LETTER

Variability-based constraint on ocean primary production models

B. B. Cael *ID* *

National Oceanography Centre, Southampton, UK

Scientific Significance Statement

Primary production by plankton is essential to ocean ecology and biogeochemistry, so satellite models that estimate primary production from remote sensing data are indispensable in numerous scientific applications. However, because the corresponding in situ data are rarely measured when a satellite passes overhead, are measured on a much smaller spatial scale, and are highly variable, it is very difficult to compare different satellite models, evaluate how accurate they are, or to constrain their parameters. Here, a different approach to evaluating, comparing, and constraining these models is described, which accounts for or avoids all of these issues with the standard approach. This approach finds excellent agreement overall with a recent satellite model, while also identifying room for improvement.

Abstract

Primary production (PP) is fundamental to ocean biogeochemistry, but challengingly variable. Satellite models are unique tools for investigating PP, but are difficult to compare and validate because of the scale separation between in situ and remote measurements, which also are rarely coincident. Here, I argue that satellite estimates should be log-skew-normally distributed, because of this scale separation and because PP measurements are log-normally distributed. Whether they conform to this distributional shape is therefore a powerful variability-based constraint on such models. Satellite models that do follow a log-skew-normal may then also be concisely characterized by three parameters (log-mean, log-standard deviation, and log-skewness). I show that the output from a recent satellite model (CAFE) over 2019 agrees excellently with the log-skew-normal, globally and for most spatiotemporal subsets investigated here. The exception is the Northern Hemisphere winter, which may suggest future model improvements. PP by plankton is essential to ocean ecology and biogeochemistry, so satellite models that estimate PP from remote sensing data are indispensable in numerous scientific applications. However, because the corresponding in situ data are rarely measured when a satellite passes overhead, are measured on a much smaller spatial scale, and are highly variable, it is very difficult to compare different satellite models, evaluate how accurate they are, or to constrain their parameters. Here, a different approach to evaluating, comparing, and constraining these models is described, which accounts for or avoids all of these issues with the standard approach. This approach finds excellent agreement overall with a recent satellite model, while also identifying room for improvement.

*Correspondence: cael@noc.ac.uk

In the open ocean, primary production (PP) by phytoplankton forms the basis of the food web. As such, PP controls the flow of carbon, nutrients, and energy into ocean ecosystems. Phytoplankton are responsible for ~ 60 Pg C yr^{-1} of net PP (NPP = gross PP minus autotrophic respiration) (Buitenhuis et al. 2013) and therefore have an appreciable effect on global biogeochemical cycles and thereby climate, including by regulating a large carbon flux to the deep ocean via the biological pump. While total NPP is fairly well-constrained, understanding the variability and spatial and temporal patterns in NPP is a key oceanographic pursuit, with many applications and implications.

Satellite-based models of NPP that utilize remote sensing data to estimate NPP from space have revolutionized the study of ocean productivity thanks to their synoptic coverage. Such models have been refined and employed to myriad ends for decades (Behrenfeld and Falkowski 1997). Dozens of satellite NPP models have been proposed (Carr et al. 2006), whose structural differences evince the complexity of depth-integrated NPP and the challenge of estimating it from radiance. Improvements over the satellite ocean color era have been substantive (Silsbe et al. 2016) but there is still a great deal of room for improvement in terms of constructing, selecting, and tuning a NPP model that yields credible and accurate NPP patterns and variations.

The continued uncertainty in satellite-derived NPP is due in no small part to the challenge of comparing and validating NPP models with in situ measurements. While these measurements come with their own uncertainties (Marra 2009) and are biased in space and time, one should still in principle prefer the satellite model that corresponds best to in situ data. The problem is that there are few colocated in situ and satellite measurements for NPP. It is challenging enough to assemble a sufficient database of matched-up instantaneous concentration measurements, but making a depth-integrated rate measurement is a much more complex endeavor. This lack of in situ data has hampered model comparison, despite substantial efforts (Carr et al. 2006 and references therein). (Note here that while co-located in situ data are sparse, tens of thousands of NPP measurements have been made and assembled in the global ocean; Buitenhuis et al. 2013.) Where comparisons have been possible, such as at subtropical time series stations, those models that do capture the mean state do not capture the variability around that state (Siegel et al. 2001).

This last point has been used as a criticism of NPP models, but it is not necessarily reasonable to expect that satellite models reproduce the variability of in situ measurements, or even agree well pointwise, given that they consider very different spatial scales. In situ measurements are made from bottles with volumes on the order of liters, whereas satellites sample pixels that are several square kilometers large. Directly comparing these assumes that bottle measurements have a spatial footprint, that is, are representative, of a comparable size to a satellite pixel. Extensive recent research into submesoscale

(0.1–10 km) dynamics suggests the opposite. It has been argued instead that the large nutrient-transporting vertical velocities caused by motions at these spatial scales and on phytoplankton growth timescales are particularly important for phytoplankton productivity, creating localized productivity-enhancing environments and enormous small-scale heterogeneity (Mahadevan 2016). This may help explain why in situ NPP measurements are so variable (Cael et al. 2018), but also suggests that satellite NPP models integrating over these scales should have dampened variability and not match individual in situ NPP measurements exactly. It would be desirable to compare satellite NPP models and in situ NPP data in a way that (1) accounts for this scale separation, (2) does not penalize models for not matching individual in situ measurements, and ideally (3) leverages the many in situ measurements that have been taken at times and places other than when and where a satellite is passing overhead with a clear sky.

One way to achieve all of the above is to evaluate the probability distribution of outputs from a NPP model. As satellite models do not use space or time information explicitly, and as the probability distribution of a variable contains all of the non-spatiotemporal information about that variable, the strictest possible non-spatiotemporal criterion by which one can evaluate a satellite model is by evaluating the probability distributions it generates (Cael et al. 2018). If a NPP model is expected to generate output values that are distributed according to some probability density function p(NPP) over some region and interval, then the mismatch between p(NPP) and the actual model output's distribution is a powerful means of evaluating that model's performance. In other words, the fit of a log-skew-normal (LSN) distribution constitutes a strict test of satellite NPP models' variability. Departures of a given NPP model's distribution from LSN (at places and times where in situ NPP measurements are log-normal) are indicative of failings of that model to accurately represent the drivers of NPP. This does not require pairwise agreement between colocated in situ measurements and satellite pixels, addressing (2) and (3) above and the hypothesized distribution p(NPP) can incorporate information about the scale separation of satellite and in situ measurements, addressing (1). The purpose of this manuscript is to illustrate such an approach to evaluating NPP models.

Theory: LSN

In Cael et al. (2018), we showed that in situ NPP measurements were robustly well-described by a log-normal distribution, globally and over large regions in space and time. This observation builds upon the classic observation that chlorophyll measurements are also log-normally distributed (Campbell 1995). We hypothesized that this is because carbon fixation by phytoplankton occurs as a result of many distinct requirements (i.e., certain light, nutrient, temperature, absence-

of-predation, etc., conditions must be met). One might naively assume that satellite model NPP should therefore be log-normal as well—but as I will explain below, scale separation changes the expected distributional form. Figure 1 shows very clearly that the NPP model used here, CAFE (Silsbe et al. 2016), is indeed on the whole far from log-normal: a log-normal variable has zero log-skewness (the skewness, i.e., third standardized moment, of the logarithm of that variable, for which I use the symbol θ) whereas CAFE's NPP estimates range from positively log-skewed with up to $\theta \sim 1$ to negatively log-skewed with as low as $\theta \sim -1$. (I return to these spatial patterns in θ below.)

What distributional form should we expect for NPP models? If in situ measurements are log-normally distributed, and satellite pixels measure spatial scales larger than the footprint of in situ measurements, then satellite NPP should be an average of log-normals. In other words, if a single satellite pixel is N times larger than the characteristic footprint of a log-normally distributed in situ NPP measurement, then a satellite pixel is measuring the average of log-normal variables, that is, $\text{NPP}_{\text{sat}} = \frac{1}{N}\sum_{i=1}^{N}\text{NPP}_{\text{in situ}}$ with each $\text{NPP}_{\text{in situ}}$ being log-normal. Satellite NPP therefore should be proportional to the sum of log-normal variables. Note that this does not require N to be uniform in space or time, only that $N > 1$. (Also, here I focus on horizontal scale separation, but equivalently there is a vertical scale separation by virtue of the fact that satellite NPP models estimate depth-integrated NPP.)

Sums of log-normals arise in wireless communications problems when one is interested in total interference. In this case, it has been shown that sums of log-normal random variables, even correlated ones and even for small values of N, are excellently approximated by a LSN distribution (Hcine and Bouallegue 2015):

$$p(\text{NPP};\xi,\omega,\alpha) = \frac{2}{\omega\text{NPP}}\phi\left(\frac{\ln\text{NPP}-\xi}{\omega}\right)\varphi\left(\alpha\frac{\ln\text{NPP}-\xi}{\omega}\right)$$

where ln is the natural logarithm, p is a probability density function, and ϕ and φ respectively are the probability and cumulative density functions of a standard Gaussian random

variable. This is a generalization of the log-normal distribution where the relationship between the log-skewness θ of the distribution is confined to the range $(-1, 1)$. The location, scale, and shape parameters (ξ, ω, α) are related to the log-mean, log-standard deviation, and log-skewness (μ, σ, θ) of a LSN random variable according to the equations:

$$\mu = \xi + \sqrt{\tfrac{2}{\pi}}\omega\delta, \text{ where } \delta = \frac{\alpha}{\sqrt{1+\alpha^2}}, \; \sigma = \omega\sqrt{1-\tfrac{2}{\pi}\delta^2}, \text{ and } \theta = \frac{(4-\pi)\left(\sqrt{\tfrac{2}{\pi}}\delta\right)^3}{2\left(1-\tfrac{2}{\pi}\delta^2\right)^{3/2}}.$$

The log-skewness θ (or equivalently the shape parameter α; Fig. 2 illustrates how the shape of the distribution changes with θ) is influenced by the number of log-normals being summed over, and hence the spatial footprint of NPP measurements in this context, as well as the correlation between these summed log-normals (Hcine and Bouallegue 2015). The problem here is analogous to that in the communications literature: if in situ NPP measurements are log-normal, then NPP models using data from large satellite pixels should be LSN. This becomes a powerful test of these models, but also a concise way of characterizing the variability in NPP over space and time in terms of only three parameters.

Here, I demonstrate the application of the LSN to satellite NPP model output, using daily 4 km data from 2019 for the recent Carbon, Absorption, and Fluorescence Euphotic-resolving (CAFE) model (Silsbe et al. 2016). Overall CAFE matches the LSN distribution extremely well, and also does robustly over various subsets in space and/or time. The one exception to this is the Northern Hemisphere winter (NHW), which is caused by a heavy tail of low productivity values. The results here are intended to be illustrative—there are numerous NPP models that could be compared in this way, and a challenging inverse problem of relating the LSN parameters to those of the underlying log-normal distributions that needs addressing before the estimated LSN parameters can be leveraged. These results nonetheless give strong support to the CAFE model's ability to resolve a sensible distribution for NPP variability over much of the ocean, and indicate a potential
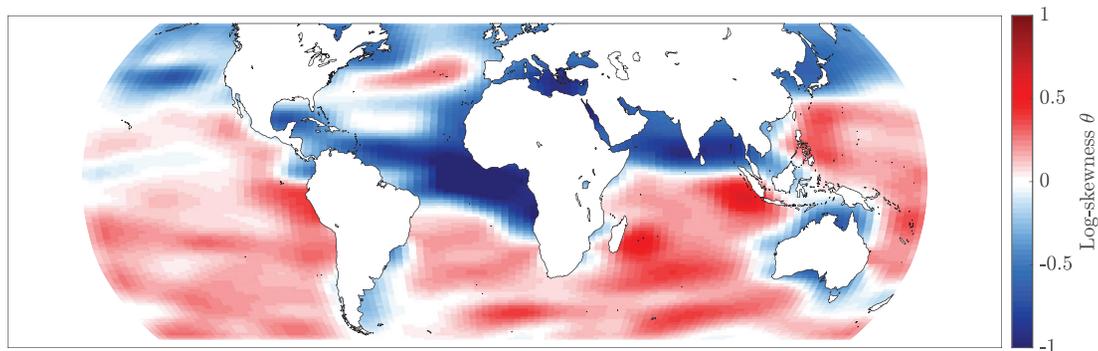
Fig. 1. Log-skewness θ of CAFE model. Output used is daily 4 km resolution over 2019, estimated from 100,000 randomly chosen measurements from $5° \times 15°$ boxes, and interpolated.
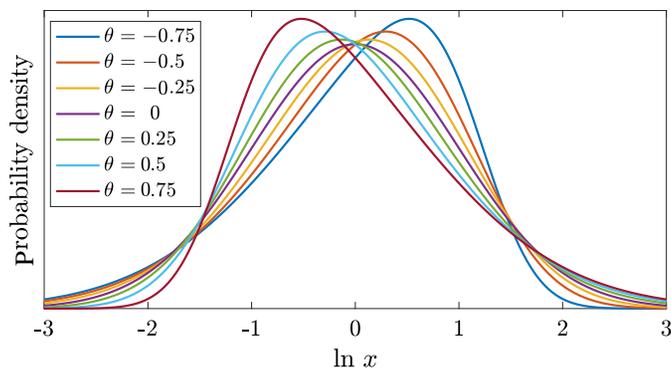
Fig. 2. Different log-skew normal distributions for a variable *x* with log-mean 0, log-standard deviation 1, and varying log-skewnesses θ. The limit case of $\theta = \pm1$ is the half-log-normal: just the right or left half of the $\theta = 0$ case.

direction for model improvement, while also evincing the utility of the LSN for characterizing NPP variability on kilometer spatial scales.

Methods

The CAFE model is a mechanistic carbon-based approach to NPP that uses inherent optical properties derived from ocean color measurements and improves upon previous approaches by addressing key physiological and ecological attributes of phytoplankton. CAFE has been shown to explain the greatest variance and have the lowest bias out of a suite of 22 NPP models (Silsbe et al. 2016). Here, I analyze daily output at a nominal 4 km spatial resolution from CAFE over the course of 2019, based on MODIS data inputs (https://doi.org/10.7910/DVN/DQ6K7R). (N.B. These MODIS inputs are kriged to fill gaps in measurements before the CAFE model is run on its input data; this could potentially confound the probability distribution p(NPP) if the complete dataset was used, but in this case as I analyze sparse random samples of the model field, it is less of a concern. Regardless, the LSN test is agnostic of how a given NPP field is generated.) To test for robustness of the LSN in space and/or time, I also analyze both the full-year global dataset, as well as five spatiotemporal subsets: (1) Northern and Southern Hemisphere seasons; (2) eight evenly spaced days throughout the year (days 45, 90, …, 360); (3) five large $40° \times 40°$ boxes in the North Atlantic (60–20°W, 10–50°N), South Atlantic (35°W–5°E, 0–40°S), North Pacific (160°E–160°W, 5–45°N), South Pacific (140–180°W, 15–55°S), and Indian (55–95°E, 0–40°S) oceans; (4) five small 5° (latitude) $\times$ 15° (longitude – size corresponding to the sampling boxes used to generate Fig. 1) boxes at the center of these larger boxes; and (5) a space-and-time subsetting case, of just the North Atlantic $40° \times 40°$ box in Spring (20 March to 21 June 2019).
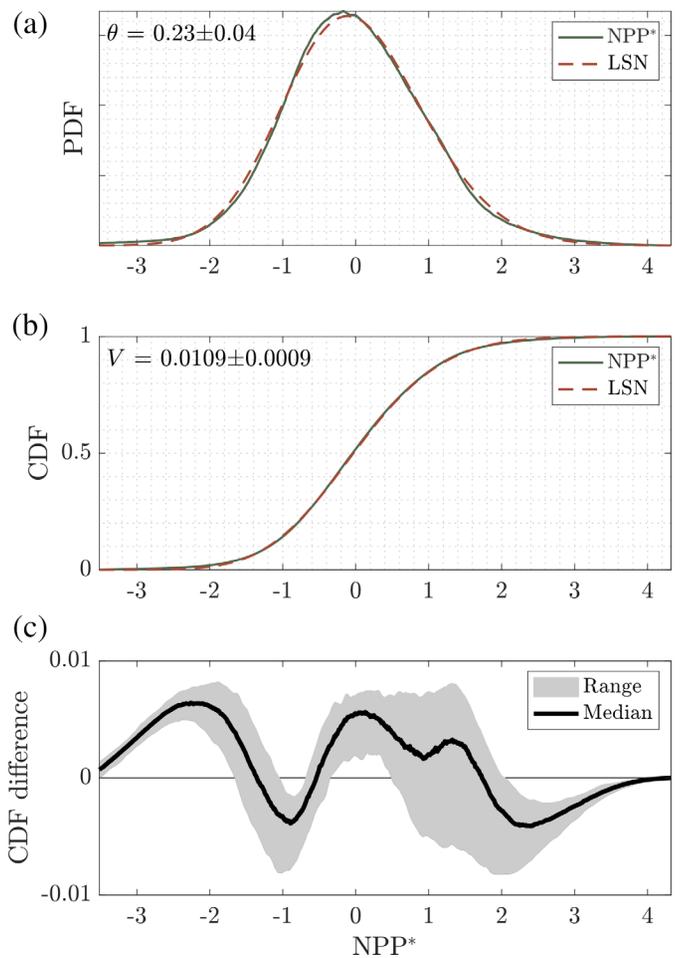
Fig. 3. (**a**) Probability density function and (**b**) cumulative distribution function of global CAFE output vs. log-skew-normal fit. (**c**) Difference between model and log-skew-normal cumulative distribution functions (i.e., departure of data from the theorized distribution, or equivalently the difference between lines in **b**). Range is over 100 iterations of 100,000 randomly chosen locations and times. NPP* is the standardized logarithm of net primary production.

The 4 km daily resolution is intractably large to analyze full datasets of. I instead randomly sample 100,000 values from each space–time subset (including the global ocean) and analyze these. I repeat this process 100 times to estimate the uncertainty associated with the random sampling. A 100 iteration is more than sufficient to estimate this uncertainty in this case (using 1000 iterations in some cases yielded identical results). The choice of 100,000 random samples also is inconsequential: using 10,000 or 1,000,000 random samples yielded similar results, with larger sample sizes (encouragingly) corresponding to slightly better correspondences between model output and the hypothesized distribution.

I fit a LSN distribution to these modeled NPP distributions by finding the parameter combinations that minimize Kuiper's statistic V (Kuiper 1960):

$$V = \max(F(\text{NPP}) - F_h(\text{NPP})) + \max(F_h(\text{NPP}) - F(\text{NPP}))$$

where F is the cumulative distribution function (CDF) of the CAFE output and F_h is a hypothesized distribution such as a LSN, hence smaller Vs indicate better fits. Kuiper's statistic V is generally preferred to the more standard Kolmogorov–Smirnov statistic as it weighs all the data equally, rather that weighting toward the median (Press 1992). The fitting was performed on standardized log-transformed NPP values, NPP*:

$$\text{NPP}^* = \frac{\ln(\text{NPP}) - \text{mean}(\ln(\text{NPP}))}{\text{standard deviation}(\ln(\text{NPP}))}$$

for computational efficiency (this narrows the size of the parameter space that must be sampled). Note that standardizing a variable by definition does not change the skewness of its distribution, the skewness being the third standardized moment. The code used in these analyses is available at (manuscript submitted).

Results

Figure 3 shows the global CAFE distribution and the closest LSN. The agreement is striking, with $V = 0.0109 \pm 0.0009$ and the CDFs for any of the iterations never differ by more than 0.009 (n.b. this is also true using 1000 iterations). The estimated log-skewness is $\theta = 0.12 \pm 0.04$, substantially larger than zero. Globally, then, CAFE output appears to be excellently characterized by a LSN distribution. (Note that while the LSN can be an extremely heavy-tailed distribution, the global distribution's small value for $\sigma = 0.325 \pm 0.001$ means that total productivity is not dominated not by extreme tail values and is instead determined by the probability distribution overall.)

Figure 4 shows the value of Kuiper's statistic V and log-skewness θ for the spatial and temporal subsets considered (I focus on these because the mean state and variance characteristics of CAFE are described in Silsbe et al. 2016). In general, the spatial and temporal subsets are almost all well-described as LSN-distributed. It is challenging to ascribe p-values to these fits because of the large sample sizes (Stephens 1974) but all of
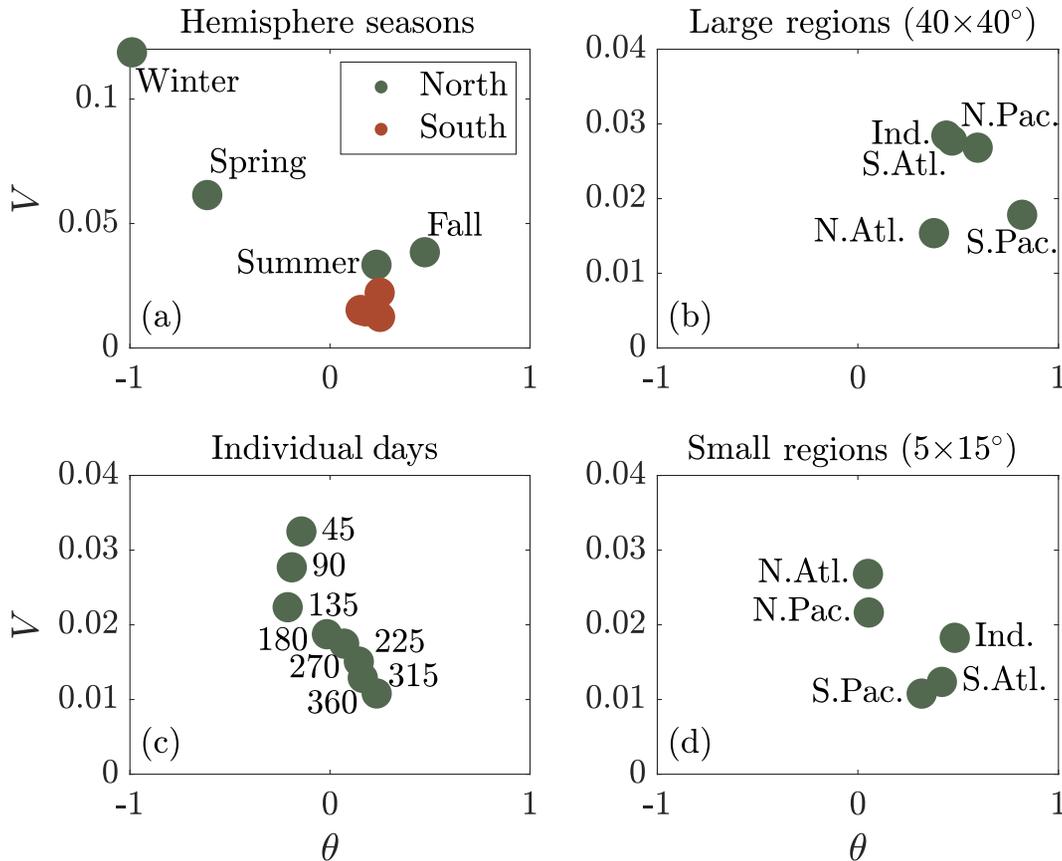
Fig. 4. Goodness-of-fit statistic V and log-skewness θ for (**a**) each season in the Northern and Southern Hemisphere, (**b**) five large (40° × 40°) spatial regions, (**c**) eight evenly-spaced individual days, and (**d**) five smaller (5° × 15°) regions in the center of the larger regions in panel (**b**). Note the different y-scale in panel (**a**). Error bars are smaller than scatter point sizes in each case.
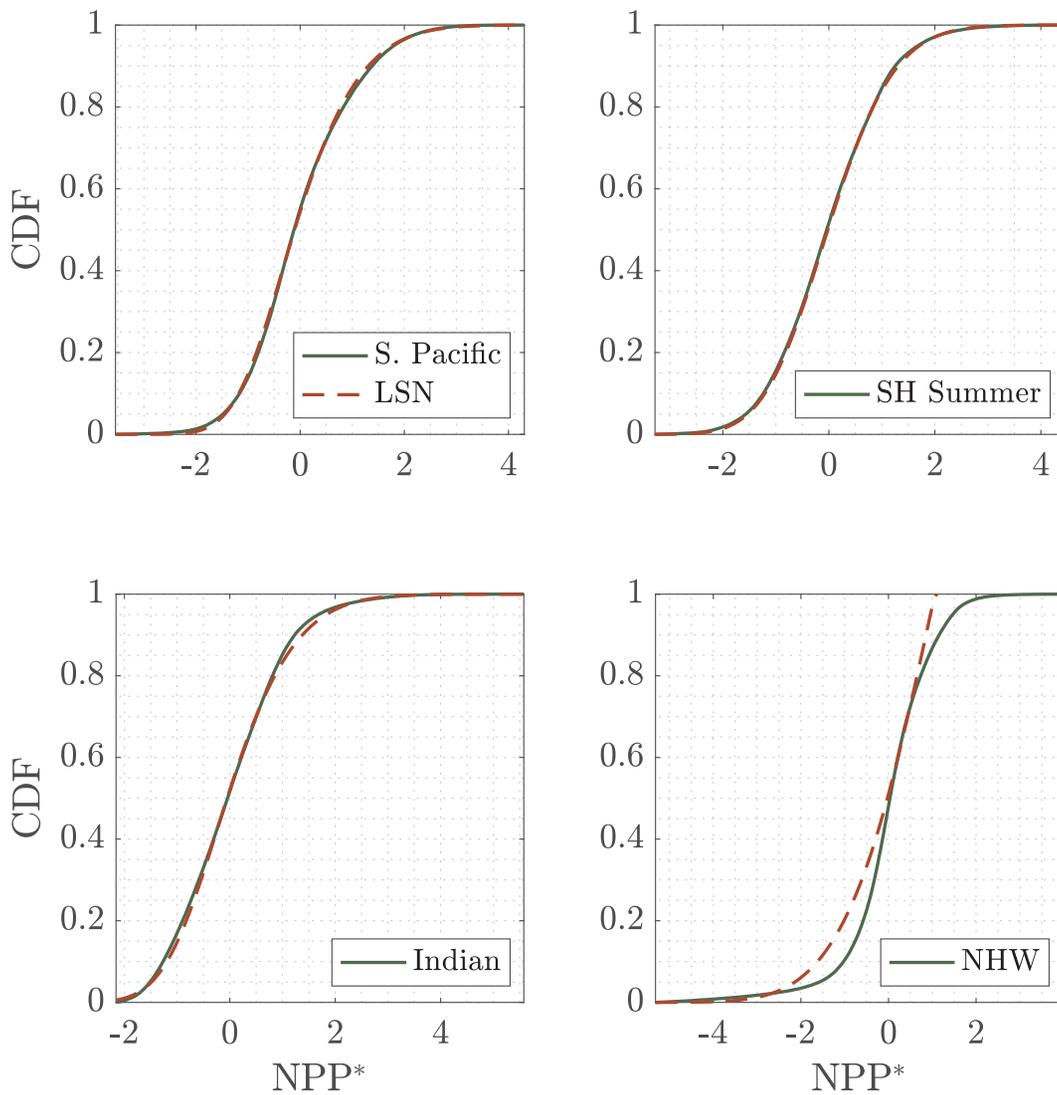
Fig. 5. As in Fig. 2(b), but for particular regions (Indian Ocean box and South (S.) Pacific box) and seasons (Southern Hemisphere [SH] Summer, Northern Hemisphere Winter (NHW)).

these in Panels 3b-3d are have V values corresponding to unambiguously good fits. Figure 5 gives spatial and temporal examples with high and low V and θ values; here one can see that the Indian Ocean $40° \times 40°$ box, the worst-fitting of the large spatial regions, with $V = 0.028 \pm 0.002$, is visually a comparably good fit to the global case with $V = 0.0109 \pm 0.0009$—the difference is an almost indistinguishable underestimation of the CDF around NPP* ~ −1 and NPP* ~ 1.5. The large spatial regions all have $V < 0.03$ and $\theta > 0$, the latter likely related to times of year when blooms occur. The smaller spatial regions behave similarly, with $V < 0.03$ and $\theta > 0$, indicating that the LSN description is not only applicable at large basin-type scales but also at smaller regional scales. Interestingly, when individual days are considered, the $\theta \sim 0$ and V decreasing monotonically with time from $V \approx 0.03$ to $V \approx 0.01$. This is curious but

may just as likely be a coincidence for the year in question as a meaningful trend. The Southern Hemisphere seasons all have small V values and slightly positive θ values. The space-and-time subset of the North Atlantic Spring behaves similarly to these spatial or temporal subsets, with $V = 0.0337 \pm 0.0006$ and $\theta = 0.27 \pm 0.01$.

The exception in Figs. 4 and 5 is the NHW (and to a lesser extent the spring, presumably for similar reasons). In this case, θ is close to its minimum for a LSN variable at $\theta \sim -1$, and V is more than three times larger than for any case shown in Panels 3b-3d. These CAFE output values are clearly not LSN-distributed. From the CAFE and LSN CDF's for the NHW (Fig. 5), it is clear that this poor fit is driven by their being a very heavy tail on low end of the CAFE NPP distribution in addition to relatively heavy tail on the high end, whereas the

LSN corresponds to either zero ($\theta \sim 0$) or one heavy tail ($\theta \sim \pm 1$). This poor fit is not due to values at particular times of year or latitudes: isolating tropical or subpolar regions or particular months yielded similarly poor fits. This mismatch can either be interpreted as a failing of the LSN to capture a real phenomenon in the NHW, or a failing of CAFE model to accurately model NPP in the NHW. Because the log-normal distribution holds for in situ measurements in the NHW (Cael et al. 2018 Fig. 5) and this mismatch is primarily caused by the very heavy low tail, it seems most likely that the smallest $\sim 5\%$ of CAFE NHW NPP estimates are artificially low. Note that when this lower tail is ignored, the remaining distribution conforms to a LSN similarly to the other subsets considered here, for example, by discarding all NPP values < 200 mg C m^{-2} d^{-1}, the good-ness-of-fit V is similar to other spatio-temporal subsets considered here ($V = 0.032 \pm 0.003$) and $\theta > 0$. These low productivity values may be due to the sensitivity of CAFE (like other NPP models) to mixed layer depth; erroneously overestimated mixed layer depths in the NHW impose photoacclimation in such models that artificially lowers NPP (Milutinović et al. 2009). Another possibility is that these low productivity values are due to optical constituents such as colored dissolved organic matter not being accurately accounted for, so that the phytoplankton properties used in the CAFE model are inaccurately estimated, leading to underestimated NPP (Siegel et al. 2005; Nelson and Siegel 2013). That this mismatch is driven by such a small fraction of NPP estimates further underscores the power of this distributional test for evaluating NPP models, which in turn underscores that CAFE appears to reproduce NPP variability excellently on satellite-relevant spatial scales. Resolving these NHW issues should improve not only NPP estimates not just in the NHW, but globally.

Altogether, it appears that the LSN distribution is an excellent and robust description of variability in a state-of-the-art satellite-based net primary productivity model. LSN is derived from the underlying log-normality of in situ measurements and the scale separation of in situ vs. satellite measurements. This in turn helps explain why satellite NPP measurements fail to capture the variability of in situ NPP measurements. The fit of a LSN distribution constitutes a strict test of satellite net PP models' variability, in this case identifying a potential issue of artificially low estimates in the NHW. It would be instructive to compare how LSN the outputs of different net PP models are. Note also that the same LSN test should apply equally to other bio-optical variables whose in situ measurements are log-normal, particularly chlorophyll (Campbell 1995), and that even if more matched-up measurements were available for these variables (as is the case with chlorophyll), the advantages of such a test—accounting for scale separation, not penalizing models for mismatches in individual measurements, and leveraging of more measurements than matchups alone—still apply. The log-skewness of the NPP model analyzed here also has interesting spatial patterns, with a clear hemispheric asymmetry and

some additional second-order features. These patterns merit further investigation, for which the LSN is a useful lens as it condenses the variations in net PP into three parameters. In principle, these three parameter values can be used to estimate not only the two parameters of the underlying log-normal distribution that in situ measurements conform to, but also the spatial footprint of these bottle measurements. However, this inverse problem is complicated by the potential correlations between nearby measurements due to mesoscale and submesoscale phenomena. Resolving this would make it possible to utilize satellite measurements to subpixel-scale information about NPP variability, a problem of great interest in the submesoscale dynamics community. Critically, better understanding these parameters and using independently estimated values for them, instead of allowing them to be free fitting parameters as I have done here, would also be a much stricter and therefore preferable version of this distributional test.

References

Behrenfeld, M. J., and P. G. Falkowski. 1997. A consumer's guide to phytoplankton primary productivity models. Limnol. Oceanogr. **42**: 1479–1491. doi:10.4319/lo.1997.42.7.1479

Buitenhuis, E. T., T. Hashioka, and C. Le Quéré. 2013. Combined constraints on global ocean primary production using observations and models. Global Biogeochem. Cycles **27**: 847–858. doi:10.1002/gbc.20074

Cael, B. B., K. Bisson, and C. L. Follett. 2018. Can rates of ocean primary production and biological carbon export be related through their probability distributions? Global Biogeochem. Cycles **32**: 954–970. doi:10.1029/2017GB005797

Campbell, J. W. 1995. The lognormal distribution as a model for bio-optical variability in the sea. J. Geophys. Res. Oceans **100**: 13237–13254. doi:10.1029/95JC00458

Carr, M.-E., and others. 2006. A comparison of global estimates of marine primary production from ocean color. Deep-Sea Res. II Top. Stud. Oceanogr. **53**: 741–770. doi:10.1016/j.dsr2.2006.01.028

Hcine, M. B., and R. Bouallegue. 2015. Highly accurate log skew normal approximation to the sum of correlated log-normals. arXiv preprint arXiv:1501.02347.

Kuiper, N. H. 1960. Tests concerning random points on a circle. Nederl. Akad. Wetensch. Proc. Ser. A **63**: 38–47.

Mahadevan, A. 2016. The impact of submesoscale physics on primary productivity of plankton. Ann. Rev. Mar. Sci. **8**: 161–184. doi:10.1146/annurev-marine-010814-015912

Marra, J. 2009. Net and gross productivity: Weighing in with 14C. Aquat. Microb. Ecol. **56**: 123–131.

Milutinović, S., M. J. Beherenfeld, J. A. Johannessen, and T. Johannessen. 2009. Sensitivity of remote sensing–derived phytoplankton productivity to mixed layer depth: Lessons

from the carbon-based productivity model. Global Biogeochem. Cycles **23**: GB4005. doi:10.1029/2008GB003431

Nelson, N. B., and D. A. Siegel. 2013. The global distribution and dynamics of chromophoric dissolved organic matter. Ann. Rev. Mar. Sci. **5**: 447–476. doi:10.1146/annurev-marine-120710-100751

Press, W. H., and others. 2002, *Numerical recipes in C++. The art of scientific computing*, 2nd Edition. Cambridge: Cambridge Univ. Press, p. 1002.

Siegel, D. A., and others. 2001. Bio-optical modeling of primary production on regional scales: The Bermuda BioOptics project. Deep-Sea Res. II Top. Stud. Oceanogr. **48**: 1865–1896. doi:10.1016/S0967-0645(00)00167-3

Siegel, D. A., S. Maritorena, N. B. Nelson, M. J. Behrenfeld, and C. R. McClain. 2005. Colored dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere. Geophys. Res. Lett. **32**: L20605. doi: 10.1029/2005GL024310

Silsbe, G. M., M. J. Behrenfeld, K. H. Halsey, A. J. Milligan, and T. K. Westberry. 2016. The CAFE model: A net production model for global ocean phytoplankton. Global Biogeochem. Cycles **30**: 1756–1777. doi:10.1002/2016GB005521

Stephens, M. A. 1974. EDF statistics for goodness of fit and some comparisons. J. Am. Stat. Assoc. **69**: 730–737.