yield efficiency index was again highest in upwelling systems, those with seasonal or intermittent upwelling, and some western boundary current systems, particularly when the working group used the microplankton-based metric. In some areas where satellite-derived estimates of primary production may be underestimated, the yield efficiency ratio will be inflated. Results for several high-latitude LMEs appeared to reflect this issue. In contrast, high levels of non-organic particulate matter in the water column may result in overestimates of primary production and underestimates of yield efficiency in immediate nearshore regions. For example, in Australia, high iron and dust content in the water column can affect satellite-derived estimates.

FIGURE 25
Ratio of landings to total phytoplankton primary production (nano-picoplankton and microplankton production) (upper), and the ratio of landings to microplankton primary production (lower panel)



FISHERY PRODUCTION POTENTIAL

Estimates of fishery production potential depend on: the available production at different trophic levels; the proportion of the production comprising species suitable for harvest (including considerations of species composition, marketability, and economic efficiency of harvesting operations); and the determination of sustainable exploitation levels. The working group provided estimates of the overall available production by ecotype and functional group for potentially harvestable components of the LMEs considered in this report. In this study, the working group applied the proposed limiting exploitation level set by the fraction of microplankton production. Assumptions concerning the suitability for harvest at different trophic levels for species comprising the functional groups were necessary to complete these estimates.

In the following, the working group assumed that about 10 percent of the benthic production consists of species suitable for harvest (principally molluscs, crustaceans and a restricted number of echinoderm species). The working group based this assumption on the estimated biomass of meio-, macro- and mega-benthos fauna by Peters-Mason (2007) with the additional assumption that 50 percent of the megabenthic species biomass would be available for harvest. Based on inspection of the set of Ecopath models used earlier in this analysis to define transfer efficiency rates, the working group assumed that 75 percent of the production of the benthivore production and that 90 percent of the piscivore production was potentially harvestable. For planktivores, the working group assumed that 50 percent of the production comprised species of current or potential economic importance. In reality, this value could be higher or lower, depending on how markets and economically efficient harvesting strategies develop for mesopelagic and other underutilized planktivorous fishes. Jennings *et al.* (2008) conducted a size-based analysis using macroecological principles that assumed 50 percent of the biomass of organisms greater than 10 g comprised fish.

Under this set of assumptions, the working group estimated an overall potential yield of about 180 million tonnes for the benthivore, planktivore and piscivore functional groups for the LMEs considered here, and about 50 million tonnes of benthic organisms if up to 10 percent of the benthic production is suitable for harvest (Table 14). Although this level of benthic fishery yield may not be fully attainable by capture fisheries under current market preferences and economic conditions, the working group notes that the energetic pathways supporting natural benthic production could also potentially support enhanced mariculture production for molluscs in particular. Aquaculture production has been rapidly increasing (FAO, 2012) and, although freshwater aquaculture remains dominant, important increases in mariculture are possible and would require adequate environmental controls. The estimate of potential fish production is strongly influenced by the assumption that 50 percent of planktivore production is potentially suitable for harvest. If this estimate is high, the fishery production potential will be correspondingly lower.

TABLE 14
Estimated potential yield (in thousand tonnes) by ecotype and functional group

Ecotype	Benthos	Benthivore	Planktivore	Piscivore
Subarctic-boreal	6 628.6 (4 619.4, 8 639.9)	5 801.8 (3 833.8, 7 911.1)	24 195.7 (17 131.7, 31 036.0)	4 290.5 (3 077.2, 5 493.2)
Temperate	4 519.2 (3 525.6, 5 595.7)	3 522.4 (2 498.8, 4 600.9)	14 077.8 (10 805.4, 17 587.8)	3 026.6 (2 257.1, 3 882.1)
Subtropical	1 969.2 (1 280.3, 2 728.2)	1 315.9 (727.4, 2 034.0)	5 178.7 (3 321.2, 7 345.0)	875 (527.1, 1 291.5)
Tropical	6 557.7 (4 215.7, 9 317.9)	6 096.7 (3 420.8, 9 266.7)	24 184.3 (16 277.6, 33 264.1)	4 777.3 (3 022.3, 6 919.2)

Note: The 95 percent confidence intervals are in parentheses.

RATIO OF CATCH TO AVAILABLE PRODUCTION

Benthivorous, piscivorous and planktivorous fish represent the dominant component of landings in the LMEs considered in this study. As noted above, many exploited fish species can occupy all of these compartments owing to ontogenetic shifts in diet and generalist feeding strategies. For example, Atlantic cod (*Gadus morhua*), typically classified as a piscivore, is planktivorous in the larval stage, is principally benthivorous in the juvenile stage and follows a mixed feeding strategy including benthos and fish as an adult. Full information required to partition the existing catches into different life-history stanzas for individual species is not available on a global scale. The working group examined a combined benthivore–piscivore category in the assessment of the ratio of catch to available production in an attempt to circumvent some of the difficulties associated with generalist feeding strategies for demersal fish, particularly in assessing harvesting levels by trophic category. The working group retained the original designation of planktivores as the second major category for exploited fish species in the following analysis.

The working group focused on exploited fish species in the assessment of the catch-to-production ratio. Catches of benthic organisms remained relatively low in relation to the potentially available production and are not considered further in this section. Christensen *et al.* (2008, 2009) and Jennings *et al.* (2008) have also recently focused on exploited fish species in their evaluations of production potential. Benthic production in relation to yield were largely absent in early analyses of fishery production potential (but see Moiseev, 1971, 1994).

The working group used the global analysis of discards by LME (Kelleher, 2005), as noted above, to determine catch levels (landings plus discards). Discard data were not available for all LMEs; the working group therefore applied the average discard rate over all LMEs where discard estimation was possible. The catch-to-production ratio was high relative to the proposed reference point in the subarctic-boreal ecotype for benthivorous and piscivorous fish, substantially exceeding the limit set by microplankton production (Table 15). The estimated ratio in temperate systems also exceeded the proposed limit reference point. The working group estimated that subtropical and tropical ecotypes have lower catch-to-production ratios for benthivorous and piscivorous fish with means near or below the proposed reference levels. The highly productive upwelling systems were uniformly low relative to proposed limits for benthivorous and piscivorous fish. However, the high ratio of catch to available production in these systems may reflect the difficulties in estimating production at higher latitudes using satellite data. The estimates for this ratio for planktivorous fish were generally low. In general, this suggests that fisheries do not currently realize the production potential of potentially important fisheries in pelagic systems.

TABLE 15
Estimated ratios of catch to available production for benthivorous and piscivorous fish and for planktivorous fish

Ecotype	Benthivorous and piscivorous fish	Planktivorous fish
Subarctic-boreal	49.3 (37.8, 62.8)	7.6 (5.5, 10.1)
Temperate	31.2 (24.4, 38.6)	7.3 (5.6, 9.2)
Subtropical	23 (15.2, 31.9)	8.91 (5.69, 12.6)
Tropical	16.8 (11.2, 23.3)	8 (5.4, 11.0)
Upwelling	3.5 (2.2, 5.1)	4.8 (3.1, 7.0)

Note: The 95 percent confidence intervals are in parentheses.

As a whole, the relatively low ratios of catch to available production for planktivorous assemblages found here do not imply that traditionally exploited planktivorous species were exploited at low levels (Christensen *et al.*, 2009). If fishery production is to be expanded, it will require identification and sustainable exploitation of a broader array of planktivorous species.

7. DISCUSSION

In order to meet the demands of a burgeoning human population, it is necessary to understand how to increase fishery production sustainably. The United Nations projects that the global population will grow from 7 billion to 11 billion by 2050.[8] Currently, 3 billion people obtain almost 20 percent of their dietary animal protein needs from aquatic sources, and 4.3 billion obtain about 15 percent of these requirements from fishery and aquaculture products (FAO, 2012). Global per capita consumption of aquatic food products has increased steadily in recent years (FAO, 2012), and sharp regional differences between availability and utilization and between developed and developing countries make this an important issue in global food security. Here, the working group provided updated estimates of global fishery production potential from marine capture fisheries to contextualize the prospects for meeting human demands for protein and essential micronutrients.

The working group has developed the first application of a new approach to estimating fishery production potential. Earlier such analyses (e.g. Graham and Edwards, 1962; Schaefer, 1965; Ricker, 1969; Ryther, 1969; Gulland, 1970, 1971) relied on a combination of methods including temporal and spatial extrapolations of catch trends and simple food chain models. The latter considered overall phytoplankton primary production, ecological transfer efficiencies (typically a single value applied to all trophic levels) and the designation of a single MTL, at which catch was extracted. The approach broadened the consideration of energetic pathways through the classical grazing and microbial food webs. The working group allowed for differential ecological transfer efficiencies for individual trophic levels and for extraction of catches at multiple levels in the food web. The working group attempted to strike a balance between the simple earlier models and more-complex ecological network models that often require specified parameter estimates for a large number of nodes representing different species or species groups. The approach involved a projection through this simplified food web starting with phytoplankton production. It therefore explicitly considered bottom-up forcing of the food web to be the dominant factor in the production dynamics of these systems. This interplay between bottom-up and top-down controls can be important in many food webs.

Ryther (1969) was the first to partition ocean provinces into fishery production domains in his analysis of simple food chain models. The approach the working group adopted expands Ryther's method (Ryther, 1969) using LMEs as strata (see also Christensen *et al.*, 2008; Christensen *et al.*, 2009). Pauly (1995) suggested that drawing on multiple methods of estimation and spatial domains can help provide more robust overall determinations of fishery production potential. In the absence of other information, the working group assumed that exchange rates between LMEs were in balance. The working group also implicitly assumed that the overall analysis captured spatial dynamics when the working group integrated over LMEs to generate estimates over broader geographical scales.

The working group proposed that the fraction of microplankton production in the system could be one possible measure for an ecosystem-based reference point. Moiseev (1994) recommended that exploitation rates not exceed 20–25 percent, although he did not specify the exact rationale for this level. However, his recommendation is broadly consistent with the microplankton production reference level for the LMEs considered in this study.

It is also important to consider directly the energetic requirements of other ecosystem components. Cury *et al.* (2011) noted that seabird fledging success was significantly impaired when pelagic prey were reduced to below one-third of their presumed maximum levels in their consideration of forage fish management to meet the needs of a broad suite of predators, including mammals and seabirds. Pikitch *et al.* (2012) recommended establishing precautionary exploitation rates that halved the values assigned under conventional single-species management. Their simulations indicated this reduction in exploitation rates would result in increased overall economic returns and reduced impacts on upper-trophic-level predators. Christensen (1996) noted that estimates of the consumption of groundfish were often higher relative to catches by a factor of three. Collectively, these independent recommendations and observations suggested that exploitation rates generally should not exceed

[8] www.un.org/popin/

25 percent of available production and were consistent with the recommendation for a reference point.[9]

The first-order estimates of fishery production potential based on this new approach suggested a potential yield of about 180 million tonnes for planktonic and nektonic organisms within the LMEs considered if up to 50 percent of the production of planktivorous species were potentially suitable for harvest. The estimate of the fishery production potential for benthic organisms was about 50 million tonnes if up to 10 percent of the benthic production were suitable for harvest. Perhaps a more likely scenario for benthic production would entail a combination of expanded capture fisheries and some form of sustainable mariculture, principally for molluscs. The working group suggested that an overall diversification of the complex of harvested species should be attained if these potential yields are to be realized. It will also be necessary to reduce exploitation rates on overfished species to increase overall yields. Moiseev (1994) estimated that up to 20 percent of potential yield is lost by "non-rational exploitation".

It is clear that the greatest prospect for potential increase in fish yield is for planktivorous species. Expanding this yield entails consideration of the forage needs of other species in the system and must recognize that many of these species (e.g. mesopelagic fish) will be processed for fishmeal and fish oils, not used for direct human consumption. These mesopelagic fish may contribute to the expansion of the mariculture industry for upper-trophic-level species and can potentially support terrestrial farm industries by serving as feed supplements. Moiseev (1994) estimated the global fishery production potential for conventionally harvested species to be between 120 and 150 million tonnes. He further suggested that underutilized species including krill, deep-sea squids and mesopelagic species could potentially add between 60 and 80 million tonnes to this estimate for conventional resource species.

More recently, Jennings *et al.* (2008) estimated global fish production based on size-spectrum models to be of the order of 790 million tonnes/year, with a corresponding biomass estimate of about 900 million tonnes. Christensen *et al.* (2009) provided a fish biomass estimate of 1.1 gigatonnes for the LMEs of the world ocean. Wilson *et al.* (2009) applied the method of Christensen *et al.* (2008, 2009) to the global ocean and estimated fish biomass to be of the order of 2 gigatonnes. The geographic coverage of the analysis in Wilson *et al.* (2009) is more directly comparable with the assessment by Jennings *et al.* (2008), resulting in higher overall biomass estimates by a factor of two. However, specific estimates of potential fish yield were not provided in these analyses. The geographical coverage and model structure of the estimate is most directly comparable with Christensen *et al.* (2009) in that both were limited to LMEs for which satellite-derived estimates of primary production were available.

The working group provided these first-order estimates with the recognition that inputs to the analytical framework presented here will need to be continually refined. The international scientific community is making significant advances in satellite oceanography that will improve estimates of size-fractionated chlorophyll concentrations. Corrections for potential biases in chlorophyll concentration in nearshore waters due to particulate matter other than phytoplankton in the surface layer are under constant development, resulting in improved estimates. However, attention to specialized issues in some LMEs (e.g. Australian shelf systems) will require further consideration. The estimates of trophic transfer efficiency and energetic pathways through the benthos and mesozooplankton can be re-evaluated by examining additional food webs constructed for marine systems. A critically important need is to refine the estimation of the harvestable component of the benthic and planktivorous components of the food web. These improvements and further developments in the size-based macroecological approach of Jennings *et al.* (2008), the data-driven EwE approach of Christensen *et al.* (2009) and others will eventually allow for a multimodel approach that will permit evaluation of the effects of alternative assumptions and model structures on estimates of fishery production potential.

[9] For alternative qualitative recommendations concerning ecosystem overfishing definitions, see Murawski (2000) and Tudela, Coll and Palomera (2005).

Potential impacts of climate change will be particularly important in assessing issues related to global food production and security. Increased stratification, particularly in mid-latitude regions that are currently characterized by relatively high fishery potential will probably result in changes in phytoplankton community composition and a potential increase in the importance of energetic pathways through the microbial food web. A decrease in ice cover and other factors in high-latitude systems will potentially increase overall fishery production potential for coldwater species. It will be increasingly important to couple the outputs from general circulation models (e.g. models that forecast changing phytoplankton communities and associated physical variables) with food web models in order to assess potential impacts on global food supplies from marine ecosystems.

8. REFERENCES

Abramowitz, M. & Stegun, I.A. 1972. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York, USA, Dover.

Alverson, D.L., Freeberg, M.H., Murawski, S.A. & Pope, J. 1994. *A global assessment of fisheries bycatch and discards.* Rome, FAO. 233 pp.

Anticamara, J.A., Watson, R., Gelchu, A. & Pauly, D. 2011. Global fishing effort (1950-2010): trends, gaps, and implications. *Fisheries Research*, 107(1–3): 131–136.

Behrenfeld, M.J. & Falkowski, P.G. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, 42 (1): 1–20.

Berger, W. & Parker, F. 1970. Diversity of planktonic foraminifera in deep-sea sediments. *Science*, 168: 1345–1347.

Berkson, J., Barbieri, L., Cadrin, S., Cass-Calay, S., Crone, P., Dom, M., Friess, C. & Kobayashi, D. 2011. *Calculating acceptable biological catch for stocks that have reliable catch data only (Only Reliable Catch Stocks - ORCS). 56. NOAA Technical Memorandum NMFS-SEFSC-616*. Miami, USA, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southeast Fisheries Science Center.

Branch, T.A., Jensen, O.P., Ricard, D., Ye, Y.M. & Hilborn, R. 2011. Contrasting global trends in marine fishery status obtained from catches and from stock assessments. *Conservation Biology*, 25(4): 777–786.

Caddy, J.F. & Gulland, J.A. 1983. Historical patterns of fish stocks. *Marine Policy*, 7: 267–278.

Casella, G. & Berger, R.L. 2002. *Statistical inference*. Second edition. Pacific Grove, USA, Duxbury Thomson Learning.

Charnov, E.L. & Gillooly, J.F. 2004. Size and temperature in the evolution of fish life histories. *Integrative and Comparative Biology*, 44(6): 494–497.

Christensen, V. 1996. Managing fisheries involving predator and prey species. *Review of Fish Biology and Fisheries*, 6: 417–443.

Christensen, V. & Pauly, D. 1992. ECOPATH II – a software for balancing steady-state ecosystem models and calculating network characteristics. *Ecological Modelling*, 61: 169–185.

Christensen, V., Walters, C.J. & Pauly, D. 2005. *Ecopath with Ecosim: a user's guide*. Vancouver, Canada, Fisheries Centre, University of British Columbia.

Christensen, V., Walters, C.J., Ahrens, R., Alder, J., Buszowski, J., Christensen, L.B., Cheung, W.W.L., Dunne, J., Froese, R., Karpouzi, V., Kaschner, K., Kearney, K., Lai, S., Lam, V., Palomares, M.L.D., Peters-Mason, A., Pirodi, C., Sarmiento, J.L., Steenbeek, J., Sumaila, R., Watson, R., Zeller, D. & Pauly, D. 2008. *Models of the world's large marine ecosystems*. GEF/LME global project Promoting Ecosystem-based Approaches to Fisheries Conservation and Large Marine Ecosystems IOC Technical Series No. 80. Paris, UNESCO.

Christensen, V., Walters, C.J., Ahrens, R., Alder, J., Buszowski, J., Christensen, L.B., Cheung, W.W.L., Dunne, J., Froese, R., Karpouzi, V., Kaschner, K., Kearney, K., Lai, S., Lam, V., Palomares, M.L.D., Peters-Mason, A., Piroddi, C., Sarmiento, J.L., Steenbeek, J., Sumaila, R., Watson, R., Zeller, D. & Pauly, D. 2009. Database-driven models of the world's large marine ecosystems. *Ecological Modelling*, 220(17): 1984–1996.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O. & Lester, S.E. 2012. Status and solutions for the world's unassessed fisheries. *Science*, 338(6106): 517–520.

Cury, P.M., Boyd, I.L., Bonhommeau, S., Anker-Nilssen, T., Crawford, R.J.M., Furness, R.W., Mills, J.A., Murphy, E.J., Oesterblom, H., Paleczny, M., Piatt, J.F., Roux, J.-P., Shannon, L. & Sydeman, W.J. 2011. Global seabird response to forage fish depletion – one-third for the birds. *Science*, 334(6063): 1703–1706.

De'ath, G. & Fabricius, K. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81(11): 3178–3192.

Dick, E.J. & MacCall, A.D. 2011. Depletion-based stock reduction analysis: a catch-based method for determining sustainable yields for data-poor fish stocks. *Fisheries Research*, 110(2): 331–341.

Eppley, R.W. 1972. Temperature and phytoplankton growth in the sea. *Fishery Bulletin*, 70(4): 1063–1085.

Essington, T.E., Beaudreau, A.H. & Wiedenmann, J. 2006. Fishing through marine food webs. *Proceedings of the National Academy of Sciences of the United States of America*, 103(9): 3171–3175.

FAO. 2010. *The State of World Fisheries and Aquaculture 2010*. Rome. 197 pp.

FAO. 2012. *The State of World Fisheries and Aquaculture 2012*. Rome. 209 pp.

Froese, R. & Kesner-Reyes, K. 2002. *Impact of fishing on the abundance of marine species*. Copenhagen, ICES CM 12/L:12.

Froese, R. & Pauly, D. 2000. *FishBase 2000: Concepts, design, and data sources*. Los Baños, Philippines, ICLARM.

Froese, R., Zeller, D., Kleisner, K. & Pauly, D. 2012. What catch data can tell us about the status of global fisheries. *Marine Biology*, 159(6): 1283–1292.

Garcia, S.M., Kolding, J., Rice, J., Rochet, M.J., Zhou, S., Arimoto, T., Beyer, J.E., Borges, L., Bundy, A., Dunn, D., Fulton, E.A., Hall, M., Heino, M., Law, R., Makino, M., Rijnsdorp, A.D., Simard, F. & Smith, A.D.M. 2012. Reconsidering the consequences of selective fisheries. *Science*, 335(6072): 1045–1047.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. 2004. *Bayesian data analysis*. Boca Raton, USA, Chapman & Hall/CRC.

Gislason, H., Pope, J.G., Rice, J.C. & Daan, N. 2008. Coexistence in North Sea fish communities: implications for growth and natural mortality. *ICES Journal of Marine Science: Journal du Conseil*, 65(4): 514–530.

Gjøsæter, J. & Kawaguchi, K. 1980. *A review of the world resources of mesopelagic fish*. FAO Fisheries Technical Paper No. 193. Rome, FAO. 151 pp.

Graham, H.W. & Edwards, R.L. 1962. The world biomass of marine fishes. *In* E. Heen & R. Kreuzer, eds. *Fish in nutrition*. London, Fishing New Books.

Gulland, J.A., comp. 1970. *The fish resources of the ocean*. FAO Fisheries Technical Paper No. 97. Rome, FAO. 418 pp.

Gulland, J.A. 1971. *The fish resources of the ocean*. West Byfleet, UK, Fishing New Books.

Hilborn, R. & Walters, C.J. 1991. *Quantitative fisheries stock assessment: choice, dynamics, and uncertainty*. Norwell, USA, Springer.

International Council for the Exploration of the Sea (ICES). 2012a. *Report of the Workshop on the Development of Assessments based on LIFE history traits and Exploitation Characteristics (WKLIFE), 13–17 February 2012, Lisbon, Portugal*. Lisbon, ICES.

International Council for the Exploration of the Sea (ICES). 2012b. *Working Group on Methods of Fish Stock Assessments (WGMG)*. Lisbon, ICES.

Iverson. R.L. 1990. Control of marine fish production. *Limnology and Oceanography*, 35(7): 1593–1604.

Jennings, S., Melin, F., Blanchard, J.L., Forster, R.M., Dulvy, N.K. & Wilson, R.W. 2008. Global-scale predictions of community and ecosystem properties from simple ecological theory. *Proceedings of the Royal Society B: Biological Sciences*, 275: 1375–1383.

Kell, L.T., Mosqueira, I., Grosjean, P., Fromentin, J.M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F. & Scott, R.D. 2007. FLR: an open-source framework for the evaluation and development of management strategies. *ICES Journal of Marine Science*, 64(4): 640–646.

Kelleher, K. 2005. *Discards in the world's marine fisheries: an update*. FAO Fisheries Technical Paper No. 470. Rome, FAO. 152 pp.

Kesteven, G.L. & Holt, S.J. 1955. *A note on the fisheries resources of the North West Atlantic*. FAO Fisheries Paper No. 7. Rome, FAO. 12 pp.

Kimura, D.K. & Tagart, J.V. 1982. Stock reduction analysis, another solution to the catch equations *Canadian Journal of Fisheries and Aquatic Sciences*, 39(11): 1467–1472.

Kimura, D.K., Balsiger, J.W. & Ito, D.H. 1984. Generalized stock reduction analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 41: 1325–1333.

Kirk, J.T.O. 1994. *Light and photosynthesis in aquatic ecosystems.* Cambridge, UK, Cambridge University Press.

Kleisner, K., Zeller, D., Froese, R. & Pauly, D. 2013. Using global catch data for inferences on the world's marine fisheries. *Fish and Fisheries*, 14(3): 293–311.

MacCall, A.D. 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. *ICES Journal of Marine Science*, 66: 2267–2271.

May, R. M. 1975. Patterns of species abundance and diversity. *In* M. L. D. Cody, J. M. Diamond eds. *Ecology and Evolution of Communities*. Cambridge, Mass. Harvard University Press.

Martell, S. & Froese, R. 2013. A simple method for estimating MSY from catch and resilience. *Fish and Fisheries*, 14(4): 504–514.

McAllister, M.K. & Kirchnet, C. 2002. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. *Bulletin of Marine Sciences*, 70: 499–540.

McAllister, M.K. & Pikitch, E.K. 1997. A Bayesian approach to choosing a design for surveying fisheries resources: application to the eastern Bering Sea trawl survey. *Canadian Journal of Fisheries and Aquatic Sciences*, 54: 301–311.

McAllister, M.K., Pikitch, E.K., Punt, A.E. & Hilborn, R. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences*, 51: 2673–2687.

Medley, P., Cheung, W., Fulton, B. & Minte-Vera, C.V. 2009. *Multispecies and ecosystem indicators, and biomass-fleet dynamics stock assessment: an initial evaluation.* FAO Fisheries and Aquaculture Circular No. 1045. Rome, FAO. 28 pp.

Moiseev, P.A. 1969. *The living resources of the world ocean Pishchevaiia prmyshlannost.* Moscow. 338 pp. (translated edition 1971, Jerusalem, by Israel Program for Scientific Translations).

Moiseev P.A. 1971. *The living resources of the world ocean*. Jerusalem, Israel Program for Scientific Translations.

Moiseev, P.A. 1994. Present fish productivity and bioproduction potential of the world aquatic habitats. *In* C.W. Voigtlander, ed. *The state of the world's fisheries: Proceedings of the World Fisheries Congress*. Oxford, UK, and New Delhi, India, IBH.

Moler, C. 1986. Matrix computation on distributed memory multiprocessors. *In* M.T. Heath, ed. *Hypercube multiprocessors*. Philadelphia, USA, Society for Industrial and Applied Mathematics.

Mora, C., Myers, R.A., Coll, M., Libralato, S., Pitcher, T.J., Sumaila, R.U., Zeller, D., Watson, R., Gaston, K.J. & Worm, B. 2009. Management effectiveness of the world's marine fisheries. *PLoS Biol*, 7(6): e1000131 [online]. [Cited 5 November 2013]. www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1000131

Morel, A. 1991. Light and marine photosynthesis: a spectral model with geochemical and climatological implications. *Progress in Oceanography*, 96: 263–306.

Murawski, S.A. 2000. Definitions of overfishing from an ecosystem perspective. *ICES Journal of Marine Science*, 57(3): 649–658.

Musick, J. 1999. Criteria to define extinction risk in marine fishes. *Fisheries*, 24: 6–14.

Myers, R.A. 2001. Stock and recruitment: generalizations about maximum reproductive rate, density dependence, and variability. *ICES Journal of Marine Science*, 58: 937–951.

Myers, R.A., Bowen, K.G. & Barrowman, N.J. 1999. Maximum reproductive rate of fish at low population sizes. *Canadian Journal of Fisheries and Aquatic Sciences*, 56: 2404–2419.

National Research Council. 1998. *Improving fish stock assessments*. Washington, DC, National Academy Press.

O'Reilly, J.E., Evans-Zetlin, C. & Busch, D.A. 1987. Primary Production. *In* R.H. Backus, ed. *Georges Bank*, pp. 220–233. Cambridge, USA, MIT Press.

Pan, X., Mannino, A., Russ, M.E., Hooker, S.B. & Harding, Jr, L.W. 2010. Remote sensing of phytoplankton pigment distribution in the United States northeast coast. *Remote Sensing of Environment*, 114(11): 2403–2416.

Pan, X., Mannino, A., Marshall, H.G., Filippino, K.C. & Mulholland, M.R. 2011. Remote sensing of phytoplankton community composition along the northeast coast of the United States. *Remote Sensing of Environment*, 115(12): 3731–3747.

Pauly, D. 1995. Anecdotes and the shifting baseline syndrome of fisheries. *Trends in Ecology and Evolution*, 10(10): 430.

Pauly D. 1996. One hundred million tonnes of fish, and fisheries research. *Fisheries Research*, 25(1): 25–38.

Pauly, D. 2007. The Sea Around Us project: documenting and communicating global fisheries impacts on marine ecosystems. *Ambio*, 34: 290–295.

Pauly, D. & Christensen, V. 1995. Primary production required to sustain global fisheries. *Nature*, 376(6537): 279–279.

Pauly D., Hilborn R. & Branch T.A. 2013. Does catch reflect abundance? *Nature*, 494(7437): 303–306.

Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. 1998. Fishing down marine food webs. *Science*, 279(5352): 860–863.

Peters-Mason, A. 2007. *The conservation applications of global scale modeling: a look at climate change and benthic biomass distribution and abundance*. Duke University, USA, Durham.

Pikitch, E.K., Rountos, K.J., Essington, T.E., Santora, C., Pauly, D., Watson, R., Sumaila, U.R., Boersma, P.D., Boyd, I.L., Conover, D.O., Cury, P., Heppell, S.S., Houde, E.D.,

Mangel, M., Plagányi, É., Sainsbury, K., Steneck, R.S., Geers, T.M., Gownaris, N. & Munch, S.B. 2012. The global contribution of forage fish to marine fisheries and ecosystems. *Fish and Fisheries*: DOI: 10.1111/faf.12004 [online]. [Cited 5 November 2013]. http://onlinelibrary.wiley.com/doi/10.1111/faf.12004/abstract

Plummer, M., Best, N., Cowles, K. & Vines, K. 2009. *CODA: Output analysis and diagnostics for MCMC*.

Polovina, J.J. 1984. Model of a coral-reef ecosystem. 1. The Ecopath model and its application to French Frigate Shoals. *Coral Reefs*, 3(1): 1–11.

Punt, A.E. & Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Review of Fish Biology and Fisheries*, 7: 35–63.

Pyper, B.J. & Peterman, R.M. 1998. Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Canadian Journal of Fisheries and Aquatic Sciences*, 55: 2127–2140.

Raftery, A.E. & Bao, L. 2010. Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling. *Biometrics*, 66: 1162–1173.

Ricard, D., Minto, C., Jensen, O.P. & Baum, J.K. 2012. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish and Fisheries*, 13(4): 380–398.

Ricker, W.E. 1969. Food from the sea. *In* U.S. National Academy of Sciences. *Resources and man*. San Francisco, USA, W.H. Freeman.

Ryther, J.H. 1969. Photosynthesis and fish production from the sea. *Science*, 166(3901): 72–76.

Schaefer, M. 1954. Some aspects of the dynamics of populations important to the management of commercial marine fisheries. *Bulletin of the International American Tropical Tuna Commission*, 1(2): 26–56.

Schaefer, M.B. 1965. The potential harvest of the sea. *Transactions of the American Fisheries Society*, 94(2): 123–128.

Sela, R.J. & Simonoff, J.S. 2012. RE-EM trees: a data mining approach for longitudinal and clustered data. *Machine Learning*, 86(2): 169–207.

Sherman, K. 1991. The large marine ecosystem concept - research and management strategy for living marine resources. *Ecological Applications*, 1(4): 349–360.

Sherman, K. & Alexander, L. 1986. *Variability and management of large marine ecosystems*. Boulder, USA, Westview Press.

Slobodkin, L.B. 1961. *Growth and regulation of animal populations*. New York, USA, Holt, Rinehart, and Winston.

Steeman-Nielsen, E. 1951. Measurement of production of organic matter in sea by means of carbon-14. *Nature*, 267(4252): 684–685.

Straile, D. 1997. Gross growth efficiencies of protozoan and metazoan zooplankton and their dependence on food concentration, predator-prey weight ratio, and taxonomic group. *Limnology and Oceanography*, 42(6): 1375–1385.

Swartz, W., Sala, E., Tracey, S., Watson, R. & Pauly, D. 2010. The spatial expansion and ecological footprint of fisheries (1950 to present). *PLoS ONE*, 5(12): e15143 [online]. [Cited 5 November 2013]. www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0015143

Therneau, T., Atkinson, B. & Ripley, B. 2012. rpart: Recursive Partitioning, R package version 3.1-55.

Thorson, J.T., Branch, T.A. & Jensen, O.P. 2012. Using model-based inference to evaluate global fisheries status from landings, location, and life history data. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(4): 645–655.

Thorson, J.T., Minto, C., Minte-Vera, C.V., Kleisner, K. & Longo, C. (forthcoming). A new role for effort dynamics in the theory of harvested populations and data-poor stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences.*

Tudela, S., Coll, M. & Palomera, I. 2005. Developing an operational reference framework for fisheries management on the basis of a two-dimensional index of ecosystem impact. *ICES Journal of Marine Science*, 62(3): 585–591.

Ulanowicz, R.E. 1993. Ecosystem trophic foundation: Lindeman exonerate. *In* B.C. Patten & S.E. Jorgensen, eds. *Complex ecology*. Englewood Cliffs, USA, Prentice-Hall.

Vasconcellos, M. & Cochrane, K. 2005. Overview of world status of data-limited fisheries: inferences from landing statistics. *In* G.H. Kruse, V.F. Gallucci, D.E. Hay, R.I. Perry, R.M. Peterman, T.C. Shirley, P.D. Spencer, B. Wilson & D. Woodby, eds. *Fisheries assessment and management in data-limited situations*, pp. 1–20. Fairbanks, USA, Alaska Sea Grant College Program.

Vert-Pre, K.A., Amoroso, R.O., Jensen, O.P. & Hilborn, R. 2013. Frequency and intensity of productivity regime shifts in marine fish stocks. *Proceedings of the National Academy of Sciences*, 110(5): 1779–1784.

Vidussi, F., Claustre, H., Manca, B.B., Luchetta, A. & Marty, J.-C. 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research*, 106(C9): 19939–19956.

Ware, D.M. 2000. Aquatic ecosystems: properties and models. *In* P.J. Harrison & T.R. Parsons, eds. *Fisheries oceanography: an integrative approach to fisheries and ecology and management*. Oxford, UK, Blackwell Science.

Watson, R., Kitchingman, A., Gelchu, A. & Pauly, D. 2004. Mapping global fisheries: sharpening our focus. *Fish and Fisheries*, 5: 168–177.

Watson, R., Cheung, W.W.L., Anticamara, J.A., Sumaila, R., Zeller, D. & Pauly, D. 2013. Global marine yield halved as fishing intensity redoubles. *Fish and Fisheries*, 14(4): 493–503.

Wetzel, C.R. & Punt, A.E. 2011. Model performance for the determination of appropriate harvest levels in the case of data-poor stocks. *Fisheries Research*, 110: 342–355.

Wickham, H. 2009. *ggplot2: elegant graphics for data analysis*. New York, USA, Springer.

Wickham, H. & Hofmann, H. 2011. Product plots. *IEEE Transactions on Visualization and Computer Graphics*, 17(12): 2223–2230.

Wiedenmann, J., Wilberg, M.J. & Miller, T.J. 2013. An evaluation of harvest control rules for data-poor fisheries. *North American Journal of Fisheries Management*, 33(4): 845–860.

Wilson, R.W., Millero, F.J., Taylor, J.R., Walsh, P.J., Christensen, V., Jennings, S. & Grosell, M. 2009. Contribution of fish to the marine inorganic carbon cycle. *Science*, 323(5912): 359–362.

Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., Jensen, O.P., Lotze, H.K., Mace, P.M., McClanahan, T.R., Minto, C., Palumbi, S.R., Parma, A.M., Ricard, D., Rosenberg, A.A., Watson, R. & Zeller, D. 2009. Rebuilding global fisheries. *Science*, 325(5940): 578–585.

Yu, H. 2002. Rmpi: Parallel Statistical Computing in R. *R News*, 2(2): 10–14.

Zhou, L.M., Dickinson, R.E., Dai, A.G. & Dirmeyer, P. 2010. Detection and attribution of anthropogenic forcing to diurnal temperature range changes from 1950 to 1999: comparing multi-model simulations with observations. *Climate Dynamics*, 35(7–8): 1289–1307.

APPENDIX 1
ADDITIONAL FIGURES REFERRED IN PART I

FIGURE A1.1

Deterministic simulation trajectories of rescaled stock biomass B/B_{MSY} (black line), and estimated B/B_{MSY} for CMSY for each combination of the design given in Table 6. Harvest dynamics levels (HD) and length of time series (TS) structure the columns, initial depletion (ID) and life history (LH) structure the rows.

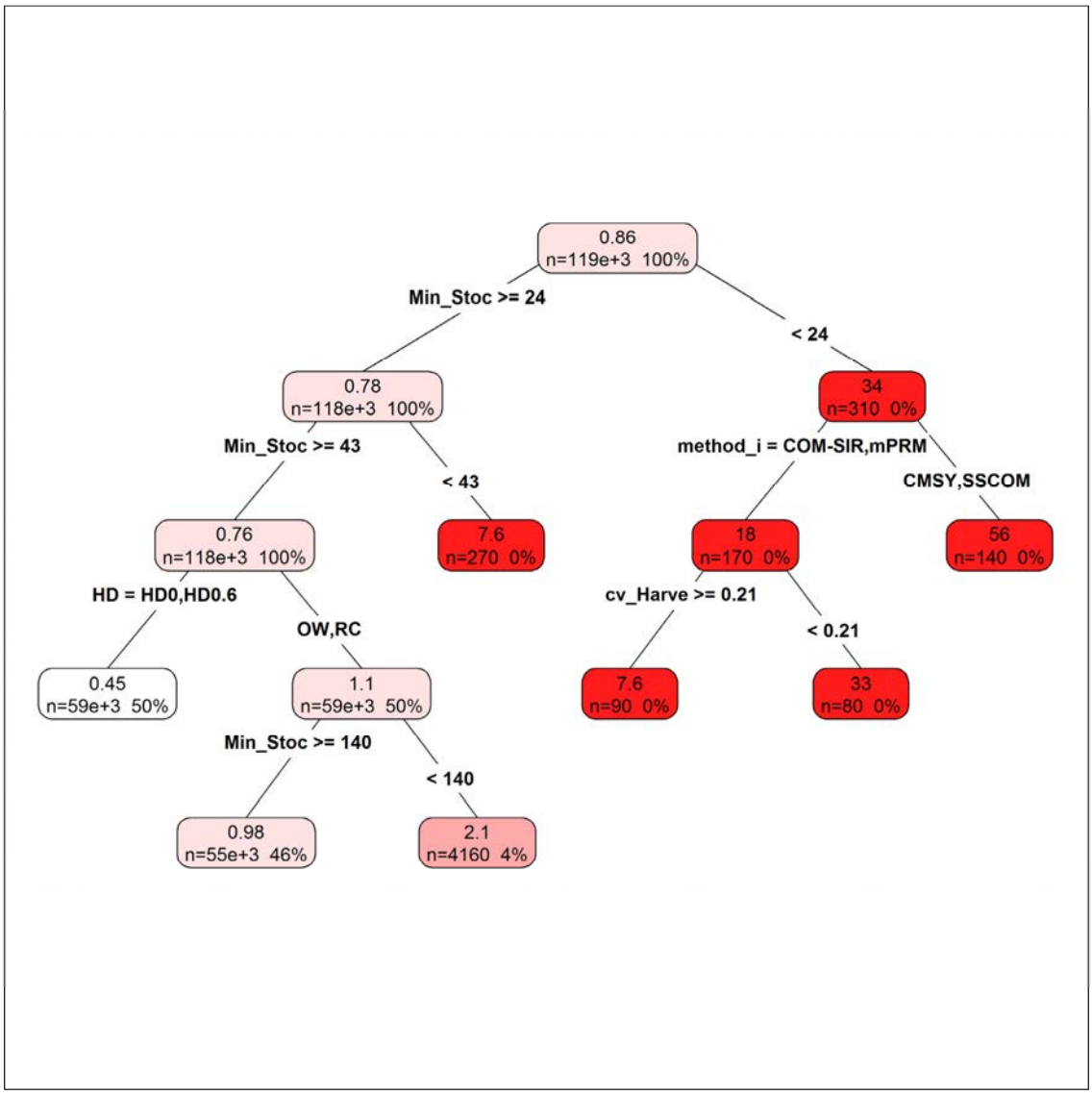


FIGURE A1.2

Deterministic simulation trajectories of rescaled stock biomass B/B_{MSY} (black line), and estimated B/B_{MSY} for COM-SIR for each combination of the design given in Table 6. Harvest dynamics levels (HD) and length of time series (TS) structure the columns, initial depletion (ID) and life history (LH) structure the rows.

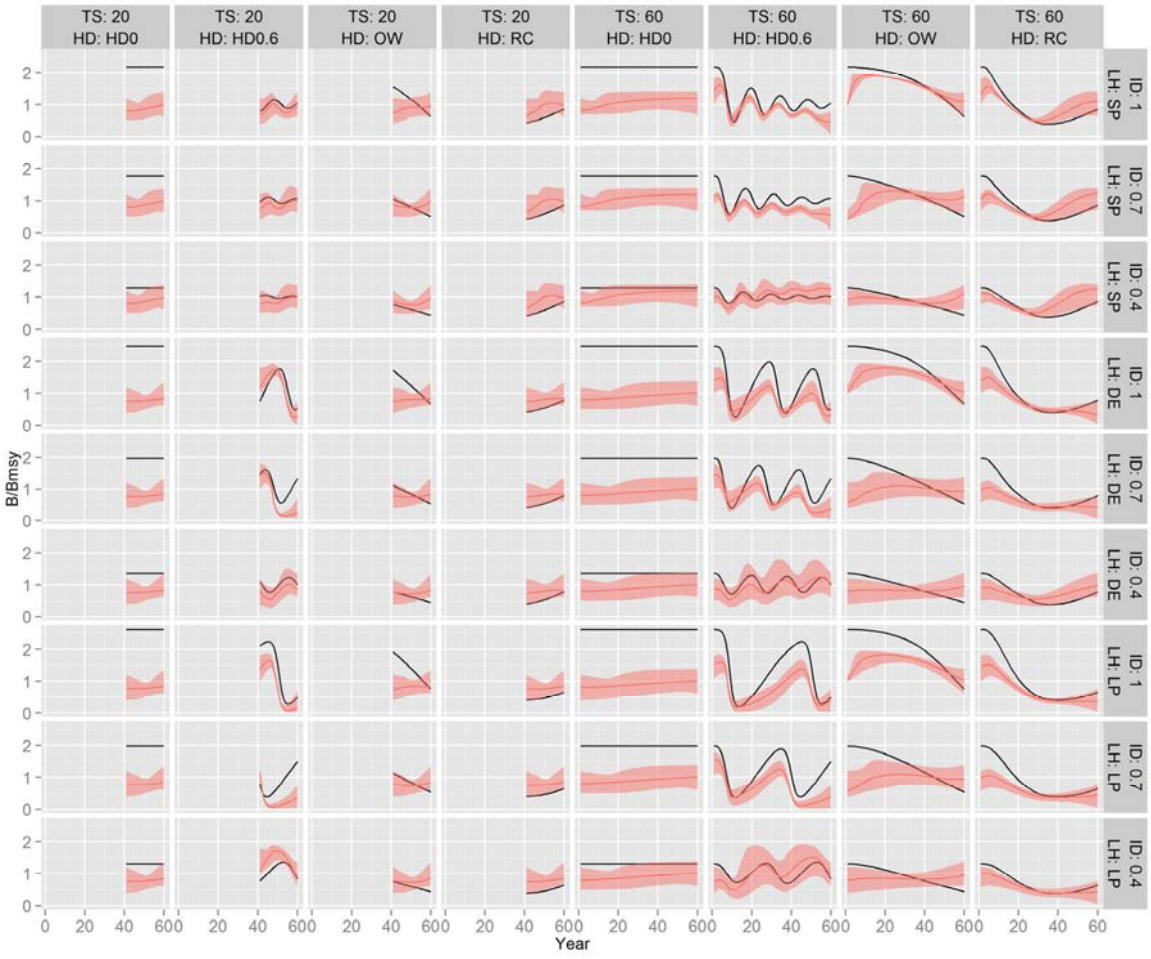


FIGURE A1.3

Deterministic simulation trajectories of rescaled stock biomass B/B_{MSY} (black line), and estimated B/B_{MSY} for mPRM for each combination of the design given in Table 6. Harvest dynamics levels (HD) and length of time series (TS) structure the columns, initial depletion (ID) and life history (LH) structure the rows.

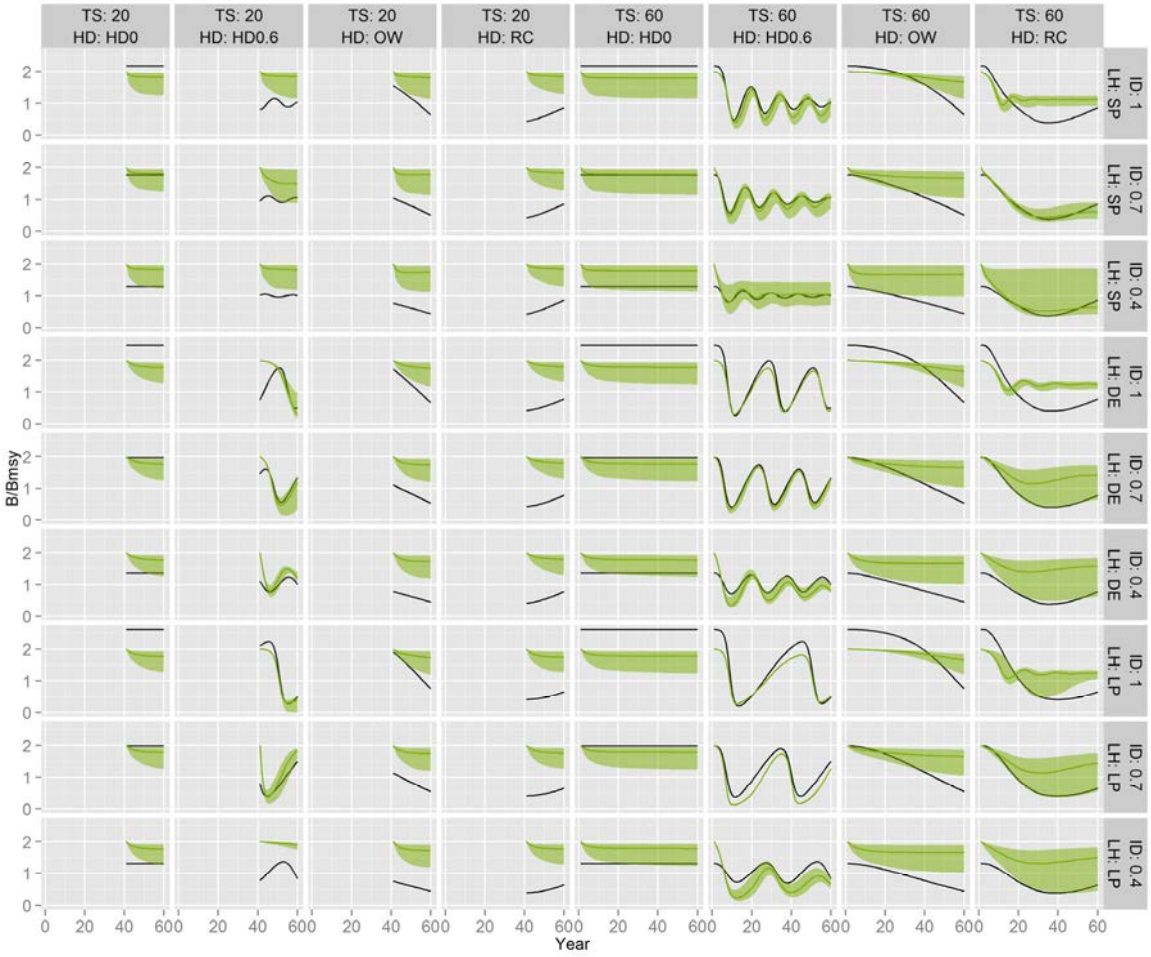


FIGURE A1.4

Deterministic simulation trajectories of rescaled stock biomass B/B_{MSY} (black line), and estimated B/B_{MSY} for SSCOM for each combination of the design given in Table 6. Harvest dynamics levels (HD) and length of time series (TS) structure the columns, initial depletion (ID) and life history (LH) structure the rows.

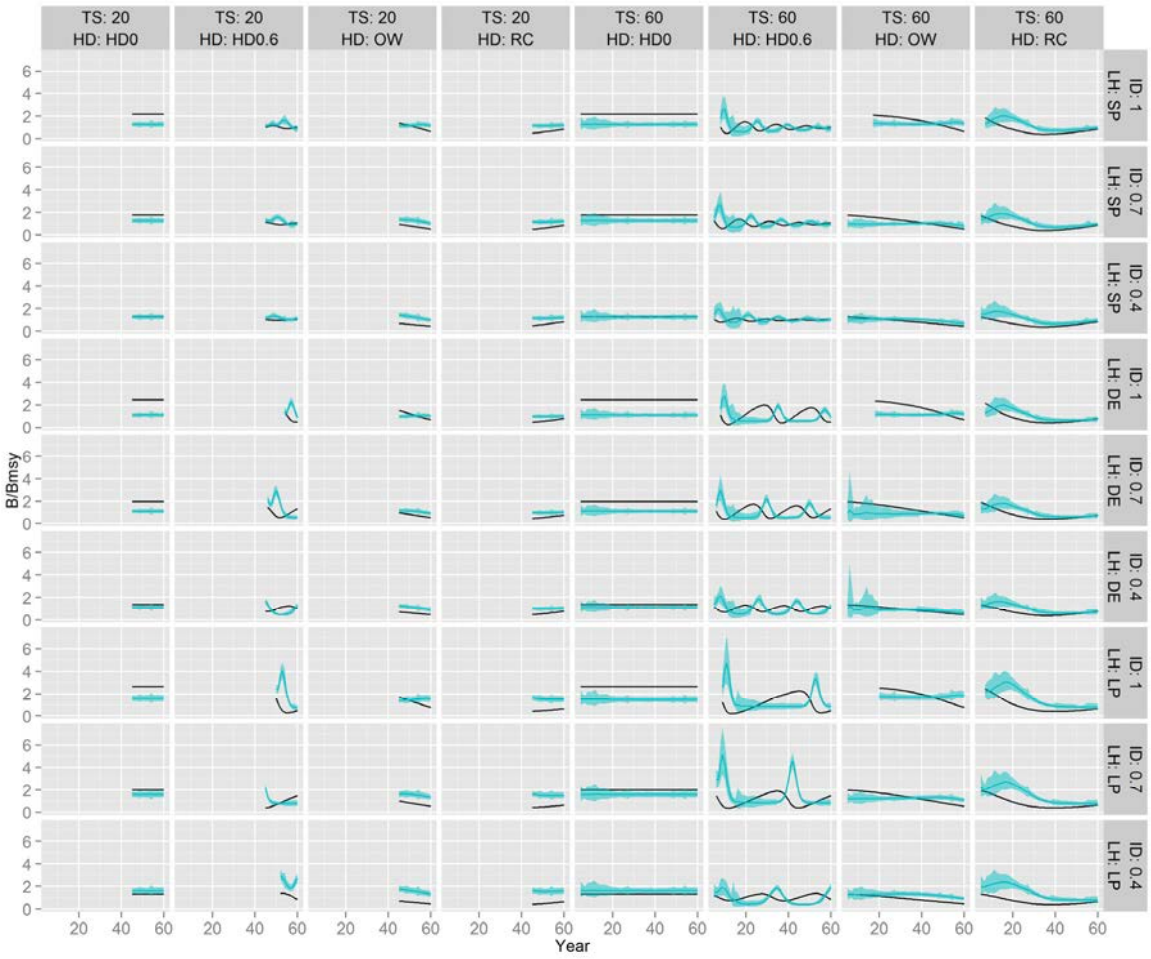


FIGURE A1.5

Mean Proportional Error (MPE) estimated according to equation 21 for the 4 models (CMSY, COM-SIR, mPRM and SSCOM) and each factor in the full factorial design. Harvest dynamics levels (HD) and length of time series (TS) structure the columns, initial depletion (ID) and life history (LH) structure the rows. The value of MPE provides the average bias of the estimate (B/B_{MSY}) and for example a value of 0.25 indicates that the model is overestimating B/B_{MSY} by 25%.

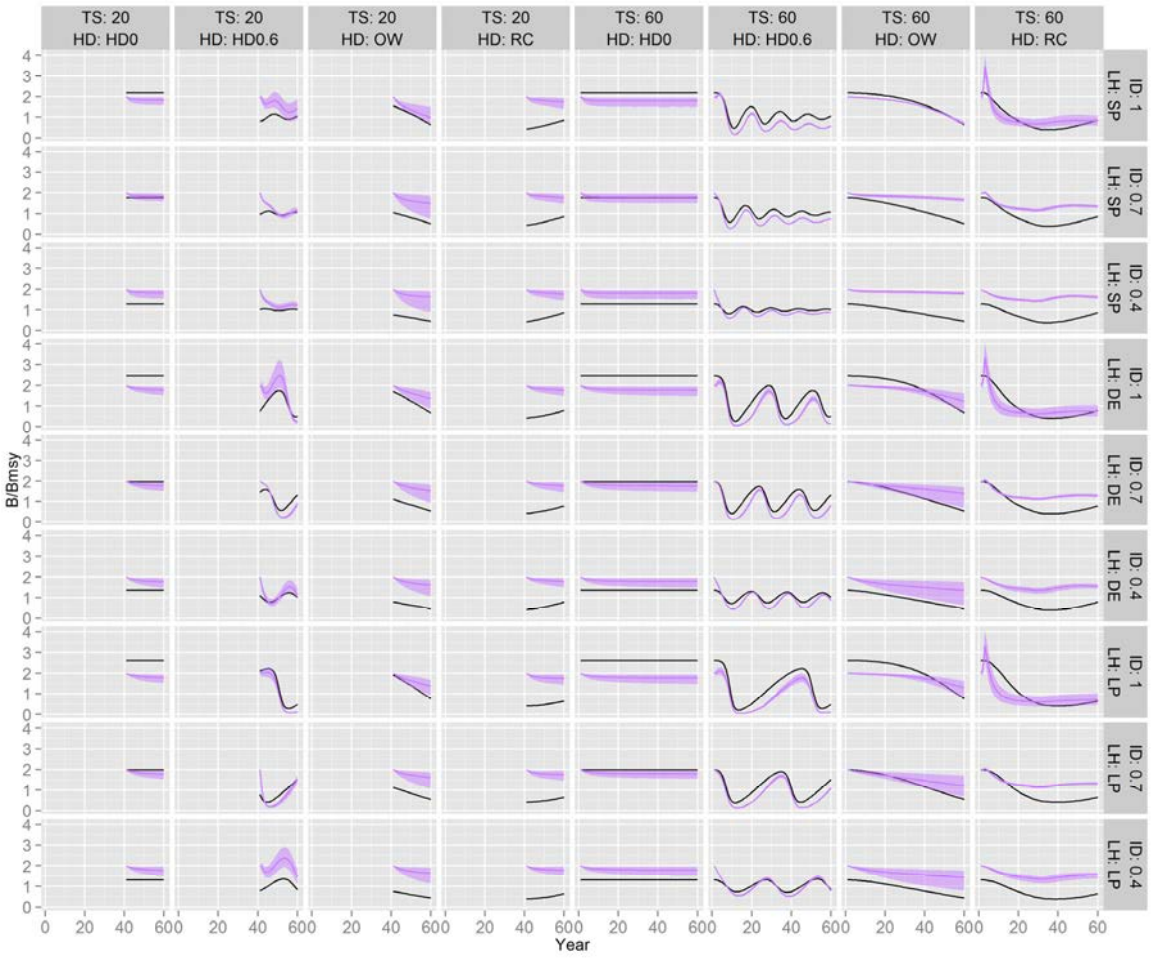


FIGURE A1.6

Iteration 1 of the stochastic simulation trajectories of rescaled stock biomass B/B_{MSY} (black line), and estimated B/B_{MSY} for each model (CMSY, COM-SIR, mPRM and SSCOM) for each combination of the design given in Table 8. Harvest dynamics levels (HD) and length of time series (TS, only 60 years), level of recruitment variability (sigmaR) and measurement error in catch (sigmaC) structure the columns. Initial depletion (ID), life history (LH) and autocorrelation on recruitment residuals (AR) structure the rows. The figure displays only the first iteration of each stochastic run out of the 10 available for readability.

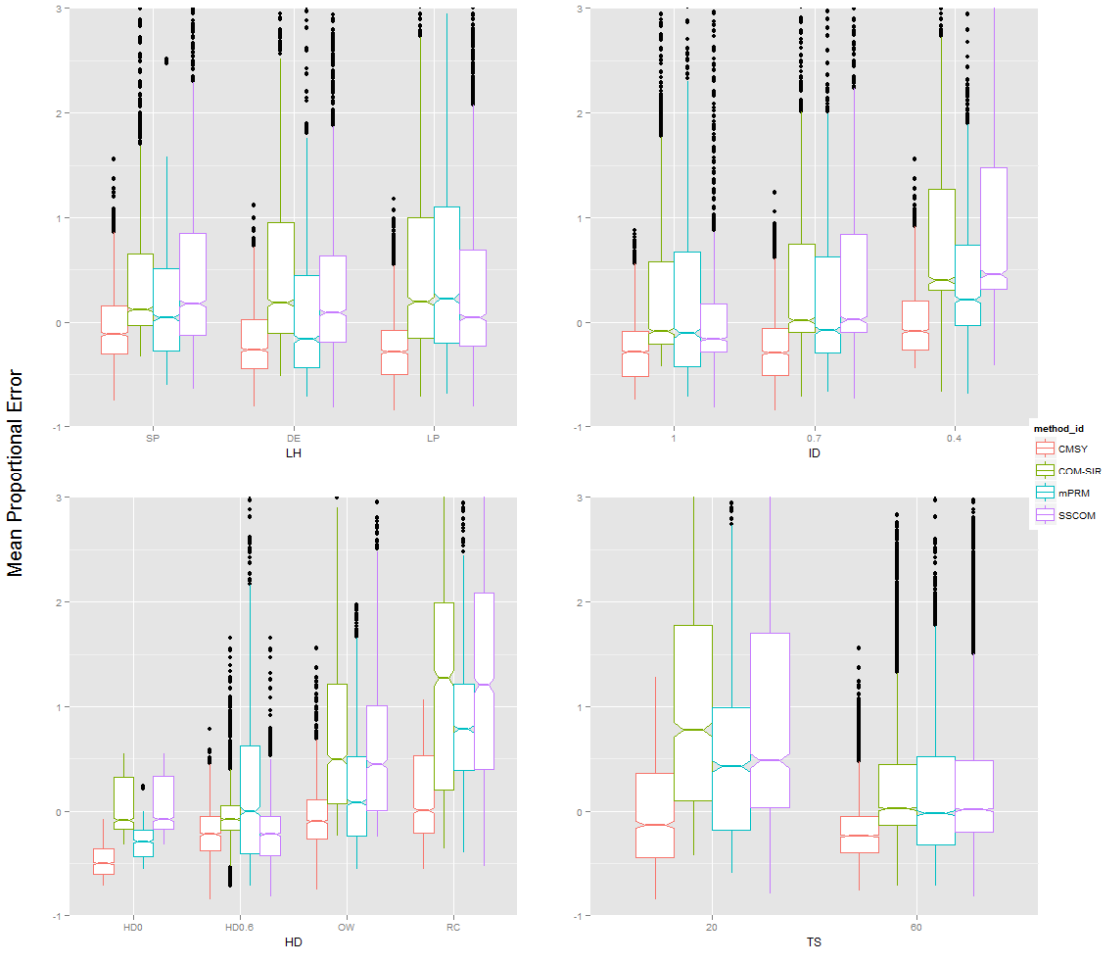


FIGURE A1.7

Iterations 1-10 of the stochastic simulation trajectories of rescaled stock biomass B/B_{MSY} (black line) and estimated B/B_{MSY} by each model (CMSY, COM-SIR, mPRM and SSCOM) for each combination of the design given in Table 8. Harvest dynamics levels (HD) and length of time series (TS, only 60 years), level of recruitment variability (sigmaR) and measurement error in catch (sigmaC) structure the columns. Initial depletion (ID), life history (LH) and autocorrelation on recruitment residuals (AR) structure the rows.

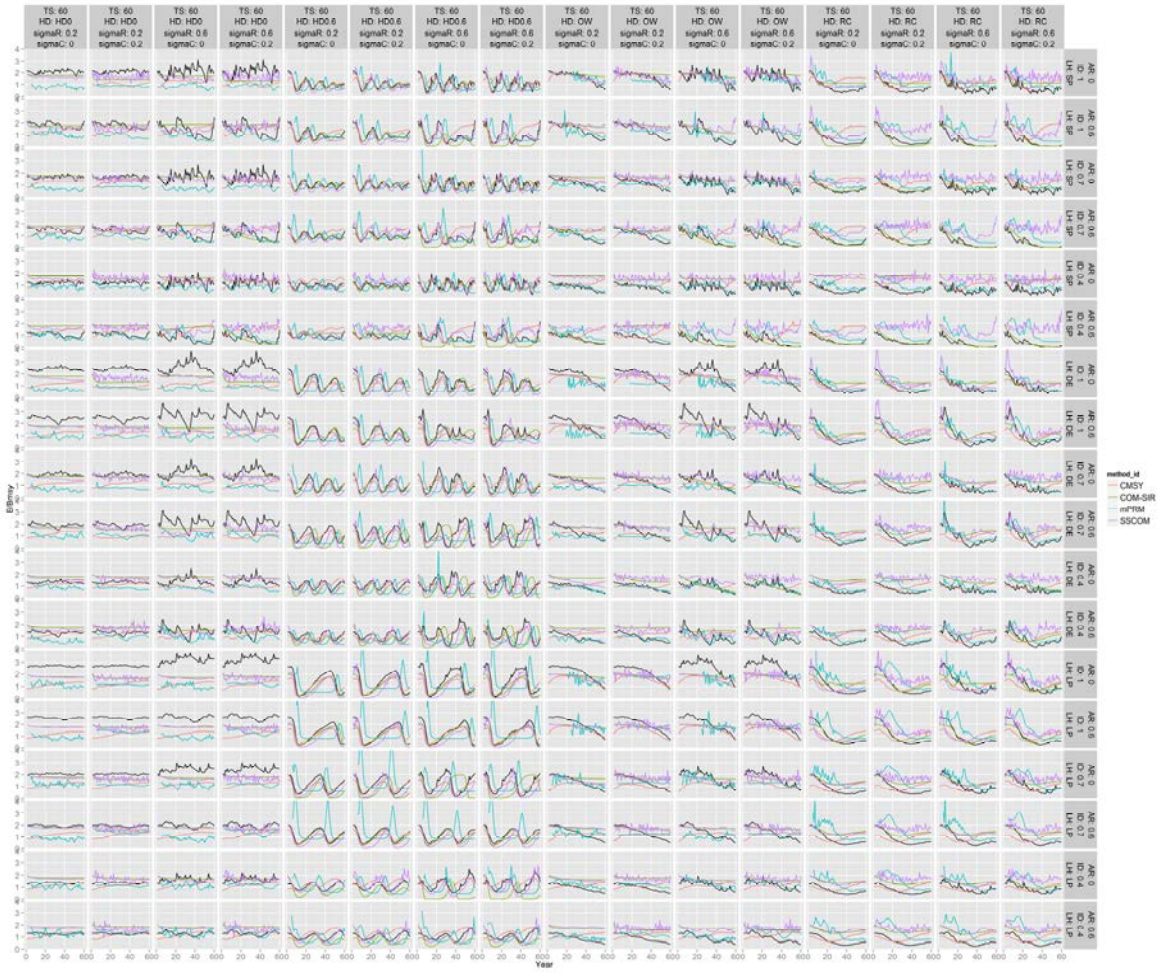


FIGURE A1.8

Mean Proportional Error between true and estimated B/B_{MSY} in all the 10 iterations of the stochastic runs by recruitment variability (sigmaR), measurement error in catch (sigmaC), autocorrelation on recruitment residuals (AR) and model for all years available.

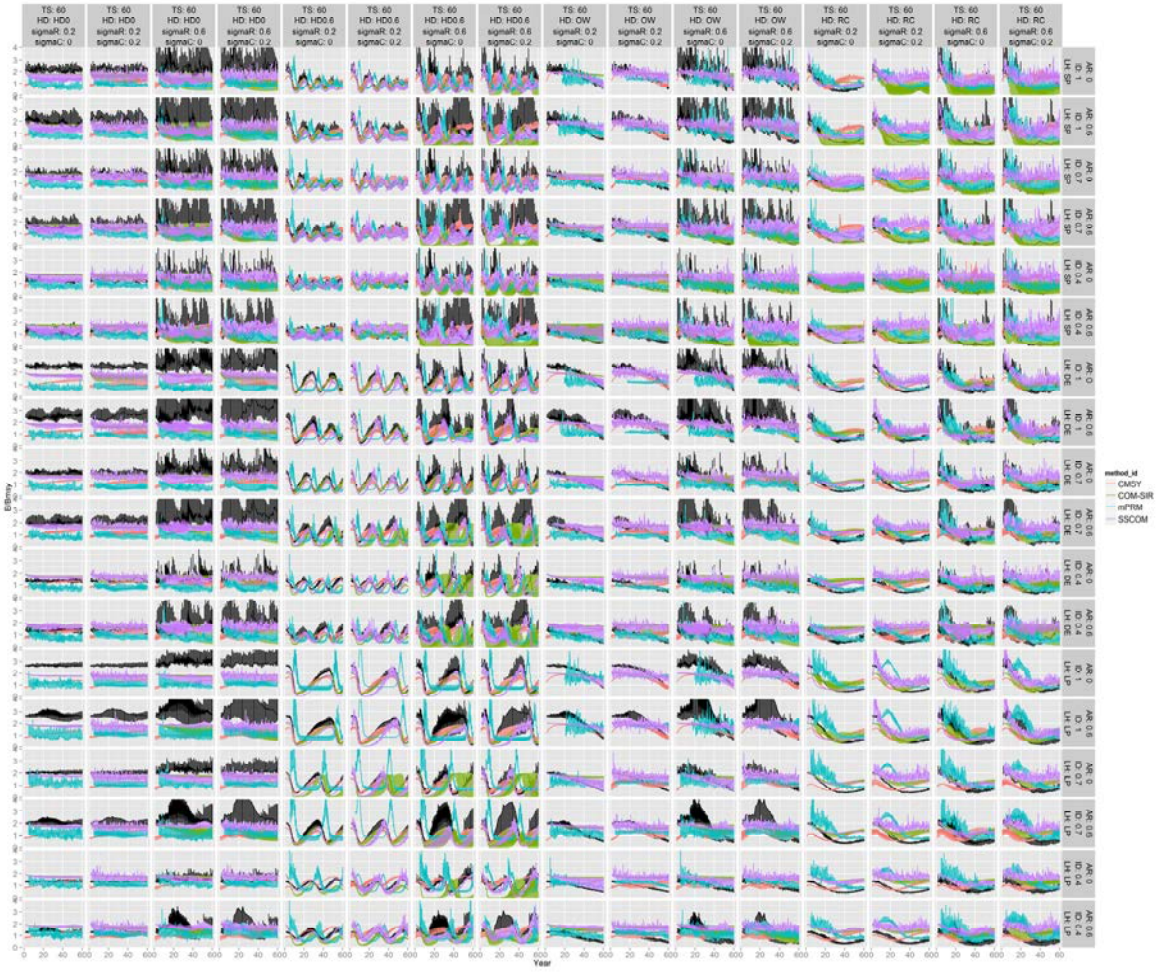


FIGURE A1.9

Mean Proportional Error between true and estimated B/B_{MSY} in all the 10 iterations of the stochastic runs by life history (LH), initial depletion (ID), harvest dynamics (HD), length of the time series (TS) and model for all years available.

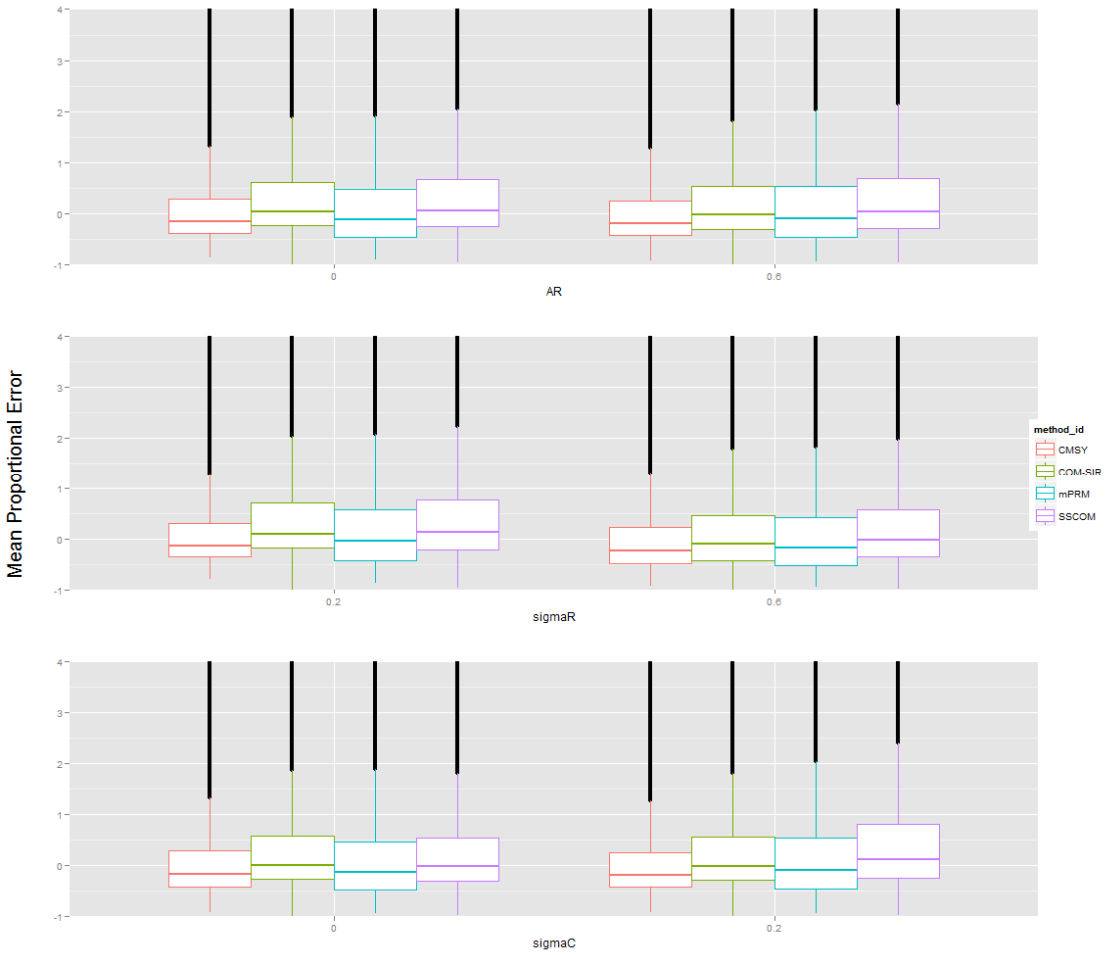


FIGURE A1.10

Regression tree of proportional error (PE) across all methods for all years for main factors and emergent properties variables. The top number in each box is the average PE for a set of simulation scenarios (i.e. the averaged PE across all methods and simulations was 0.29 or 29%). The numbers in the second row of the boxes list the number of data points and percentage of simulation scenarios in that set (i.e. the top box has 9350 scenarios representing 100% of the scenarios), and each box either has no boxes below it (i.e. it is a terminal node), or has two boxes below it (i.e. it has additional branching). The percentages in each box of a single tier sum to 100% (see Tables 7, 8 and 10 for factors, levels and emergent properties variables).

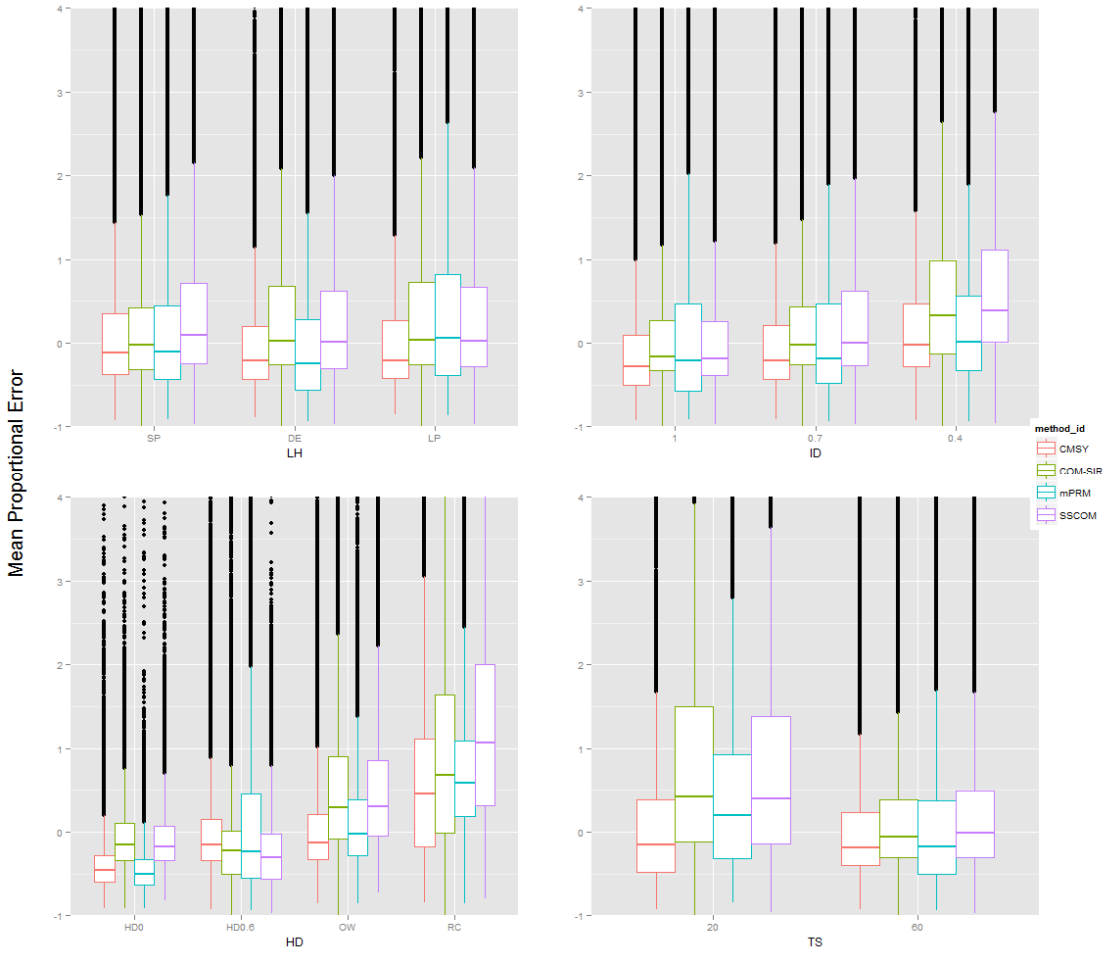


FIGURE A1.11

Regression tree of proportional error (PE) across all methods for the last five years for main factors and emergent proprieties variables (see Figure A1.10 caption).

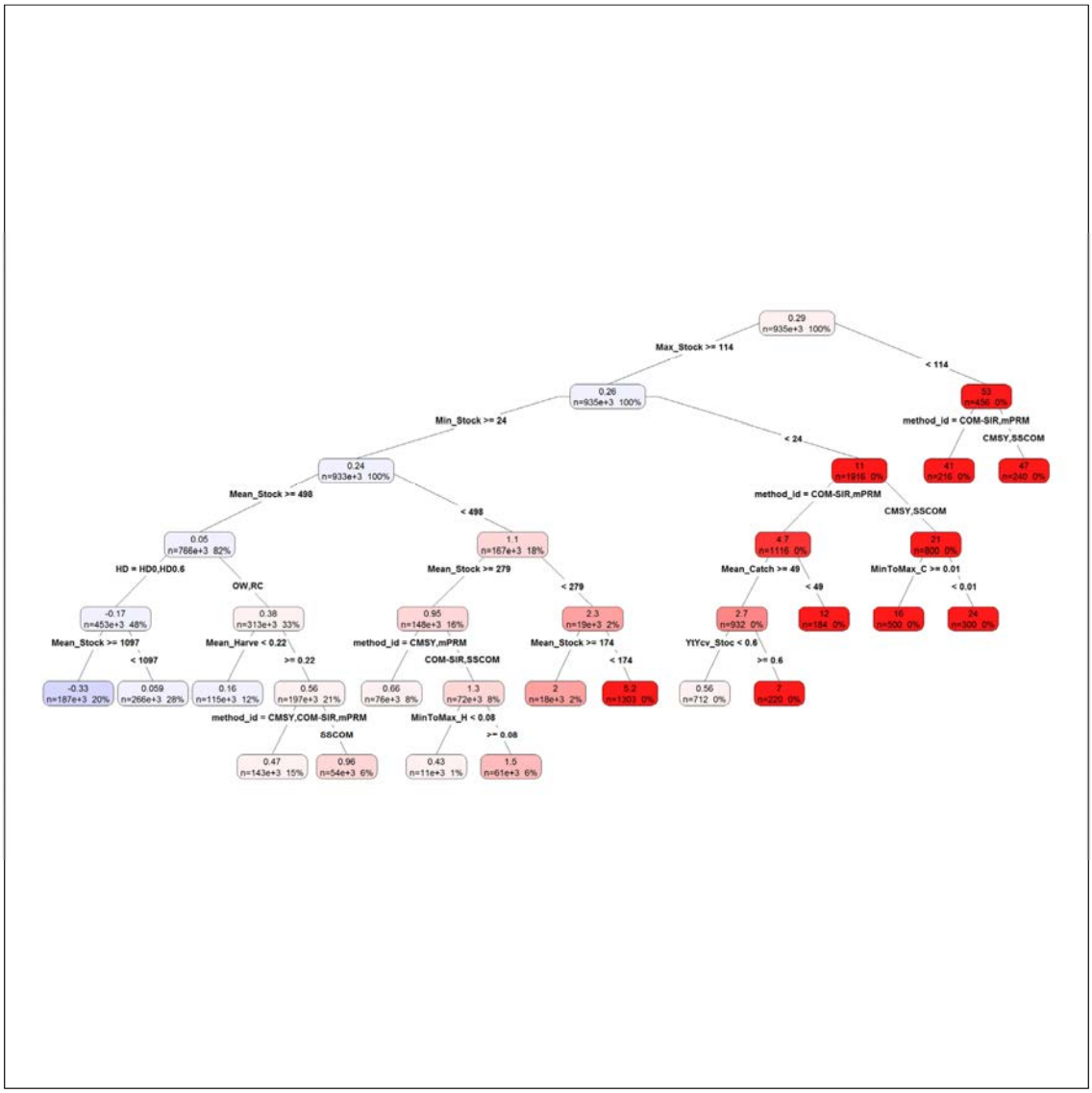


FIGURE A1.12

Regression tree of absolute proportional error (APE) across all methods for all years for main factors and emergent proprieties variables (see Figure A1.10 caption).

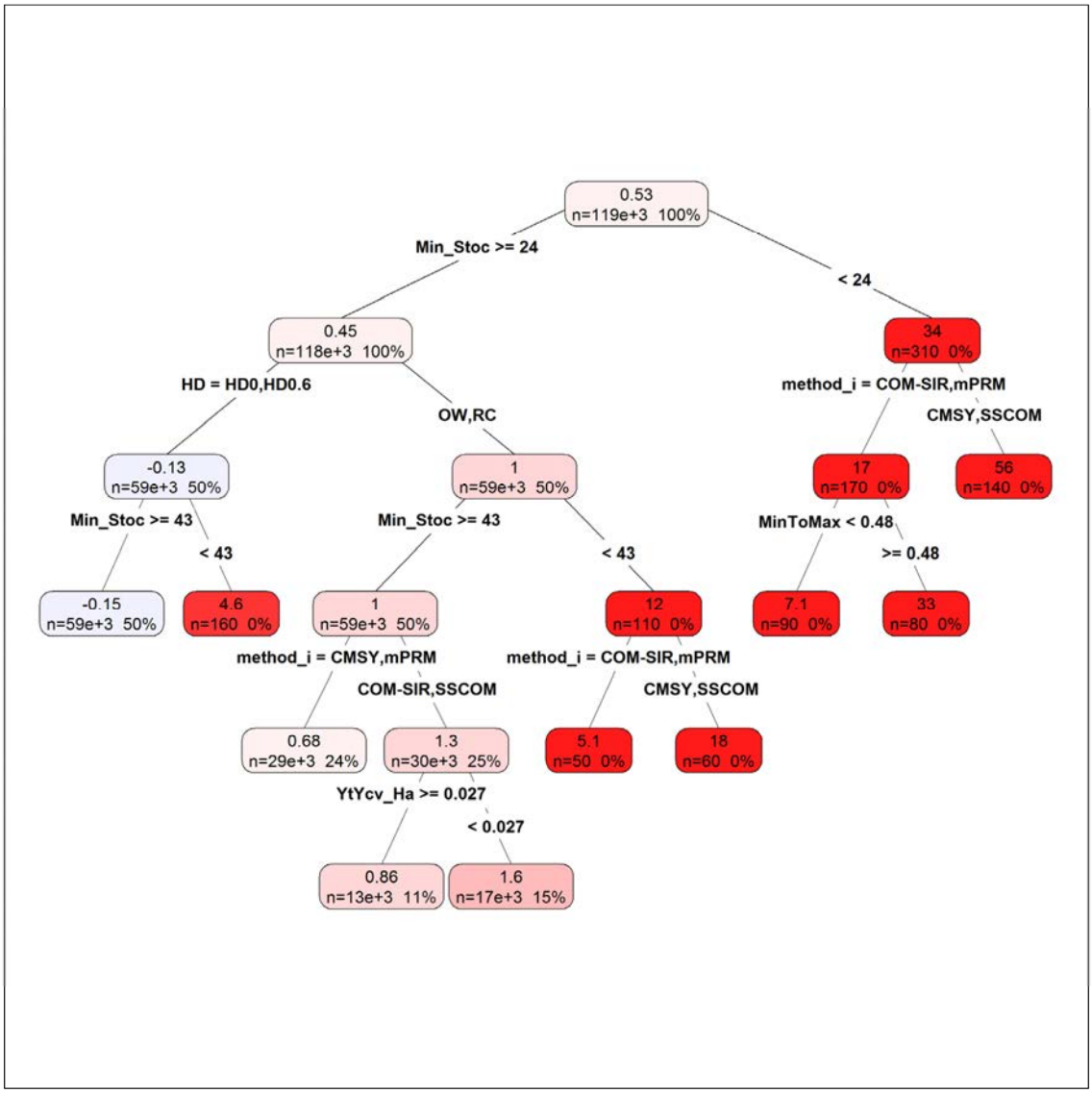
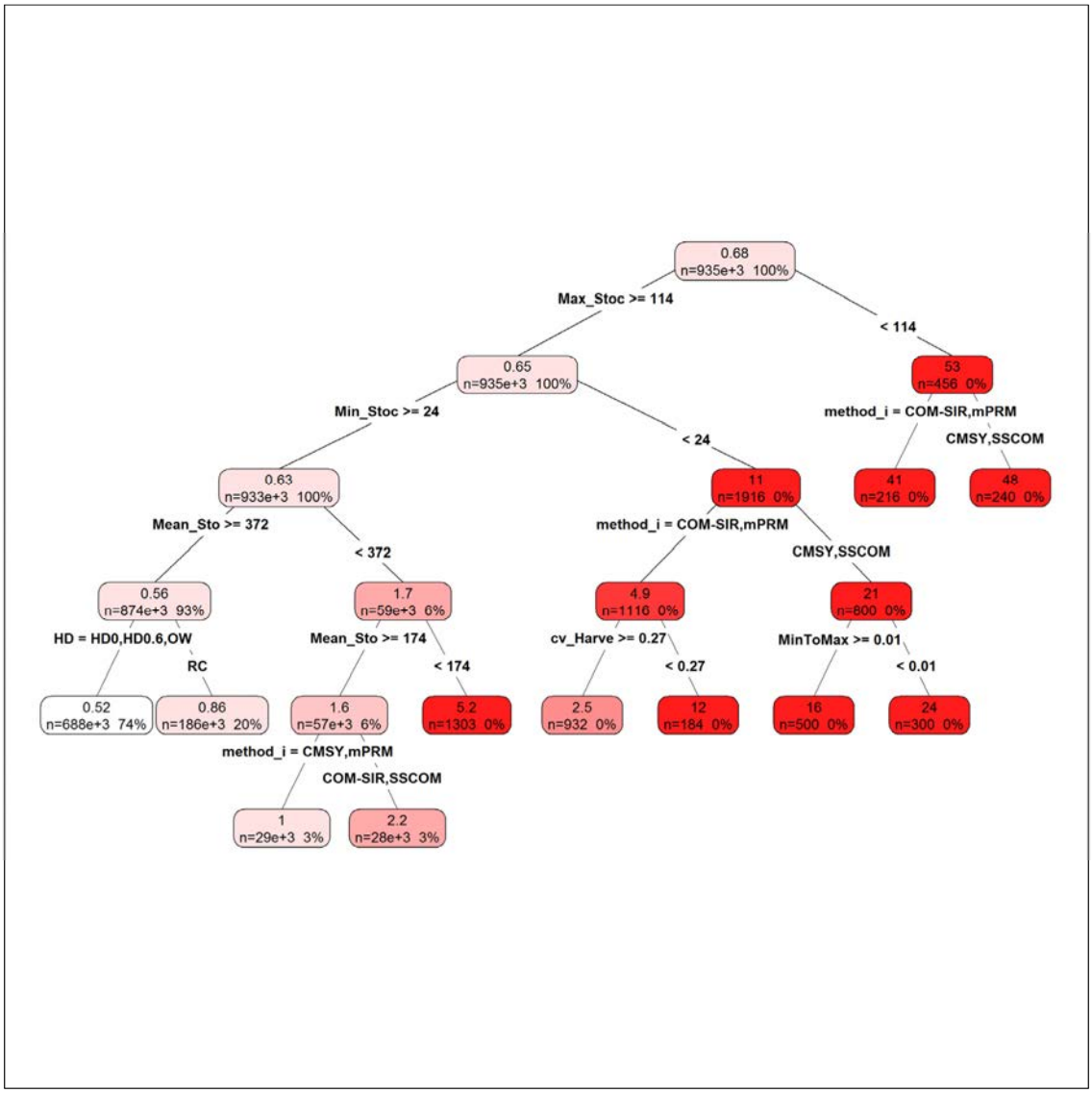


FIGURE A1.13
Regression tree of absolute proportional error (APE) across all methods for the last five years main factors and emergent proprieties variables (see Figure A1.10 caption).



APPENDIX 2
ADDITIONAL FIGURES REFERRED IN PART II

FIGURE A2.1

Changes in the proportion of taxa at each level of taxonomic resolution, low (dark blue, 1; e.g. miscellaneous fishes) to high (light blue, 6; e.g. species genus), in the landings by decade in each Large Marine Ecosystem (LME).

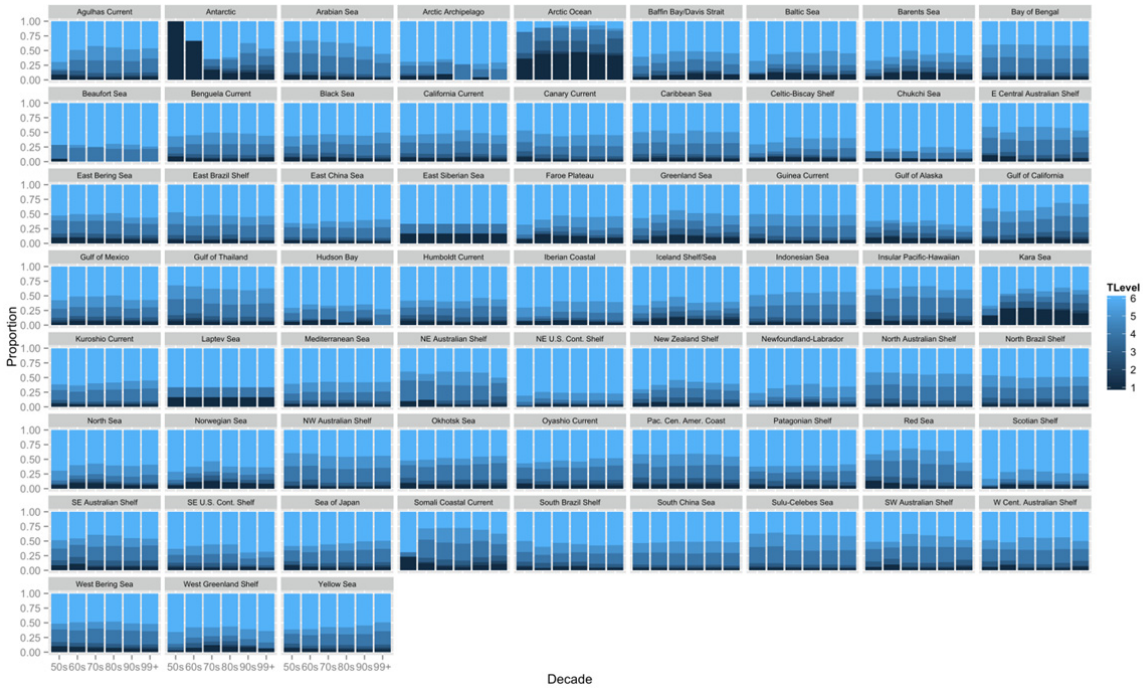


FIGURE A2.2 (next page)
The mean climatological (1998-2007) chlorophyll (CHL - left) and primary production (PP - right) on the top row; the mean microplankton and nano+picoplankton CHL and PP on the second row; and the percent CHL and PP attributed to the microplankton and nano+picoplankton size classes on the third for each LME and FAO region. The black line on each plot represents the LME boundary and the white line is the 300 m isobath. The composites also include climatological monthly and annual bar plots showing the seasonal and interannual variability of the size fractionated CHL and PP for each depth strata. Note, no depth strata data were calculated for the FAO subareas.

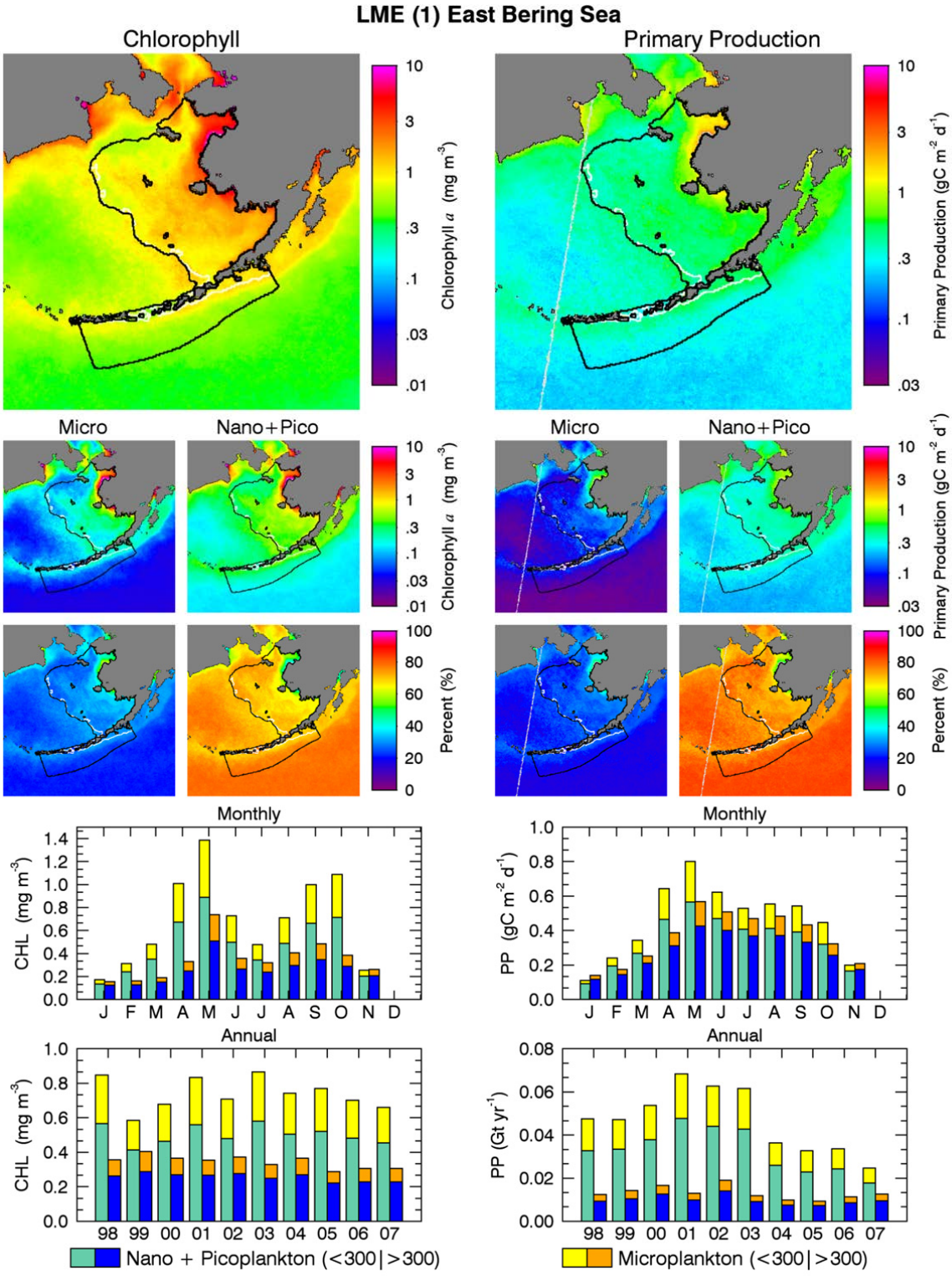

LME (1) East Bering Sea
Chlorophyll
Primary Production
Chlorophyll a (mg m-3)
Primary Production (gC m-2 d-1)
Micro
Nano+Pico
Percent (%)
Monthly
Annual
CHL (mg m-3)
PP (gC m-2 d-1)
PP (Gt yr-1)
J F M A M J J A S O N D
98 99 00 01 02 03 04 05 06 07
Nano + Picoplankton (<300|>300)
Microplankton (<300|>300)

FIGURE A2.3
See legend of Figure A2.2

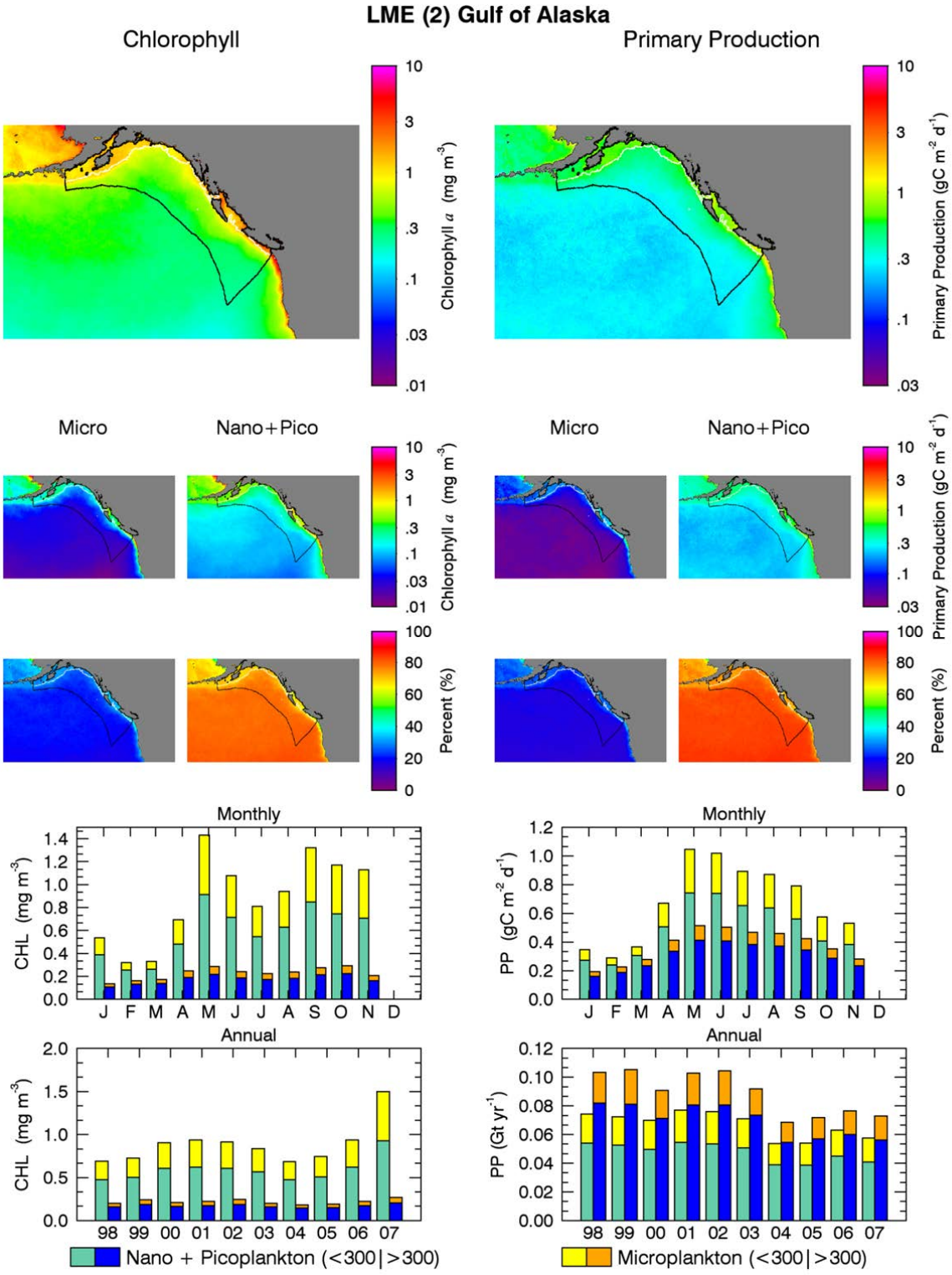


FIGURE A2.4
See legend of Figure A2.2

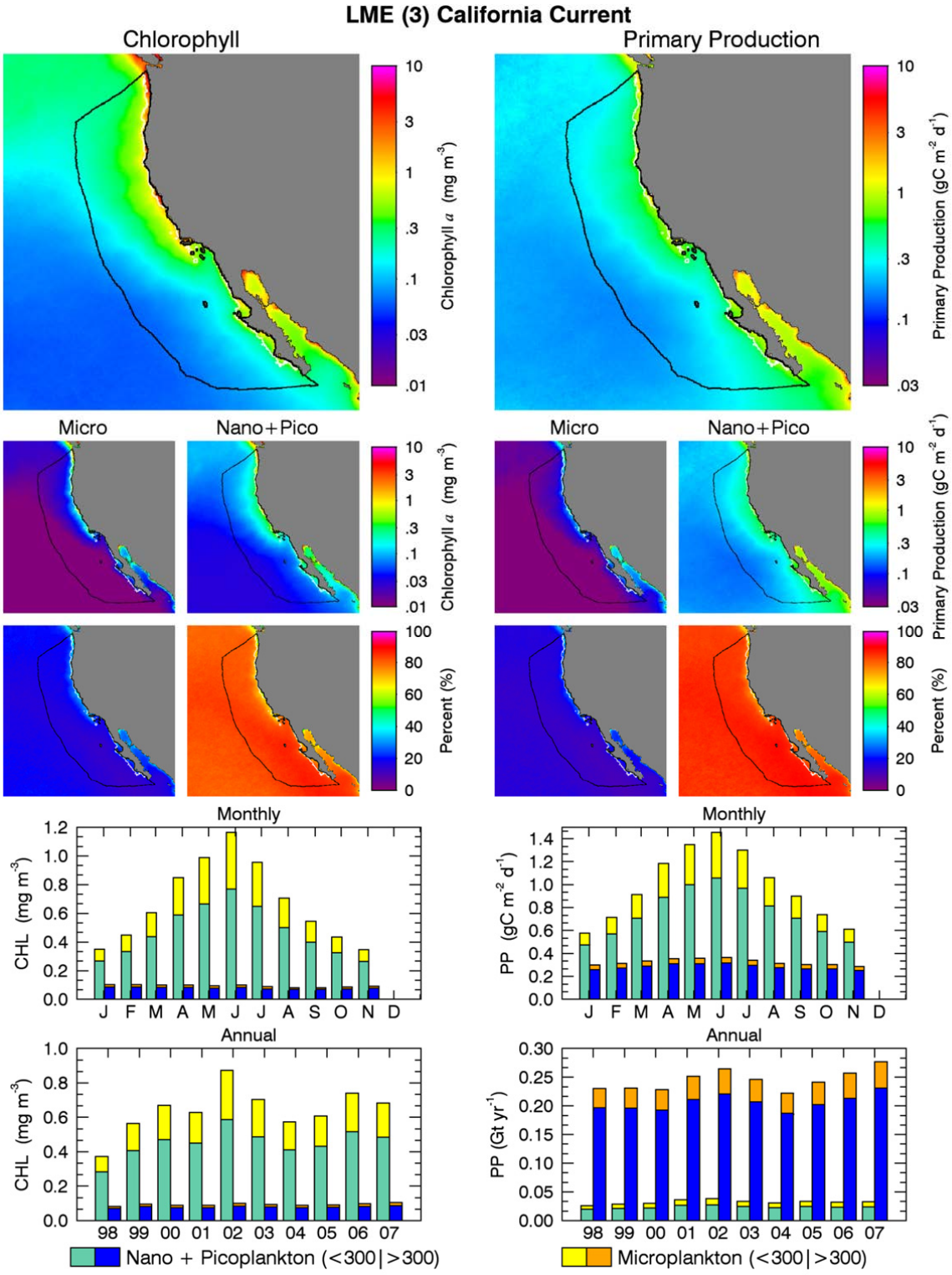


FIGURE A2.5
See legend of Figure A2.2

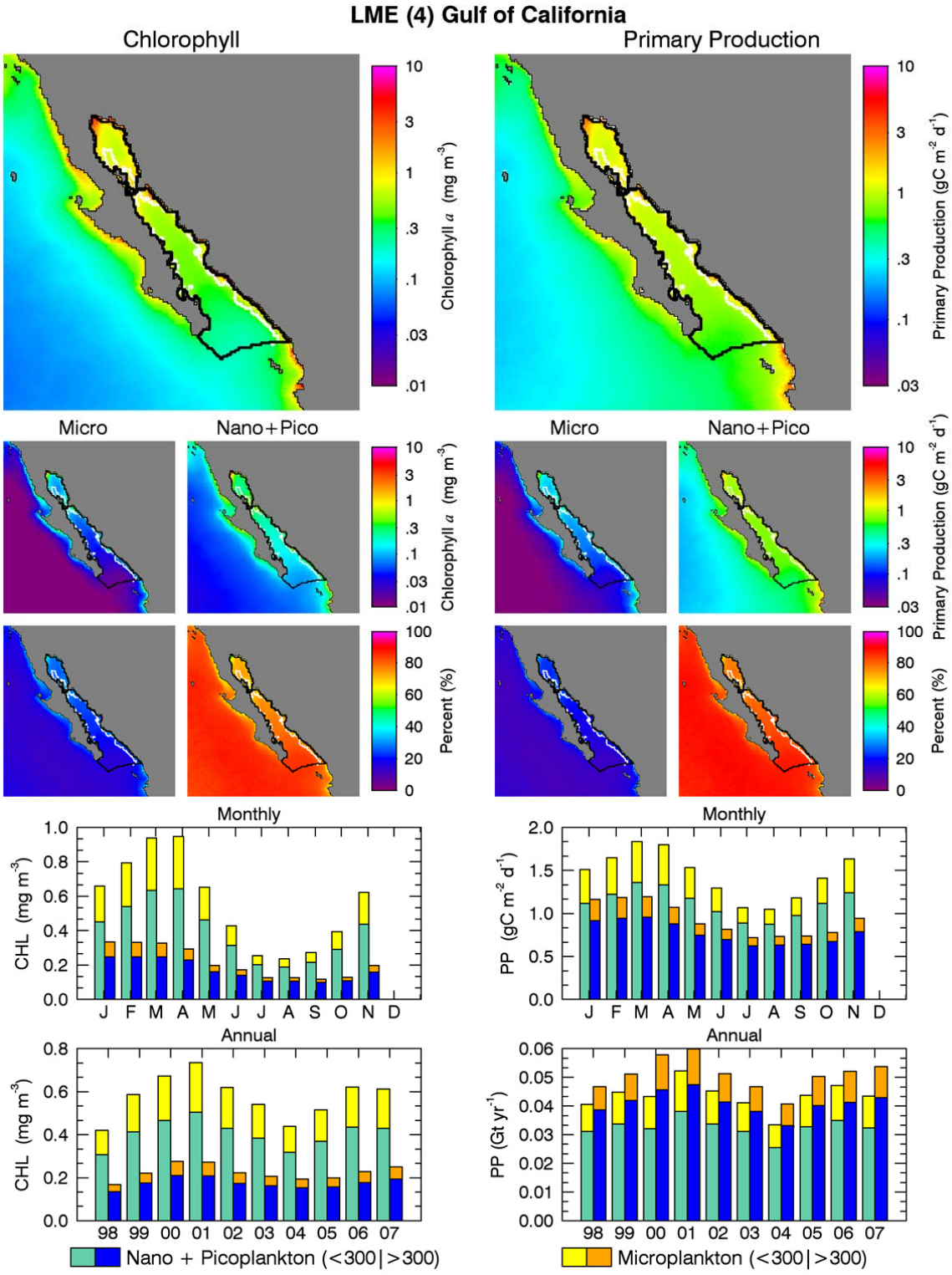


FIGURE A2.6
See legend of Figure A2.2

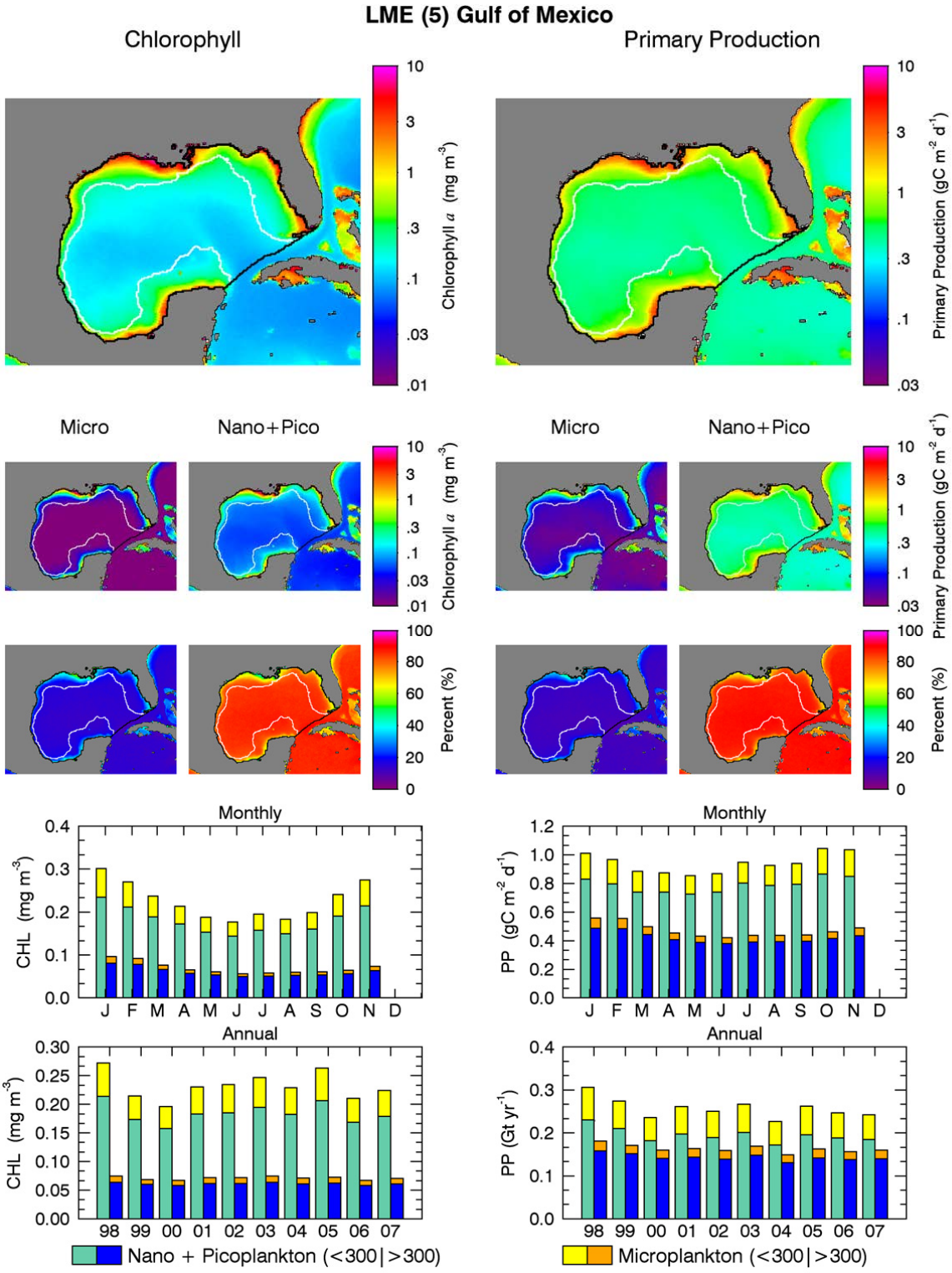


FIGURE A2.7
See legend of Figure A2.2

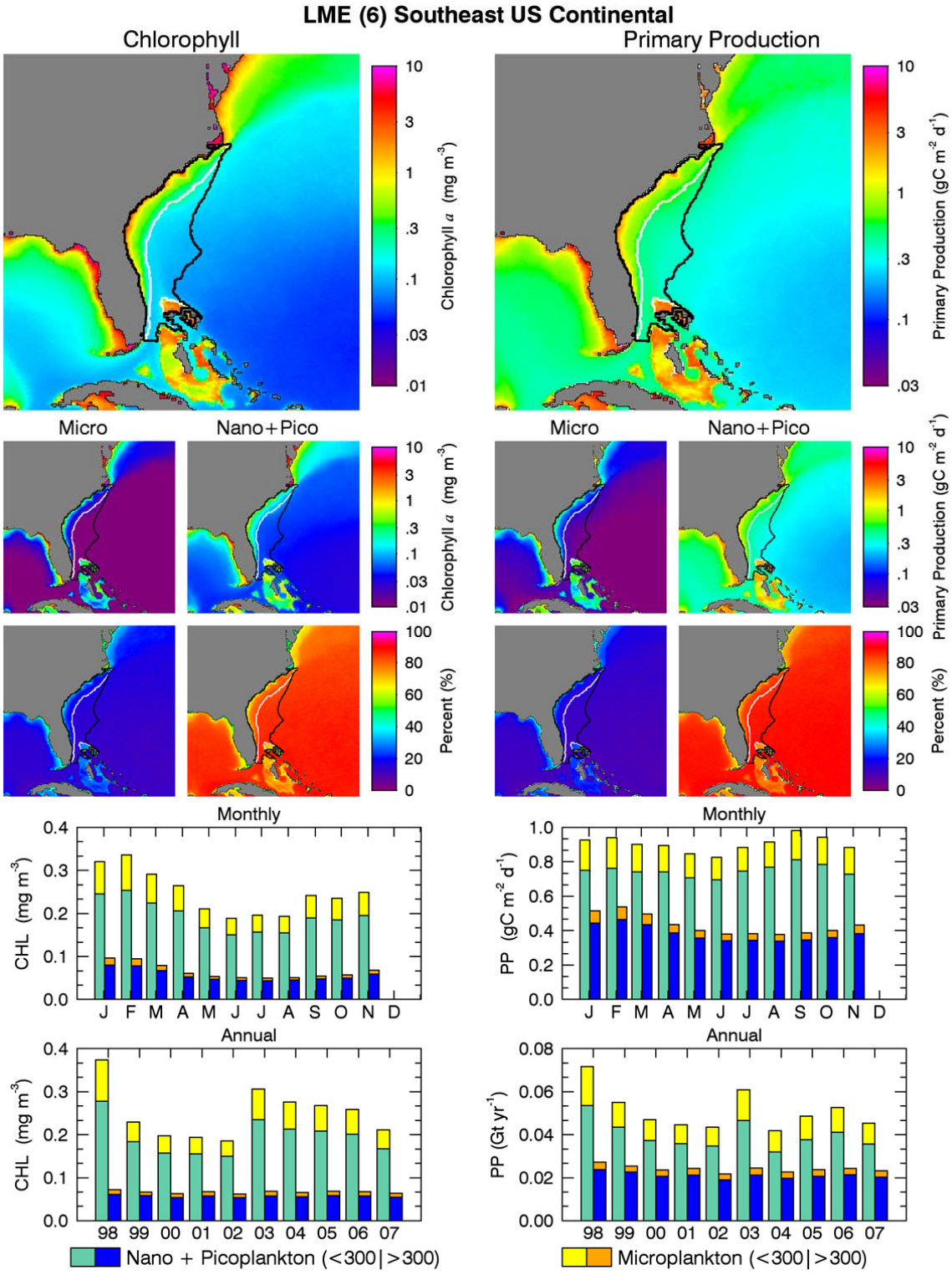


FIGURE A2.8
See legend of Figure A2.2

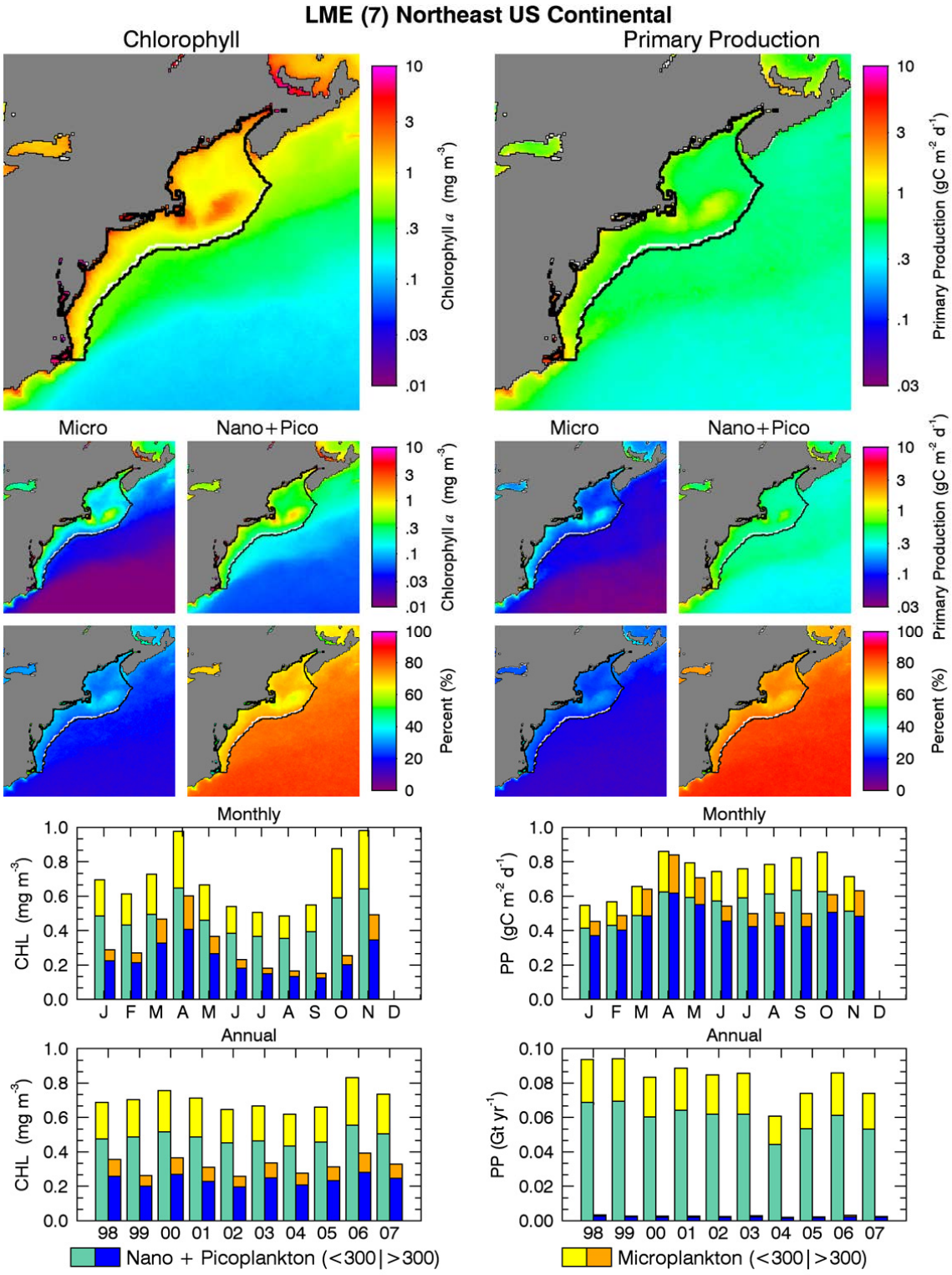


FIGURE A2.9
See legend of Figure A2.2

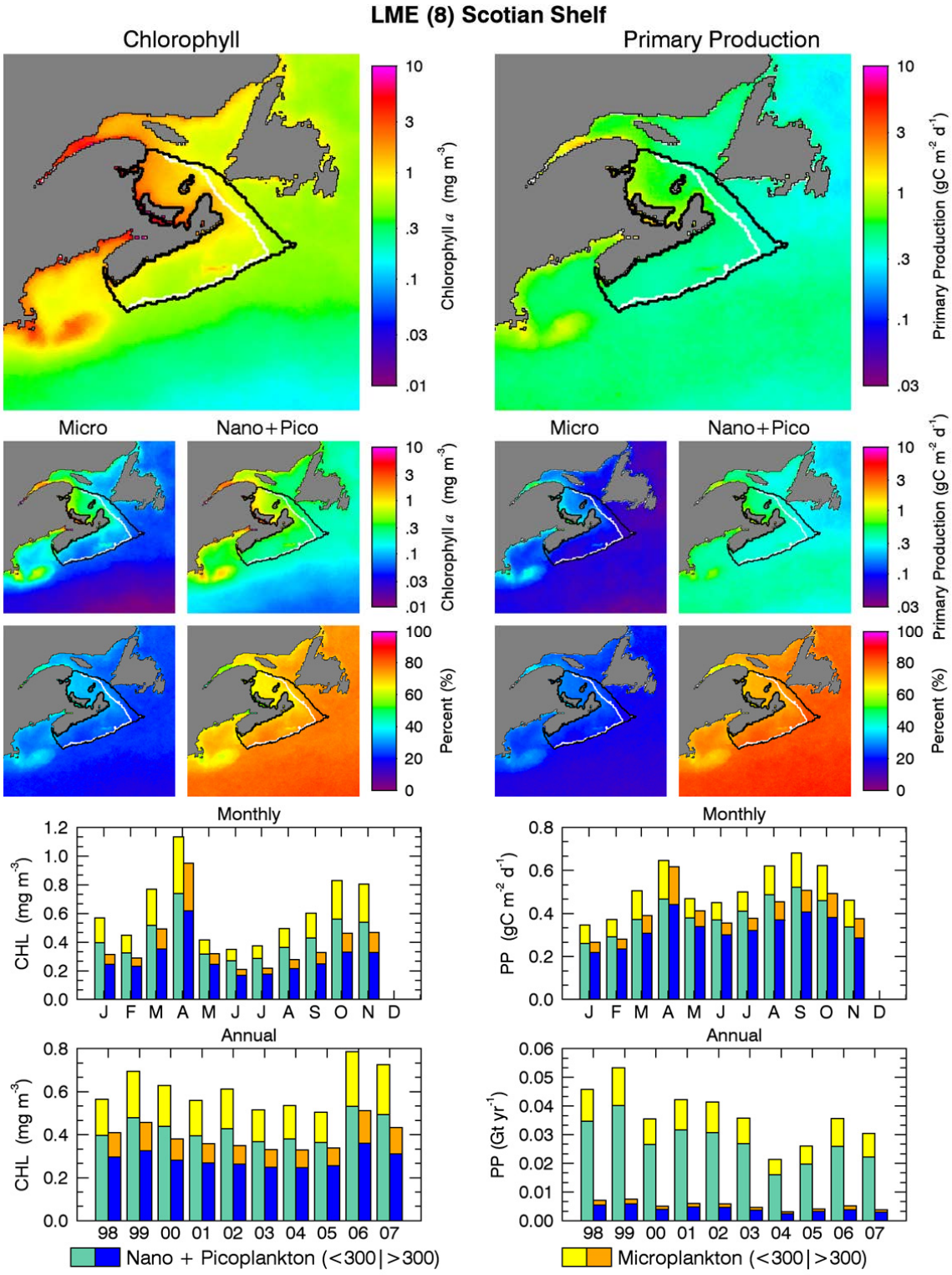


FIGURE A2.10
See legend of Figure A2.2

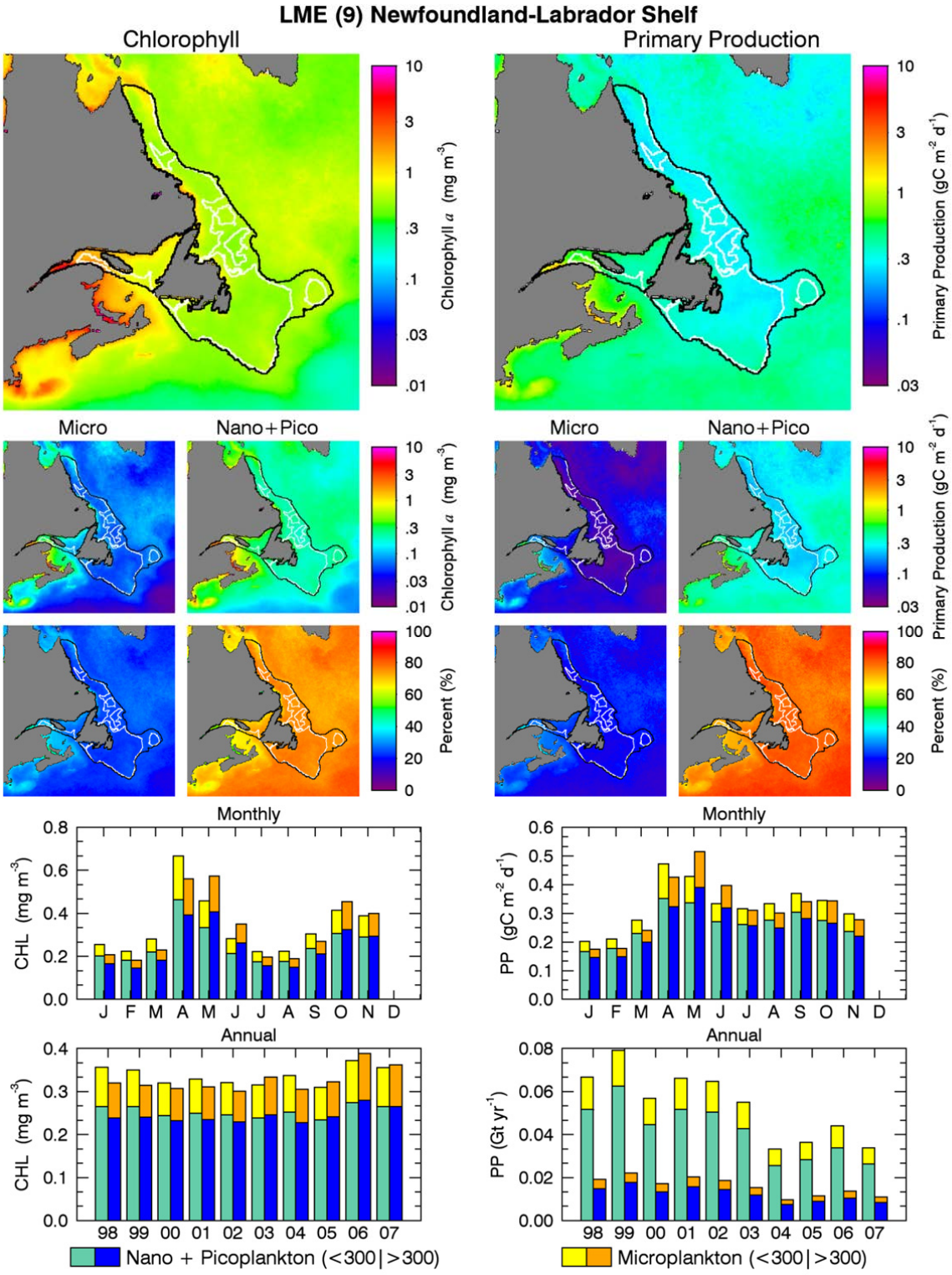


FIGURE A2.11
See legend of Figure A2.2

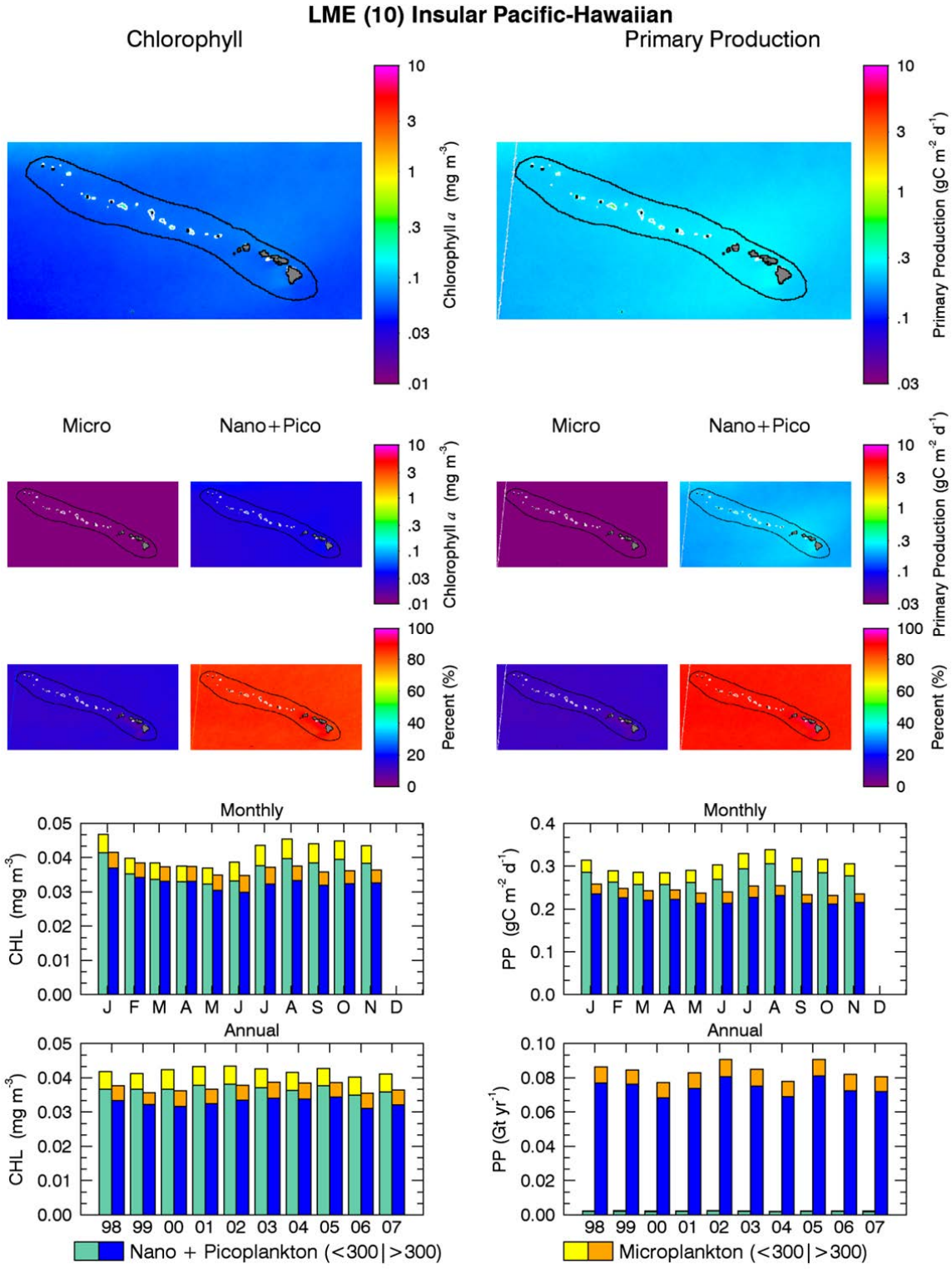


FIGURE A2.12
See legend of Figure A2.2

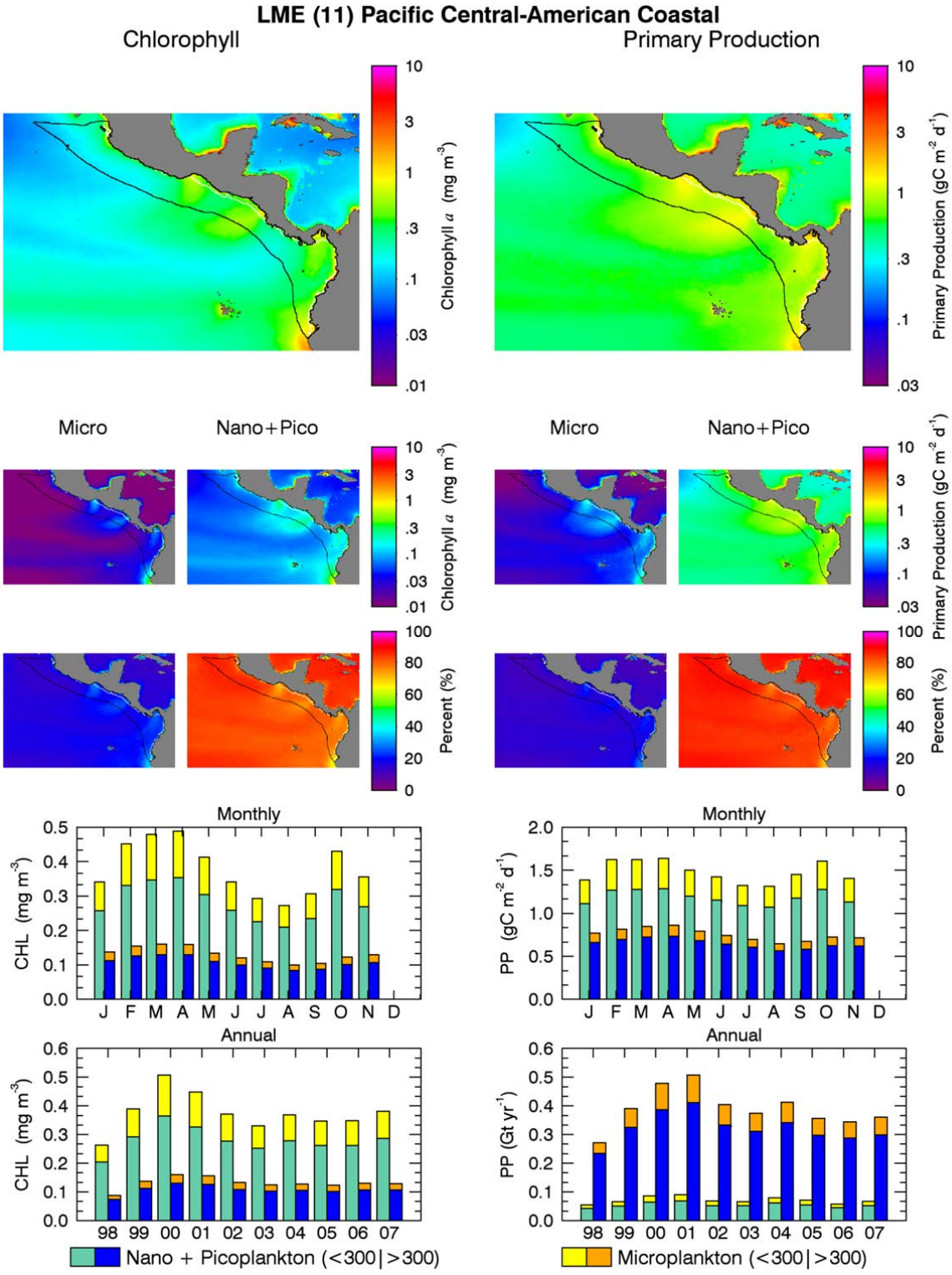


FIGURE A2.13
See legend of Figure A2.2

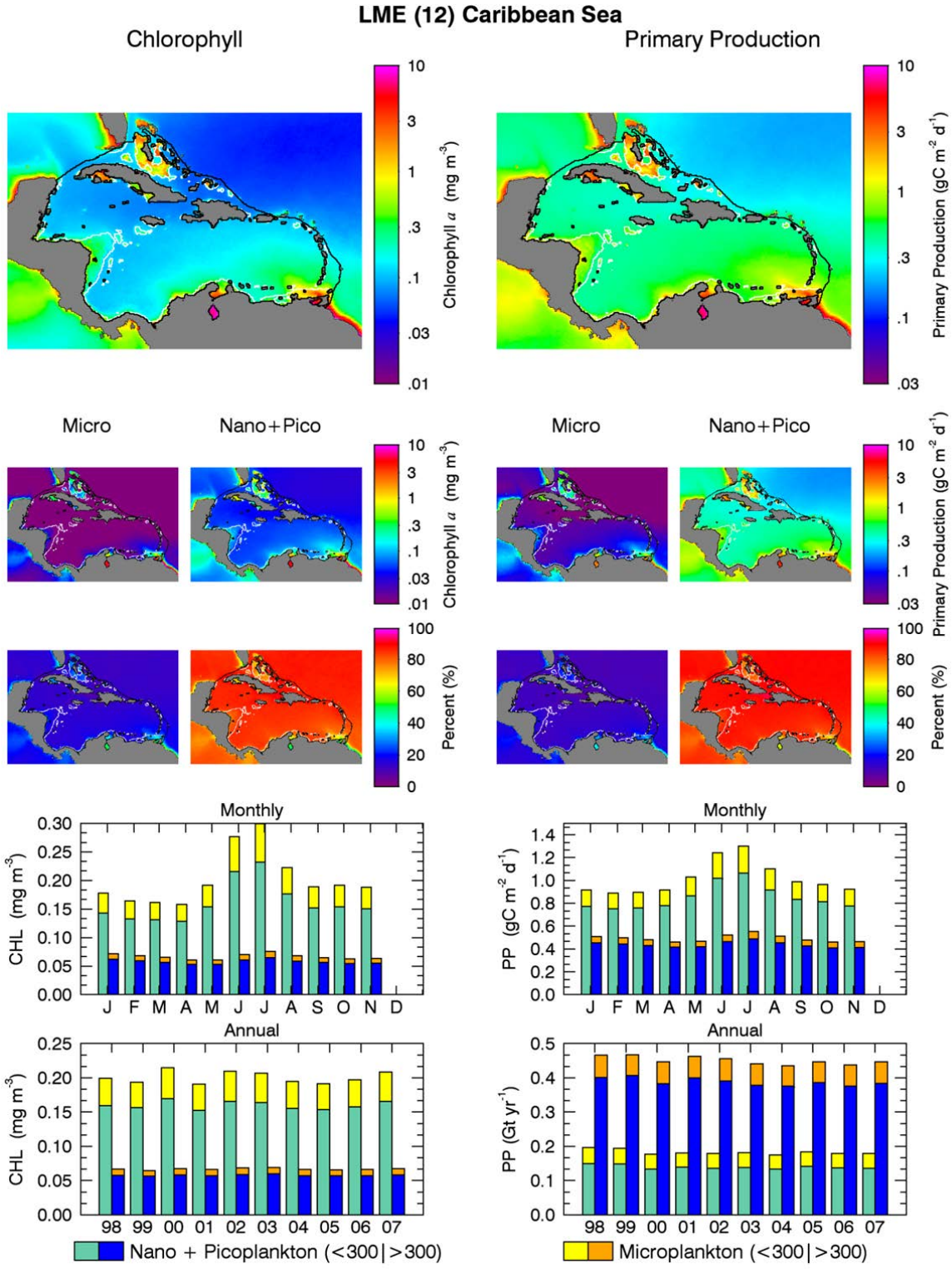


FIGURE A2.14
See legend of Figure A2.2

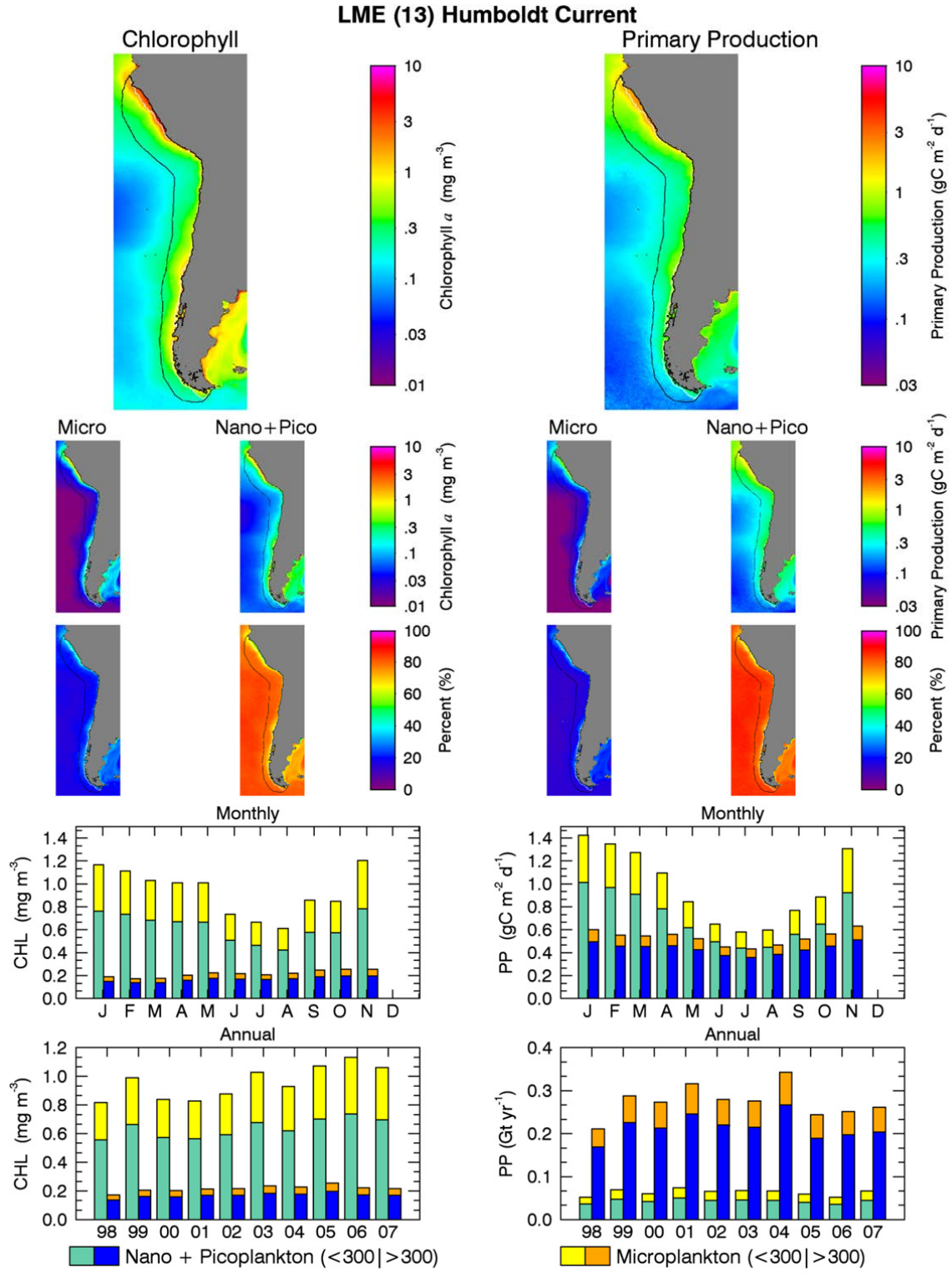


FIGURE A2.15
See legend of Figure A2.2

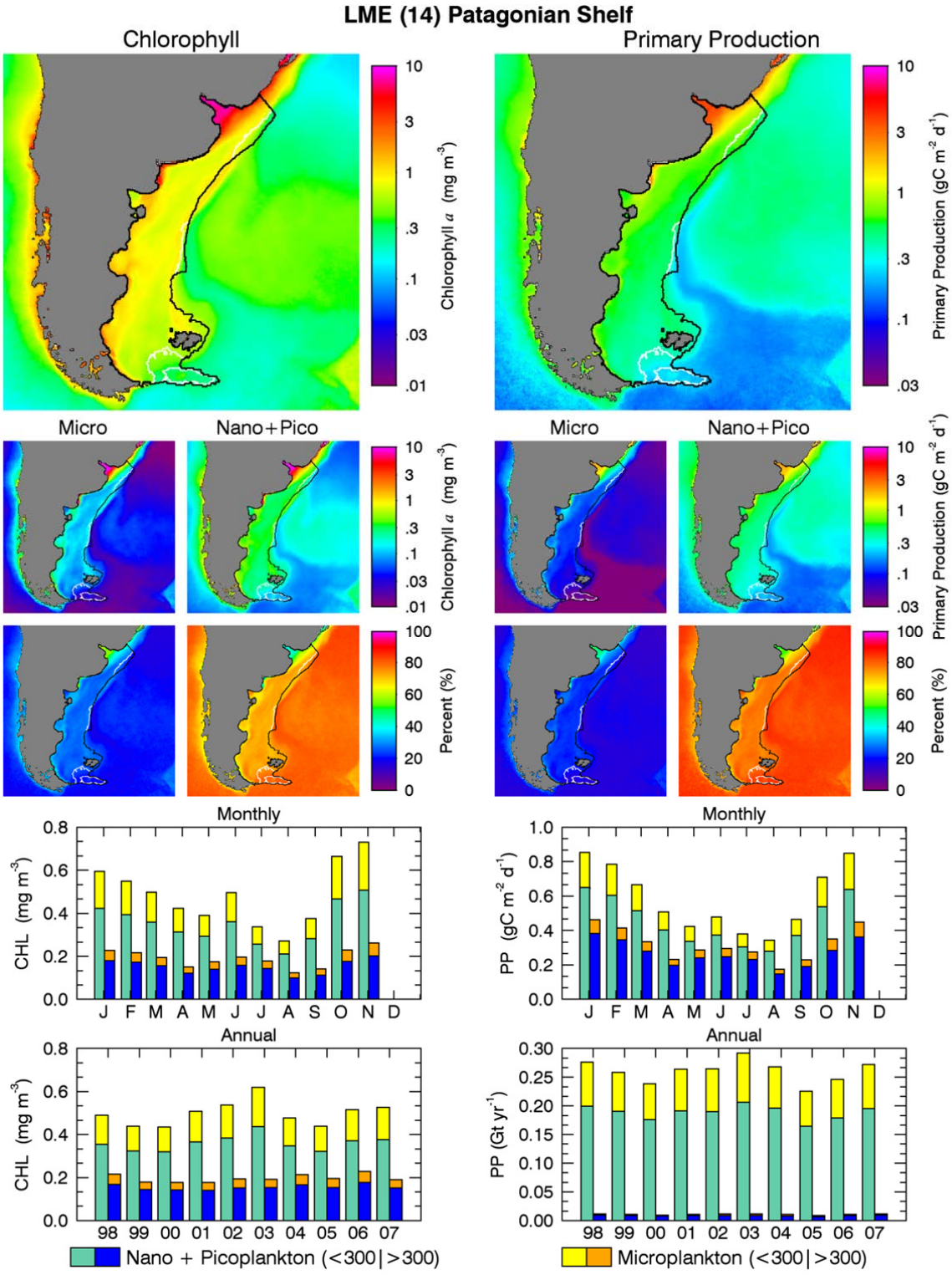


FIGURE A2.16
See legend of Figure A2.2

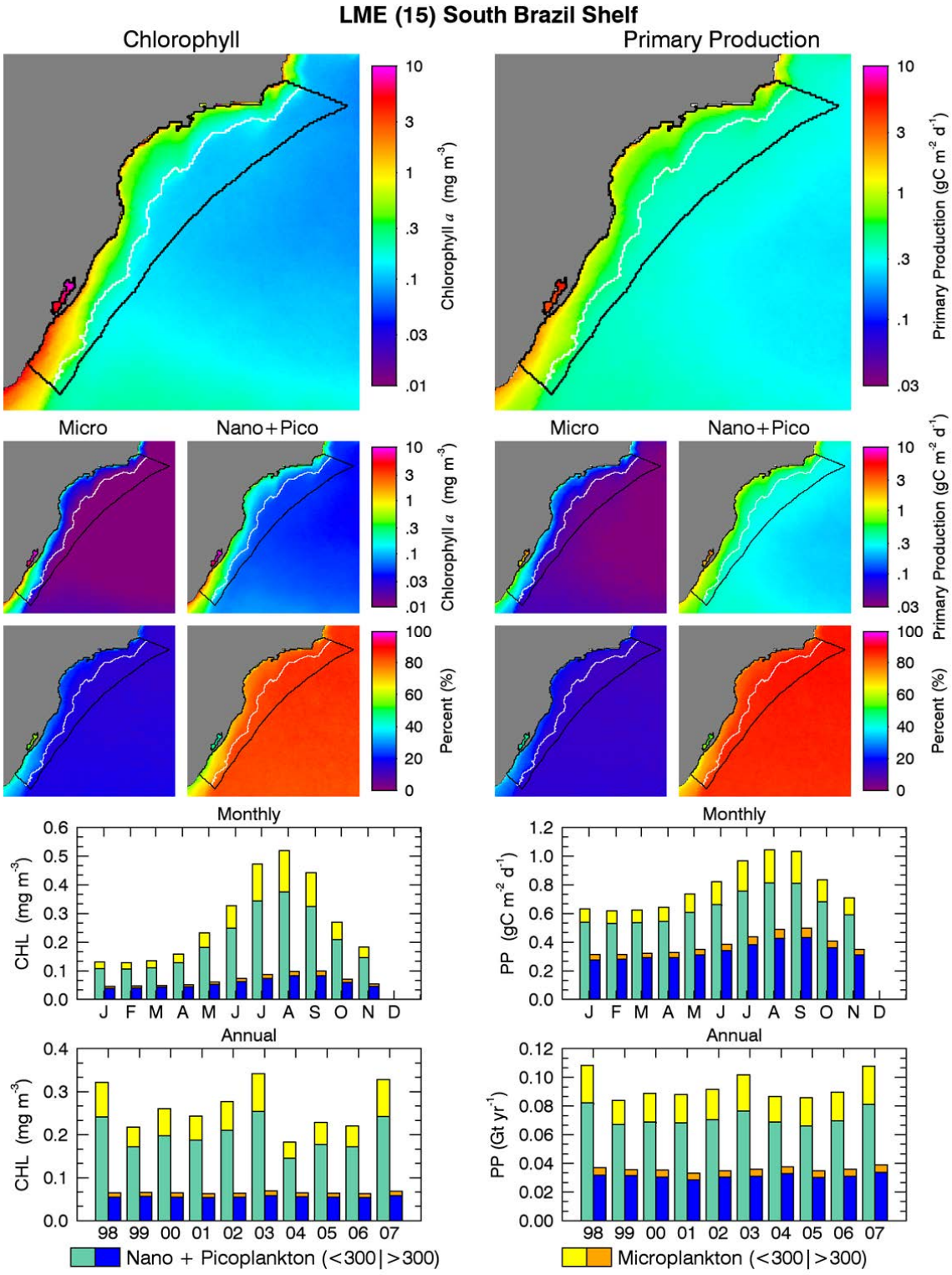


FIGURE A2.17
See legend of Figure A2.2

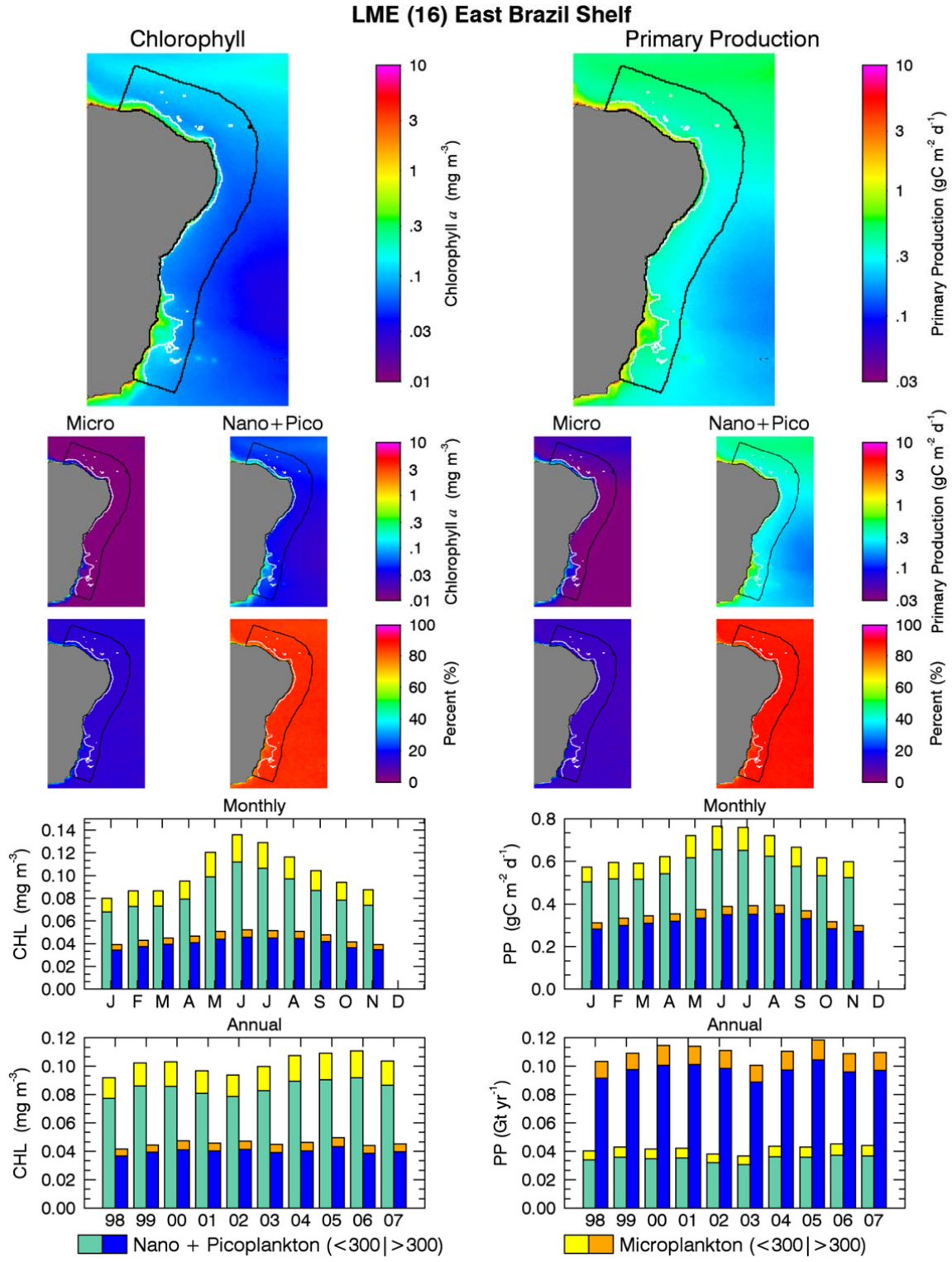


FIGURE A2.18
See legend of Figure A2.2

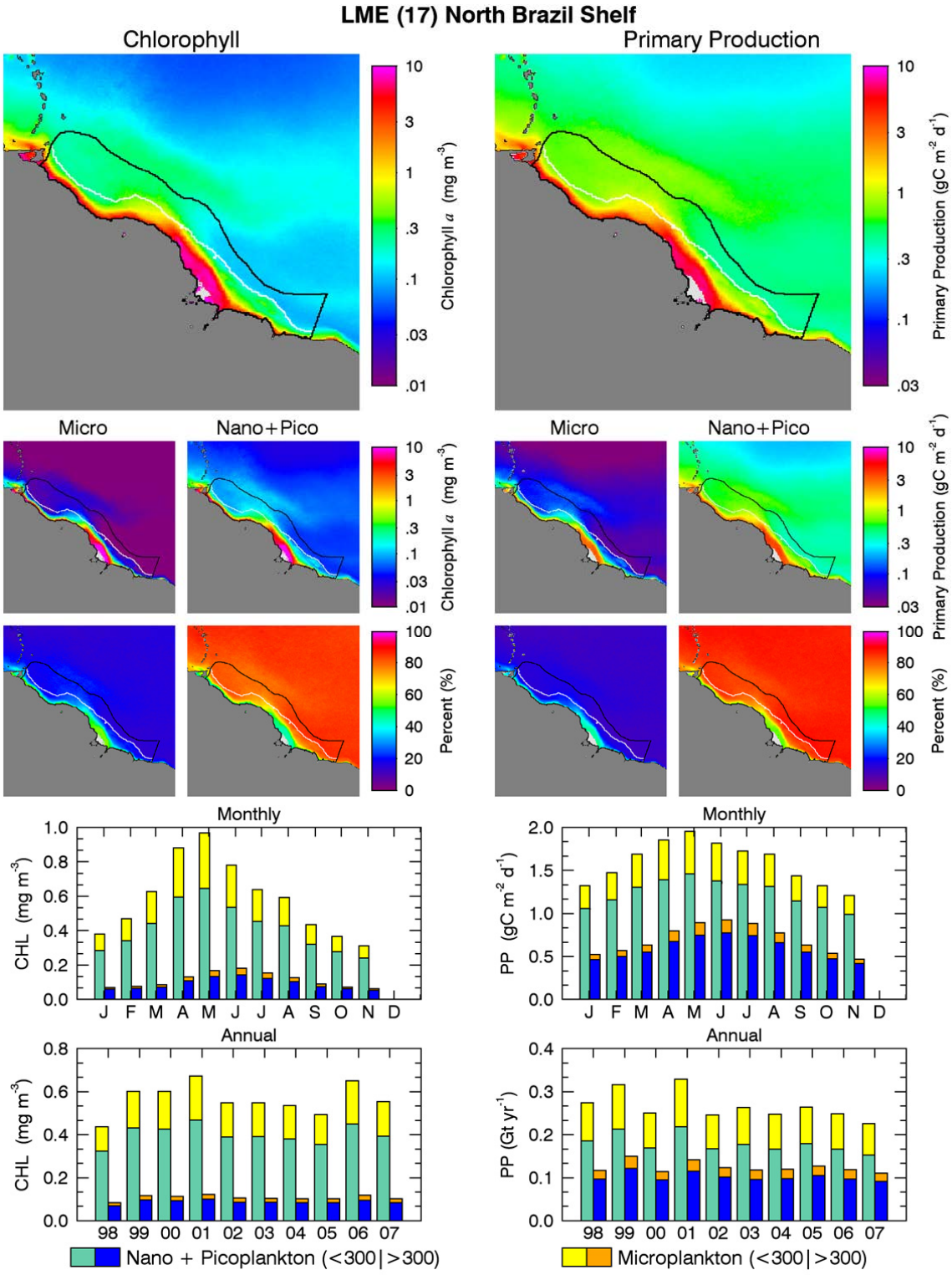


FIGURE A2.19
See legend of Figure A2.2

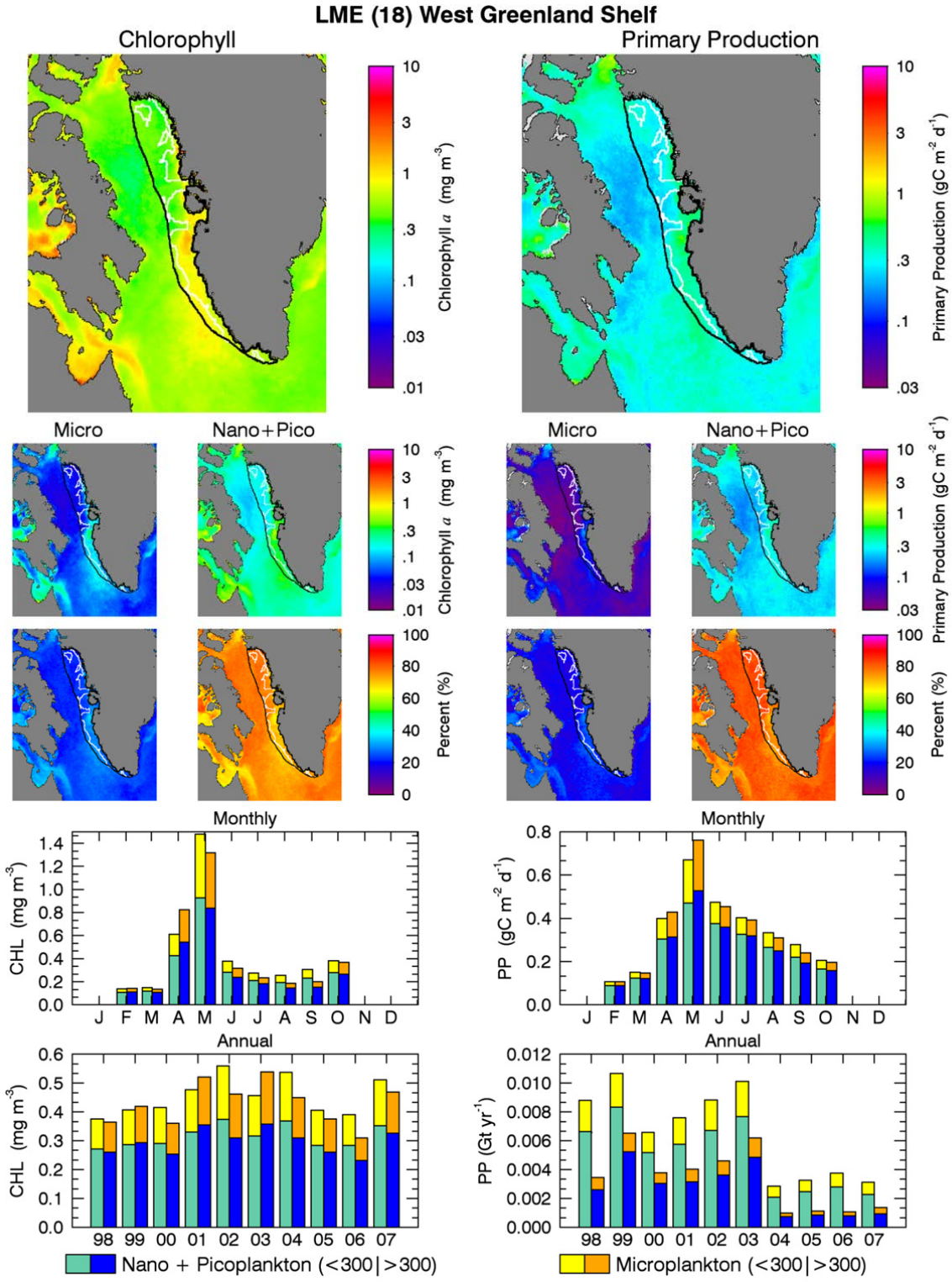


FIGURE A2.20
See legend of Figure A2.2

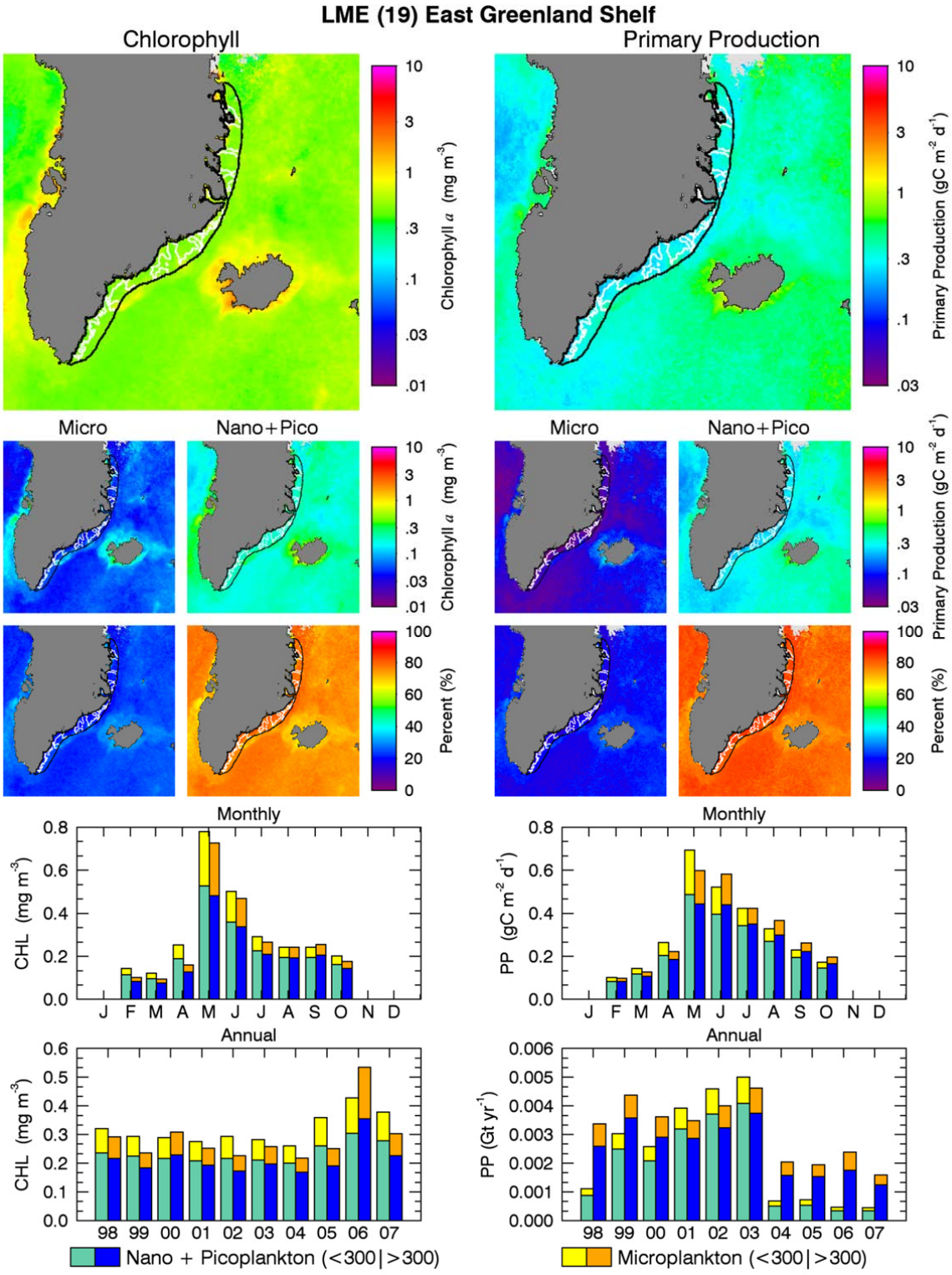


FIGURE A2.21
See legend of Figure A2.2

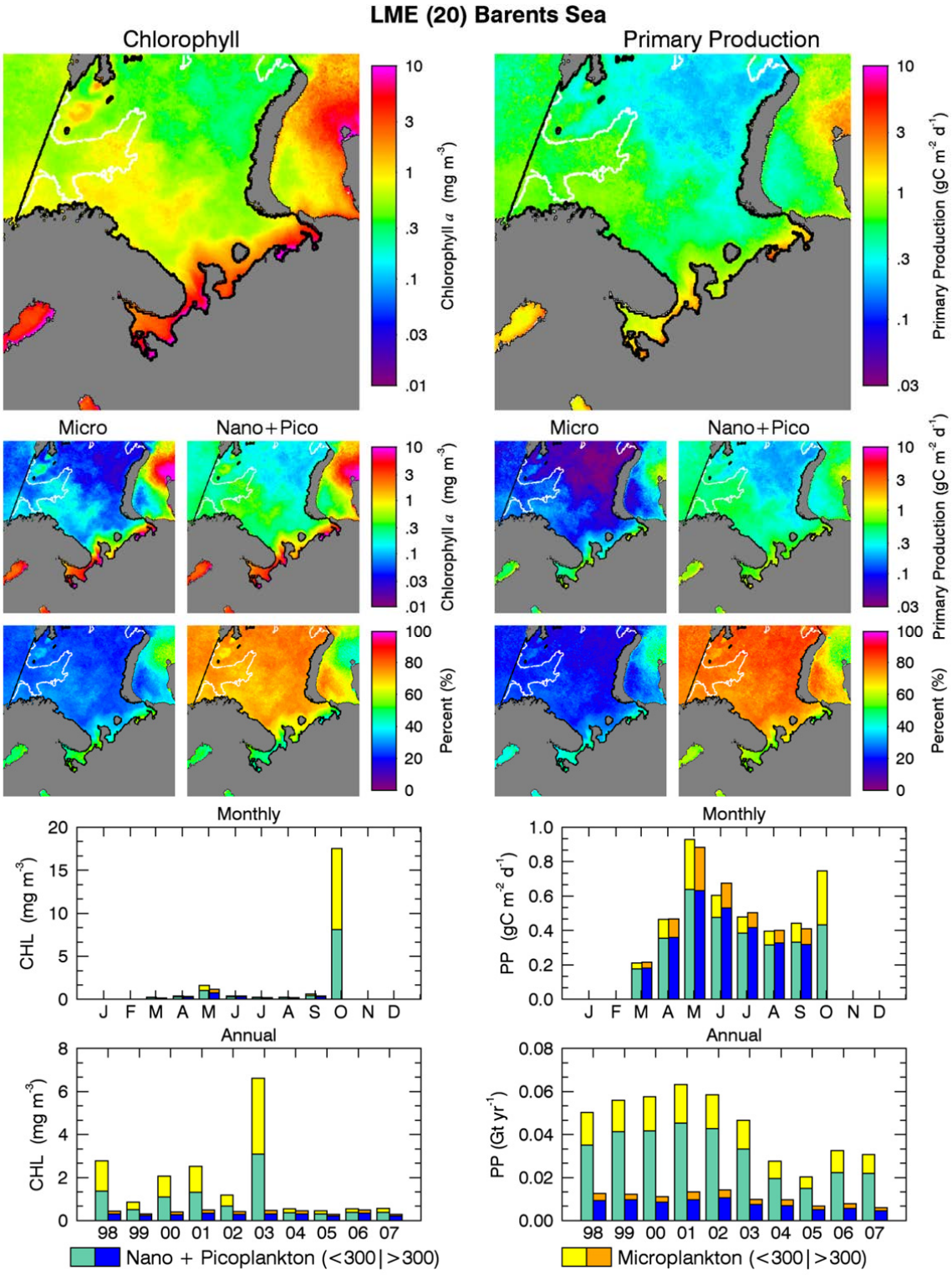


FIGURE A2.22
See legend of Figure A2.2

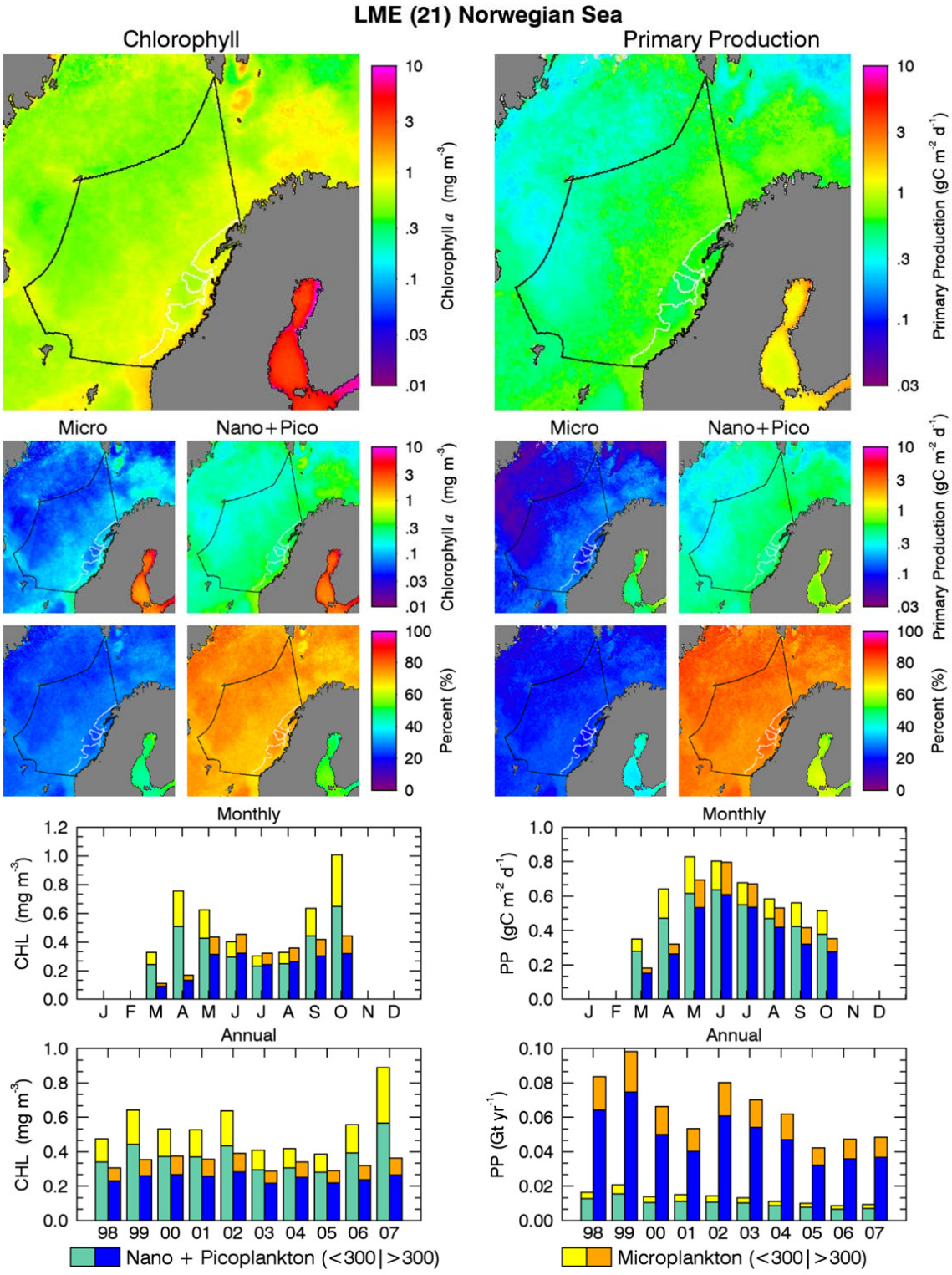


FIGURE A2.23
See legend of Figure A2.2

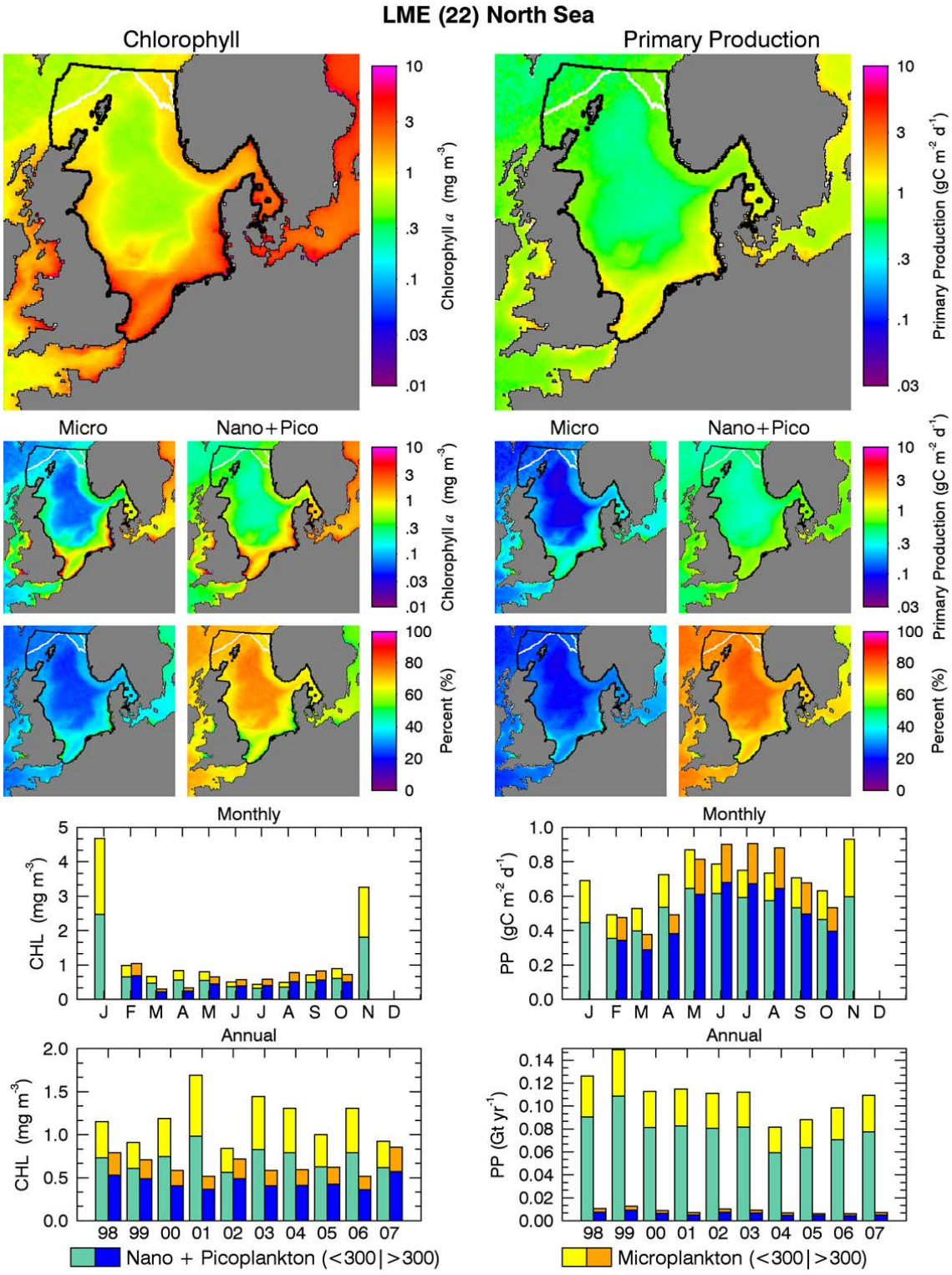


FIGURE A2.24
See legend of Figure A2.2

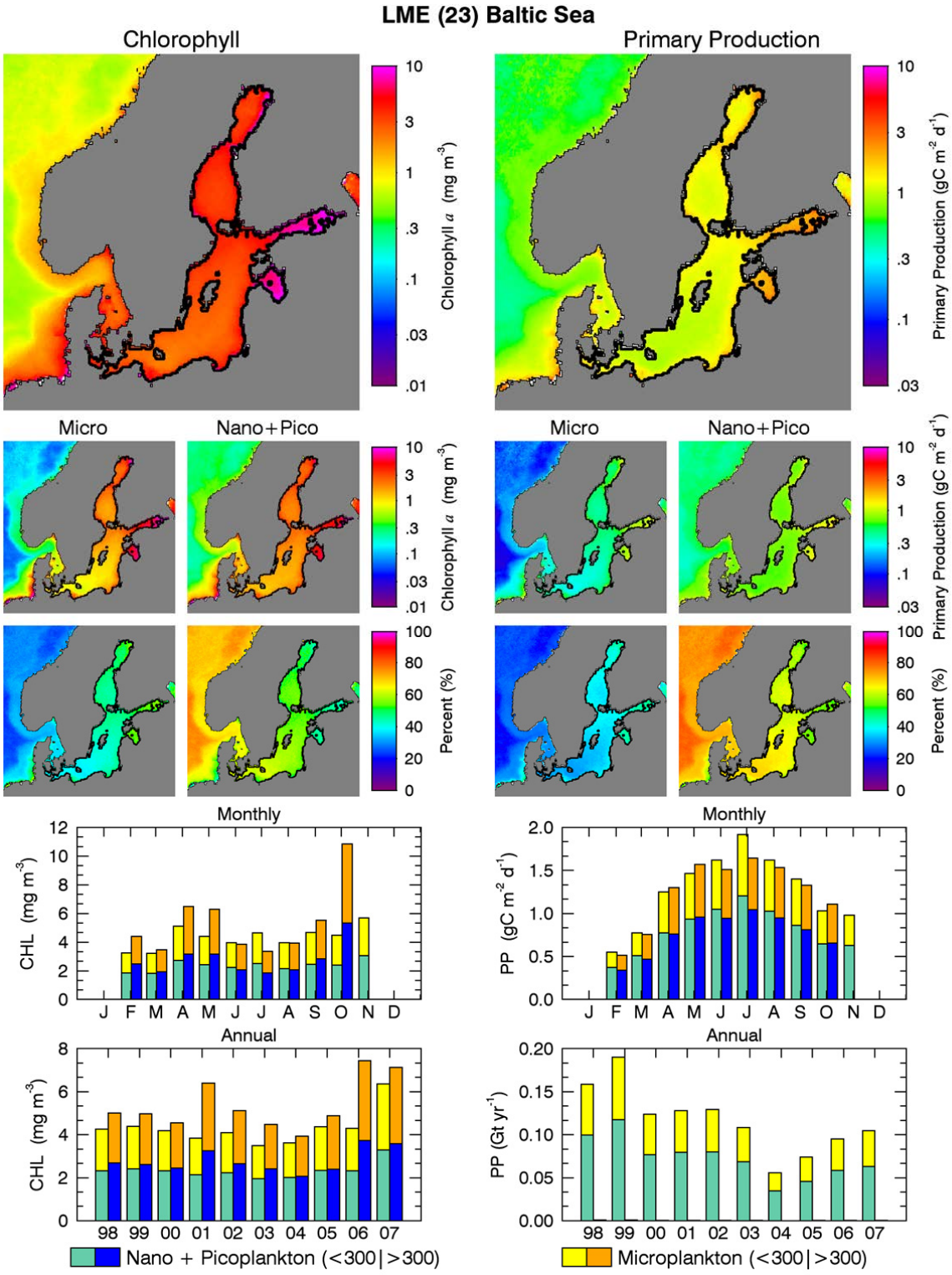


FIGURE A2.25
See legend of Figure A2.2

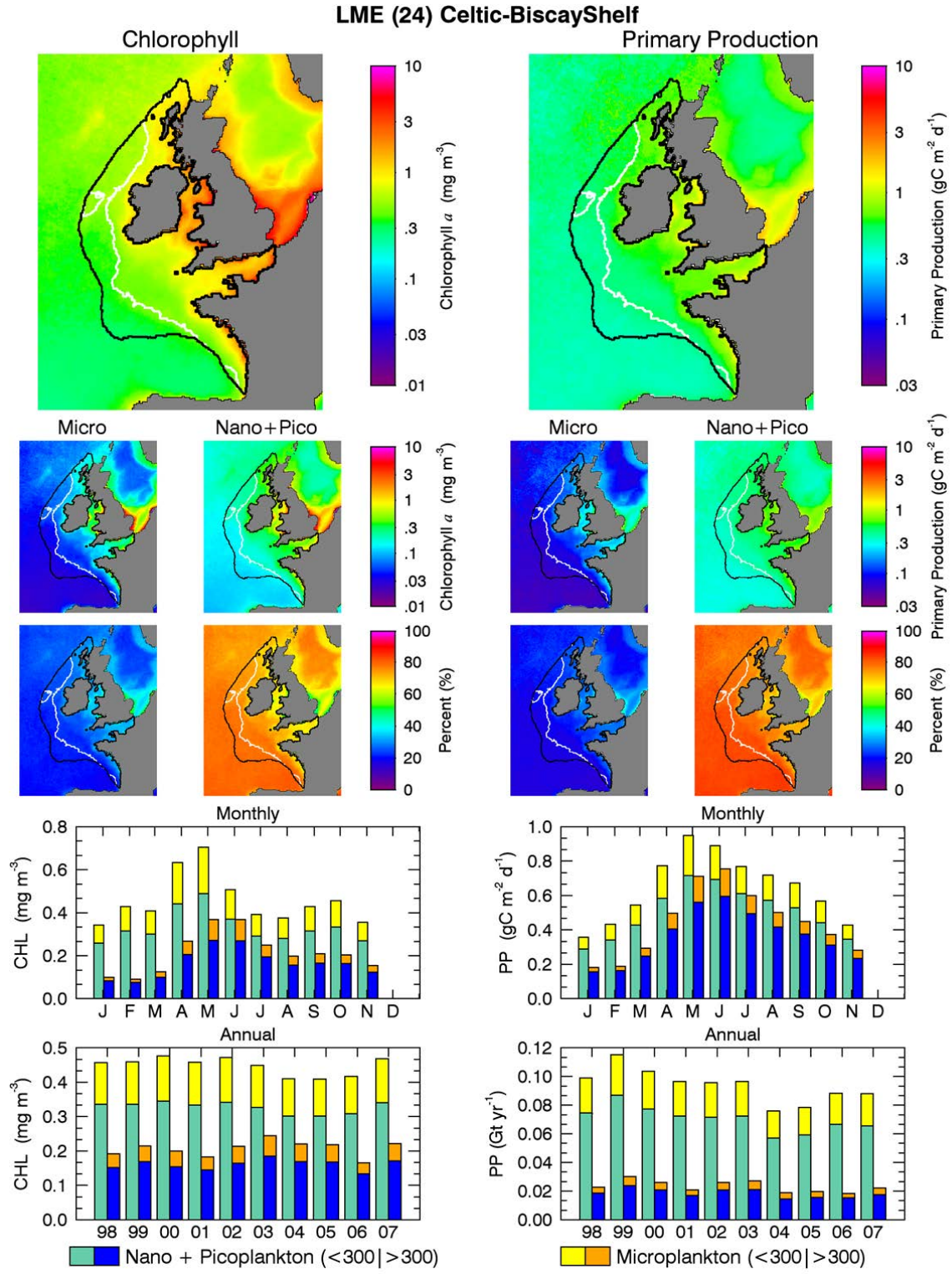


FIGURE A2.26
See legend of Figure A2.2

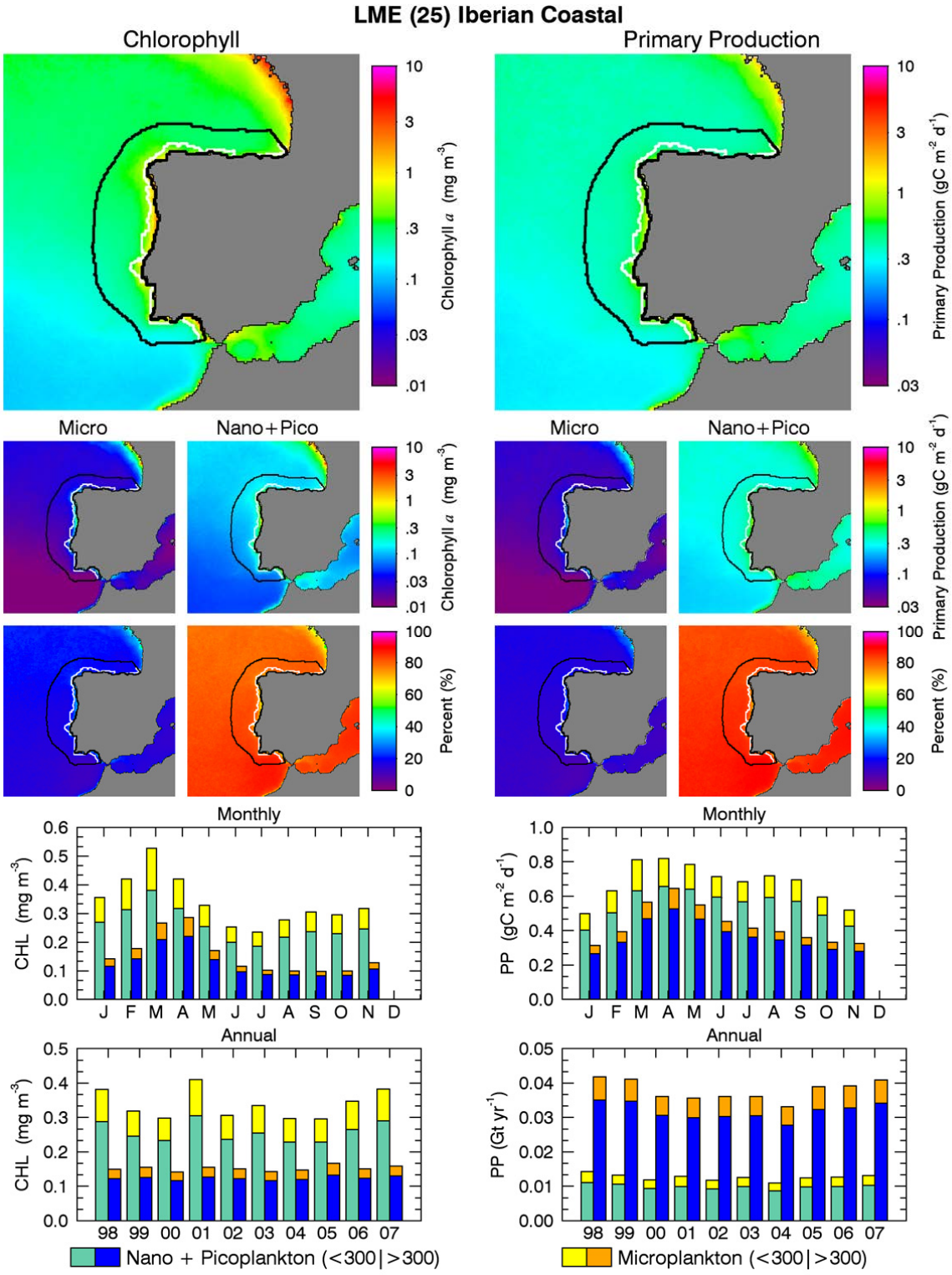


FIGURE A2.27
See legend of Figure A2.2

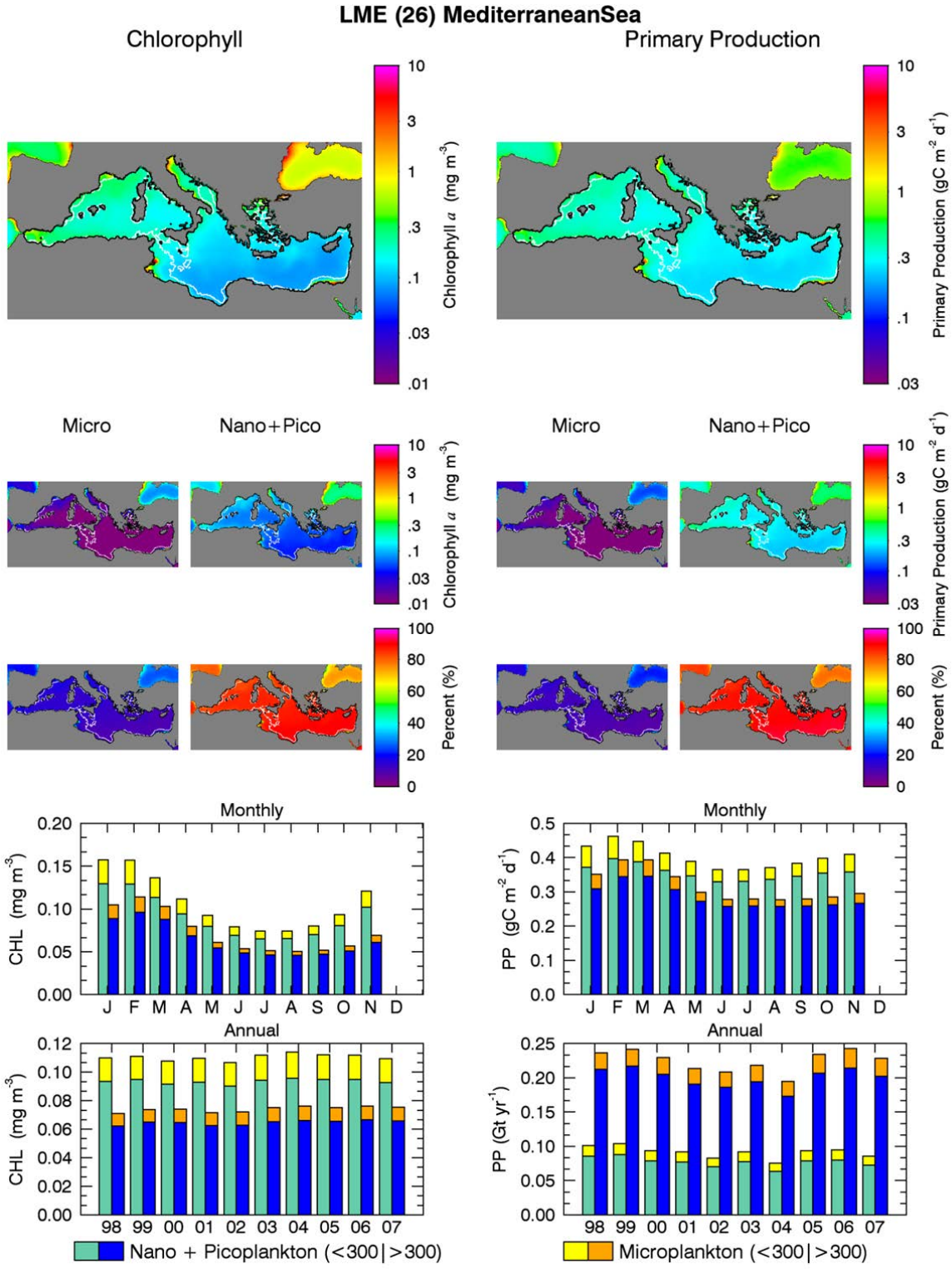


FIGURE A2.28
See legend of Figure A2.2

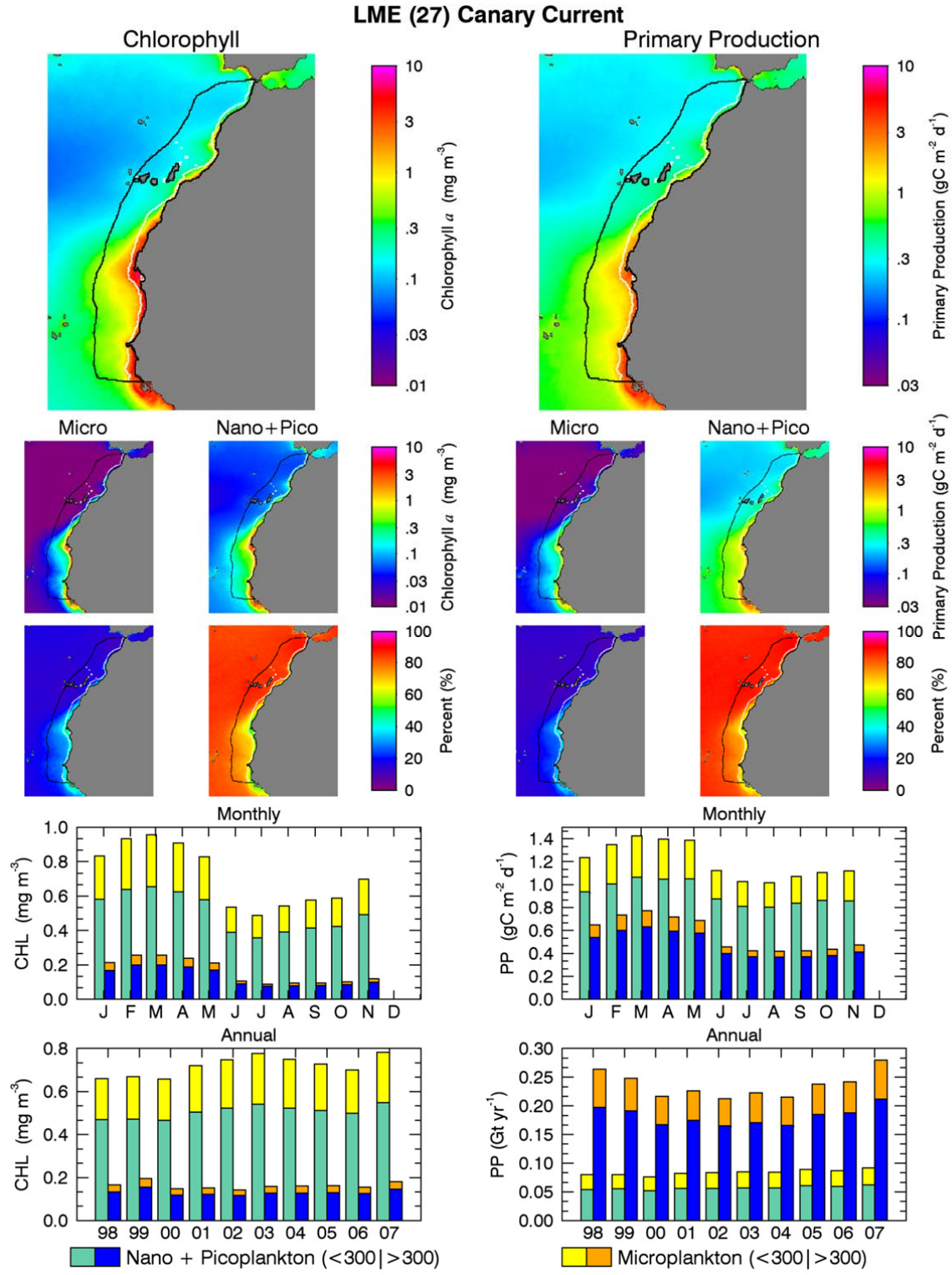


FIGURE A2.29
See legend of Figure A2.2

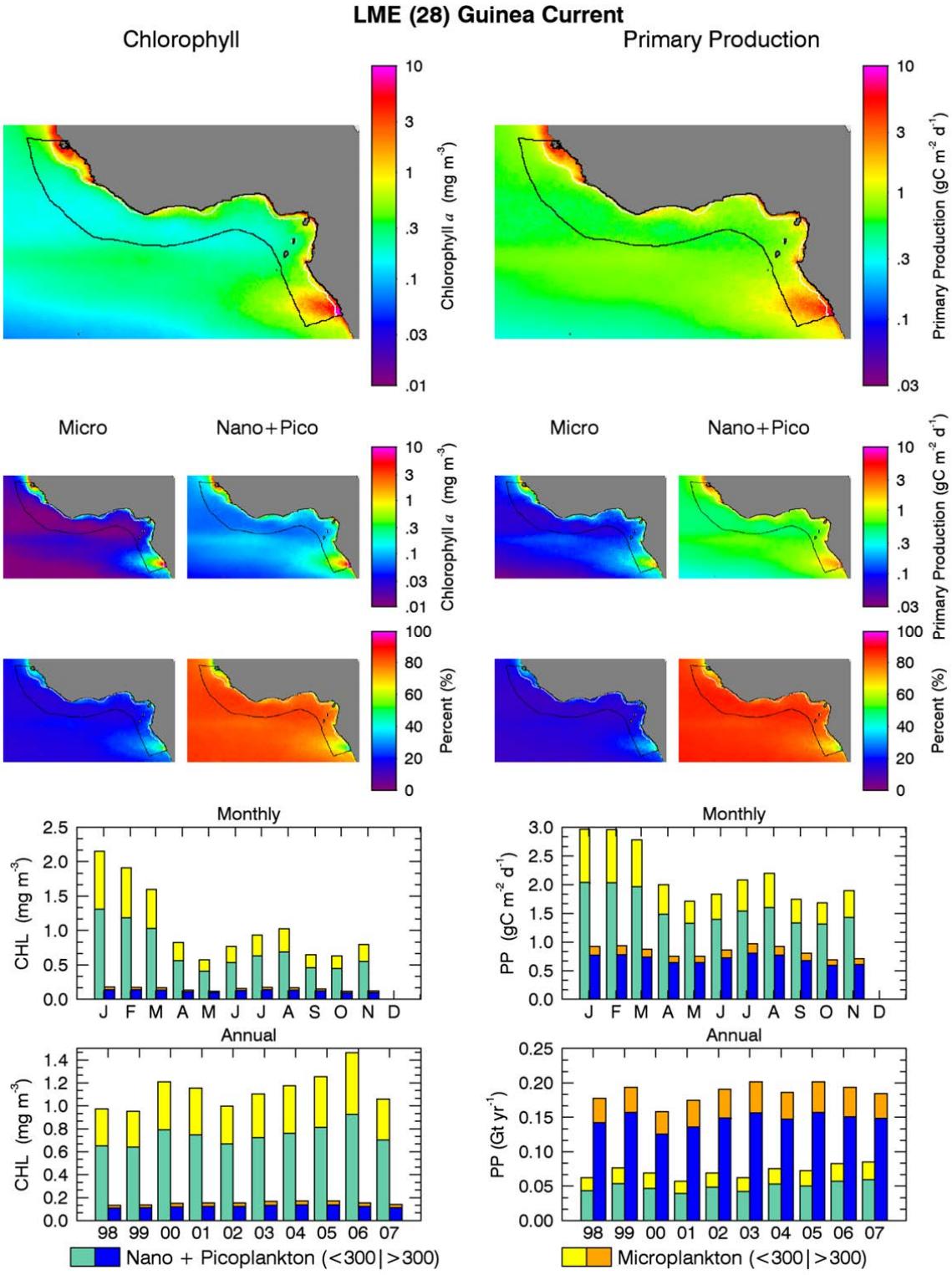


FIGURE A2.30
See legend of Figure A2.2

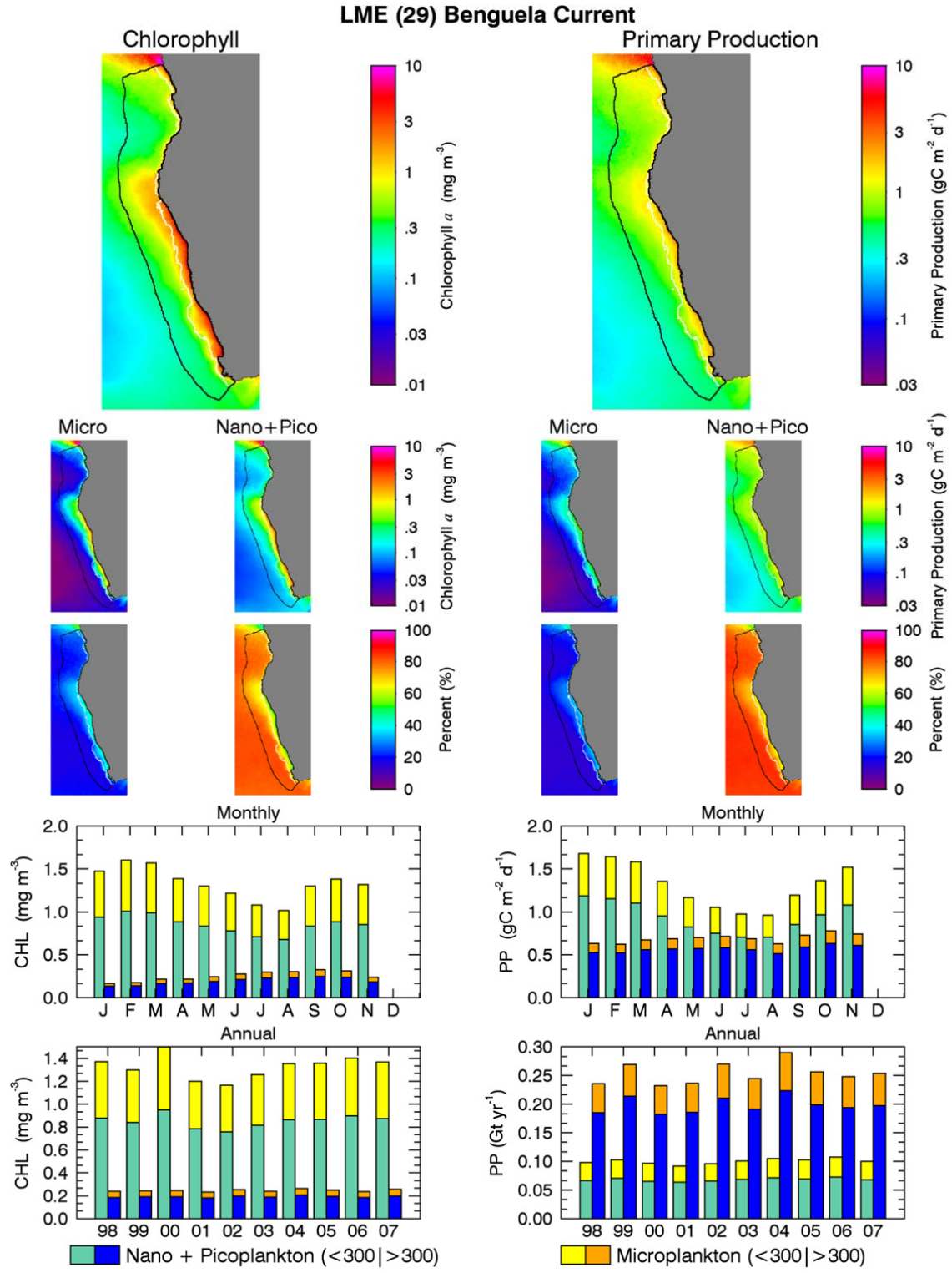


FIGURE A2.31
See legend of Figure A2.2

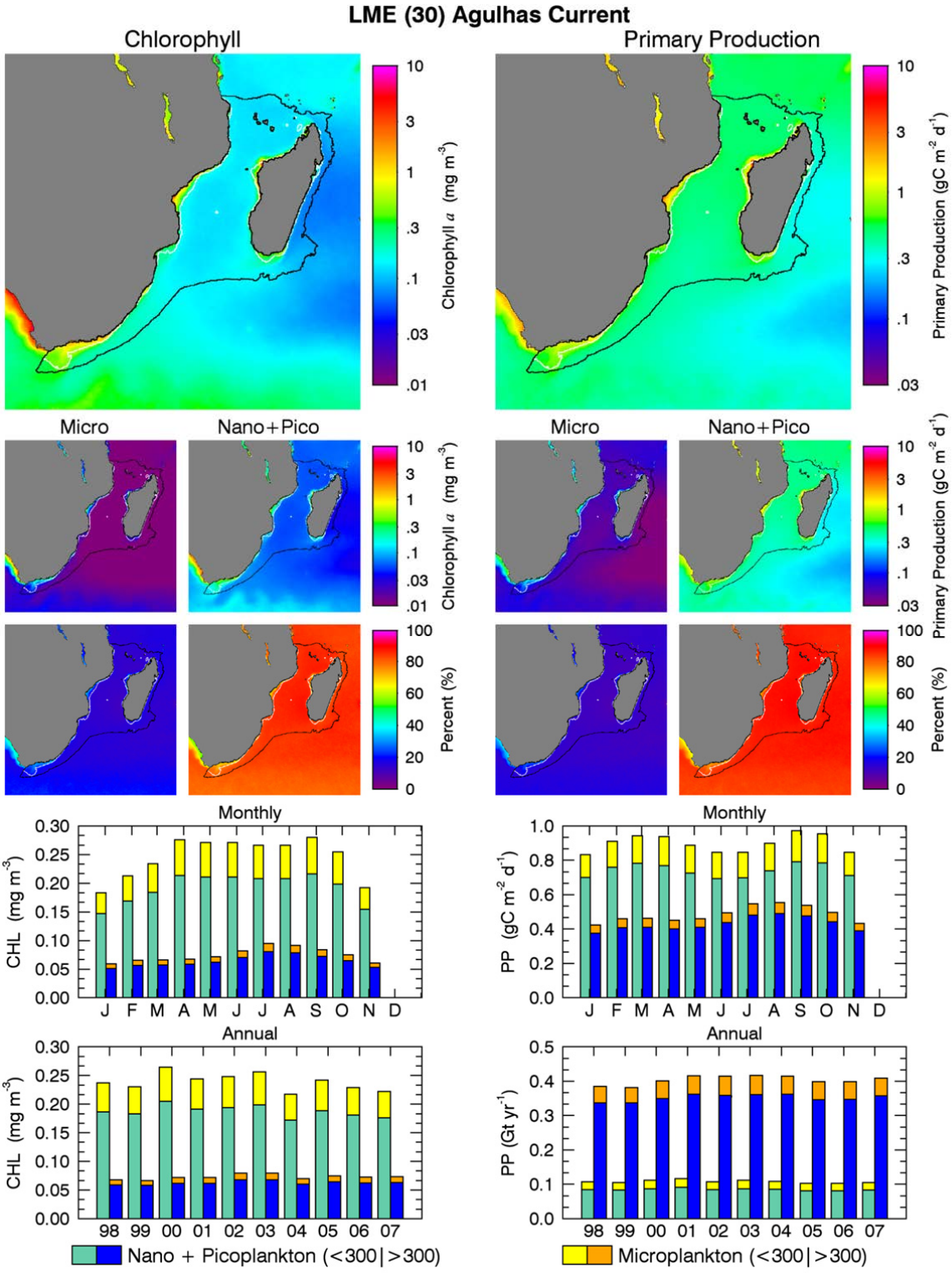


FIGURE A2.32
See legend of Figure A2.2

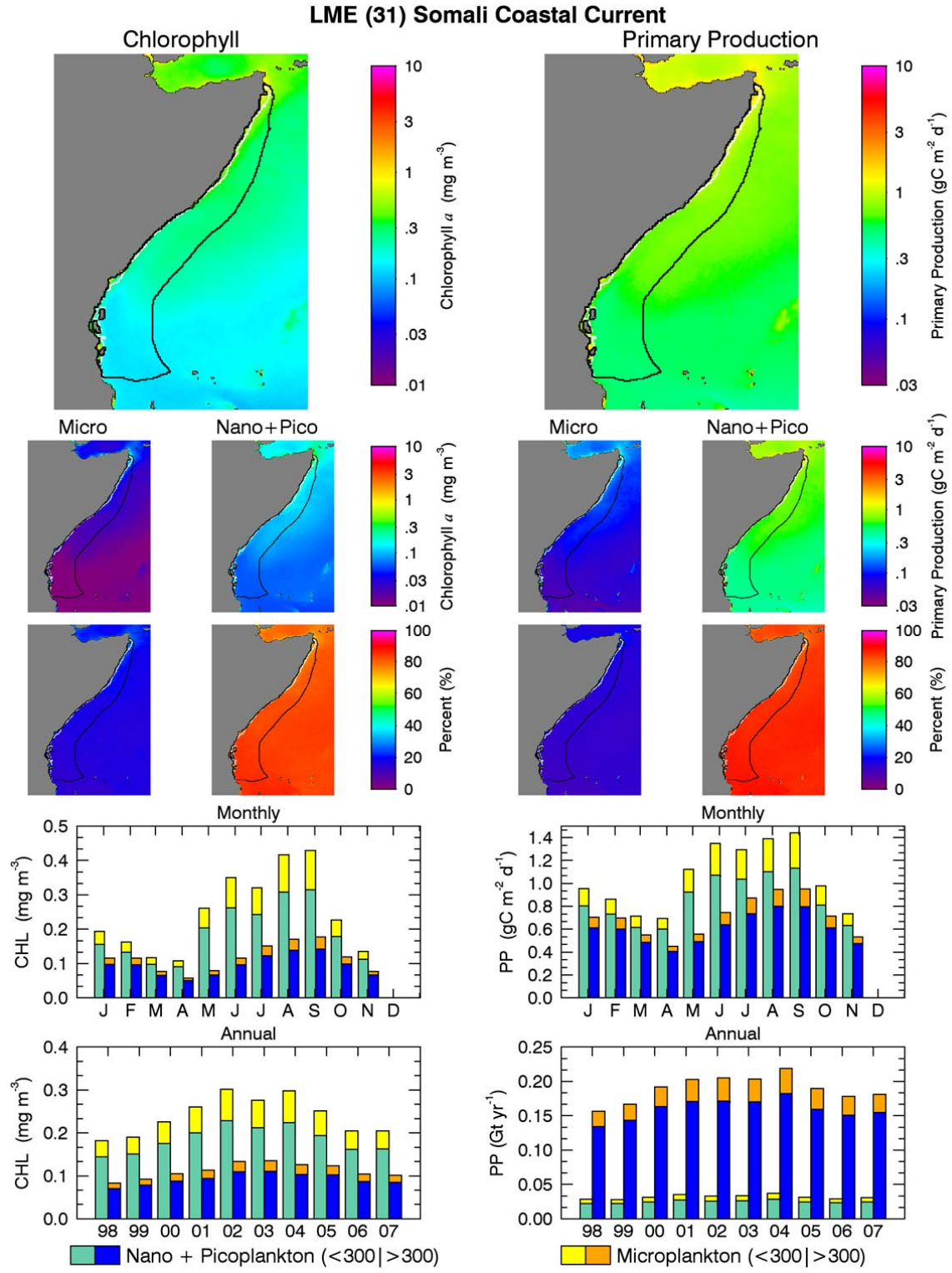


FIGURE A2.33
See legend of Figure A2.2

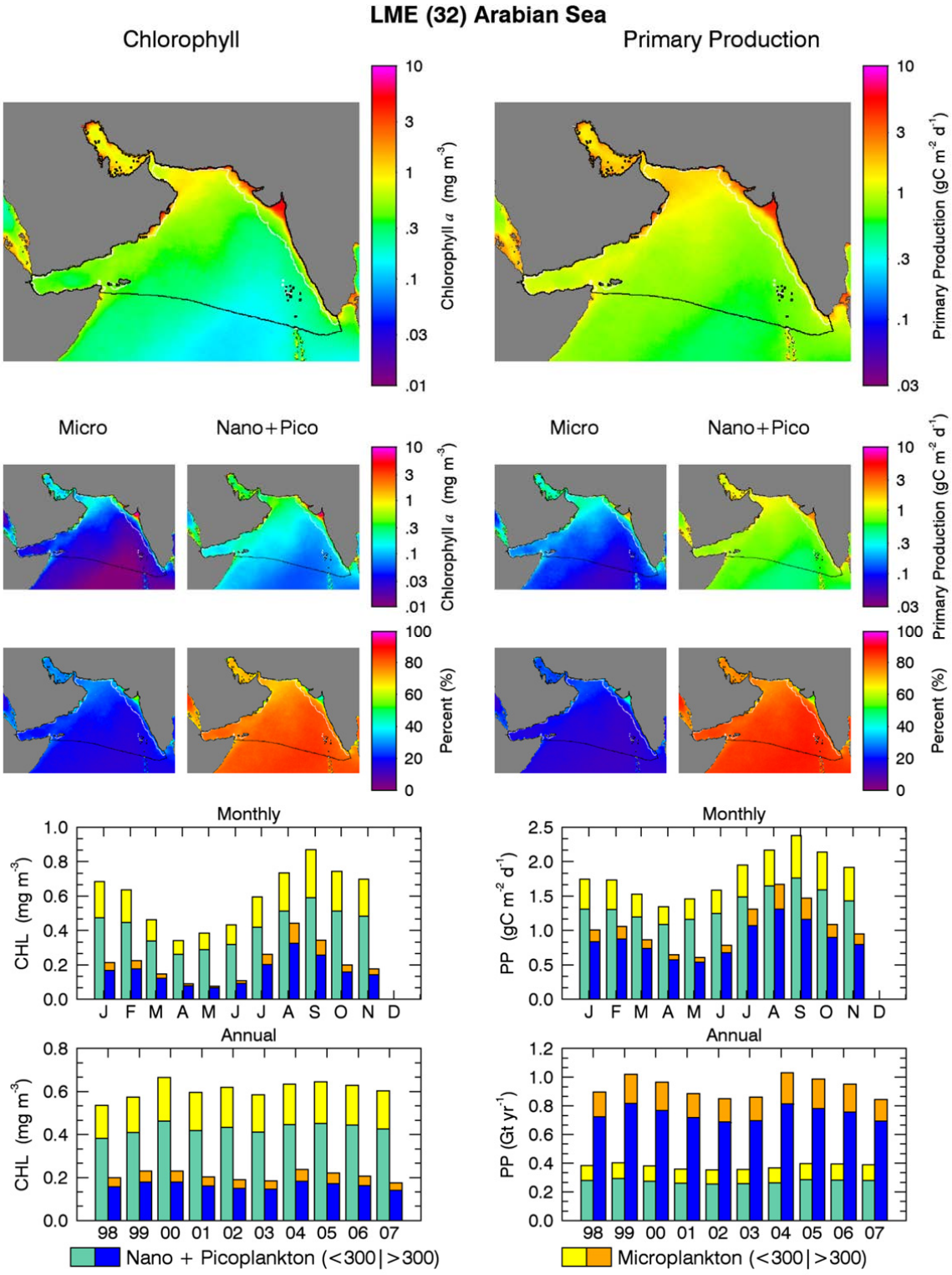


FIGURE A2.34
See legend of Figure A2.2

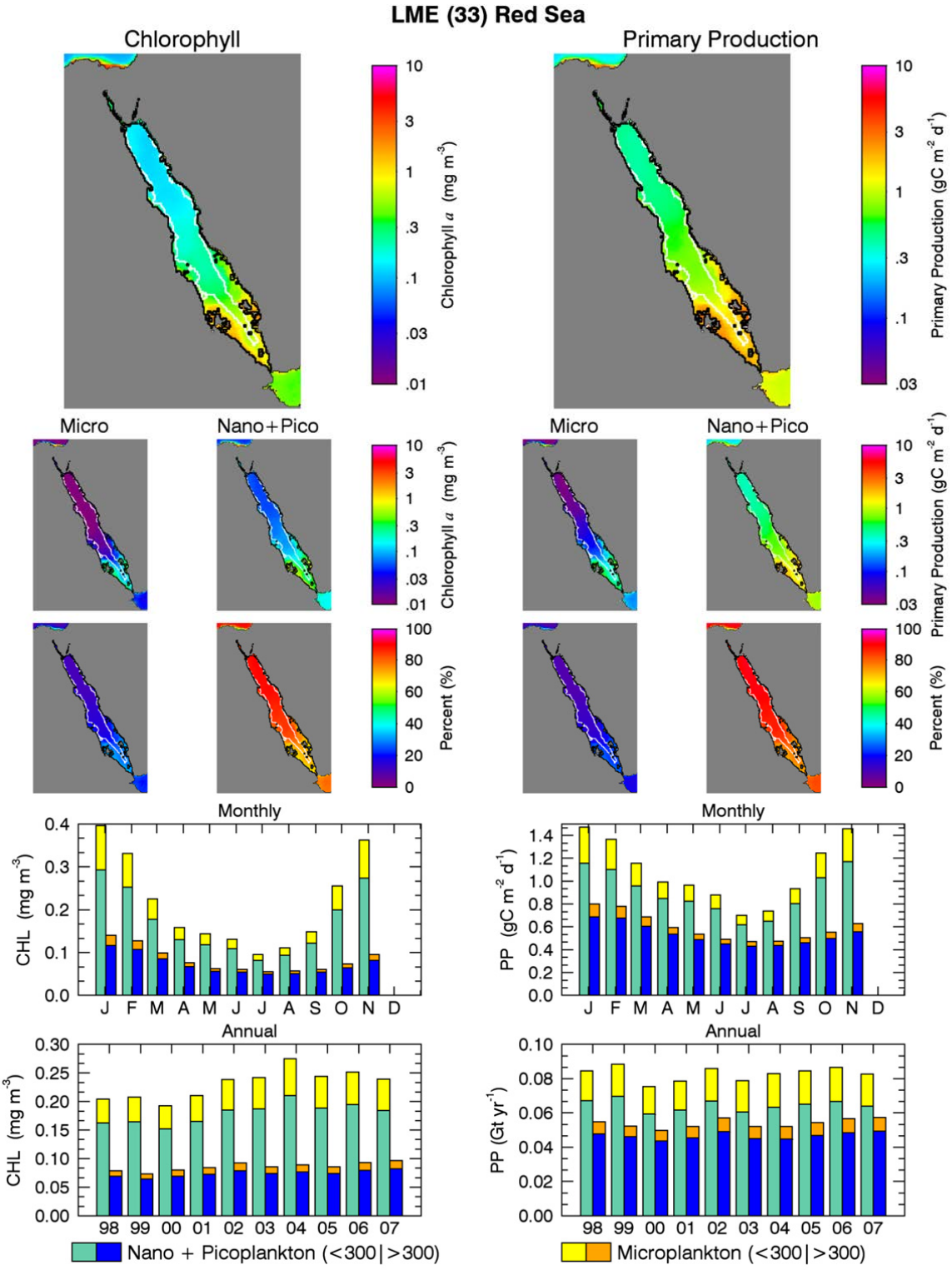


FIGURE A2.35
See legend of Figure A2.2

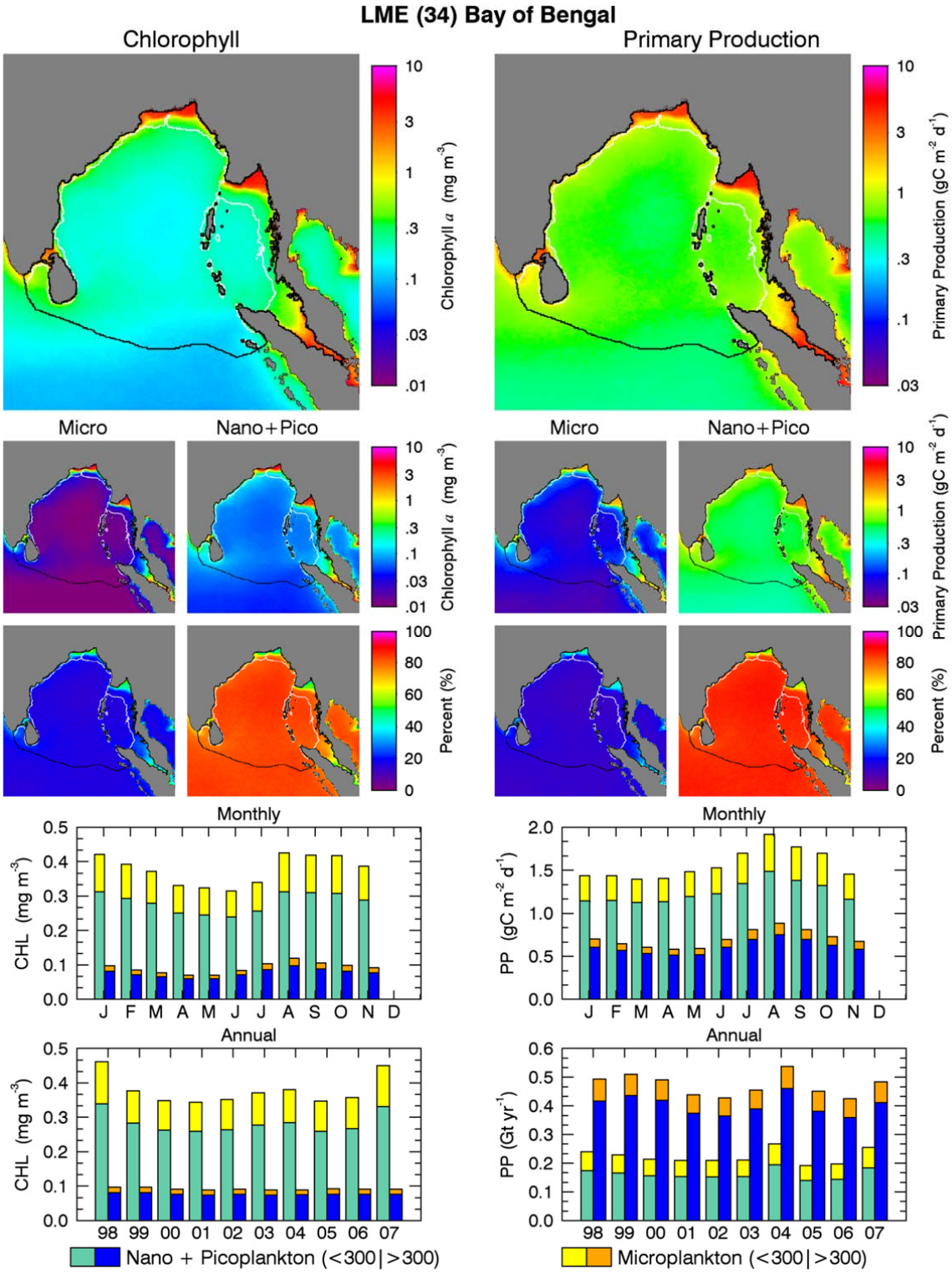


FIGURE A2.36
See legend of Figure A2.2

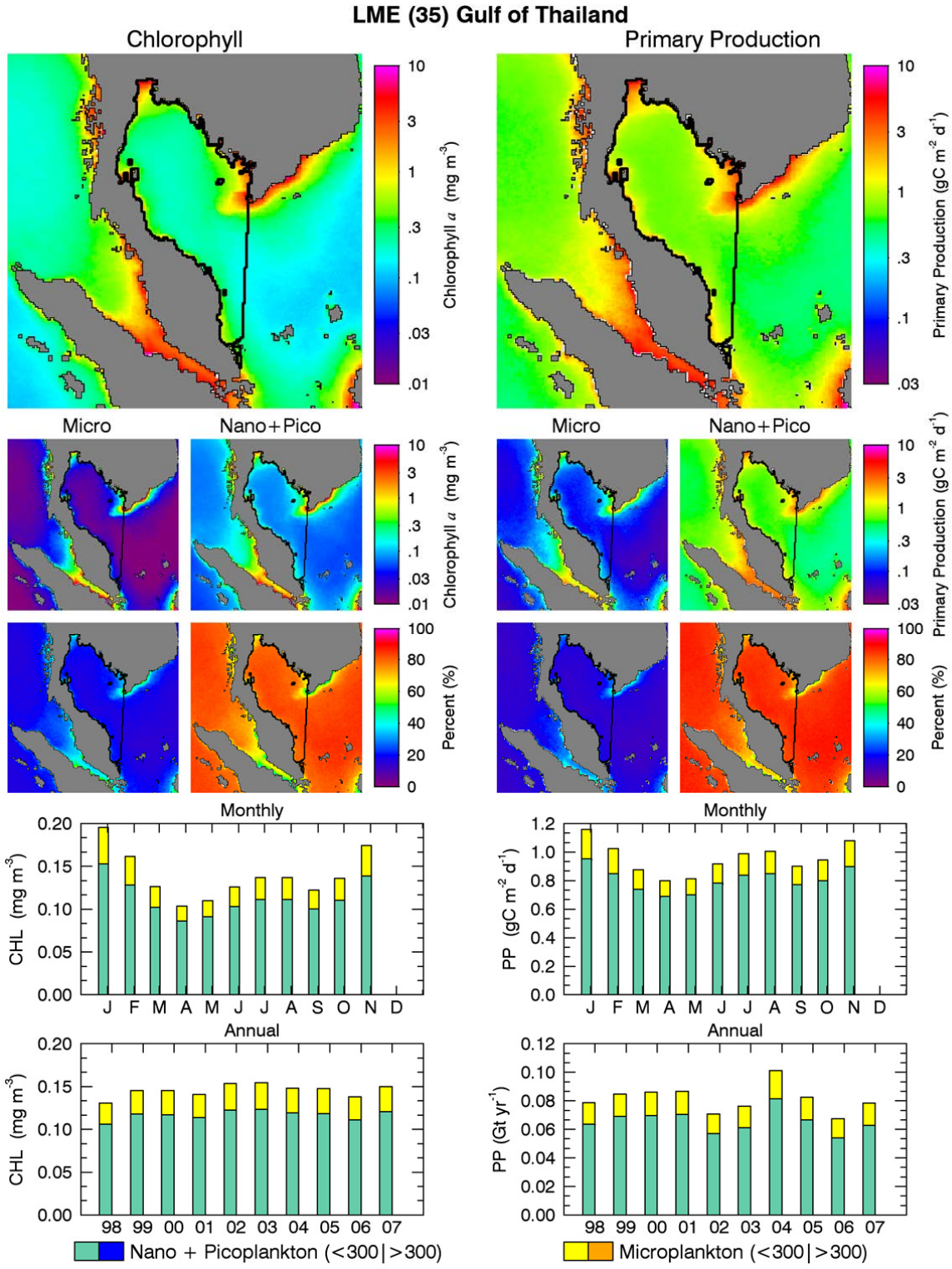


FIGURE A2.37
See legend of Figure A2.2

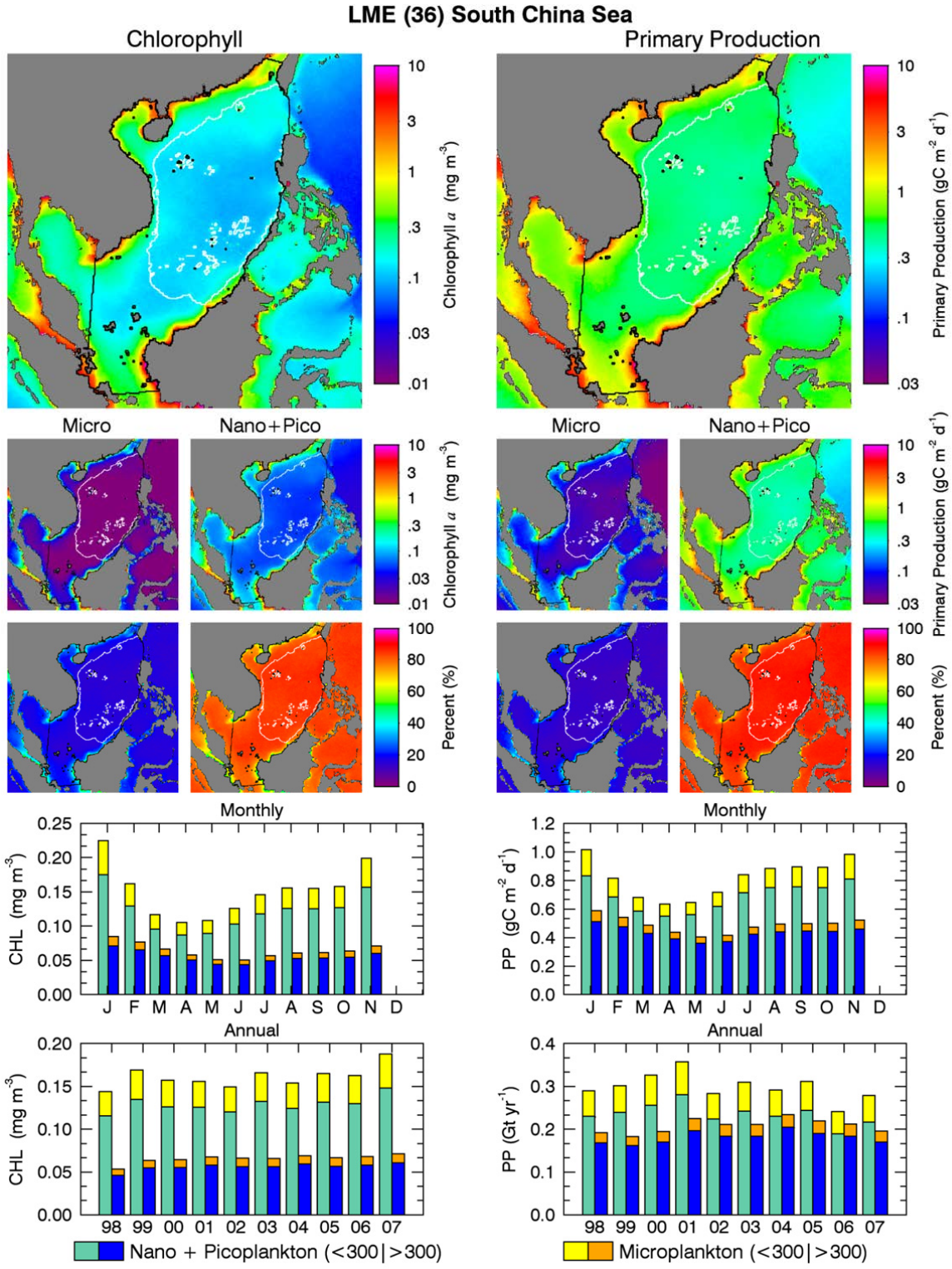


FIGURE A2.38
See legend of Figure A2.2

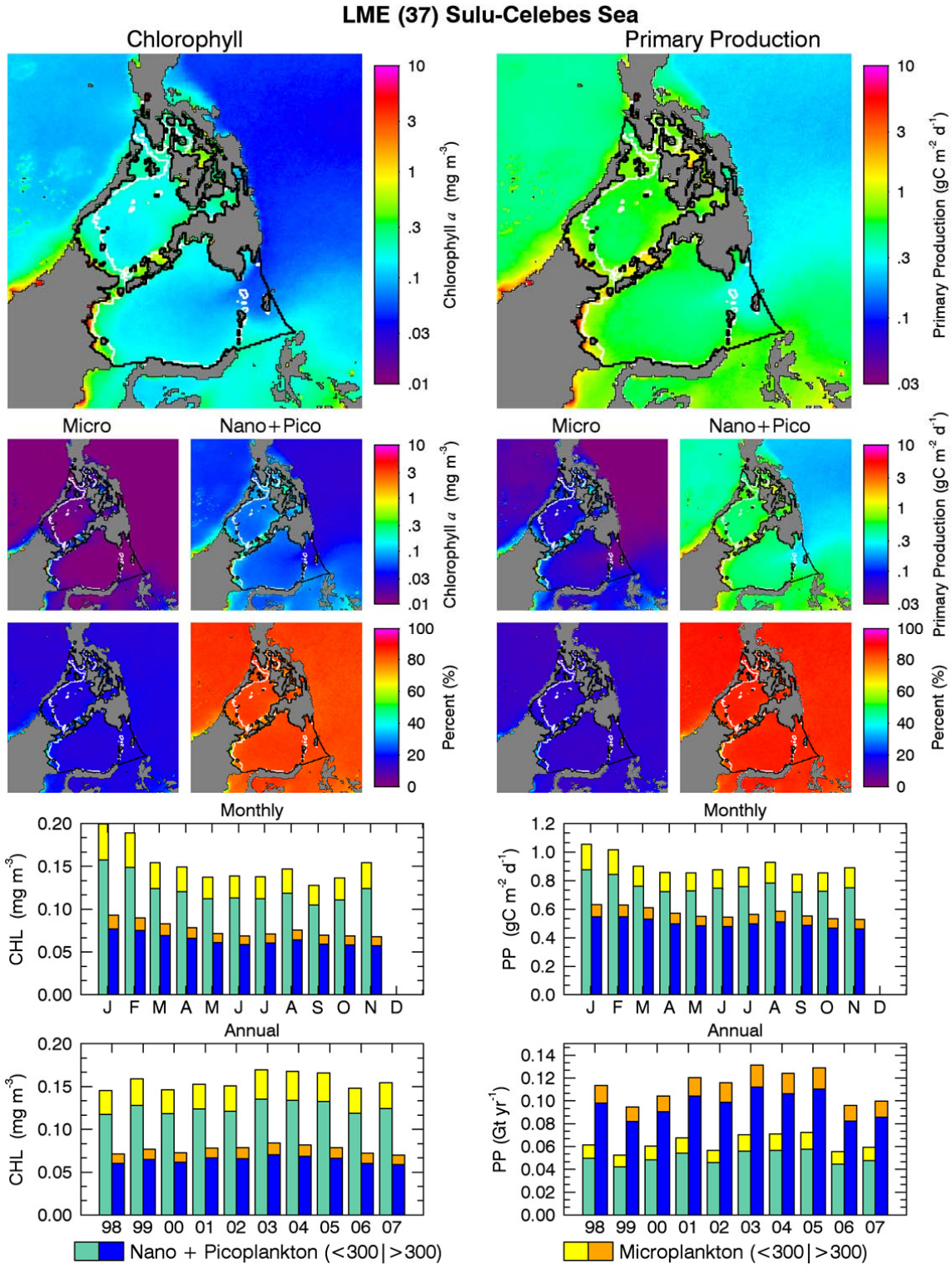


FIGURE A2.39
See legend of Figure A2.2

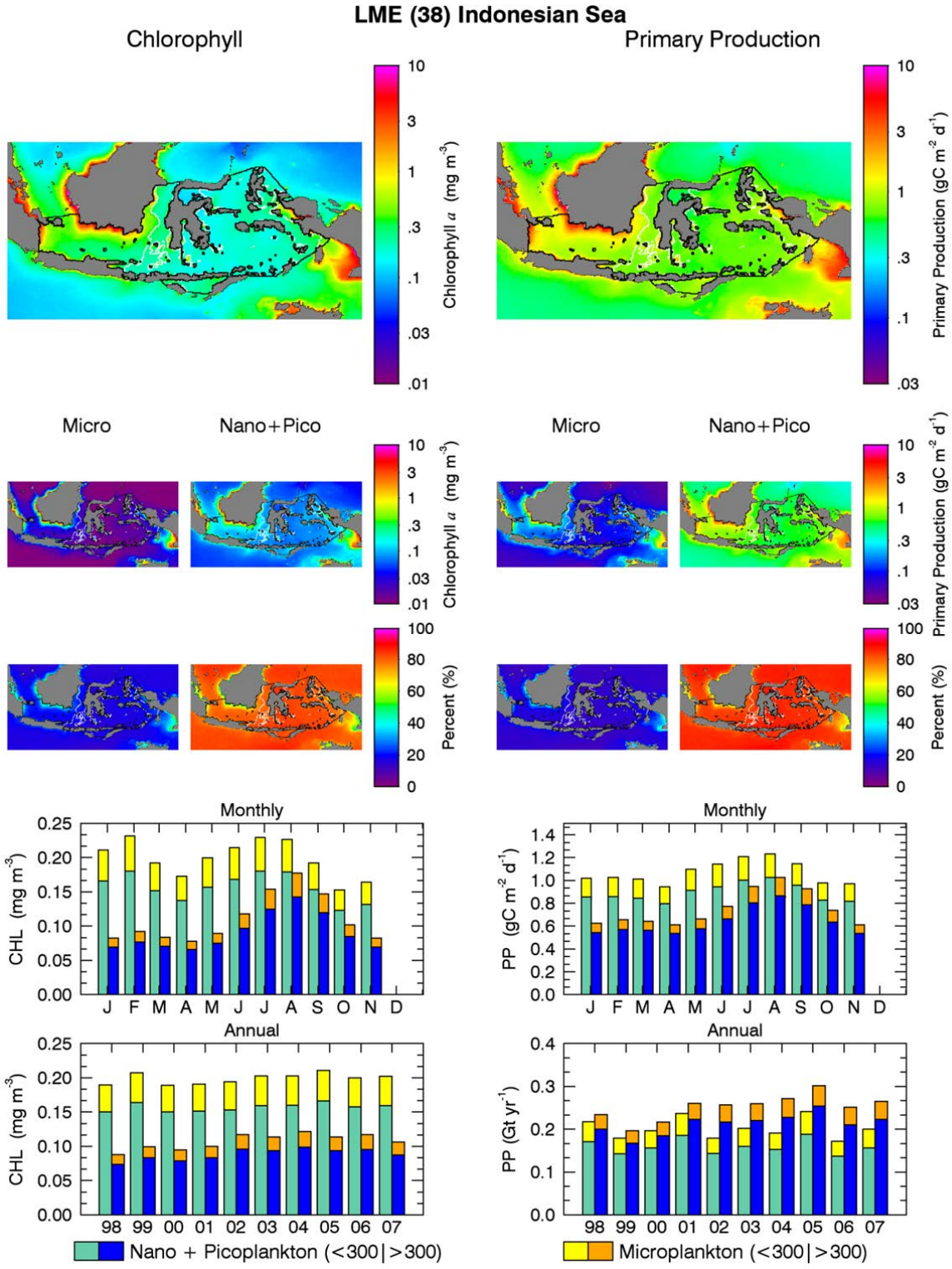


FIGURE A2.40
See legend of Figure A2.2

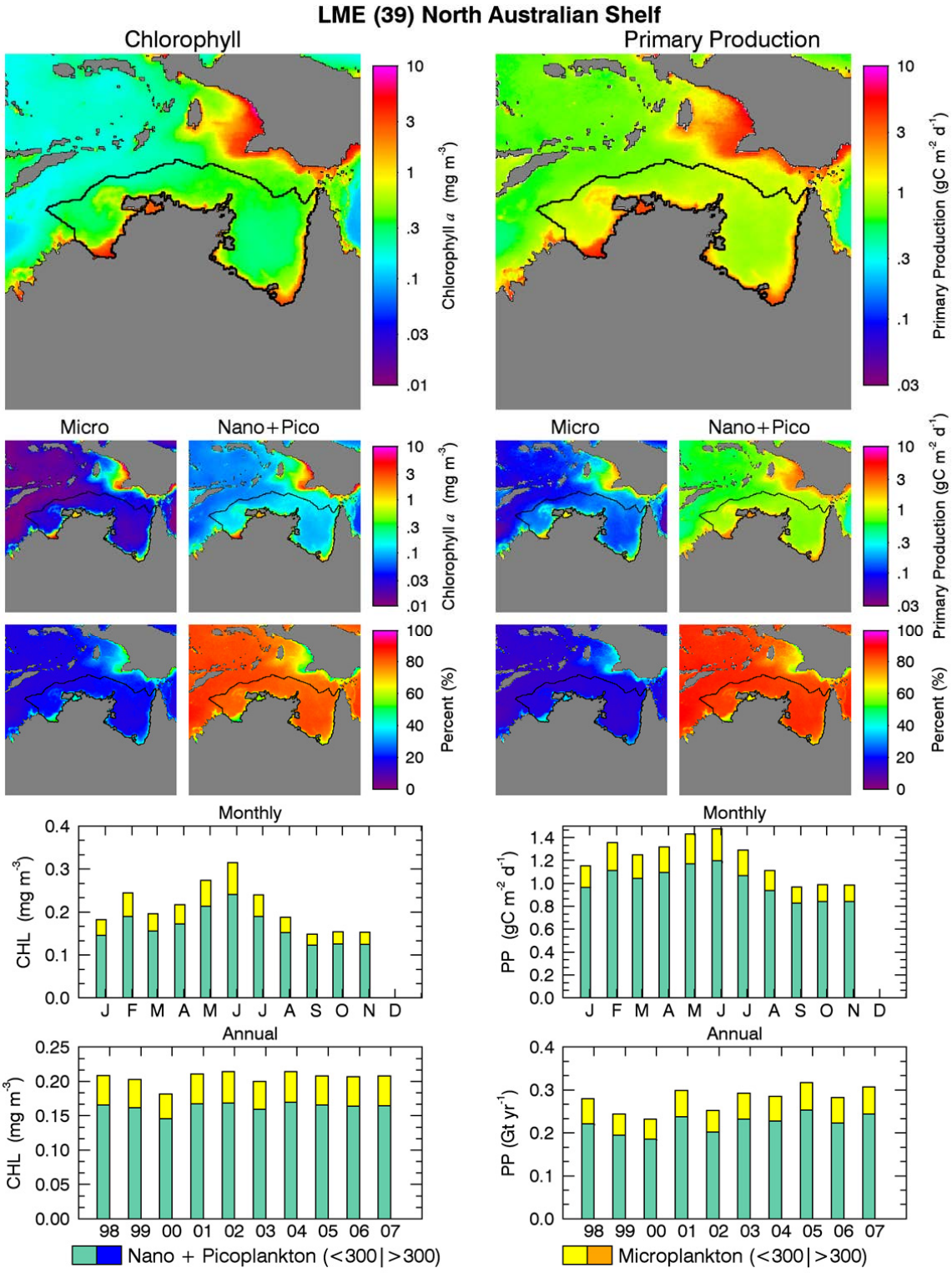


FIGURE A2.41
See legend of Figure A2.2

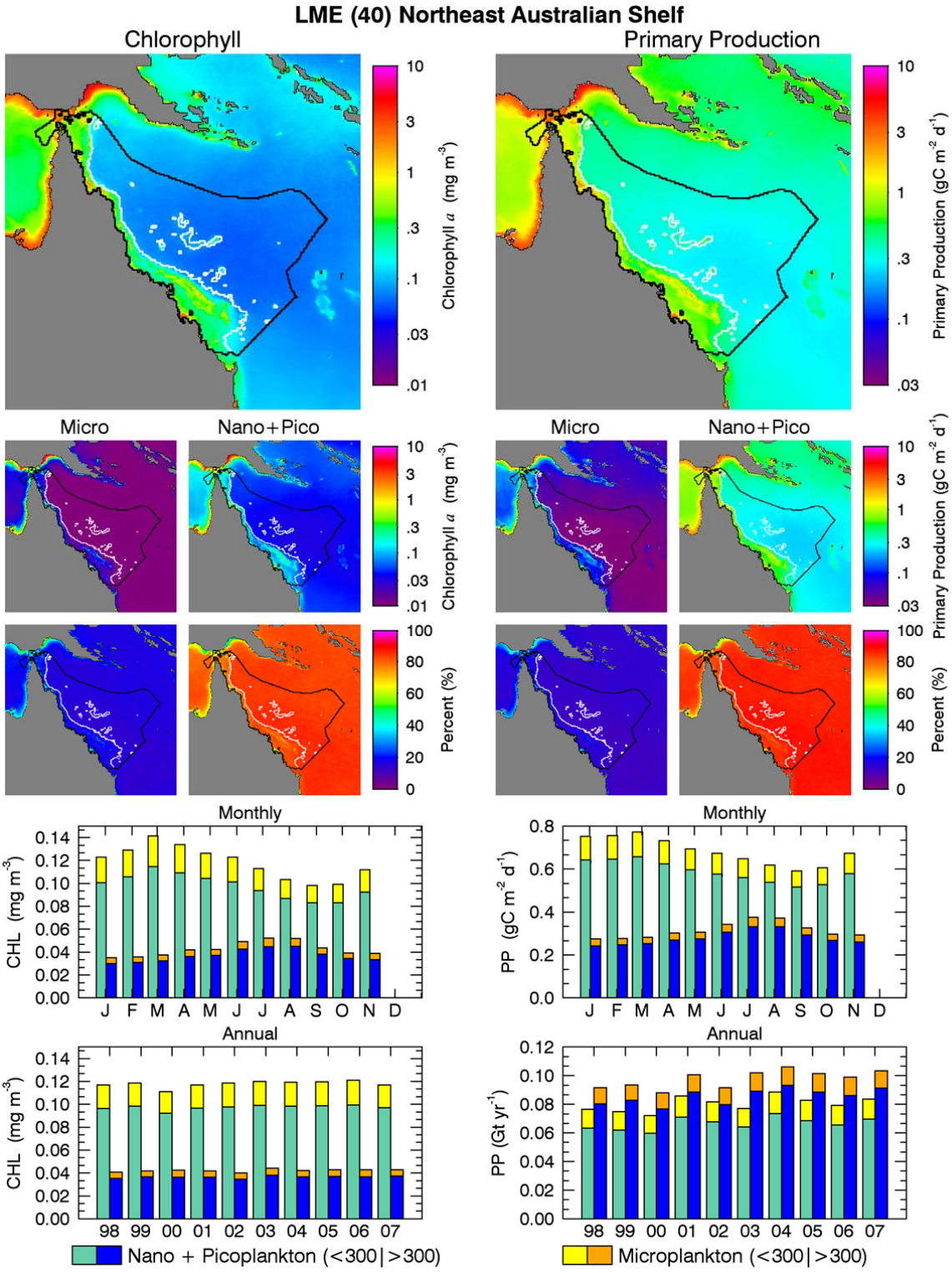


FIGURE A2.42
See legend of Figure A2.2

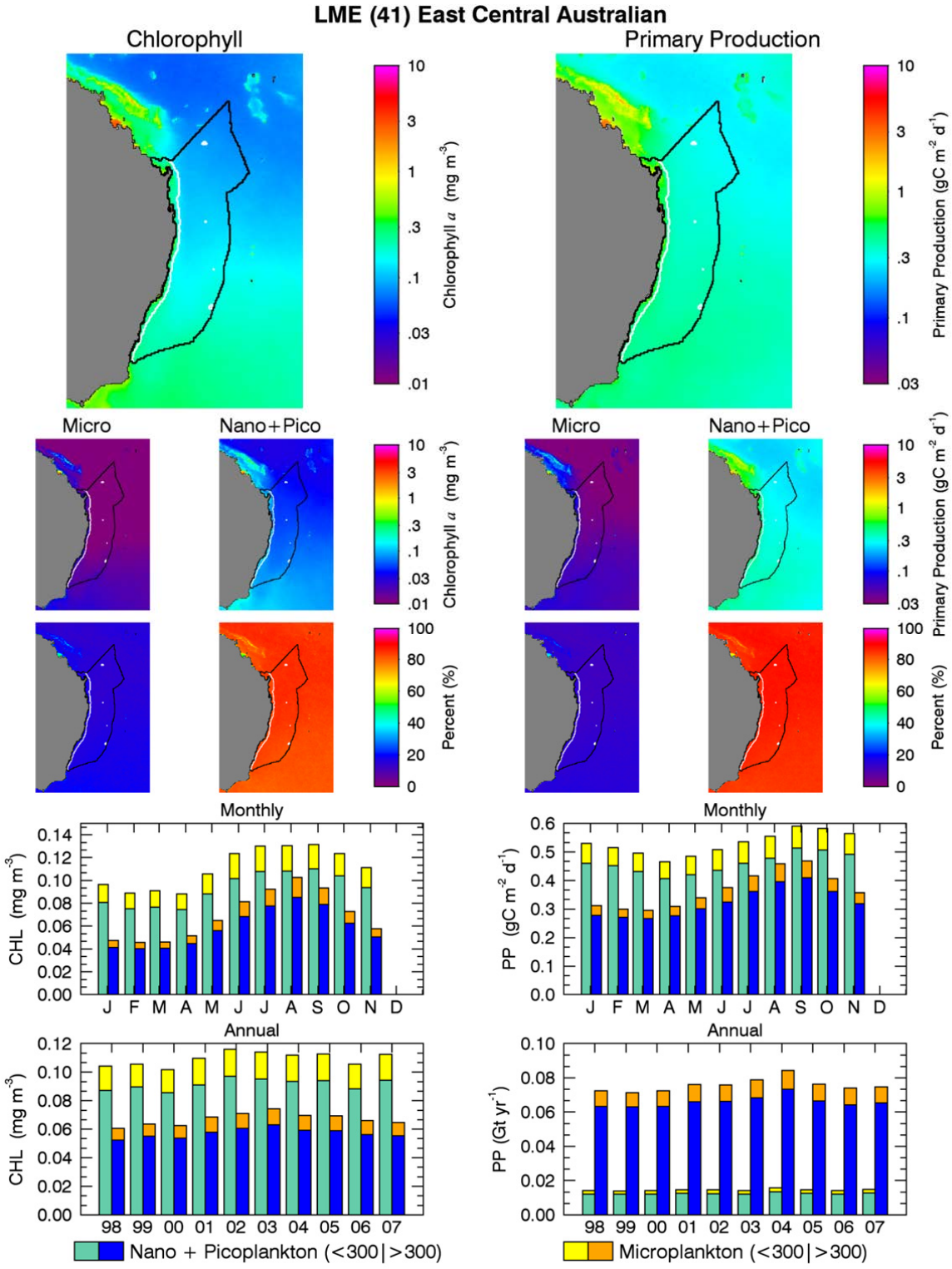


FIGURE A2.43
See legend of Figure A2.2

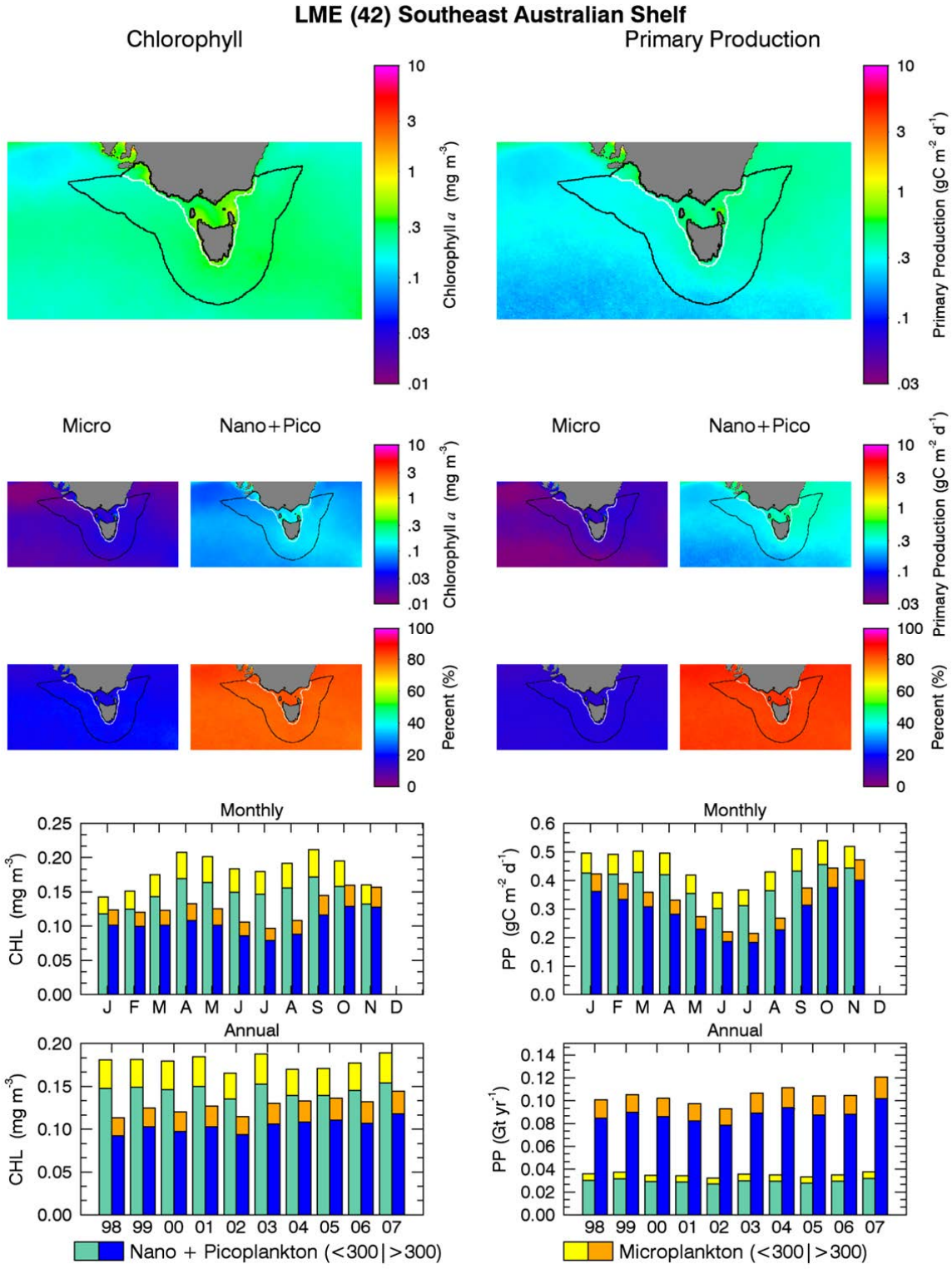


FIGURE A2.44
See legend of Figure A2.2

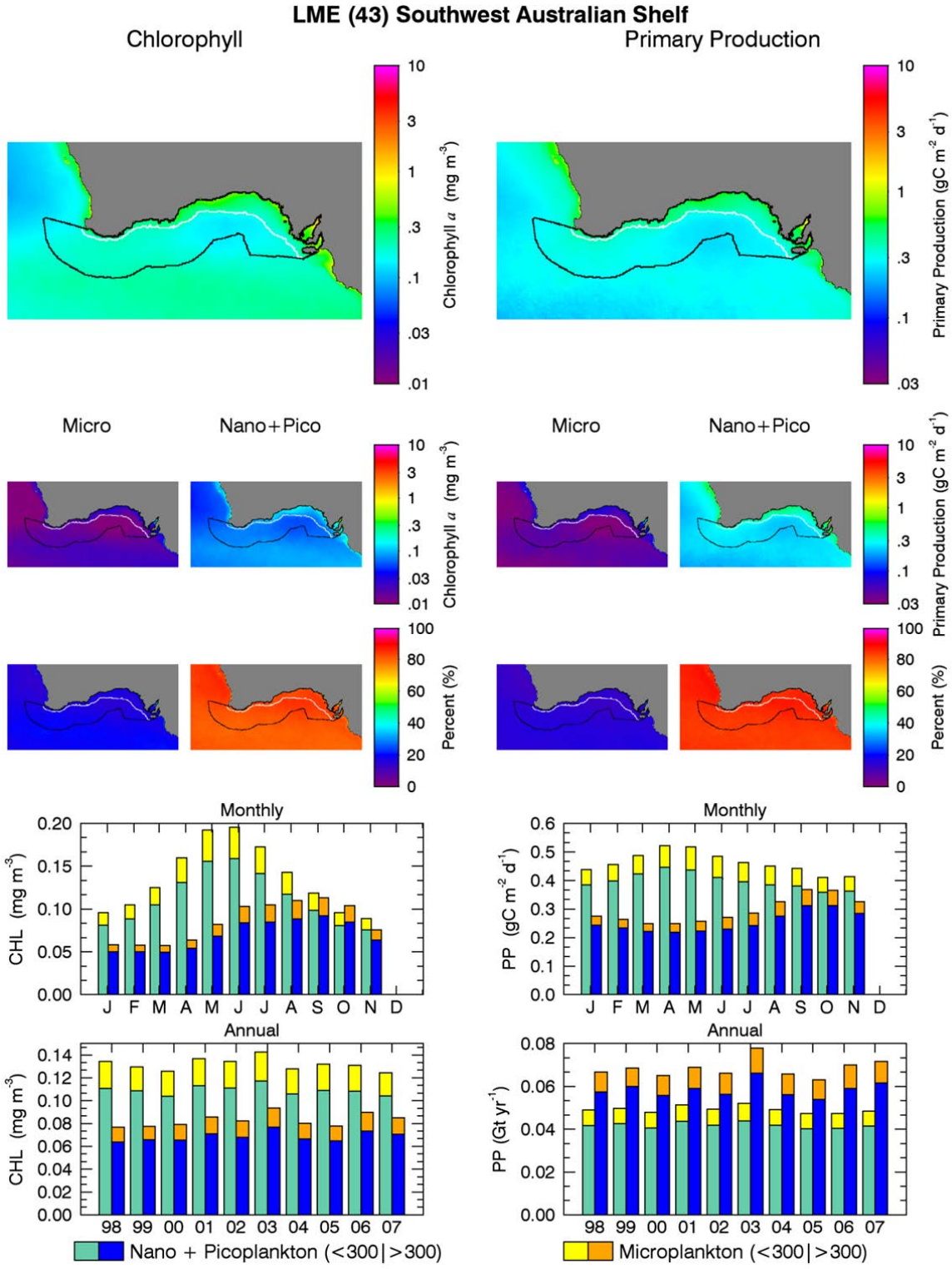


FIGURE A2.45
See legend of Figure A2.2

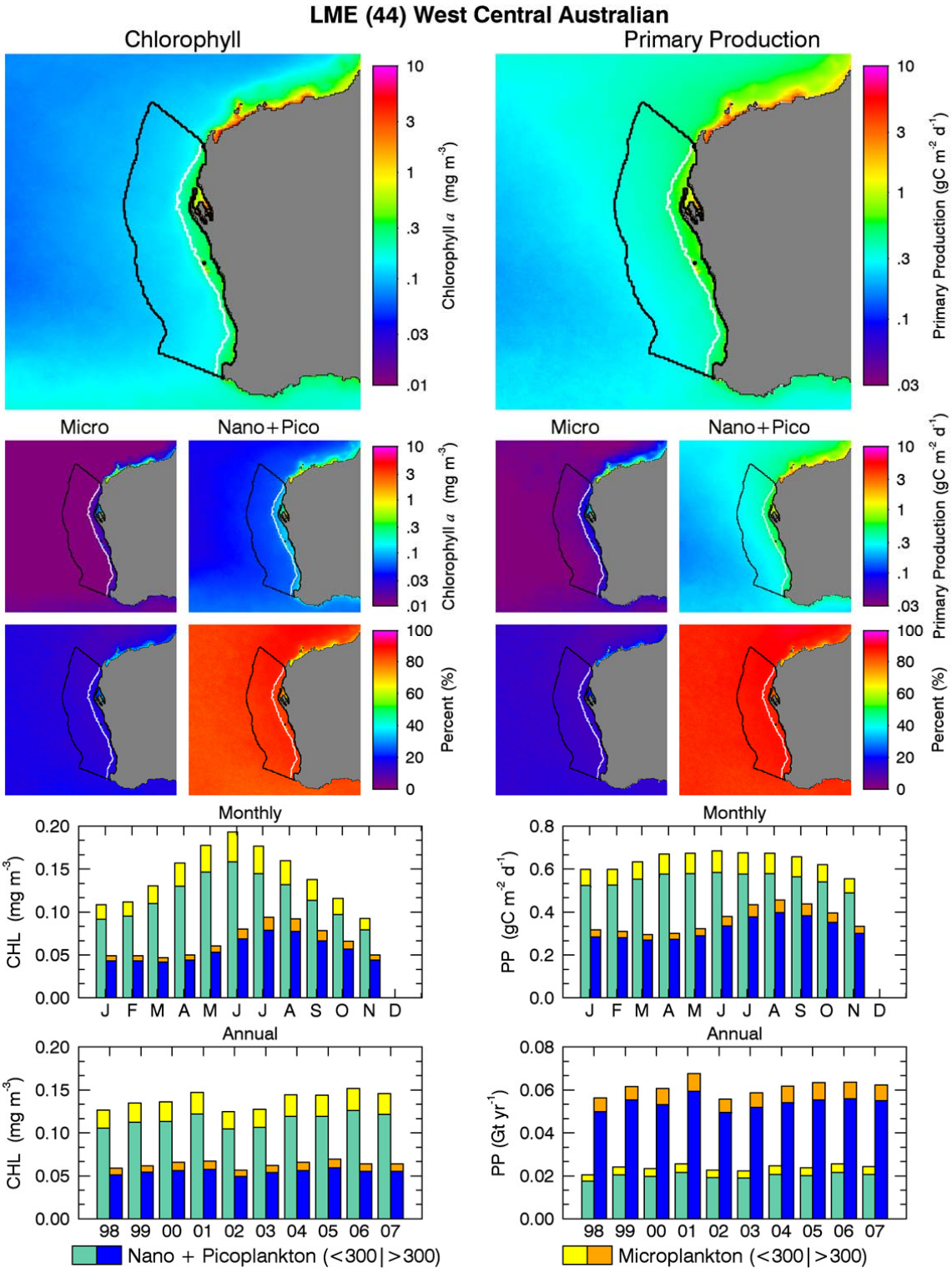


FIGURE A2.46
See legend of Figure A2.2

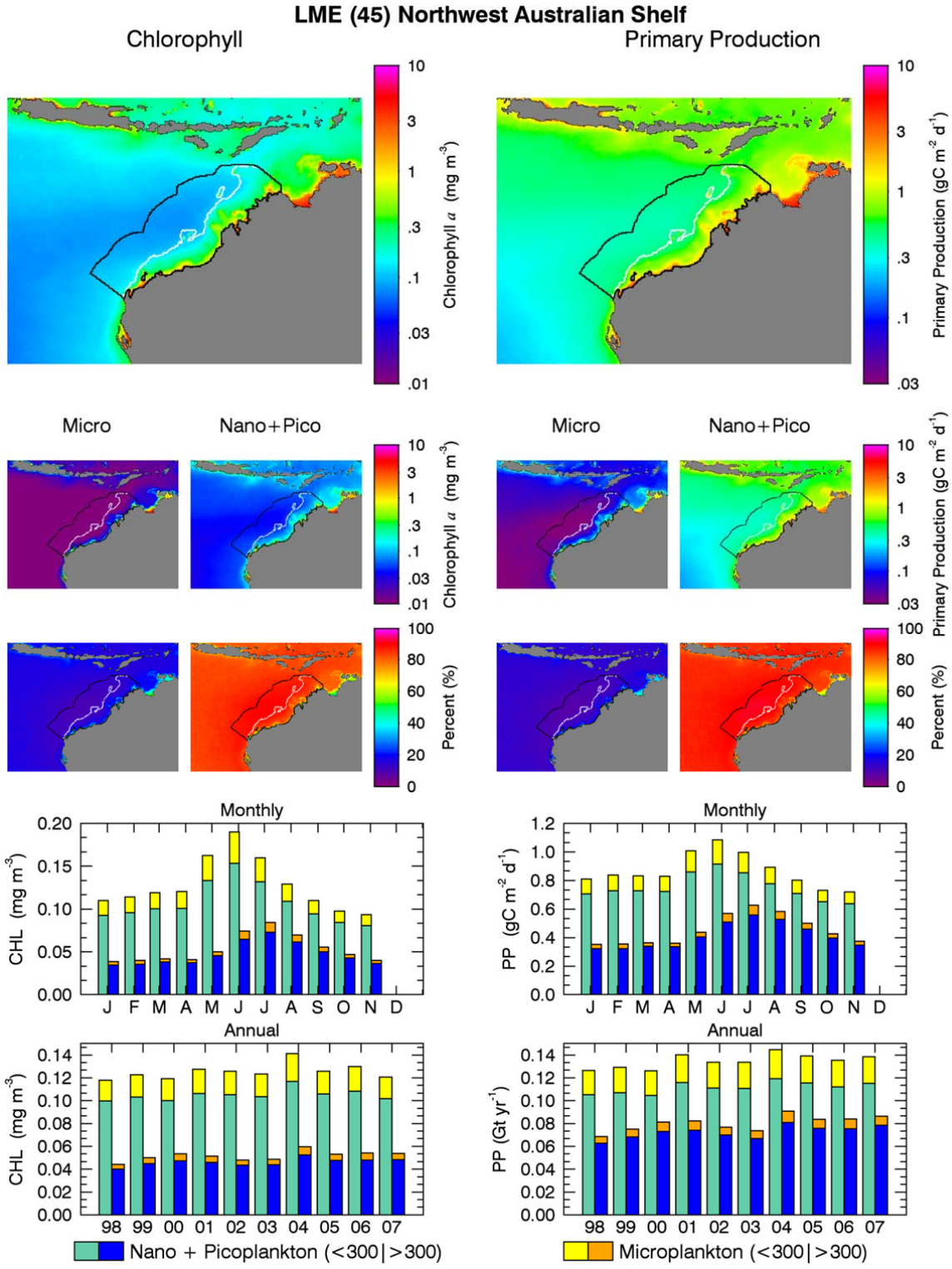


FIGURE A2.47
See legend of Figure A2.2

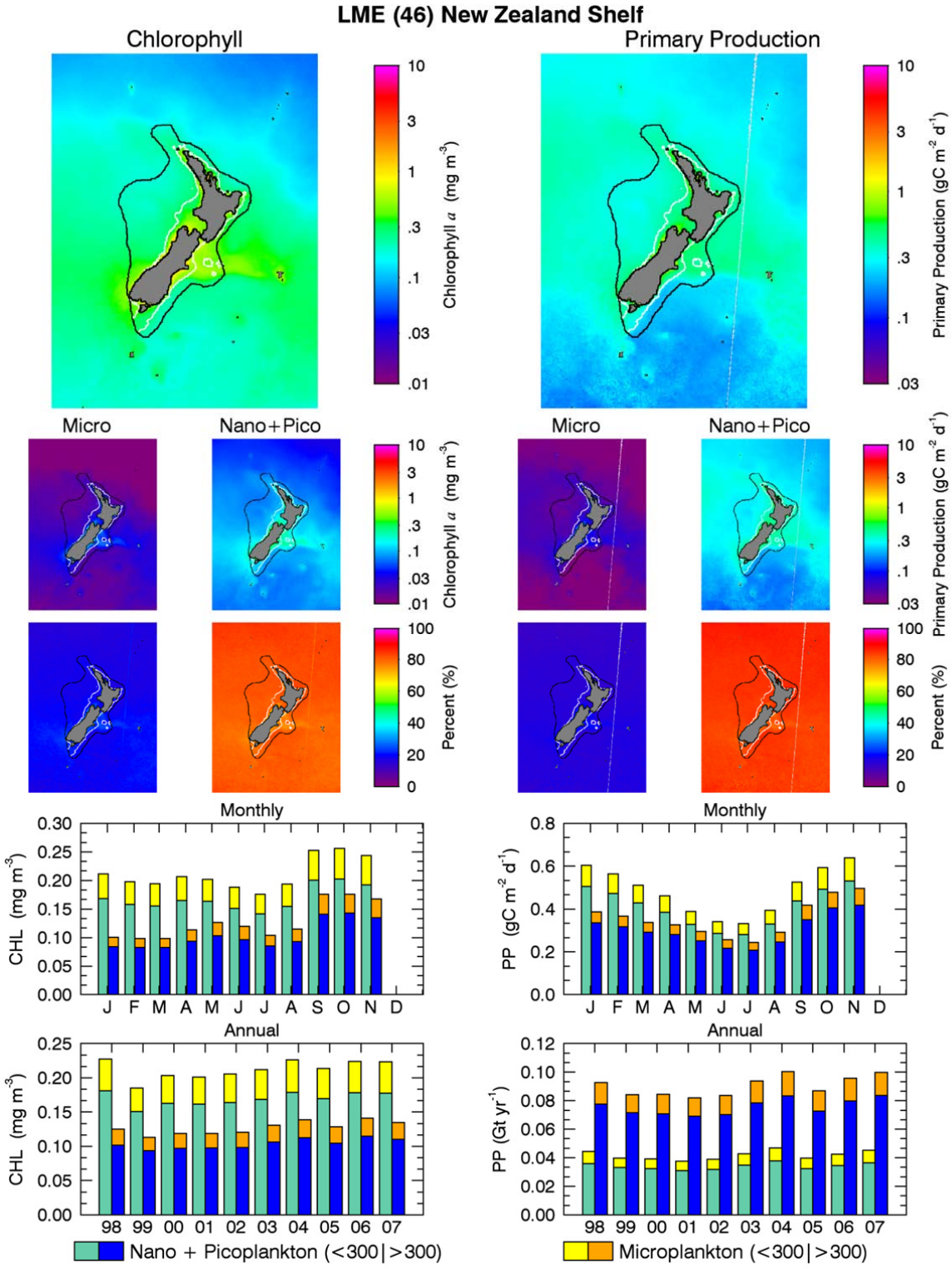


FIGURE A2.48
See legend of Figure A2.2

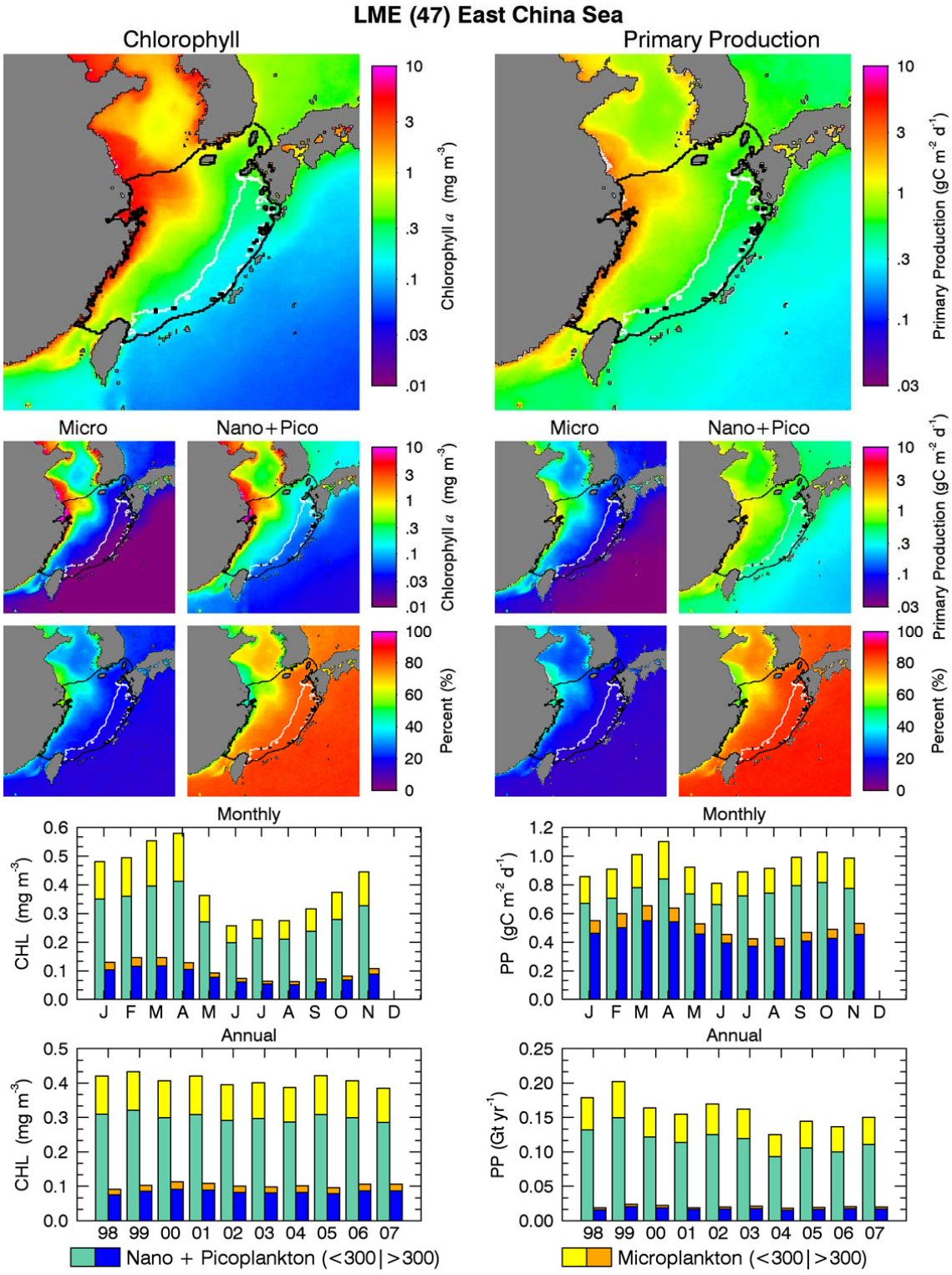


FIGURE A2.49
See legend of Figure A2.2

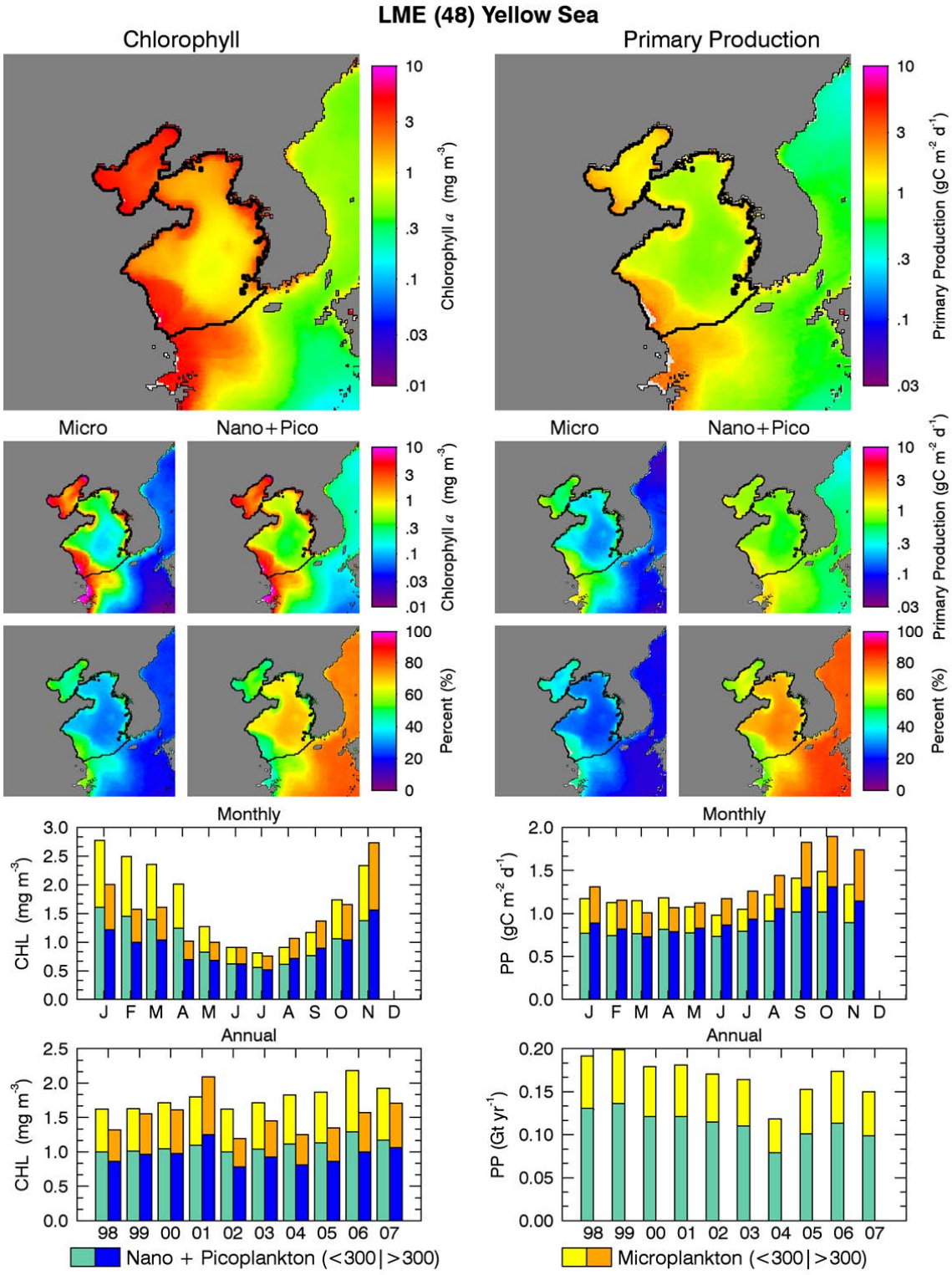


FIGURE A2.50
See legend of Figure A2.2

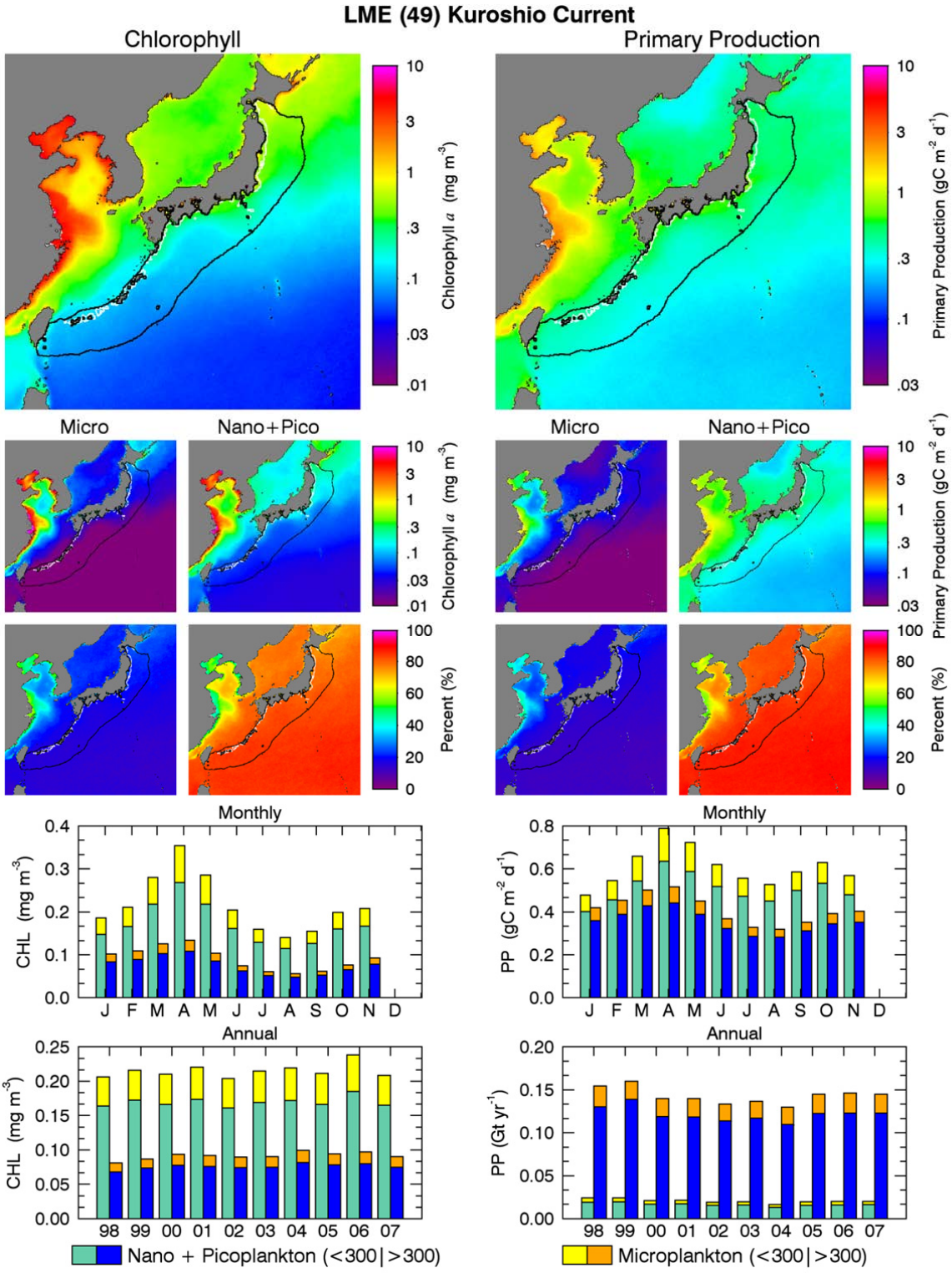


FIGURE A2.51
See legend of Figure A2.2

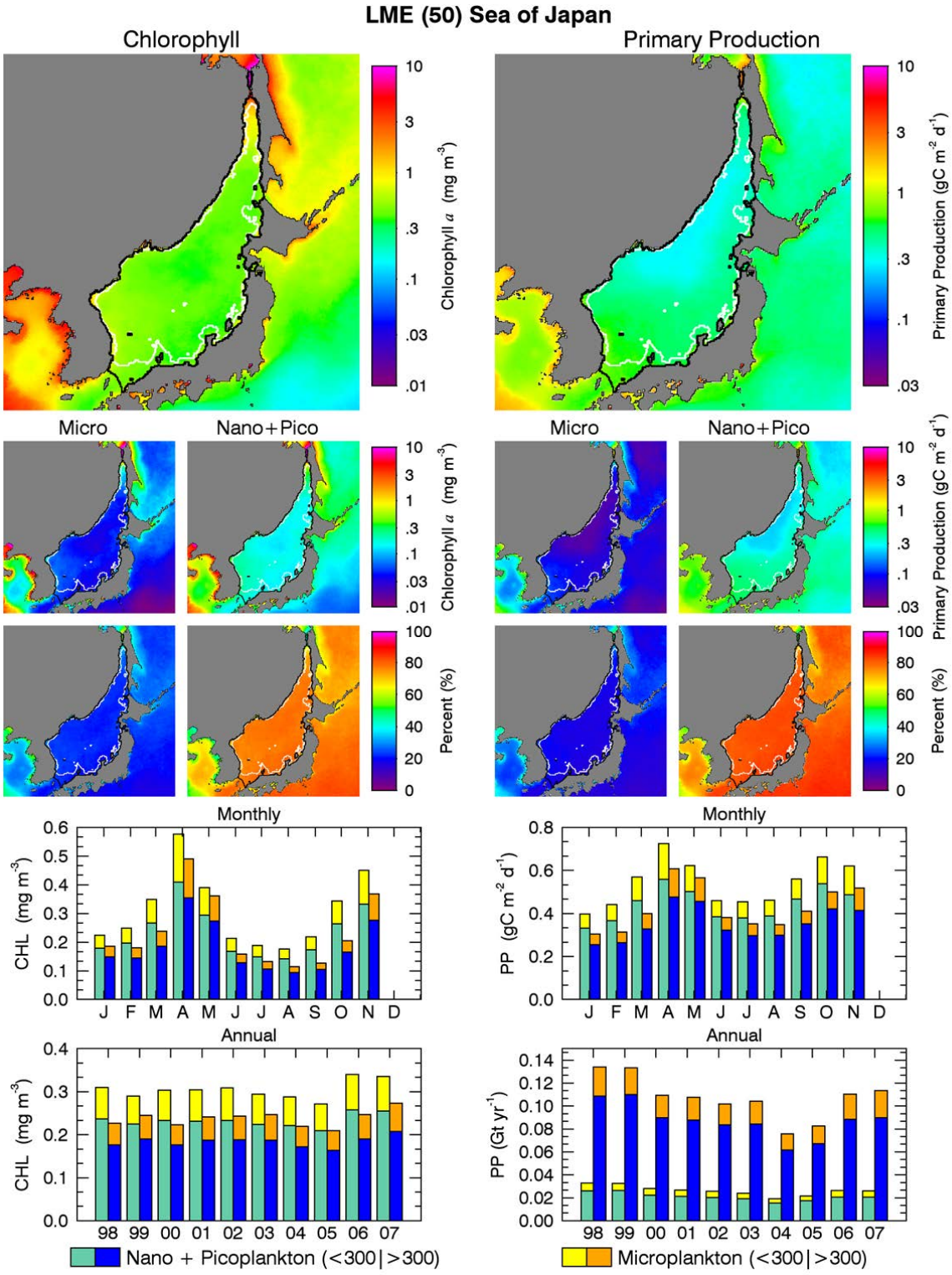


FIGURE A2.52
See legend of Figure A2.2

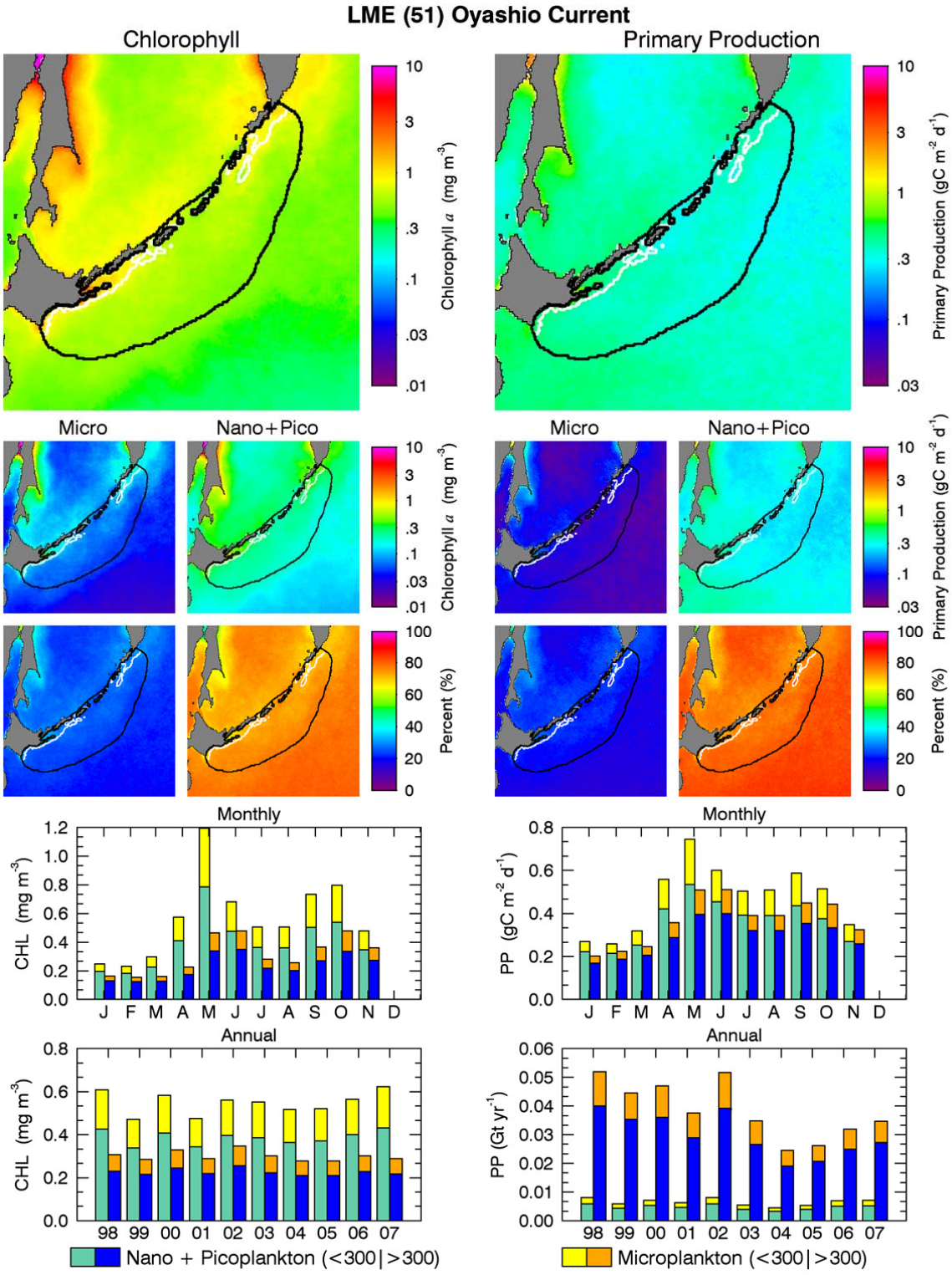


FIGURE A2.53
See legend of Figure A2.2

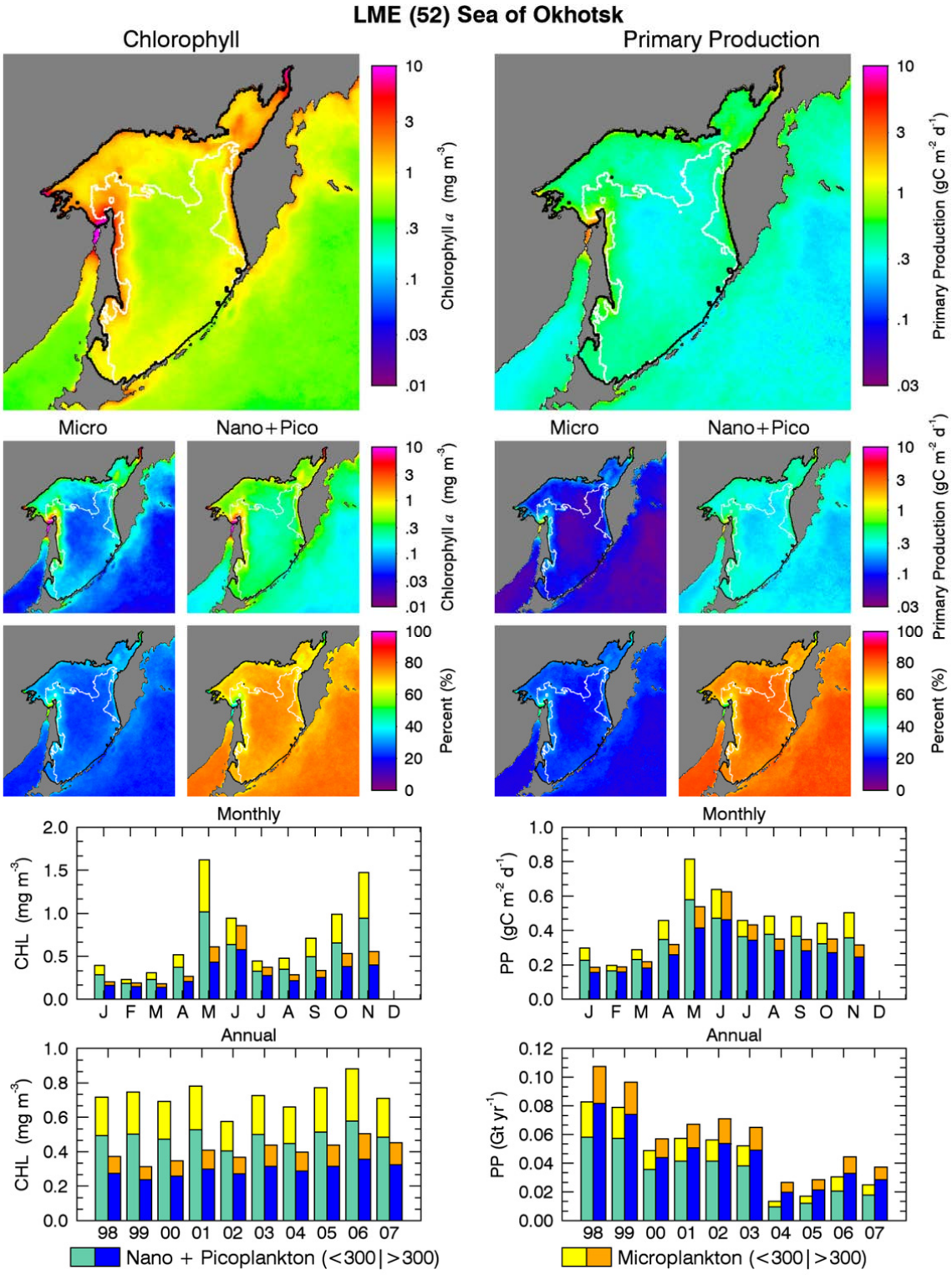


FIGURE A2.54
See legend of Figure A2.2

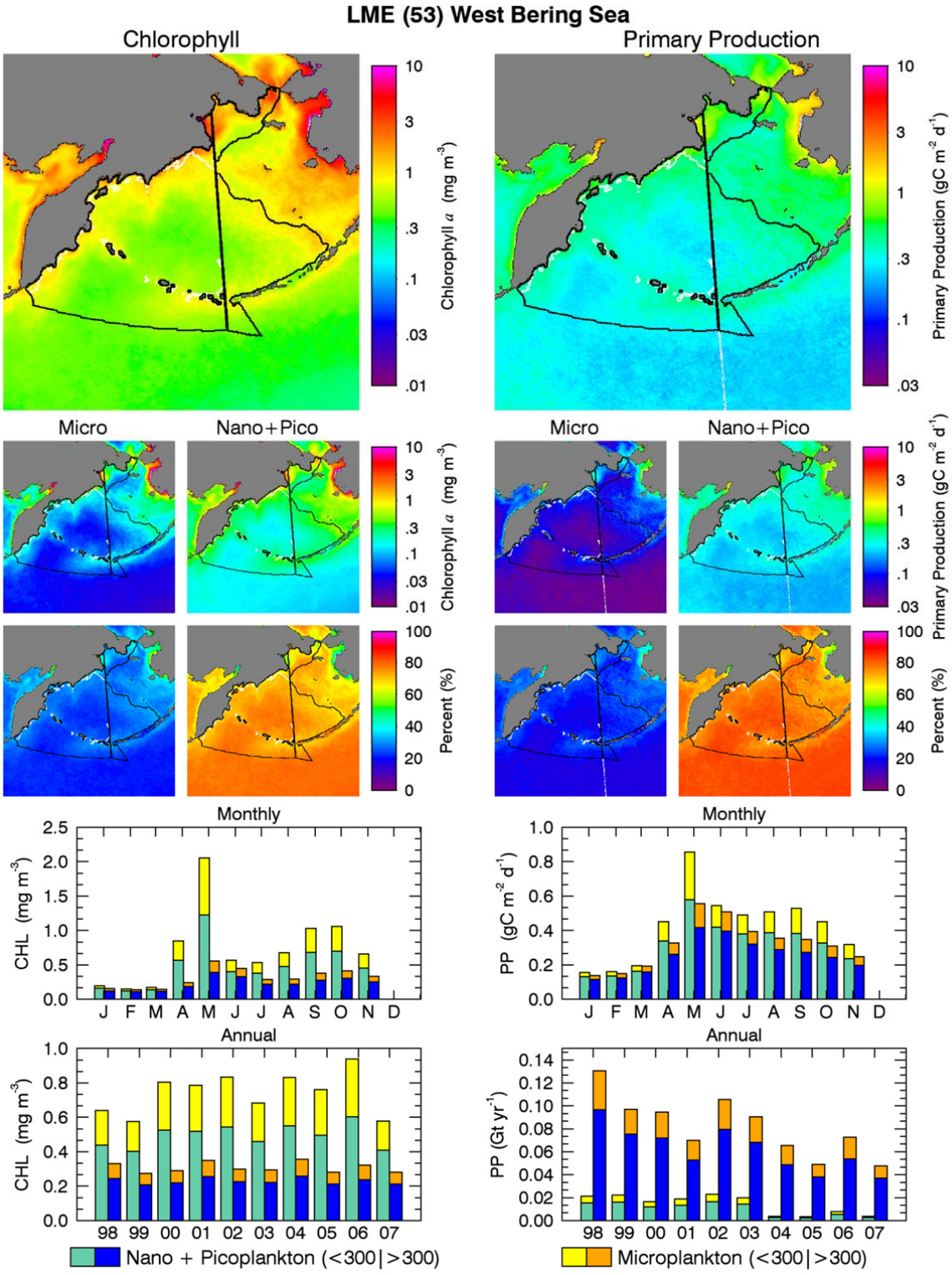


FIGURE A2.55
See legend of Figure A2.2

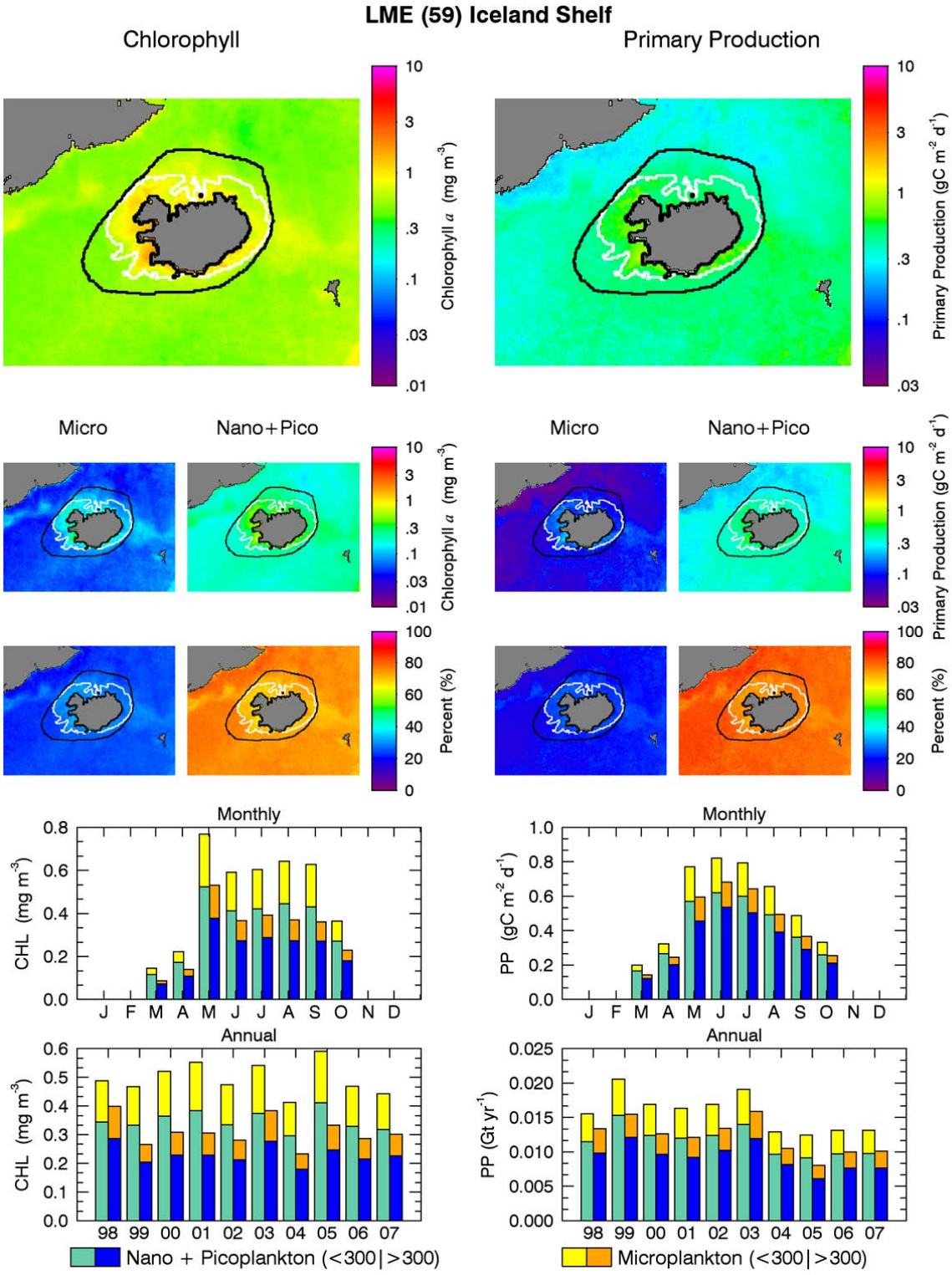


FIGURE A2.56
See legend of Figure A2.2

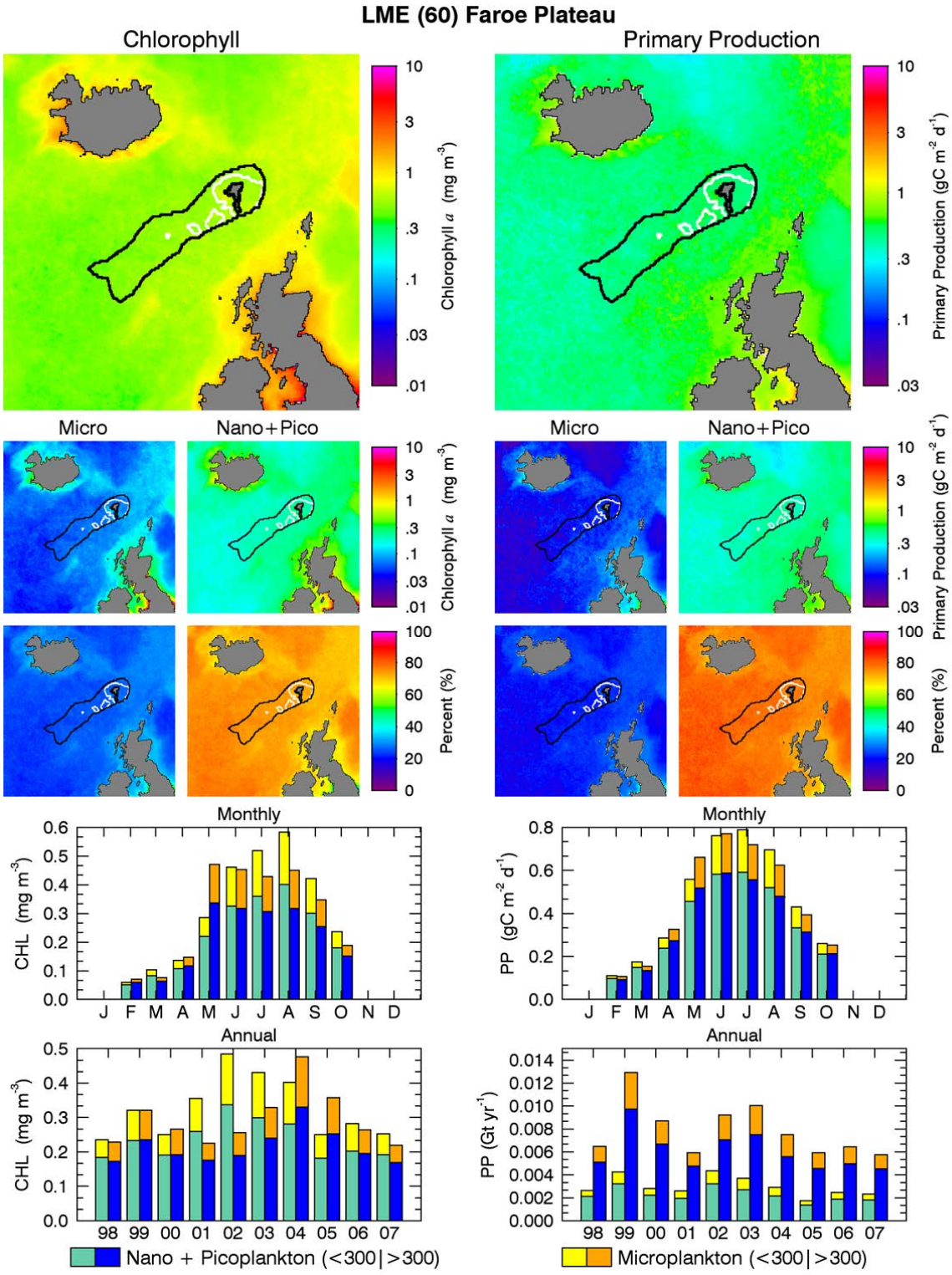


FIGURE A2.57
See legend of Figure A2.2

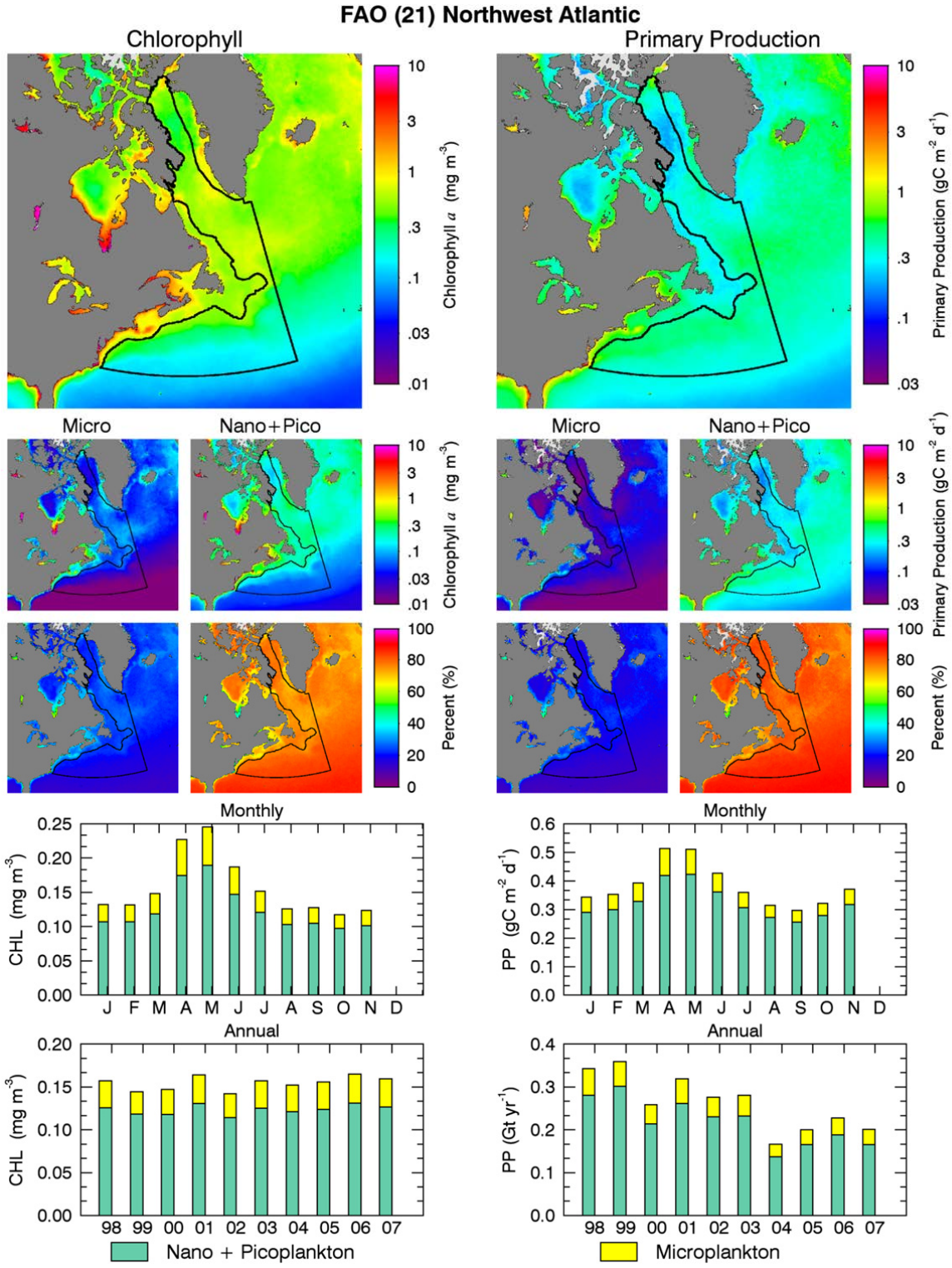


FIGURE A2.58
See legend of Figure A2.2

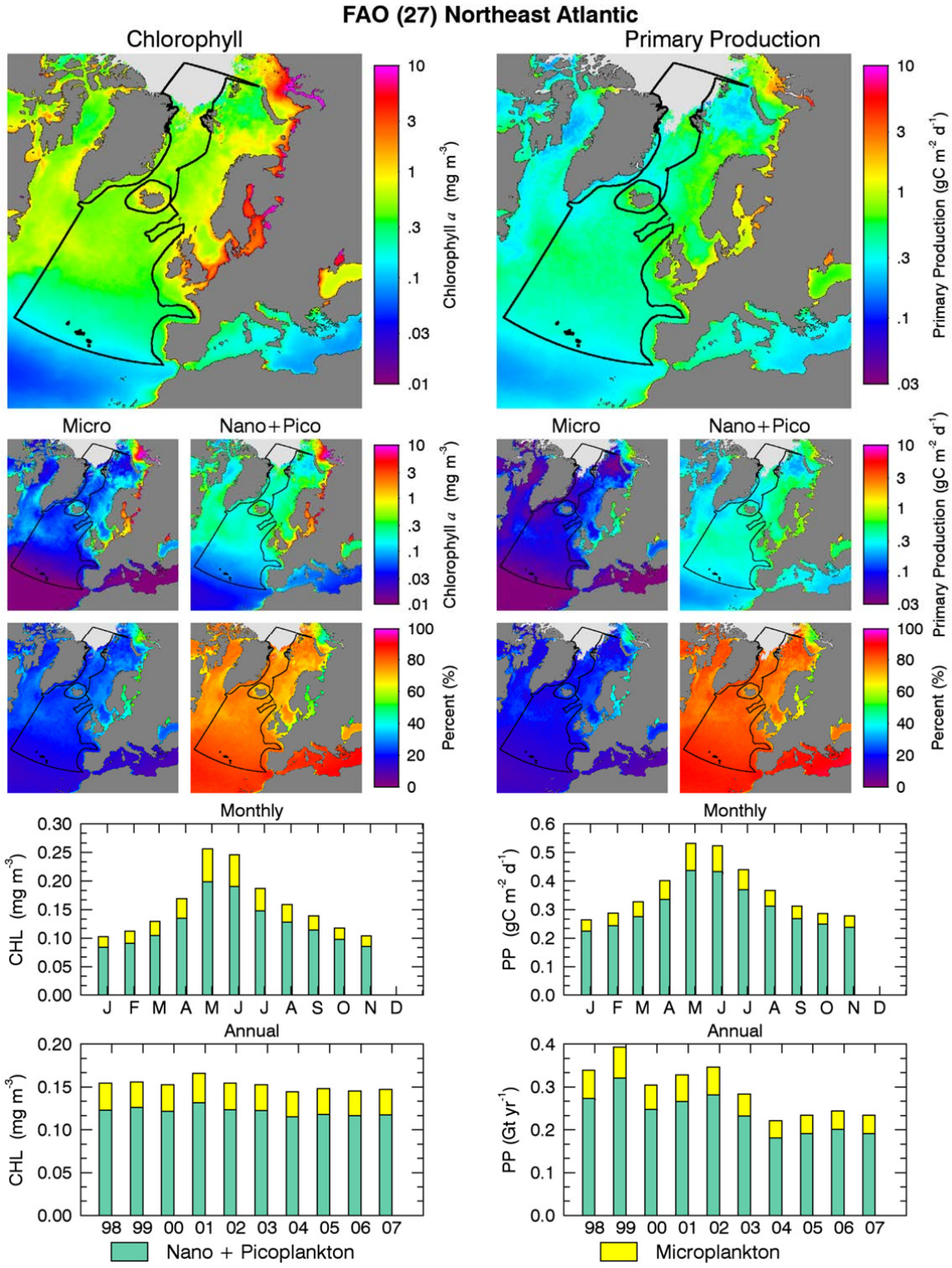


FIGURE A2.59
See legend of Figure A2.2

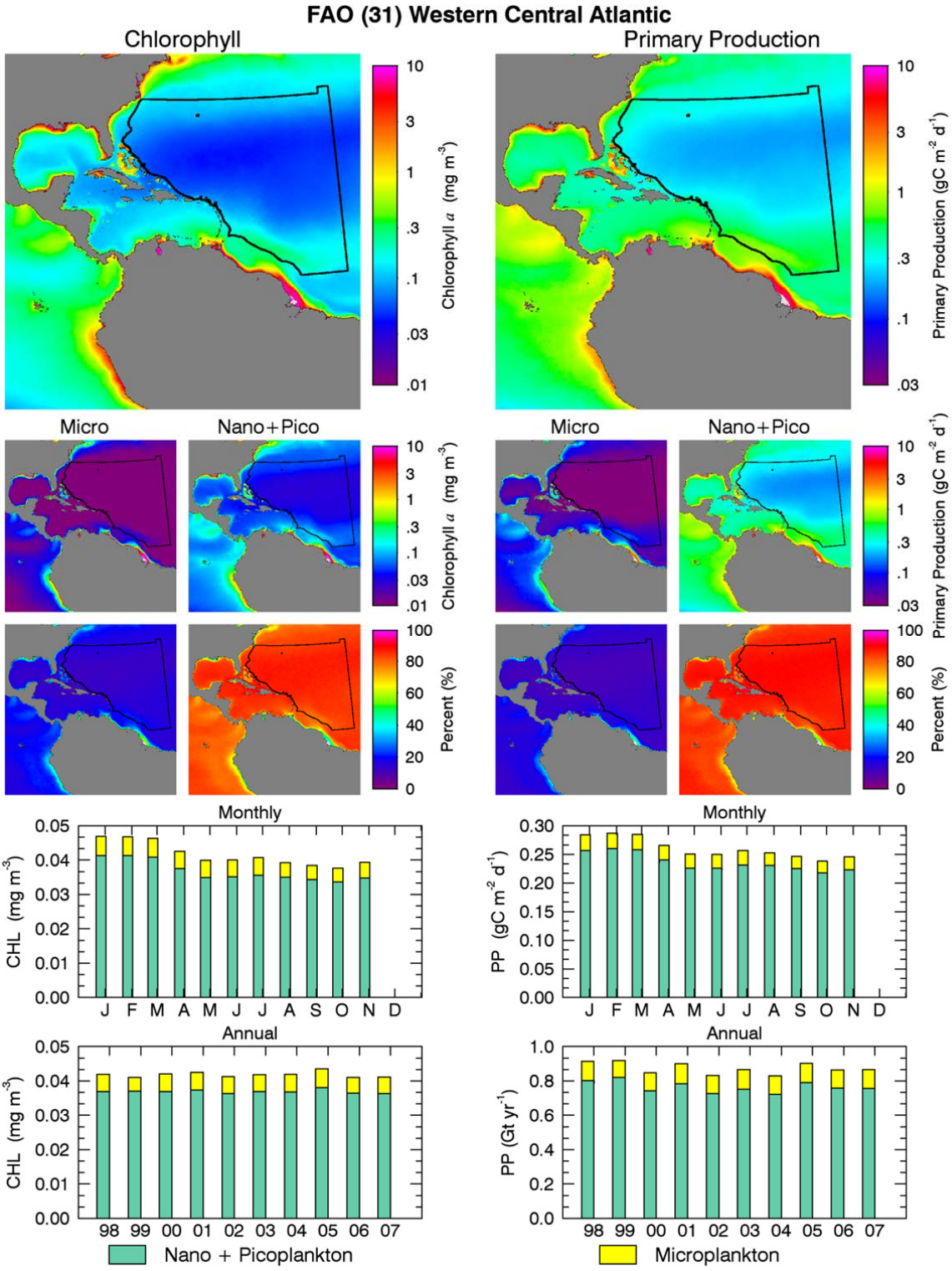


FIGURE A2.60
See legend of Figure A2.2

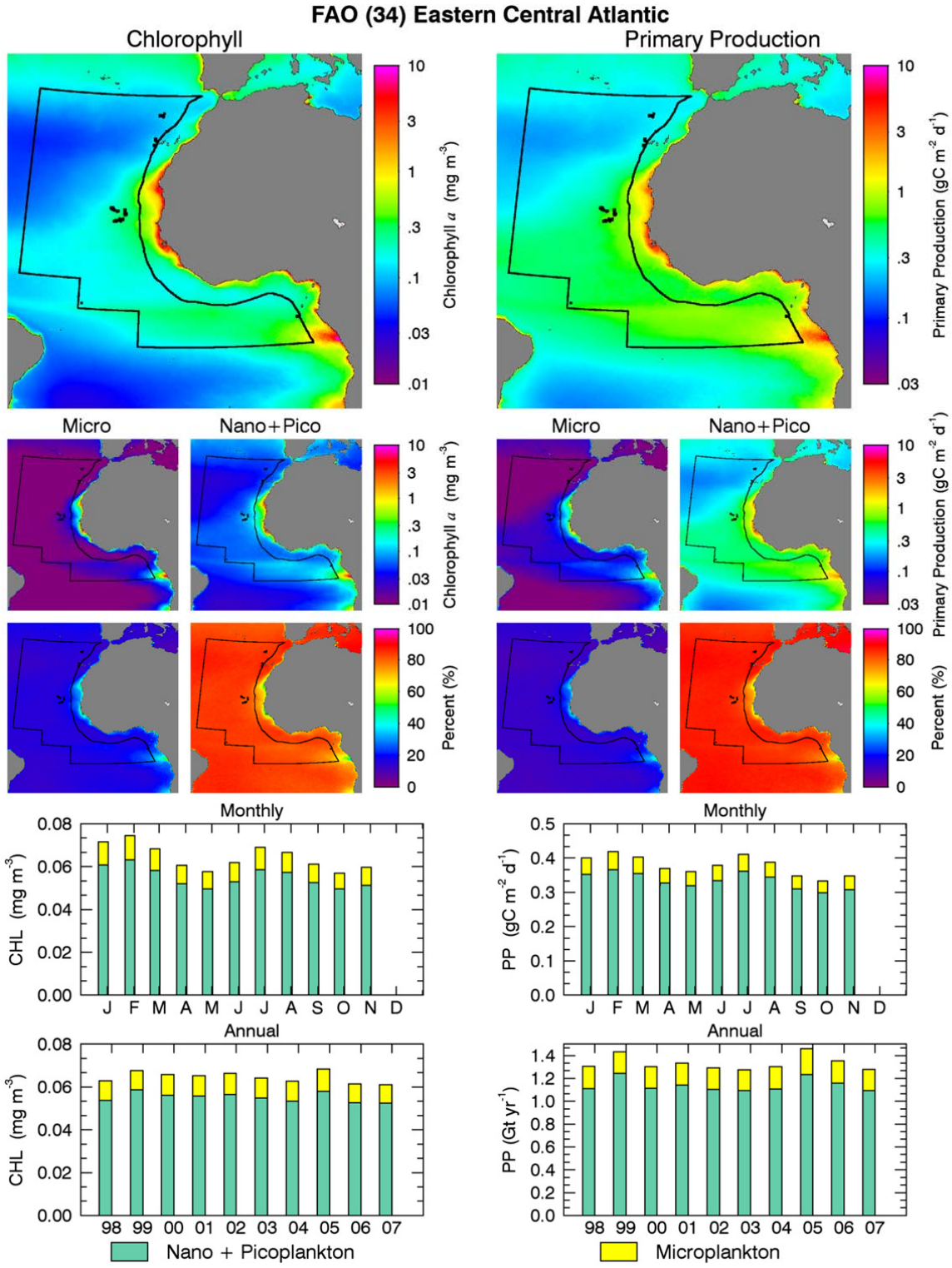


FIGURE A2.61
See legend of Figure A2.2

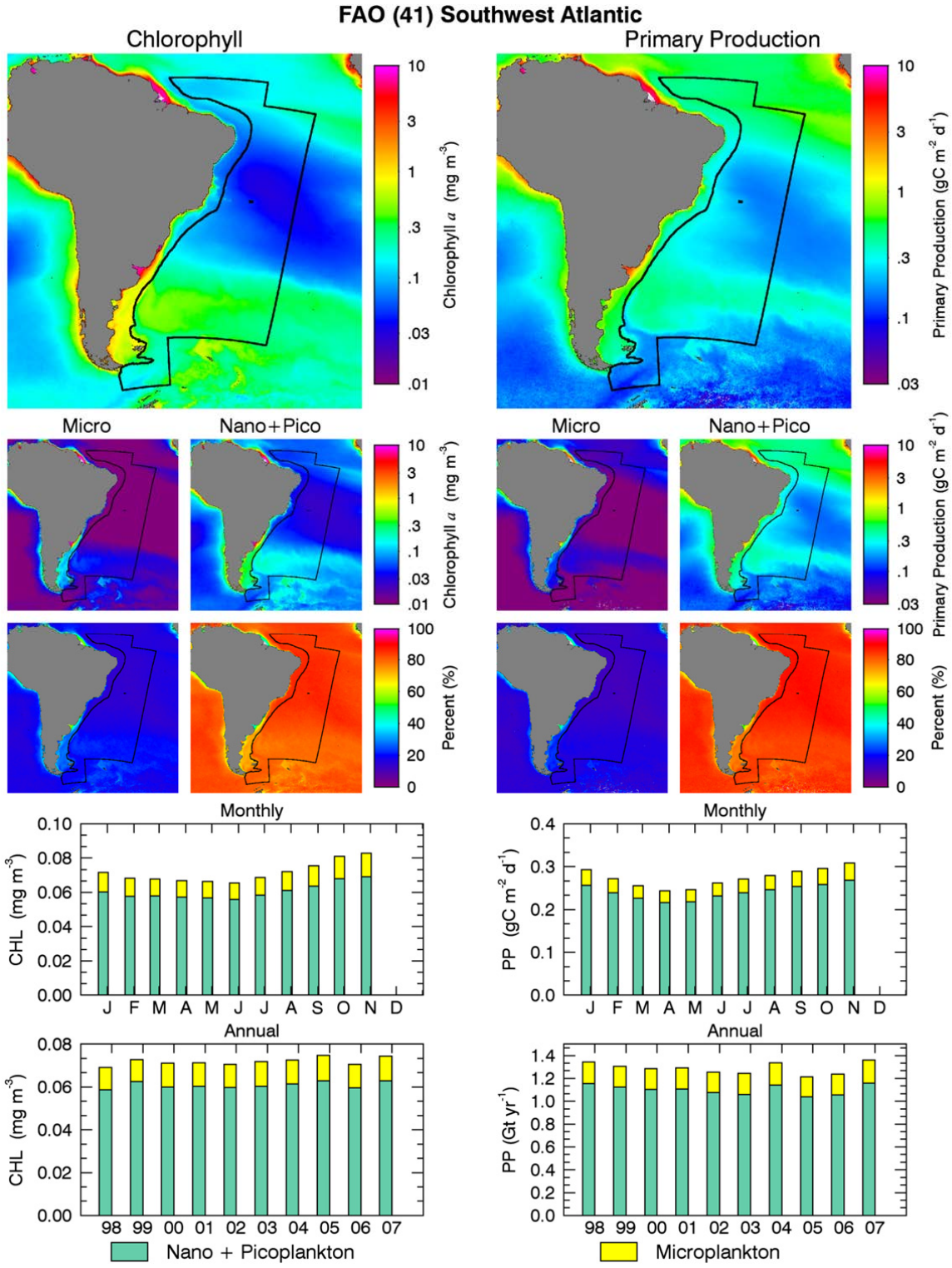


FIGURE A2.62
See legend of Figure A2.2

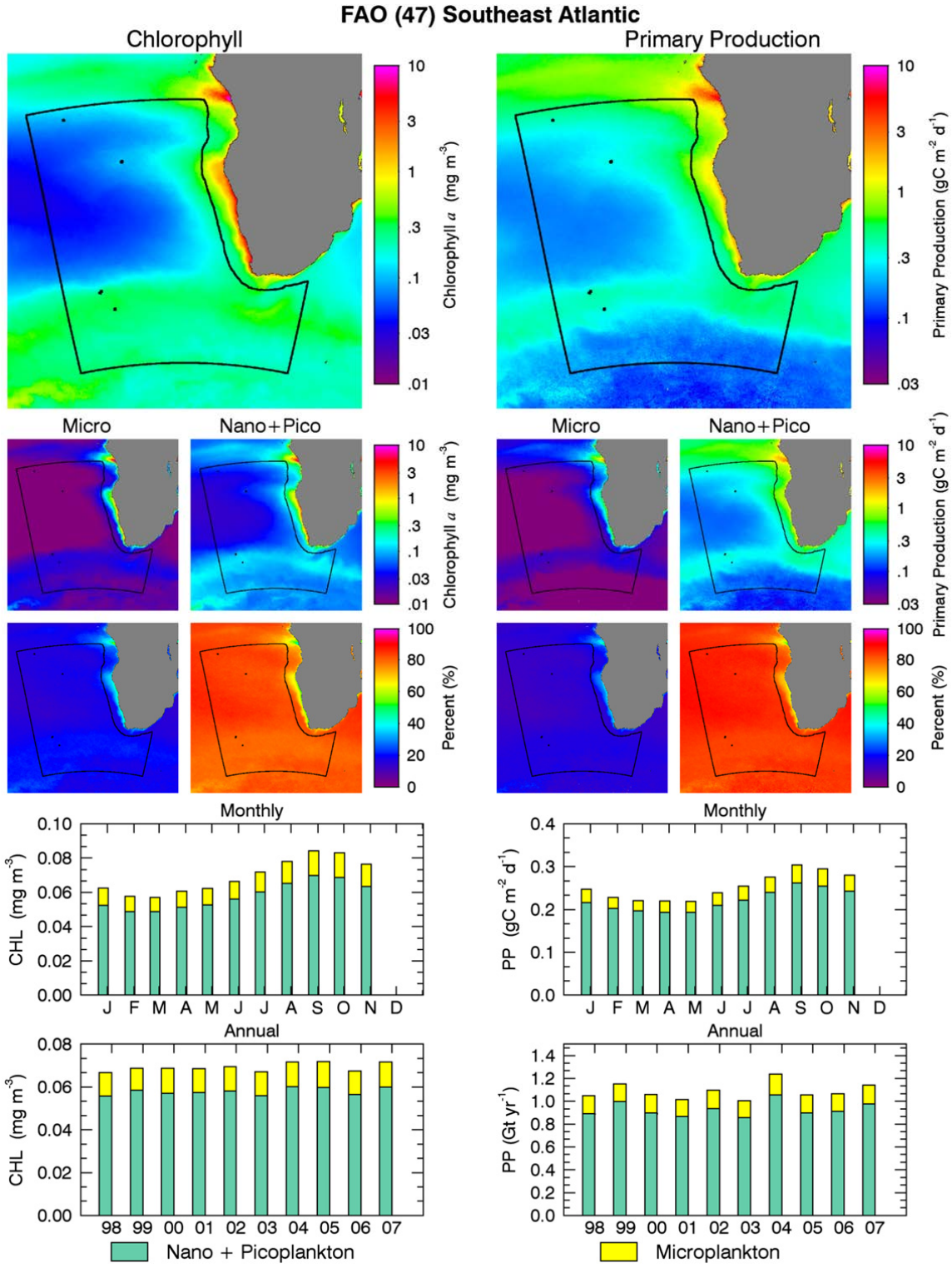


FIGURE A2.63
See legend of Figure A2.2

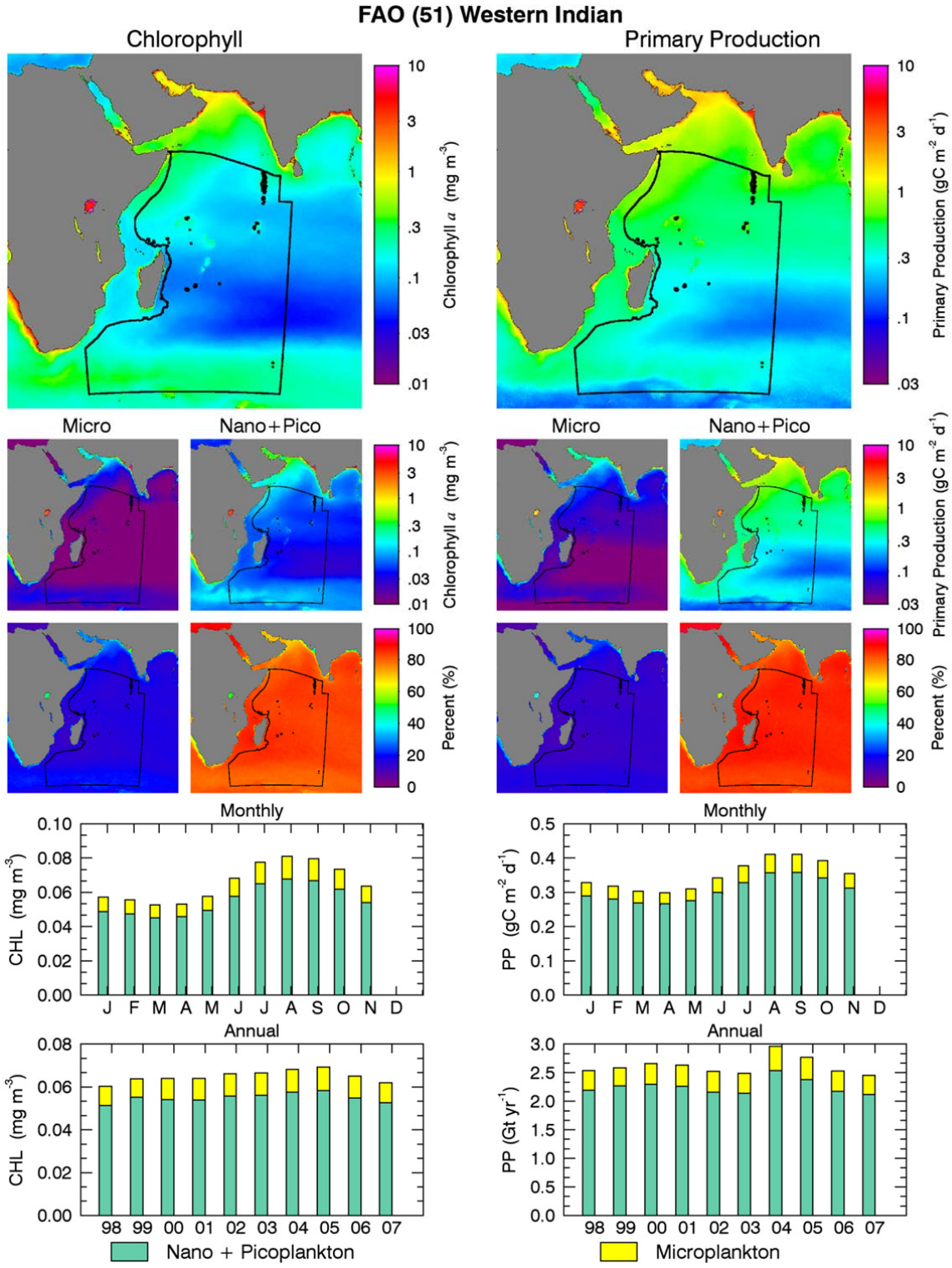


FIGURE A2.64
See legend of Figure A2.2

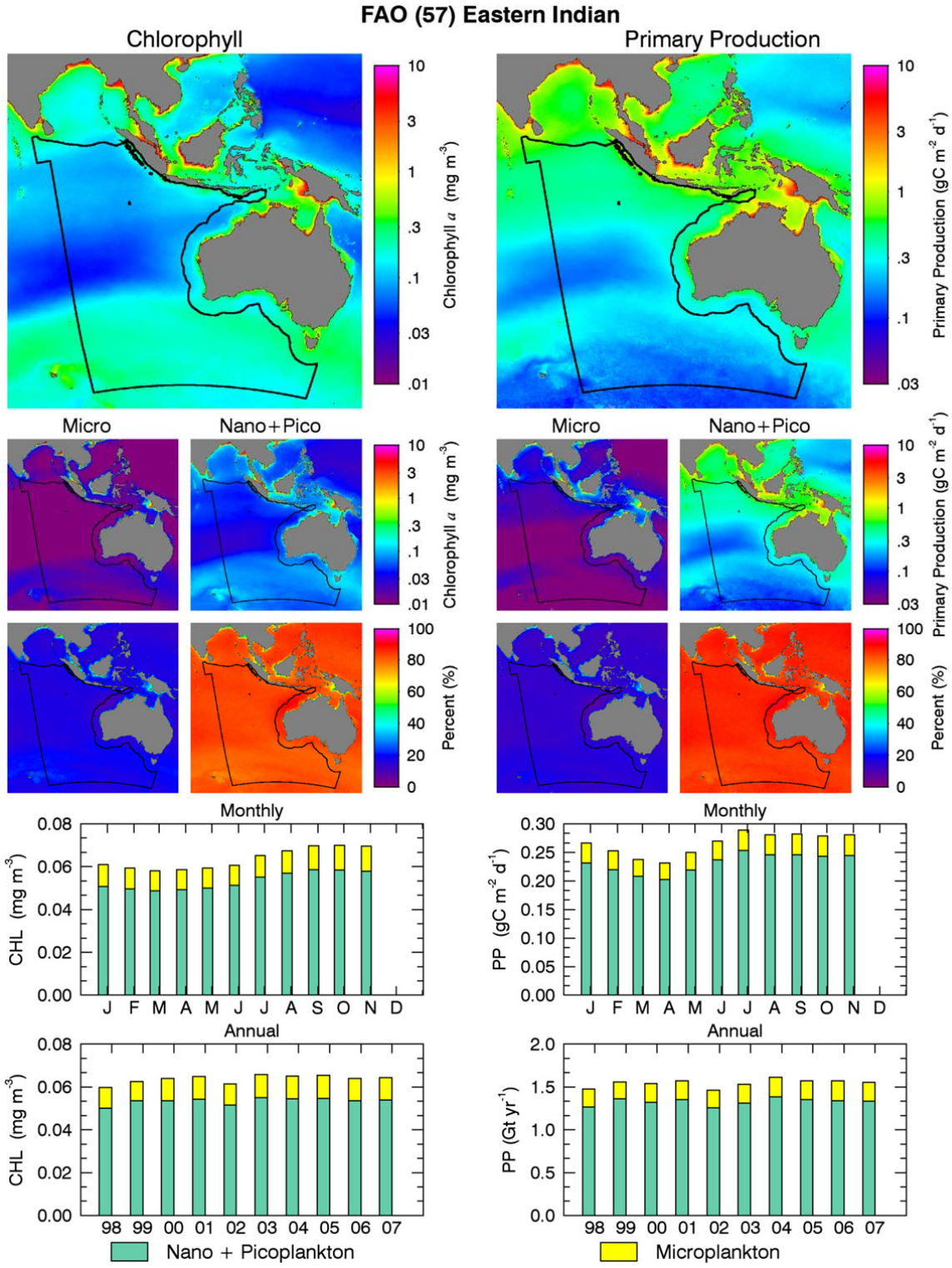


FIGURE A2.65
See legend of Figure A2.2

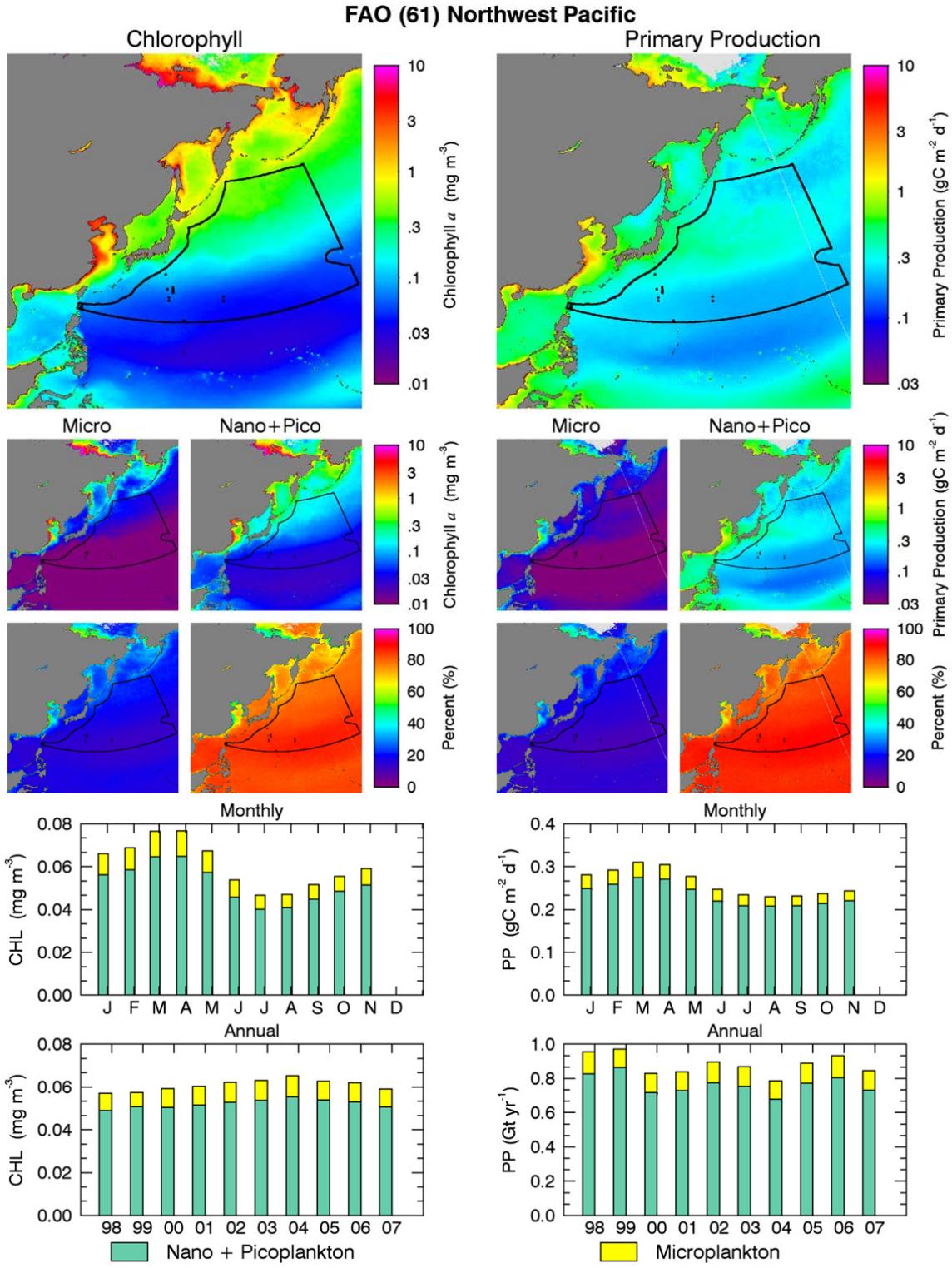


FIGURE A2.66
See legend of Figure A2.2

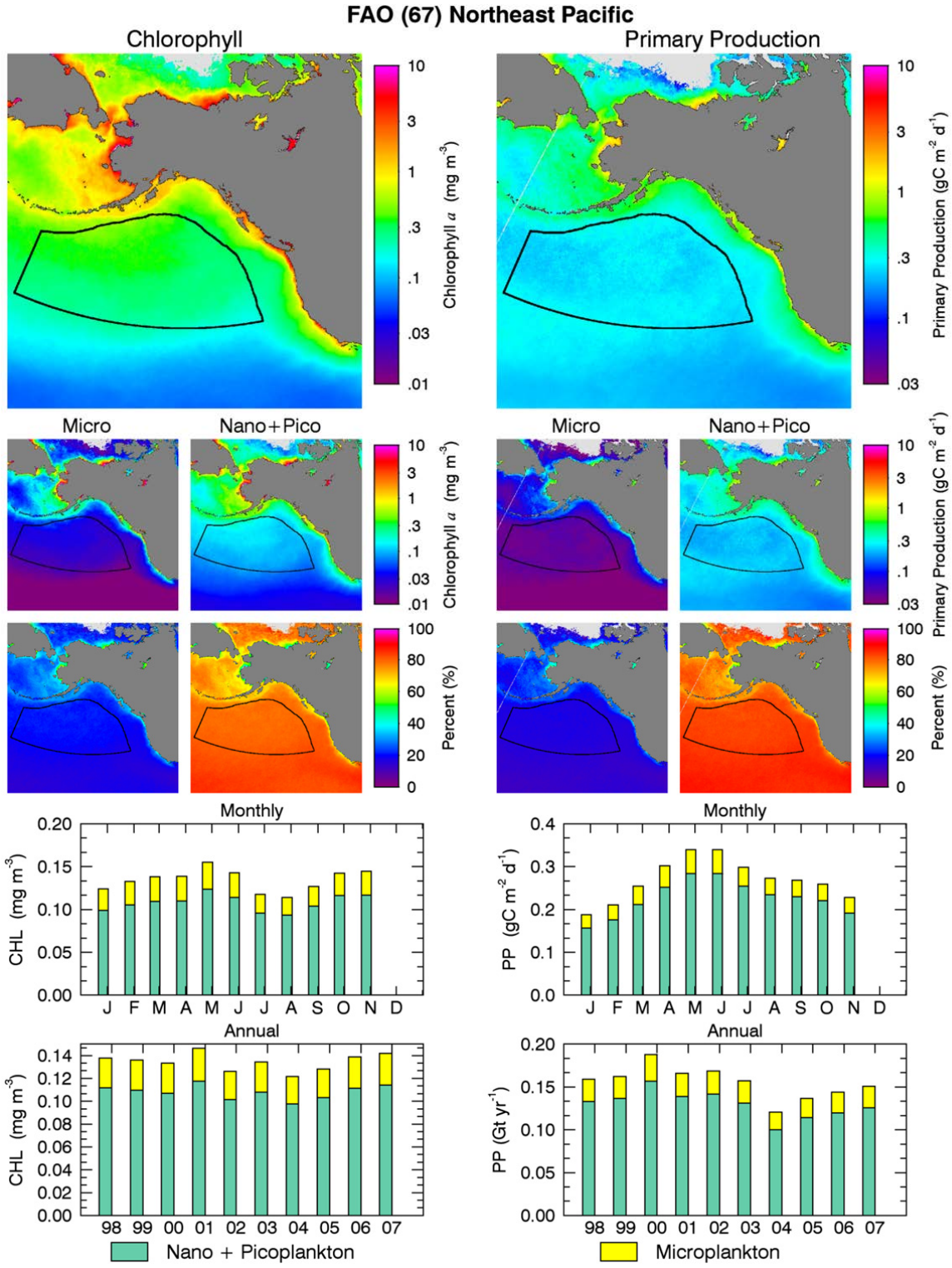


FIGURE A2.67
See legend of Figure A2.2

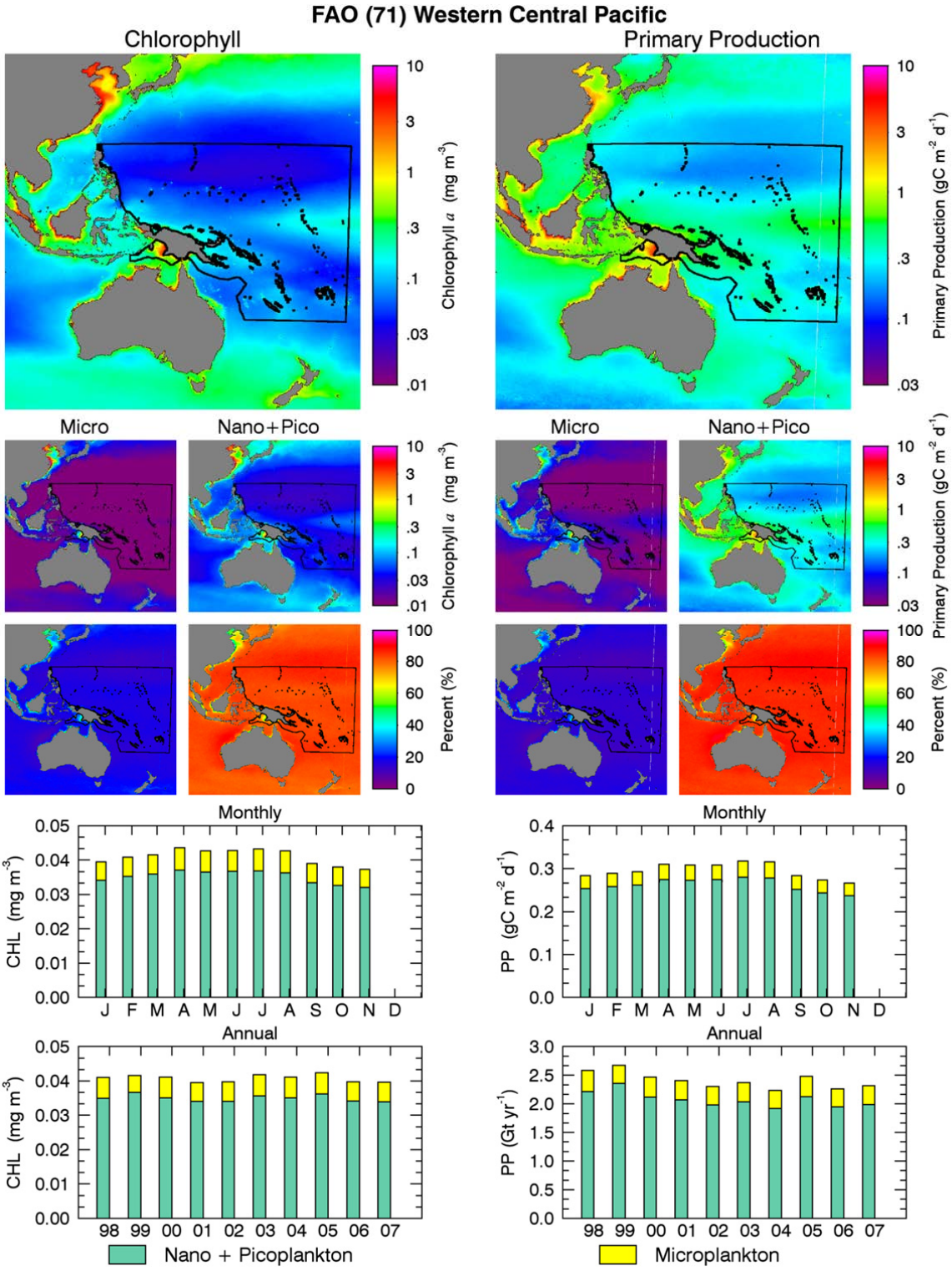


FIGURE A2.68
See legend of Figure A2.2

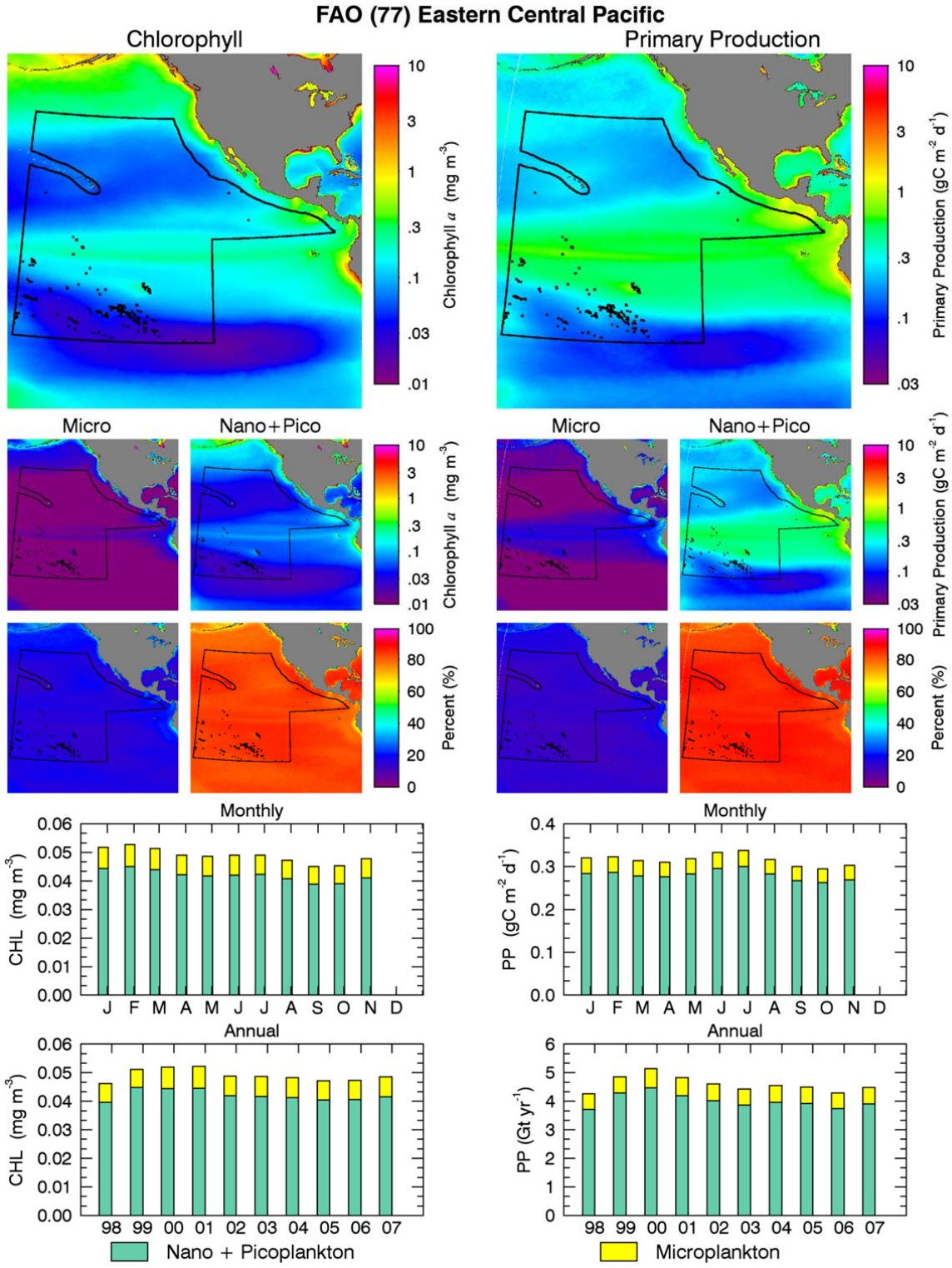


FIGURE A2.69
See legend of Figure A2.2

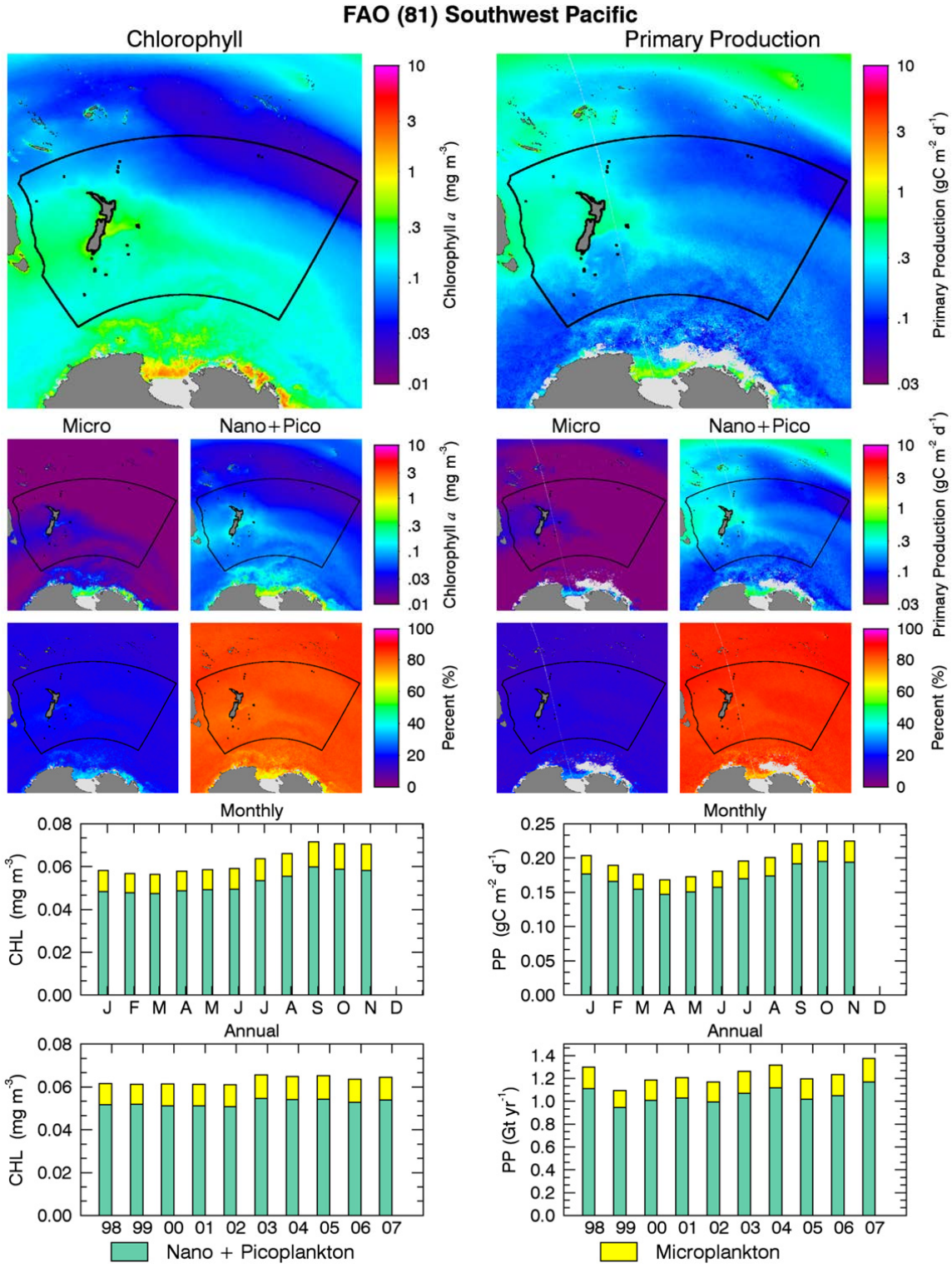


FIGURE A2.70
See legend of Figure A2.2