

1 Forecasting ocean chlorophyll in the Equatorial Pacific

2

3 **Cecile S. Rousseaux**^{1,2,*} and **Watson W. Gregg**¹

4 ¹ Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt,
5 Maryland, USA

6 ² Universities Space Research Association, Columbia, Maryland, USA

7 *Author to whom correspondence should be addressed; E-Mail: Cecile.S.Rousseaux@nasa.gov

8 Tel.: +1-301-614-5750; Fax: +1-301-614-5644

9

10 **Abstract**

11 Using a global ocean biogeochemical model combined with a forecast of physical oceanic
12 and atmospheric variables from the NASA Global Modeling and Assimilation Office, we assess
13 the skill of a chlorophyll concentrations forecast in the Equatorial Pacific for the period 2012-2015
14 with a focus on the forecast of the onset of the 2015 El Niño event. Using a series of retrospective
15 9-month hindcasts, we assess the uncertainties of the forecasted chlorophyll by comparing the
16 monthly total chlorophyll concentration from the forecast with the corresponding monthly ocean
17 chlorophyll data from the Suomi-National Polar-orbiting Partnership Visible Infrared Imaging
18 Radiometer Suite (S-NPP VIIRS) satellite. The forecast was able to reproduce the phasing of the
19 variability in chlorophyll concentration in the Equatorial Pacific, including the beginning of the
20 2015-2016 El Niño. The anomaly correlation coefficient (ACC) was significant ($p < 0.05$) for
21 forecast at 1-month ($R=0.33$), 8-month ($R=0.42$) and 9-month ($R=0.41$) lead times. The root mean
22 square error (RMSE) increased from $0.0399 \mu\text{g chl L}^{-1}$ for the 1-month lead forecast to a maximum
23 of $0.0472 \mu\text{g chl L}^{-1}$ for the 9-month lead forecast indicating that the forecast of the amplitude of
24 chlorophyll concentration variability was getting worse. Forecasts with a 3-month lead time were
25 on average the closest to the S-NPP VIIRS data (23% or $0.033 \mu\text{g chl L}^{-1}$) while the forecast with
26 a 9-month lead time were the furthest (31% or $0.042 \mu\text{g chl L}^{-1}$). These results indicate the potential
27 for forecasting chlorophyll concentration in this region but also highlights various deficiencies and
28 suggestions for improvements to the current biogeochemical forecasting system. This system
29 provides an initial basis for future applications including the effects of El Niño events on fisheries
30 and other ocean resources given improvements identified in the analysis of these results.

31

32

33 **Introduction**

34 Forecast models of atmospheric conditions have considerably improved over the past few
35 decades and are routinely used to predict weather patterns including hurricanes, winds and other
36 potentially threatening conditions. Natural processes in the atmosphere, ocean and land can each
37 influence climate in sometimes predictable ways. Developing forecasting systems for ocean
38 biogeochemical processes is a scientific challenge that has important implications in the
39 management of marine ecosystems and resources. One of the challenges of improving subseasonal
40 to seasonal forecasting skill is to identify and characterize sources of subseasonal to seasonal
41 natural modes of variability (e.g. El Niño Southern Oscillation), slowly varying processes (e.g.
42 ocean biogeochemistry), and external forcing (e.g. winds, radiation).

43 Most oceanographic forecasts emphasize physical conditions (e.g. temperature, mixing),
44 ocean biogeochemical forecasts are less common and have mostly focused on the prediction of
45 algal blooms and hypoxia [e.g. *Evans and Scavia*, 2010; *Greene et al.*, 2009; *Stumpf et al.*, 2009;
46 *Wynne et al.*, 2005]. Various approaches have been developed to predict biogeochemical variables
47 from statistical relationships with temperature, wind speed and other variables to the use of more
48 complex numerical models. A typical application of these biogeochemical forecasts is the
49 prediction of Harmful Algal Blooms [e.g. *Raine et al.*, 2010; *Stumpf et al.*, 2009]. One example is
50 the Eastern Gulf of Mexico Harmful Algal Bloom Operational Forecast System (GOMX HAB-
51 OFS) developed by NOAA to follow the development of a toxic dinoflagellate, *Karenia brevis*,
52 that produces Neurotoxic Shellfish Poisoning, kills fishes and marine mammals and leads to health
53 and economical losses resulting from respiratory irritation in the waters off Florida. This
54 forecasting system relies on satellite ocean color and transport direction data from satellite imagery
55 combined with in situ samples. They issue semi-weekly bulletins that serve as decision support
56 tools for coastal resource managers, federal and state agencies, public officials, and academic
57 institutions [*Kavanaugh et al.*, 2016]. The forecast was expanded to other regions and the system
58 is described in several papers [e.g. *Stumpf et al.*, 2003; *Stumpf et al.*, 2009; *Tomlinson et al.*, 2004].
59 Other examples of biogeochemical forecast efforts include the forecast of hypoxia zone in the Gulf
60 of Mexico [*Scavia et al.*, 2003], net primary production in the tropical Pacific [*Séférián et al.*,
61 2014], annual salmon yields [*Scheuerell and Williams*, 2005], sardines distribution [*Kaplan et al.*,

62 2016], seasonal distributions of southern Bluefin tuna [Eveson *et al.*, 2015; Hobday *et al.*, 2011]
63 and coral bleaching [Goreau and Hayes, 2005].

64 While some of these forecasting systems rely on satellite ocean color data, others rely on
65 biochemical variables that cannot be directly derived from ocean color data or that do not have
66 statistical relationship with variables that can be derived from satellite data (e.g. nutrient, oxygen
67 concentration). Furthermore, satellite data can have large gaps (e.g. clouds, aerosols, interorbital
68 gaps, high solar zenith angles) that do not allow for a systematic and complete coverage of the area
69 of interest. Here we combine an established biogeochemical model with a seasonal forecast of
70 atmospheric and ocean conditions to provide a 9-month forecast of total chlorophyll in the
71 Equatorial Pacific for the period 2012-2015. The assimilation of satellite ocean color to provide
72 the initial conditions for the forecast ensures the best use of the data available, while the forecast
73 provides a complete coverage of the chlorophyll concentration (among other variables) for a 9-
74 month forecast. The skill of the forecasting system is assessed by comparing the total chlorophyll
75 to those from the satellite Suomi-National Polar-orbiting Partnership Visible Infrared Imaging
76 Radiometer Suite (S-NPP VIIRS).

77

78 **Material and Methods**

79 The NASA Ocean Biogeochemical Model (NOBM) is a three dimensional biogeochemical
80 model of the global ocean coupled with a circulation and radiative model [Gregg and Casey, 2007;
81 Gregg *et al.*, 2003]. NOBM has a near-global domain that spans from -84° to 72° latitude at a
82 1.25° resolution in water deeper than 200m. NOBM is coupled with the Poseidon ocean general
83 circulation model. The Poseidon model [Schopf and Lough, 1995] is a reduced gravity ocean
84 model with 14 layers in quasi-isopycnal coordinates forced by wind stress, sea surface temperature,
85 and shortwave radiation [Gregg and Casey, 2007]. The NOBM contains 4 explicit phytoplankton
86 taxonomic groups (diatoms, cyanobacteria, chlorophytes and coccolithophores), 3 detritus
87 components (silicate, nitrate/carbon and iron), 4 nutrients (nitrate, silicate, iron and ammonium)
88 and one zooplankton group. The growth of phytoplankton is dependent on total irradiance, nitrogen
89 (nitrate+ammonium), silicate (for diatoms only), iron and temperature (see Rousseaux and Gregg
90 2015 for more details). Surface photosynthetically available radiation is derived from the Ocean-
91 Atmosphere Spectral Irradiance Model [OASIM; Gregg and Casey, 2009].

92 A spin-up run of 100 years has been shown to produce stable initial conditions for biological
93 variables [Gregg and Rousseaux, 2014]. The NOBM model is then run for 14 years using ocean
94 and atmospheric variables as forcing from the Modern-Era Retrospective analysis for Research
95 and Applications [MERRA, Rienecker et al., 2011] and ocean chlorophyll data from Sea-Viewing
96 Wide Field-of-View Sensor (SeaWiFS) and Moderate-resolution imaging spectroradiometer
97 (MODIS)-Aqua in data assimilation mode [Gregg and Rousseaux, 2014]. Starting in 2012, the
98 model assimilates chlorophyll data from S-NPP VIIRS and uses transient MERRA data to force
99 the circulation model. The assimilation of satellite chlorophyll uses a multivariate methodology
100 where the nutrients are adjusted corresponding to the chlorophyll assimilation using nutrient-to-
101 chlorophyll ratios embedded in the model [Rousseaux and Gregg, 2012]. The difference between
102 the chlorophyll assimilation results and the prior chlorophyll produced by the model (the analysis
103 increments) are used to adjust the nutrient concentrations. The multivariate assimilation is applied
104 to silica and dissolved iron, as well as nitrate. These conditions are used as initial conditions for
105 each forecast (using the month prior to the start of the forecast). The forcing data used for the
106 forecast include zonal and meridional wind stress, sea surface temperature and shortwave
107 radiation. These forecast files are produced by the NASA Global Modeling and Assimilation
108 Office (GMAO) using the GEOS-5 system (https://gmao.gsfc.nasa.gov/weather_prediction/).
109 These forecasted atmospheric and ocean variables are currently provided to the North American
110 Multi-Model Ensemble (NMME) prediction project, as well as to other national (International
111 Research Institute for Climate and Society, IRI) and international (Asia-Pacific Climate Center,
112 APCC) ensemble seasonal forecasting efforts [Borovikov et al., in review].

113
114 The bias and uncertainties in the system are assessed by (1) comparing the satellite ocean
115 chlorophyll used for validation and data assimilation to in situ data, (2) comparing the chlorophyll
116 concentration from a free-run model (without data assimilation) to satellite ocean color and (3)
117 comparing the chlorophyll concentration from a run assimilating satellite chlorophyll with those
118 from the satellite (Figure 1). The in situ data used to evaluate the bias and uncertainties in the S-
119 NPP VIIRS chlorophyll include data collected from the National Oceanographic Data Center
120 [Gregg and Conkright, 2002], NASA in situ database [Werdell and Bailey, 2002; Werdell et al.,
121 2003], and Atlantic Meridional transect [Aiken et al., 2000] archives [Gregg et al., 2009]. The
122 quality of the biogeochemical system used is then assessed using a hindcast from 2012-2015 forced

123 using MERRA data (procedure 2a, b on Figure 1). The uncertainties in this system are evaluated
 124 by comparing the chlorophyll concentration in the Equatorial Pacific from this run with those from
 125 S-NPP VIIRS. To evaluate the effects of the forcing data on the chlorophyll concentration
 126 estimates, we then compare a free-run model forced by transient MERRA forcing data with one
 127 forced by climatological MERRA data. Finally we compare the monthly chlorophyll concentration
 128 from the assimilation run to the monthly concentration from S-NPP VIIRS (procedure 3 on Figure
 129 1). Bias is quantified by averaging the monthly percent difference between the chlorophyll
 130 concentration from the model (free-run and assimilating run) and the satellite chlorophyll
 131 concentration for the period 2012-2015 and the standard error is calculated. The uncertainty is
 132 quantified using a correlation coefficient. A statistically significant correlation coefficient is
 133 defined as one with a p-value smaller than 0.05.

134 The skill of the various forecasts is assessed using three metrics: (1) the percent difference
 135 between the NPP-VIIRS chlorophyll data and the forecast (bias) (procedure 4 on Figure 1), (2) the
 136 anomaly correlation coefficient (ACC) and (3) the root mean square error (RMSE). The anomaly
 137 correlation coefficient provides information on the linear association between forecast and
 138 observations but is insensitive to biases and error in variances. It is calculated as between the model
 139 prediction (p) and satellite observation (o) of chlorophyll over N months ($N=38$) and computed as:

$$140 \quad ACC = \frac{\sum(p - \bar{p})(o - \bar{o})}{\sqrt{\sum(p - \bar{p})^2 \sum(o - \bar{o})^2}}$$

141 The RMSE measures the magnitude of the error, is sensitive to large values but does not indicate
 142 the direction of the error. It is calculated as:

$$143 \quad RMSE = \sqrt{\frac{1}{N} \sum [(p - \bar{p}) (o - \bar{o})]^2}$$

144 where \bar{p} and \bar{o} are the temporal averages of chlorophyll.

145 A total of 38 retrospective forecasts were run, each for a 9-month period. The first forecast
 146 started in March 2012 and the last forecast started in April 2015. The percent difference between
 147 the satellite and the forecast chlorophyll quantifies the mean error in the forecast. It allows us to
 148 assess whether the forecast has on average a positive or a negative bias..

149

150 **Results and Discussion**

151 1. Assessing the skill of the model system

152 The first source of uncertainty reflects the inherent bias of satellite-derived chlorophyll
153 concentration and is assessed by comparing the S-NPP VIIRS chlorophyll to in situ fluorometric
154 chlorophyll data. For the period from 2012-2014, the global chlorophyll from S-NPP VIIRS
155 compared favorably to in situ chlorophyll (bias=11.8%, semi-interquartile range =27.9% and
156 $R=0.86$; Table 1).

157 The second source of uncertainty lies in how well the model simulates chlorophyll
158 concentration. This source of uncertainty is assessed by comparing the chlorophyll concentration
159 [Toggweiler *et al.*, 1991] from the free-run model (no data assimilation but uses transient forcing
160 conditions from MERRA) with the corresponding satellite ocean color data. For the period from
161 2012 until 2015, monthly chlorophyll concentration from the free-run model were significantly
162 correlated to those from the satellite ocean color (S-NPP VIIRS, $R = 0.72$, $p < 0.05$; Table 1). The
163 chlorophyll from the free-run model was on average within $27.87 \pm 1.72\%$ (average \pm standard
164 error) of the S-NPP VIIRS chlorophyll. Chlorophyll fields in the Equatorial Pacific showed
165 agreement with satellite data (Figure 2). The model reproduces the main features observed by the
166 satellite ocean color. The consistent positive bias in chlorophyll concentration in the Equatorial
167 Pacific in the free-run model suggest that the upwelling in the Equatorial Pacific in the model is
168 overestimated and therefore leads to higher chlorophyll concentration than those observed. The
169 overprediction of the upwelling in the Equatorial Pacific in models has been suggested for some
170 time [e.g. Toggweiler *et al.*, 1991; Zheng *et al.*, 2012]. In some other areas, such as along the South
171 America coastline as well as in the region of the Costa Rica Dome, the chlorophyll concentration
172 from the free-run model was underestimated. This is most likely due to the nature of the reduced
173 gravity circulation model. The model therefore does not include topographic effects, nor does it
174 allow the representation of cross-shelf advection and convection.

175 In the Equatorial Pacific, the monthly chlorophyll concentration from a run assimilating S-
176 NPP VIIRS chlorophyll data was significantly correlated ($R=0.95$, $P < 0.01$; Table 1) and on
177 average within $12.34 \pm 0.52\%$ of the S-NPP VIIRS chlorophyll concentration. The assimilation of
178 satellite chlorophyll to provide the initial conditions used for the forecast is therefore an
179 improvement over using the initial conditions provided by the free-run model without data
180 assimilation. We therefore use this set-up to provide the initial conditions for the forecasting
181 systems.

182 Finally the data used to force the model have their own inherent bias and uncertainties. While
183 this is beyond the scope of this paper, we note that the bias in the forcing data used here have been
184 assessed in other papers [e.g. *Rienecker et al.*, 2011]. By comparing the chlorophyll concentration
185 from the free-run model using climatological MERRA forcing data compared to using transient
186 MERRA data we can assess the improvements that such transient forcing data can provide to the
187 system. The chlorophyll concentration from the free-run model using transient MERRA forcing
188 data were considerably closer to the chlorophyll concentration from the S-NPP VIIRS
189 ($27.87 \pm 1.72\%$) than the free-run model using climatological MERRA data ($85.67\% \pm 2.77\%$,
190 Figure 3). This indicates the advantage of using transient forcing data to further improve the initial
191 conditions used for the forecasting system.

192

193 2. General skill of the forecasts

194 We assess the skill of our forecast by comparing each 9-month forecast to the observed
195 chlorophyll concentration in the Equatorial Pacific from S-NPP VIIRS for the corresponding
196 month. There was a consistent positive bias in the chlorophyll forecasted, as in the hindcast from
197 the free-run model compared with S-NPP VIIRS (Figure 2). Of the 38 forecasts, the average
198 percent difference between the forecasted chlorophyll and the S-NPP VIIRS chlorophyll varied
199 between 23% (3 months lead time, the equivalent of $0.033 \mu\text{g chl L}^{-1}$) and 30.7% (9 months lead
200 time, the equivalent of $0.042 \mu\text{g chl L}^{-1}$, Figure 4 and Figure 5). Except for the monthly chlorophyll
201 concentration at 5 and 6-month lead time, the chlorophyll concentration from the forecasts were
202 always significantly correlated to those from S-NPP VIIRS (data not shown). The highest
203 correlation coefficient was observed at 8-month lead time ($R=0.53$, $p<0.01$).

204 To assess the uncertainties in our forecast, we utilize two deterministic skill metrics: ACC
205 and RMSE. The ACC for the forecast was significant for the 1-month lead time ($R=0.33$, $P<0.05$)
206 as well as for the 8- and 9-month lag forecast ($R=0.42$ and $R=0.41$ respectively, Table 2). This
207 indicates that for these leads, the forecast chlorophyll had statistically the correct phasing when
208 compared to those from S-NPP VIIRS. The spatial distribution of the anomaly correlation
209 coefficient further reflects the overprediction of the upwelling in this Equatorial Pacific (Figure
210 6). While the forecasted chlorophyll concentrations at 1-month lead are significantly correlated
211 with those from S-NPP VIIRS for the majority of the Equatorial Pacific, some areas in the
212 upwelling tongue are not significant. The second skill metric, RMSE, increased from $0.040 \mu\text{g chl}$

213 L^{-1} at 1-month lead to $0.047 \mu\text{g chl } L^{-1}$ at 9-month lead forecast. These results suggest that while
214 the phasing may have been reasonable at 8- and 9-month lag forecast, the amplitude of the signal
215 was getting worse. Regardless, RMSE of $0.047 \mu\text{g chl } L^{-1}$ is still very acceptable for a 9-month lag
216 forecast. These results suggest some skill in forecasting the chlorophyll variability in the
217 Equatorial Pacific especially at 1-month lag when the ACC is significant and the RMSE is at its
218 lowest. For all forecasts, the chlorophyll concentrations were always within 30.7% of the
219 chlorophyll concentration from S-NPP VIIRS. This is similar to the uncertainties reported for this
220 instrument (semi-interquartile range of S-NPP VIIRS versus in situ chlorophyll = 27.9%).

221 3. Prediction of the 2015 El Niño

222 In the Equatorial Pacific, the El Niño Southern Oscillation is the dominant source of
223 interannual variability and has been shown to have a considerable impact of the biogeochemistry,
224 including chlorophyll concentration and recruitment of higher trophic levels, in this region [e.g.
225 *Martinez et al.*, 2009; *Strutton and Chavez*, 2000]. Forecasting El Niño events is the focus of many
226 prediction centers. While the focus of assessments such as the North American Multi-Model
227 Ensemble home has been on the skills in forecasting sea surface temperature, there has been very
228 little work on forecasting biogeochemical variables such as chlorophyll using a dynamical system.
229 The temporal evolution of the various forecasts in this study highlights the variability between the
230 forecasts and our skills in predicting the decline in chlorophyll concentration that was observed in
231 the Equatorial Pacific during the 2015 El Niño event (Figure 4). Starting in January 2015 the
232 forecast suggested a decline in chlorophyll concentration that would reach a minimum in May
233 2015 (average of the 8 forecasts available for this month of $0.13 \mu\text{g chl } L^{-1}$). The S-NPP VIIRS
234 data observed this minimum one month later in June 2015 ($0.13 \mu\text{g chl } L^{-1}$). The chlorophyll
235 concentration from S-NPP VIIRS then increased to reach a peak in August 2015 ($0.14 \mu\text{g chl } L^{-1}$).
236 This increase in chlorophyll was also reflected in the various forecasts although it was
237 overestimated. After August 2015, chlorophyll concentration declined reflecting the onset of the
238 2015 El Niño and the suppression of the upwelling in the Equatorial Pacific. This decline was also
239 observed in the chlorophyll concentration from S-NPP VIIRS. Of the four forecasts available for
240 September 2015, only one had predicted this decline. The other three forecasts predicted a decline
241 but delayed by one month (chlorophyll started to decline in October 2015). For the four forecasts,
242 September 2015 was their 6- to 9-month lead forecast which we previously showed had relatively
243 low skills compared to the 1-month lead forecasts. In the last forecast (highlighted in red in Figure

244 3), September 2015 corresponded to its 6-month lead forecast and this forecast predicted
245 particularly well the decline in chlorophyll concentration that occurred between August and
246 December 2015 in the Equatorial Pacific in response to the El Niño event. The spatial distribution
247 of the chlorophyll anomaly between December 2015 and March 2015 (first month of the last
248 forecast available) coincides well with that from S-NPP VIIRS for the corresponding month
249 (Figure 7). The area of negative anomaly in chlorophyll concentration along the South American
250 coast is distinguishable in both the forecast and the S-NPP VIIRS chlorophyll data. The
251 overestimation of the upwelling system in the forecast is also visible on this spatial representation
252 of the chlorophyll anomalies. The temporal evolution of these various forecasts highlights the
253 impacts that the atmospheric forcing data have on the forecast of chlorophyll. As the forecasts get
254 closer to the El Niño event, the forecasted atmospheric and oceanographic physical forcing data
255 have more skills and therefore lead to a better forecast in chlorophyll concentration. The forecast
256 of chlorophyll in this region therefore relies heavily on the existence of accurate forecast of
257 atmospheric forcing data. The initial conditions seem to play a more minor role in the forecasting
258 skill for predicting chlorophyll in this region.

259 4. Uncertainties of the approach

260 The uncertainties in the forecast of atmospheric and oceanic variables used to force the model
261 play a critical role in our ability to provide a successful forecast. The skill of the variables produced
262 by the GMAO forecasting system and that are used to force the model in forecast mode can also
263 be a source of uncertainties and have been assessed in Borovikov et al. [in review]. The SST
264 anomaly correlation coefficient from the forecast in the tropical Pacific has a high correlation
265 coefficient ($R > 0.8$) with the Reynolds SST for lead month 1 to 3 and remained above 0.6 by lag
266 month 9 indicating significant ($p < 0.05$) skill. A case study of the El Niño event of 2015/2016 in
267 Borovikov et al. (in review) suggested an overprediction of the magnitude in SST anomalies
268 observed during the 2015/2016 El Niño event but was overall in good agreement with the
269 conditions that were observed.

270 The forecast of chlorophyll concentration presented here is based on one single set of
271 forecasting data while the forecasting system used at GMAO provides forecasts for several
272 ensembles. Using ensemble forecasting instead of a single forecast might further improve our skill.
273 Initial conditions can be perturbed in various ways to account for initial condition uncertainty. The
274 uncertainty in the forecasted forcing data provided by GMAO could be accounted for by running

275 with the various ensembles they provide for the variables used to force the biogeochemical
276 forecast. Finally the model uncertainty could be accounted for using some stochastic
277 parametrization at the sub-grid level such as the one used by the European Centre for Medium
278 Range Weather Forecasts [Buizza *et al.*, 1999].

279 Another source of uncertainty in our forecast is the assimilation methodology, the
280 Conditional Relaxation Analysis Method used for bias correction for SST products (Reynolds,
281 1988) and applied here for chlorophyll [Gregg, 2008]. This method does not utilize ensembles
282 which can potentially improve the initial conditions for the forecast. It would also extend the
283 memory of the assimilation, which appears to survive <2 months here and assist in the skill of the
284 1-month forecast. However, there is little evidence that the 2-9 month forecasts could benefit
285 substantially from improved initial conditions, which are quite close to the S-NPP VIIRS
286 chlorophyll as suggested in Table 1.

287 5. Future improvements and applications

288 While these results suggest some skill in our ability to forecast chlorophyll concentration in
289 the Equatorial Pacific, they also highlight potential weaknesses and avenues for improvements.
290 The skill of the forecasting system relies as previously mentioned on the bias in the model's
291 representation of physical and biogeochemical processes in the oceans, and the uncertainties in the
292 forcing and assimilation data used. To further improve the forecasting system, each of these
293 sources of bias and uncertainties needs to be assessed individually for weaknesses and possibilities
294 for improvements. The range of applications of such a forecasting system, once properly set, can
295 be extended for other variables. Applications include but are not limited to the prediction of
296 Harmful Algal Blooms, fisheries, hypoxia/anoxia events, oil spills or the dispersal of pollutants.
297 Prediction of temperature, ocean currents and velocities have for example been used for
298 monitoring fisheries success, transport and spread of fish larvae, as well as seasonal fish migration
299 [Bonhommeau *et al.*, 2009; Hobday and Hartmann, 2006; Johnson *et al.*, 2005]. While the use of
300 physical variables such as temperature, salinity and currents have been successfully used as
301 covariates to explain distribution and catch rates of various species [e.g. Bigelow and Maunder,
302 2007; Cole, 1999; Herron *et al.*, 1989; Kaplan *et al.*, 2016; Zagaglia *et al.*, 2004], these
303 relationships can be limited since the behavior and recruitment of fish relies on changes in their
304 prey concentration and composition. Accurate forecasts of the resources on which fish populations
305 rely could provide the potential for strategic rather than reactive marine resource management

306 during El Niño events for example. In the Equatorial Pacific, forecast of the effects of ENSO
307 events on the physical conditions have been the subject of several studies starting in the 1980s
308 (Cane et al. 1986). In the last two decades we have witnessed the development of two major El
309 Niño events that had considerable impacts on both land and ocean conditions. The 1997-98 El
310 Niño was particularly devastating for the ocean resources and led to the collapse of several
311 fisheries and dramatic socio-economical repercussions for countries such as Peru. Anchovies, as
312 well as other fisheries collapsed during both the 1982-83 and 1997-98 El Niño events. Forecasts
313 such as the one presented here could therefore provide a framework to improve our management
314 of resources during these events. Furthermore, the forecasting system presented here may provide
315 a basis to expand the forecast from total chlorophyll to specific species including Harmful Algal
316 Blooms. This could provide support for the management of many areas that need to monitor closely
317 any development of harmful species in their waters. In the regions prone to Harmful Algal Blooms,
318 such a forecast could also be used to improve the strategies to detect and manage most efficiently
319 these events to minimize the repercussion on the human population and the associated economy.

320

321 **Acknowledgments**

322 We thank the NASA Ocean Ecology Laboratory for providing the satellite chlorophyll data and
323 the NASA Center for Climate Simulation for computational support. The GEOS-5 data used in
324 this study/project have been provided by the Global Modeling and Assimilation Office (GMAO)
325 at NASA Goddard Space Flight Center through the online data portal in the NASA Center for
326 Climate Simulation. This paper was funded by the NASA EXPORTS, MAP, PACE and S-NPP
327 Programs.

328

329 **References**

- 330 Aiken, J., N. Rees, S. Hooker, P. Holligan, A. Bale, D. Robins, G. Moore, R. Harris, and D. Pilgrim (2000),
331 The Atlantic Meridional Transect: overview and synthesis of data, *Progress in Oceanography*, 45(3), 257-
332 312.
- 333 Bigelow, K. A., and M. N. Maunder (2007), Does habitat or depth influence catch rates of pelagic
334 species?, *Canadian Journal of Fisheries and Aquatic Sciences*, 64(11), 1581-1594.
- 335 Bonhommeau, S., B. Blanke, A. TRÉGUIER, N. Grima, E. Rivot, Y. Vermard, E. Greiner, and O. Le Pape
336 (2009), How fast can the European eel (*Anguilla anguilla*) larvae cross the Atlantic Ocean?, *Fisheries
337 Oceanography*, 18(6), 371-385.
- 338 Borovikov, A., R. Cullather, R. M. Kovach, J. Marshak, G. Vernieres, Y. Vikhliaev, B. Zhao, and Z. Li (in
339 review), GEOS-5 seasonal forecast system, *Journal of Climate*.

340 Buizza, R., M. Milleer, and T. Palmer (1999), Stochastic representation of model uncertainties in the
341 ECMWF ensemble prediction system, *Quarterly Journal of the Royal Meteorological Society*, 125(560),
342 2887-2908.

343 Cole, J. (1999), Environmental conditions, satellite imagery, and clupeoid recruitment in the northern
344 Benguela upwelling system, *Fisheries Oceanography*, 8(1), 25-38.

345 Evans, M. A., and D. Scavia (2010), Forecasting hypoxia in the Chesapeake Bay and Gulf of Mexico:
346 Model accuracy, precision, and sensitivity to ecosystem change, *Environmental Research Letters*, 6(1),
347 015001.

348 Eveson, J. P., A. J. Hobday, J. R. Hartog, C. M. Spillman, and K. M. Rough (2015), Seasonal forecasting of
349 tuna habitat in the Great Australian Bight, *Fisheries Research*, 170, 39-49.

350 Goreau, T., and R. Hayes (2005), Monitoring and calibrating sea surface temperature anomalies with
351 satellite and in-situ data to study effects of weather extremes and climate changes on coral reefs, *World
352 Resource Review*, 17(2), 243-253.

353 Greene, R. M., J. C. Lehrter, and D. H. James III (2009), Multiple regression models for hindcasting and
354 forecasting midsummer hypoxia in the Gulf of Mexico, *Ecological applications*, 19(5), 1161-1175.

355 Gregg, W. W. (2008), Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global
356 ocean model, *Journal of Marine Systems*, 69(3-4), 205-225.

357 Gregg, W. W., and M. E. Conkright (2002), Decadal changes in global ocean chlorophyll, *Geophysical
358 Research Letters*, 29(15), 20-21.

359 Gregg, W. W., and N. W. Casey (2007), Modeling coccolithophores in the global oceans, *Deep-Sea
360 Research Part II*, 54(5-7), 447-477.

361 Gregg, W. W., and N. W. Casey (2009), Skill assessment of a spectral ocean-atmosphere radiative model,
362 *Journal of Marine Systems*, 76(1-2), 49-63.

363 Gregg, W. W., and C. S. Rousseaux (2014), Decadal Trends in Global Pelagic Ocean Chlorophyll: A New
364 Assessment Combining Multiple Satellites, In Situ Data, and Models, *Journal of Geophysical Research*,
365 *doi: 10.1002/2014JC010158*.

366 Gregg, W. W., P. Ginoux, P. S. Schopf, and N. W. Casey (2003), Phytoplankton and iron: validation of a
367 global three-dimensional ocean biogeochemical model, *Deep Sea Research Part II: Topical Studies in
368 Oceanography*, 50(22-26), 3143-3169.

369 Gregg, W. W., N. W. Casey, J. E. O'Reilly, and W. E. Esaias (2009), An empirical approach to ocean color
370 data: Reducing bias and the need for post-launch radiometric re-calibration, *Remote Sensing of
371 Environment*, 113(8), 1598-1612.

372 Herron, R. C., T. D. Leming, and J. Li (1989), Satellite-detected fronts and butterfly aggregations in the
373 northeastern Gulf of Mexico, *Continental Shelf Research*, 9(6), 569-588.

374 Hobday, A., and K. Hartmann (2006), Near real-time spatial management based on habitat predictions
375 for a longline bycatch species, *Fisheries Management and Ecology*, 13(6), 365-380.

376 Hobday, A. J., J. R. Hartog, C. M. Spillman, and O. Alves (2011), Seasonal forecasting of tuna habitat for
377 dynamic spatial management, *Canadian Journal of Fisheries and Aquatic Sciences*, 68(5), 898-911.

378 Johnson, D. R., H. M. Perry, and W. M. Graham (2005), Using nowcast model currents to explore
379 transport of non-indigenous jellyfish into the Gulf of Mexico, *Marine ecology. Progress series*, 305, 139-
380 146.

381 Kaplan, I. C., G. D. Williams, N. A. Bond, A. J. Hermann, and S. A. Siedlecki (2016), Cloudy with a chance
382 of sardines: forecasting sardine distributions using regional climate models, *Fisheries oceanography*,
383 25(1), 15-27.

384 Kavanaugh, K. E., K. Derner, and E. Davis (2016), Assessment of the Eastern Gulf of Mexico Harmful Algal
385 Bloom Operational Forecast System (GOMX HAB-OFS): An Analysis of Forecast Skill and Utilization from
386 May 1, 2008 to April 30, 2014 *Rep*.

387 Martinez, E., D. Antoine, F. D'Ortenzio, and B. Gentili (2009), Climate-driven basin-scale decadal
388 oscillations of oceanic phytoplankton, *Science*, 326(5957), 1253-1256.

389 Raine, R., G. McDermott, J. Silke, K. Lyons, G. Nolan, and C. Cusack (2010), A simple short range model
390 for the prediction of harmful algal events in the bays of southwestern Ireland, *Journal of Marine*
391 *Systems*, 83(3), 150-157.

392 Rienecker, M. M., M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D.
393 Schubert, L. Takacs, and G.-K. Kim (2011), MERRA: NASA's Modern-Era Retrospective Analysis for
394 Research and Applications, *Journal of Climate*, 24(14).

395 Rousseaux, C. S., and W. W. Gregg (2012), Climate variability and phytoplankton composition in the
396 Pacific Ocean, *Journal of Geophysical Research*, 117, C10006.

397 Scavia, D., N. N. Rabalais, R. E. Turner, D. Justic, and W. J. Wiseman Jr (2003), Predicting the response of
398 Gulf of Mexico hypoxia to variations in Mississippi River nitrogen load, *Limnology and Oceanography*,
399 48(3), 951-956.

400 Scheuerell, M. D., and J. G. Williams (2005), Forecasting climate-induced changes in the survival of Snake
401 River spring/summer Chinook salmon (*Oncorhynchus tshawytscha*), *Fisheries Oceanography*, 14(6), 448-
402 457.

403 Schopf, P. S., and A. Loughe (1995), A reduced-gravity isopycnal ocean model: Hindcasts of El Niño,
404 *Monthly Weather Review*, 123(9), 2839-2863.

405 Sférian, R., L. Bopp, M. Gehlen, D. Swingedouw, J. Mignot, E. Guilyardi, and J. Servonnat (2014),
406 Multiyear predictability of tropical marine productivity, *Proceedings of the National Academy of*
407 *Sciences*, 111(32), 11646-11651.

408 Strutton, P. G., and F. P. Chavez (2000), Primary productivity in the equatorial Pacific during the 1997–
409 1998 El Niño, *Journal of Geophysical Research*, 105(C11), 20089-26101.

410 Stumpf, R., M. Culver, P. Tester, M. Tomlinson, G. Kirkpatrick, B. Pederson, E. Truby, V.
411 Ransibrahmanakul, and M. Soracco (2003), Monitoring *Karenia brevis* blooms in the Gulf of Mexico using
412 satellite ocean color imagery and other data, *Harmful Algae*, 2(2), 147-160.

413 Stumpf, R. P., M. C. Tomlinson, J. A. Calkins, B. Kirkpatrick, K. Fisher, K. Nierenberg, R. Currier, and T. T.
414 Wynne (2009), Skill assessment for an operational algal bloom forecast system, *Journal of Marine*
415 *Systems*, 76(1), 151-161.

416 Toggweiler, J., K. Dixon, and W. Broecker (1991), The Peru upwelling and the ventilation of the South
417 Pacific thermocline, *Journal of Geophysical Research: Oceans*, 96(C11), 20467-20497.

418 Tomlinson, M. C., R. P. Stumpf, V. Ransibrahmanakul, E. W. Truby, G. J. Kirkpatrick, B. A. Pederson, G. A.
419 Vargo, and C. A. Heil (2004), Evaluation of the use of SeaWiFS imagery for detecting *Karenia brevis*
420 harmful algal blooms in the eastern Gulf of Mexico, *Remote Sensing of Environment*, 91(3), 293-303.

421 Werdell, P. J., and S. W. Bailey (2002), The SeaWiFS Bio-optical Archive and Storage System (SeaBASS):
422 Current architecture and implementation *Rep.*

423 Werdell, P. J., S. Bailey, G. Fargion, C. Pietras, K. Knobelspiesse, G. Feldman, and C. McClain (2003),
424 Unique data repository facilitates ocean color satellite validation, *Eos, Transactions American*
425 *Geophysical Union*, 84(38), 377-387.

426 Wynne, T. T., R. P. Stumpf, M. C. Tomlinson, V. Ransibrahmanakul, and T. A. Villareal (2005), Detecting
427 *Karenia brevis* blooms and algal resuspension in the western Gulf of Mexico with satellite ocean color
428 imagery, *Harmful Algae*, 4(6), 992-1003.

429 Zagaglia, C. R., J. A. Lorenzetti, and J. L. Stech (2004), Remote sensing data and longline catches of
430 yellowfin tuna (*Thunnus albacares*) in the equatorial Atlantic, *Remote Sensing of Environment*, 93(1),
431 267-281.

432 Zheng, Y., J. L. Lin, and T. Shinoda (2012), The equatorial Pacific cold tongue simulated by IPCC AR4
433 coupled GCMs: Upper ocean heat budget and feedback analysis, *Journal of Geophysical Research:*
434 *Oceans*, 117(C5).

435 **Table**

436 *Table 1: Summary table of bias and uncertainties of the various elements of the system used to*
 437 *forecast.*

<i>Type of bias/uncertainties</i>	<i>Bias</i>	<i>Uncertainties</i>
<i>Chlorophyll from satellite versus in situ data (Global)</i>	11.8%	R=0.86, P<0.05
<i>Chlorophyll from free-run model versus satellite chlorophyll (transient forcing data, Equatorial Pacific, 2012-2015)</i>	27.87±1.72%	R=0.72, p<0.05
<i>Chlorophyll from free-run model versus satellite chlorophyll (climatological forcing data, Equatorial Pacific, 2012-2015)</i>	85.67±2.77%	R=0.47, p<0.05
<i>Chlorophyll concentration from assimilating run versus satellite chlorophyll (Equatorial Pacific, 2012-2015)</i>	12.34±0.52%	R=0.95, P<0.05

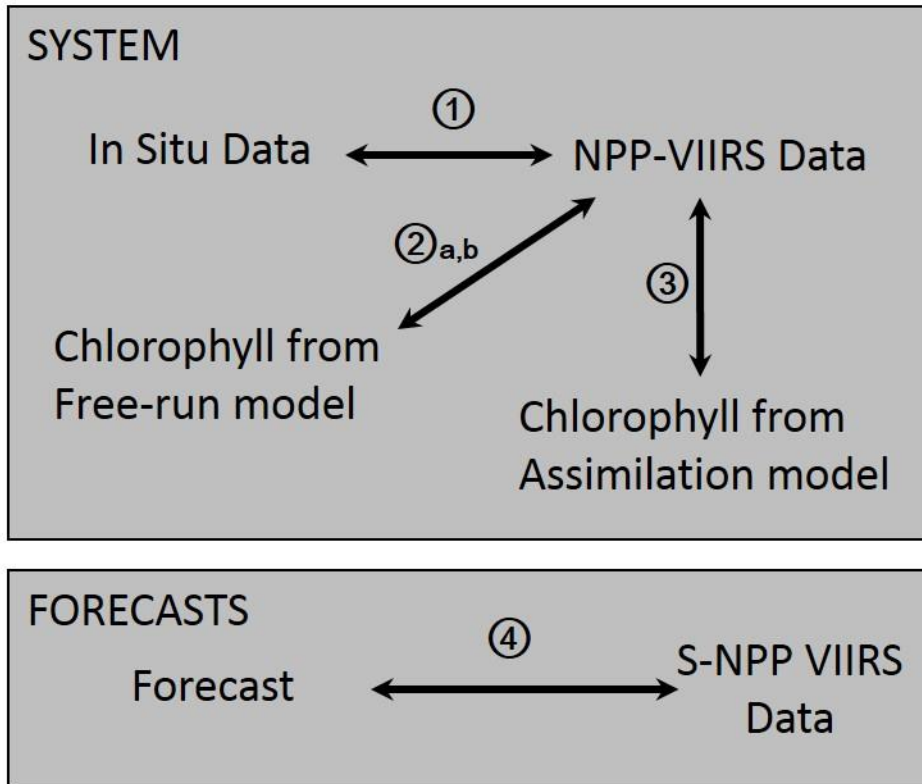
438

439 *Table 2: Anomaly Correlation Coefficient (ACC) and RMSE between the chlorophyll*
 440 *concentration in the Equatorial Pacific from the forecast at 1- to 9-month lead time and the*
 441 *corresponding monthly chlorophyll concentration from S-NPP VIIRS. *indicates that the*
 442 *anomaly correlation coefficient was significant (p< 0.05).*

<i># months lead time</i>	<i>ACC</i>	<i>RMSE</i>
1	0.329*	0.0399
2	0.272	0.0397
3	0.318	0.0411
4	0.267	0.0427
5	0.121	0.0435
6	0.153	0.0450
7	0.263	0.0470
8	0.417*	0.0471
9	0.409*	0.0472

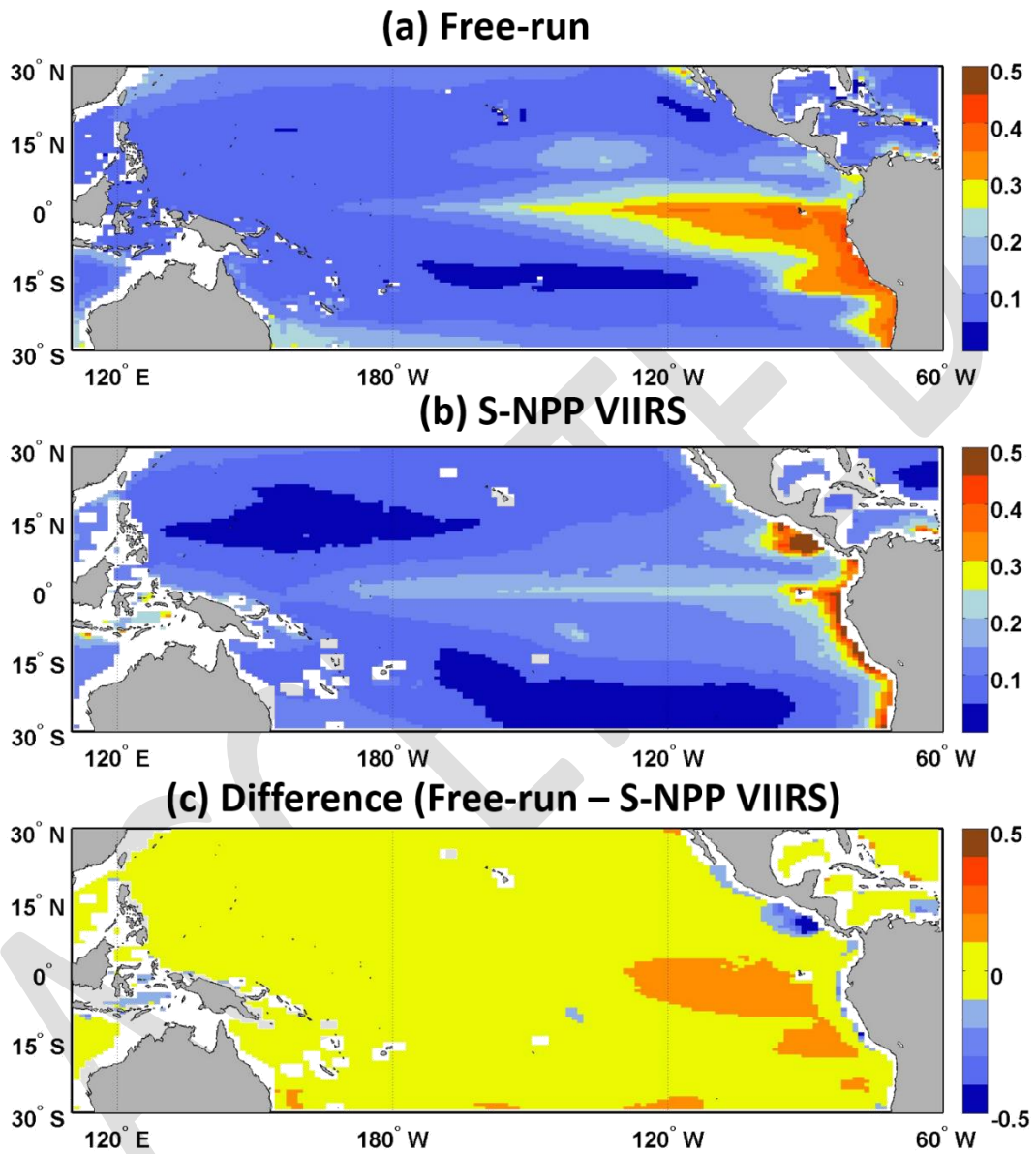
443

444



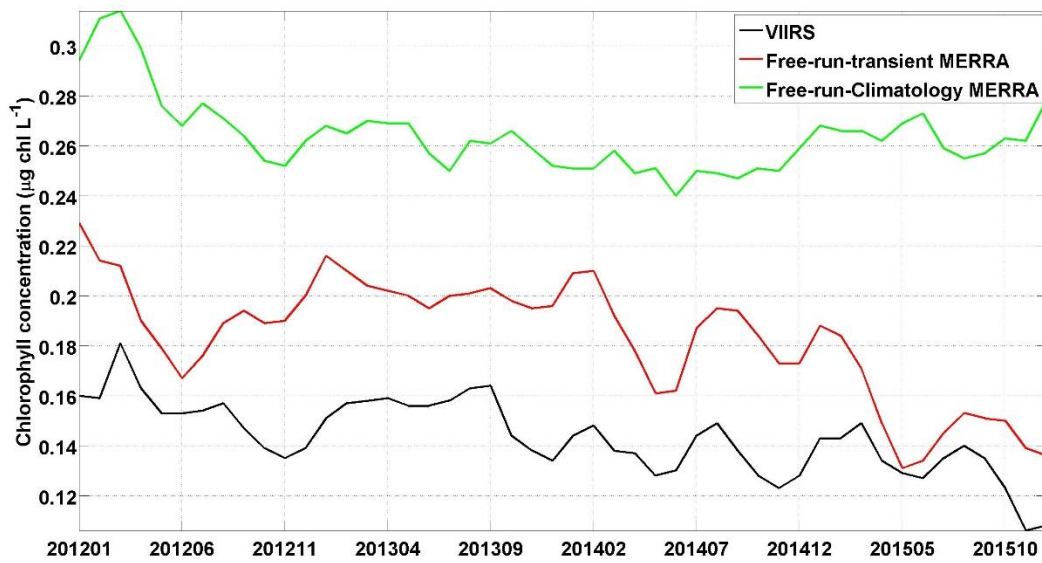
446

447 *Figure 1: Diagram describing the different procedures used to characterize bias and uncertainties*
448 *in the system and forecasts described in this study.*



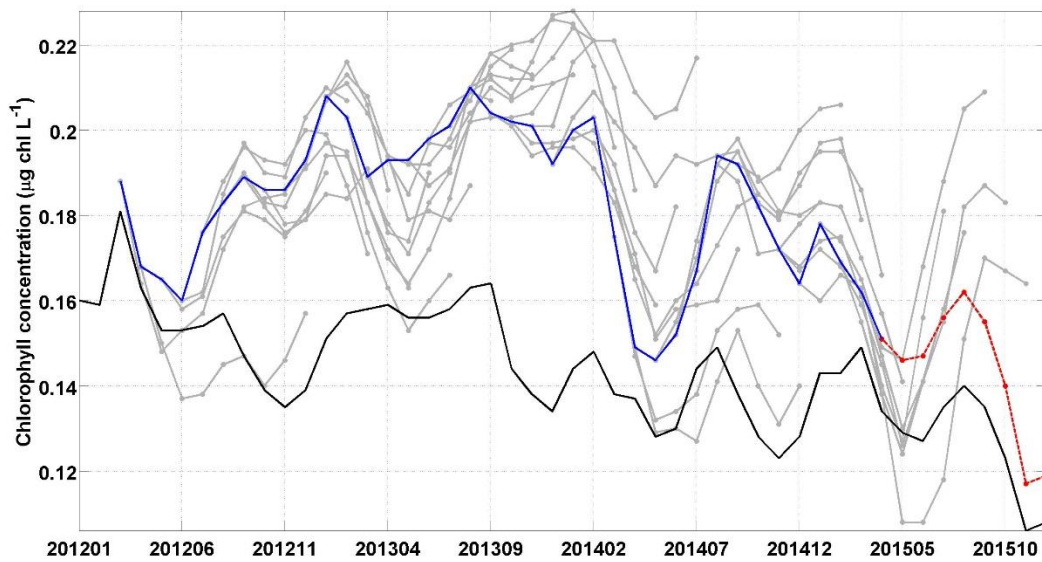
449

450 *Figure 2: Climatology of chlorophyll concentration ($\mu\text{g chl L}^{-1}$, 2012-2015) map of (a) the free-*
 451 *run model, (b) S-NPP VIIRS and (c) the difference between the free-run model and S-NPP VIIRS*
 452 *in the Equatorial Pacific.*



453

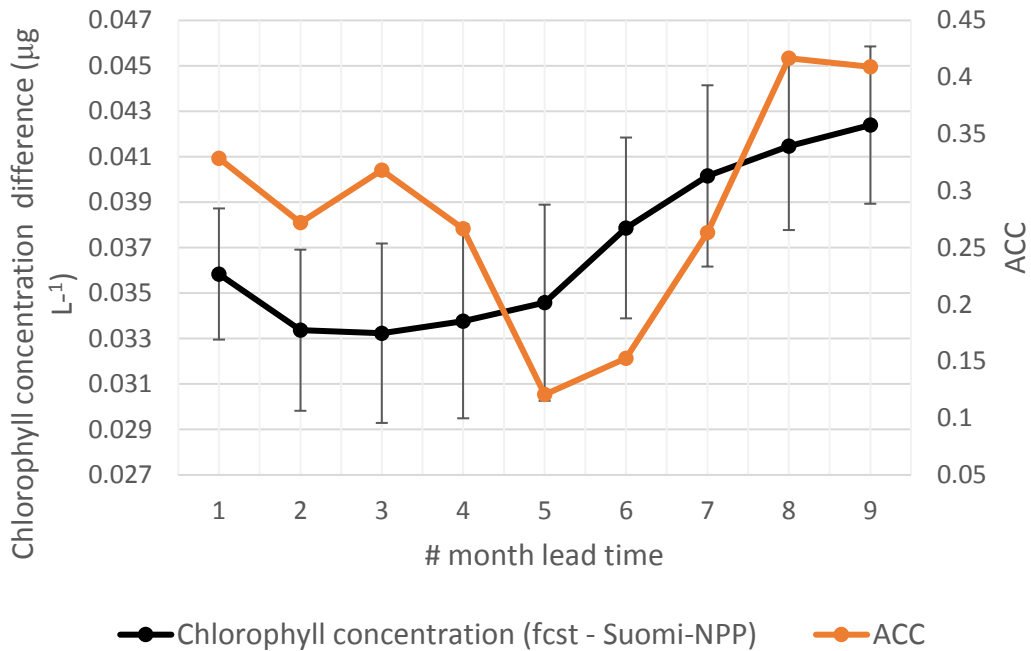
454 *Figure 3: Time series of chlorophyll concentration ($\mu\text{g chl L}^{-1}$) for NPP-VIIRS (black), free-run*
 455 *model with transient MERRA forcing data (red) and free-run model with a climatological*
 456 *MERRA forcing data (green).*



457

458 *Figure 4: Chlorophyll concentration in the Equatorial Pacific ($10^{\circ}\text{S}-10^{\circ}\text{N}$) for the period 2012-*
 459 *2015 from S-NPP VIIRS (black), individual forecasts (grey) and the 1-month lead chlorophyll*
 460 *concentration of every forecast (blue). The last forecast is highlighted in red.*

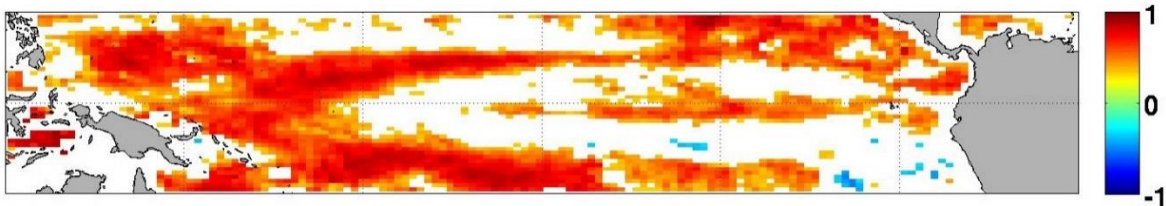
461



462

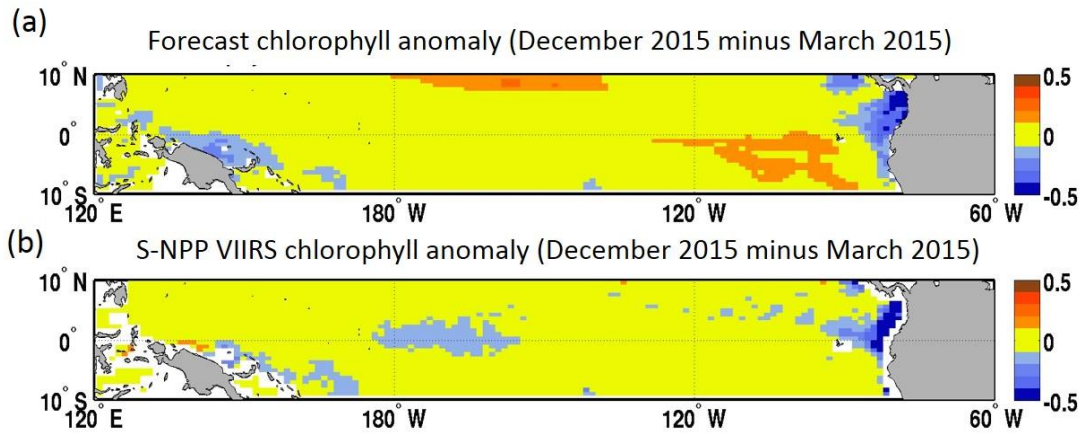
463 *Figure 5: Average difference between forecasted chlorophyll and chlorophyll from S-NPP VIIRS*
 464 *for corresponding month (left axis) and Anomaly Correlation Coefficient (ACC; right axis).*

465



466

467 *Figure 6: Anomaly correlation coefficient between the forecasted chlorophyll at 1-month lead*
 468 *and S-NPP VIIRS chlorophyll for the period 2012-2015. White indicates that the correlation was*
 469 *not significant ($p > 0.05$).*



470

471 *Figure 7: (a) Chlorophyll concentration anomaly (December 2015 minus March 2015, $\mu\text{g chl L}^{-1}$)*
 472 *from the March 2015 forecast for December 2015 and (b) chlorophyll concentration from S-NPP*
 473 *VIIRS ($\mu\text{g chl L}^{-1}$).*