



HAL
open science

Spatial and interannual variability of presettlement tropical fish assemblages explained by remote sensing oceanic conditions

Henitsoa Jaonalison, Jean-Dominique Durand, Jamal Mahafina, Hervé Demarcq, Raphaël Lagarde, Dominique Ponton

► **To cite this version:**

Henitsoa Jaonalison, Jean-Dominique Durand, Jamal Mahafina, Hervé Demarcq, Raphaël Lagarde, et al.. Spatial and interannual variability of presettlement tropical fish assemblages explained by remote sensing oceanic conditions. *Marine Biodiversity*, 2020, 50 (4), 10.1007/s12526-020-01068-6 . hal-03400681

HAL Id: hal-03400681

<https://univ-perp.hal.science/hal-03400681>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spatial and interannual variability of presettlement tropical fish assemblages explained by remote sensing oceanic conditions

Jaonalison Henitsoa ^{1,*}, Durand Jean-Dominique ², Mahafina Jamal ¹, Demarcq Herve ³, Lagarde Raphaël ⁴, Ponton Dominique ⁵

¹ Institut Halieutique et des Sciences Marines, Université de Toliara, Rue Dr. Rabesandratana, 601, Toliara, Madagascar

² MARBEC, IRD, Univ. Montpellier, CNRS, Ifremer, Montpellier, France

³ MARBEC, (IRD, Univ Montpellier, CNRS, Ifremer), Centre de Sète, France

⁴ Université de Perpignan Via Domitia-CNRS, Centre de Formation et de Recherche sur les Environnements Méditerranéens, UMR 5110, F 66860, Perpignan, France

⁵ ENTROPIE, IRD-Université de La Réunion-CNRS, Laboratoire d'Excellence CORAIL, c/o Institut Halieutique et des Sciences Marines, Université de Toliara, Rue Dr. Rabesandratana,, 601, Toliara, Madagascar

* Corresponding author : Henitsoa Jaonalison, email address : jaonasat@gmail.com

Abstract :

Understanding the interannual effect of various environmental factors on biodiversity distribution is fundamental for developing biological monitoring tools. The interannual variability of environmental factors on presettlement fish assemblages (PFAs) has been so far under investigated, especially in Madagascar. Numerous explanatory variables including local hydro-dynamic conditions recorded during the sampling night, characteristics of the benthic substrate and remotely sensed oceanic conditions (RSOC) were used to explain the spatio-temporal variability of PFAs in southwestern Madagascar. Gradient forest analyses were used to hierarchically classify the effect of these explanatory variables on the PFAs for two sites and during two different recruitment seasons. RSOC variables appeared to better explain the PFAs than the local variable and the characteristics of the benthic substrate. The PFAs caught in water masses with coastal characteristics were better explained than those with open water characteristics. This spatial variability is hypothesised to be linked to differences in feeding conditions among water masses. The gradient forest analyses also highlighted the complexity of predicting PFAs as the species for which abundances were better explained by RSOC variables varied between years. This interannual variability was mainly explained by the interannual variation of chlorophyll a (Chl a) concentration, wind and surface current, with better prediction obtained during the year with high Chl a values associated with high averaged sea surface temperature. These findings suggest the importance of forecasting Chl a concentrations, taking into account the impact of tropical storms and climate variability in order to predict PFAs in the future.

Keywords : Fish post-larvae, DNA barcoding, Structure, Environmental conditions, Hierarchical classification, Gradient forest

47 **Introduction**

48 Several studies pointed the importance of studying presettlement fish assemblages (aka post-
49 larvae) to better understand and predict how adult biomass varies (Lockwood et al. 1981;
50 Nichols et al. 1987; Hsieh et al. 2005; Koslow and Davison 2016; Chen et al. 2018) or more
51 generally to monitor marine fish communities (Koslow and Wright 2016). Indeed, post-larvae
52 represent survivors of the larval phase, and despite the important mortality that can occur after,
53 throughout the juvenile stage, they correspond to the individuals that can replenish adult fish
54 populations (Takahashi and Watanabe 2004). Presettlement fish assemblages (PFAs) also
55 represent a promising tool for monitoring coastal marine ecosystems as their structures relate
56 to environmental factors and to different anthropogenic pressures (Smith et al. 2018).
57 Developing PFAs surveys is thus considered as an urgent need in areas experiencing climate
58 change and anthropogenic pressure (Koslow and Wright 2016).

59 Furthermore, in a context of climate change it seems vital to better understand how future
60 environmental conditions will impact PFAs and thus fish communities. This highlights the
61 necessity to develop predictive models able to anticipate the PFAs dynamic and to identify the
62 hierarchical effect of explanatory variables on PFAs structure. In tropical waters, this field of
63 research has remained largely under investigated. Carassou et al. (2008) conducted one of the
64 rare studies aiming at hierarchically classifying the explanatory variables recorded at different
65 temporal and spatial scales structuring the PFAs. These authors demonstrated that abundances
66 of presettlement fish responded to large scale meteorological measured over a few days before
67 sampling, or measured the day of sampling, and small-scale water column factors measured
68 locally, according to threshold values that varied among families. However, they found no
69 significant difference between past-day and daily measures. Their study neither addressed the
70 effect of interannual variation of explanatory variables on PFAs nor incorporated the remotely
71 sensed oceanic conditions. To our knowledge, the eventual importance of remotely sensed
72 oceanic conditions (RSOC) extracted over few days preceding fish sampling in structuring
73 PFAs has never been investigated in tropical waters.

74 In Madagascar, small-scale fishermen and coastal populations highly depend on marine
75 resources for their subsistence (Cooke et al. 2000). Their increasing numbers (INSTAT and ICF
76 2016, Le Manach et al. 2012) and the declining quality of coastal habitats (IOTC 2009) may
77 lead to modifications in the taxonomic composition of fish assemblages (Folke et al. 2004). In
78 order to better record or even anticipate these modifications, PFAs need to be efficiently
79 investigated. In SW Madagascar, PFAs surveys started about a decade ago (Mahafina 2011;
80 Jaonalison et al. 2016) but employed different sampling gears and were based on different
81 taxonomic resolutions. Recently, standardized protocols for sampling and identifying
82 presettlement fish at the species level have been developed (Jaonalison et al. in review). Using
83 these protocols, these authors demonstrated that regression tree models based on remotely
84 sensed oceanic conditions (RSOC) were able to predict the richness and abundance of tropical
85 PFAs. As a follow-up, the present study aimed to further expand the approach and test if the
86 structure of PFAs described at the species level could also be explained by RSOC. More
87 precisely, the specific objective of this study was to assess the interannual variability of the
88 PFAs by characterizing: (i) the variability of the RSOC and PFAs among years, and (ii) how

89 the interannual changes in RSOC impact PFAs structure. Our hypothesis was that the PFAs
90 interannual variability was induced by the RSOC interannual variability.

91 **Materials and Methods**

92 **Study Sites, Sample Collection and Processing**

93 Presettlement fish were sampled in two sites in the Great Barrier Reef of Toliara (GRT) and in
94 the reef off Anakao (ANA) on the southwestern coast of Madagascar (**Fig.1**). A detailed
95 description of each site is provided in Jaonalison et al. (in review) but rapidly these two sites
96 mainly differ by the characteristics of their water masses. Indeed, the GRT site is mostly under
97 the influence of oceanic water masses due to the diurnal tidal inflow that renews up to 80% of
98 water in Toliara Lagoon (Chevalier et al. 2015). Oppositely, the ANA site, closer to the coast,
99 is influenced by the Onilahy River flows when the local northerly wind blows (Jaonalison et al.
100 in review). Larval collection was performed monthly from November 2016 to April 2017
101 (Campaign #1), and from November 2017 to April 2018 (Campaign #2). The monthly sampling
102 was conducted during three consecutive nights of the new moon period, three samplings were
103 collected during one night in each site (9 samples per month and per site). One light trap known
104 as SLEEP (Collet et al. 2018) was set at each of the three approx. 200 to 300m apart stations of
105 each site. As the catch of each light trap in each station correspond to a sample, a total 108
106 samples per campaign for the two sites were intended to be obtained. In the laboratory, all fishes
107 were sorted alive by morphospecies. One individual per morphospecies was photographed with
108 a “Nikon D90” camera and a piece of its caudal fin was preserved in ethanol 90° for DNA
109 barcoding.

110 **Identification Process**

111 The whole process of larval identification from DNA extraction till sequencing is detailed in
112 Collet et al. (2018). The sequences were uploaded to Barcode Of Life Datasystems (BOLD) for
113 assigning a Barcode Index Number (BIN) for each specimen sequence. Due to the ambiguity
114 in the assignment of specie name to each BIN, different labels were used for identifying the
115 species. First, if the BIN corresponded to only one species in BOLD, and this species was
116 observed in this BIN only, the specimen was identified as “Genus+species” (e.g. *Ostorhinchus*
117 *angustatus*). Second, if the BIN corresponded to different species from the same genus, or if
118 the species corresponding to this BIN was also assigned to different BINs in BOLD, the
119 specimen was identified as “Genus+BIN” (e.g. *Ostorhinchus* [BOLD:ACV9601]). Third, if the
120 BIN corresponded to species from different genera, but belonging to the same family, the
121 specimen was identified as “Family+BIN”. Note that identification such as “Genus+BIN”, or
122 “Family+BIN”, do correspond to identifications at the species level as each BIN corresponds
123 to an operational taxonomic unit, and thus to a putative species (Ratnasingham and Hebert
124 2013).

125 **Environmental Data**

126 Seven local variables were recorded: sea surface temperature (using a thermometer), water
127 transparency (using a Secchi disk), wind speed and direction (using an anemometer and a
128 compass). Due to technical problems, sea surface salinity was recorded for the second campaign

129 only and was thus not retained for the analyses. The difference between the time of sunset and
130 light-trap-setting, and between the time of sunrise and light trap retrieving were also considered
131 as explanatory variables for explaining the eventual effect of high tide which always occurred
132 around sunset during the sampling periods. The abundance of small pelagic fish species in light-
133 trap was also considered among local variables as their numbers may affect the way light traps
134 attract and retain assemblages of presettlement fish.

135 A specific campaign was conducted once a year in order to analyse the characteristics of the
136 benthic substrate at each sampling station. The approach was based on high precision photo-
137 quadrats following Dumas et al. (2009). Photo-quadrats were taken using a waterproof Olympus
138 TG 860 camera along three 20 m transects forming of star placed randomly at each sampling
139 station. In the laboratory, percentage cover of sediment, coral, seagrass, and macroalgae were
140 measured using Coral Point Count with Excel® extension (CPCe, Kohler and Gill 2006).

141 The remotely sensed oceanic conditions (RSOC) were extracted for an eight-day period
142 preceding each sampling over a 50 by 120 km area (Lon.max=43.611403, Lon.min=43.022609,
143 Lat.max=-22.900648, Lat.min=-24.115845, **Fig.1**). This spatial and temporal scale was chosen
144 as it was identified as the best predictor of presettlement fish richness and abundance at the two
145 studied sites (Jaonalison et al. in review). The RSOC included the Sea Surface Temperature
146 (SST), the concentration of chlorophyll *a* (Chla), the cross-shelf wind velocity U (Wind_U) and
147 the alongshore wind velocity V (Wind_V), and the cross-shelf current velocity U (Current_U)
148 and the alongshore current velocity V (Current_V), see **Table 1** for details.

149 **Data Analysis and Modeling**

150 Multivariate analyses were performed for depicting the spatial and interannual variability of
151 each of the three types of explanatory variables. The differences between sites in benthic
152 substrate characteristics and presettlement fish assemblages (PFAs) were visualized with
153 Mutlidimensional scaling (MDS) and tested with Permutational Multivariate Analysis of
154 Variance (Anderson 2017). The most discriminative characteristic of the benthic substrate, or
155 species within assemblages, was then identified with a Similarity of Percentage analysis (Clarke
156 et al. 1994). The spatial and interannual variability of local variables and the remotely sensed
157 oceanic conditions were analysed with Principal Component Analysis (PCA) for visualising the
158 parameter discriminating both sites and campaigns. The analysis for RSOC was run
159 independently from local variables as they concern merely the interannual variability. The
160 differences in mean variation of remotely sensed oceanic conditions between campaigns were
161 tested with analyses of variance (ANOVA).

162 A total of 19 explanatory variables were used for explaining the PFAs. Among the different
163 methods developed for identifying response of plant or animal assemblages faced to
164 environmental gradients, Gradient Forest (GF) appears to be the most robust and well suited
165 among the exploratory machine learning techniques. Indeed, GF can deal with non-linear
166 relationships (Ellis et al. 2012), and can analyse large databases with numerous species and
167 explanatory variables (Roubeix et al. 2017). GF, which is an extension of Random Forest
168 method, deal with the prediction of multi-species simultaneously (Ellis et al. 2012) by applying
169 the approach called “assemble-predict-together” defined by Ferrier and Guisan, (2006). This

170 approach consists in analysing simultaneously the response of several species to the explanatory
171 variables within a common modeling framework.

172 Given the strong expected differences in water masses characteristics between the two sites, the
173 PFAs were analysed separately. For each site, three datasets were thus created: one for the two
174 campaigns considered together, one for campaign#1 and one for campaign#2. The datasets for
175 campaign#1 and campaign#2 allowed to compare among years the explained presettlement fish
176 assemblages (PFAs), and the importance of explanatory variables. As GF analyses are sensitive
177 to the presence of rare species, only species exceeding 3% of occurrence were retained. For
178 each dataset, the GF analyses were performed following two steps. During a first step, three
179 models were trained using either local variable, characteristics of the benthic substrate, or
180 remotely sensed oceanic conditions separately. During a second step, GF analyses were
181 performed using all the explanatory variables (Table. 1). This two steps approach allowed to
182 compare the goodness of fit of each model for the explained PFAs, based on the average of R^2
183 value ($R^2_{aver.}$) of the species for which the abundances were better described as applied by
184 Roubeix et al. (2017). $R^2_{aver.}$ is the key diagnostic for Gradient Forest: the higher value of $R^2_{aver.}$
185 is the more predictable variability of the PFAs.

186 GF analyses generate four types of graphs that are of central importance for understanding the
187 results: the overall variable importance plots, the species performance plot, the split density
188 plots and the specific cumulative importance curves. The overall variable importance plot
189 corresponds to barplots presenting the explanatory variable in decreasing R^2 weighted
190 importance. Higher R^2 weighted importance values indicate important explanatory variables for
191 explaining the PFAs variability. The species performance plot which is also a barplot presents
192 the goodness of fit performance measures (R^2) for the species for which the variations in
193 abundance are better described by the GF model. The split density plots allow identifying the
194 thresholds value for each explanatory variable structuring the explained PFAs. Finally, the
195 specific cumulative importance curves are used for identifying the thresholds of the explanatory
196 variables for each species. Concerning the split density plot, the thresholds of the important
197 explanatory variables structuring the explained PFAs were obtained through the analysis of the
198 ratio between density of splits and density of data. The effect of an explanatory variable on the
199 explained PFAs is important when this ratio is greater than one. This indicates that the range of
200 values associated to ratios greater than 1 are associated to obvious change in the structure of
201 PFAs. Regarding the specific cumulative importance curve, the significant thresholds
202 influencing each of the species for which the abundances were better described were identified
203 based on the three criteria for threshold selection applied by Roubeix et al. (2017). According
204 to these authors, the threshold of a given explanatory variables is ecologically significant if (1)
205 the explanatory variable is important, based on R^2 weighted importance, for explaining the
206 abundances; (2) the threshold of this explanatory variable (indicated by the potential peak) is
207 clearly defined based on the specific cumulative importance curves; and (3) the threshold is
208 clearly highlighted for several species, based on the specific cumulative importance curves.

209 Data analyses were performed with R programming software R.5.1 (R Core Team 2018), using
210 “vegan” package (Version 2.5.4, Oksanen et al., 2019) for statistical analysis, the

211 “gradientForest” (Version 0.1-17, Ellis et al., 2012) and “extendedForest” (Version 1.6.1, Ellis
212 et al., 2012) packages for modeling the PFAs.

213 **Results**

214 **Spatio-Temporal Change of Local and RSOC Variables**

215 PCA analysis indicated that the local variables varied between sites and campaigns.
216 Campaign#1 was characterized by a higher water transparency at GRT, while cross-shelf wind
217 and SST were more important at ANA (**Fig.2b** and **c**). For campaign#2, no clear spatial
218 variation was observed.

219 Concerning the remotely sensed oceanic conditions (RSOC), their pattern differed significantly
220 among campaigns (**Fig.2f** and **Fig.3**). The difference among campaigns were mostly associated
221 to Chla and the SST (**Fig.2e** and **Fig.2f**) which were significantly higher (ANOVA, $p < 0.01$)
222 during campaign#1 (average \pm SD of $0.24 \pm 0.06 \text{ mg m}^{-3}$ and $28.06 \pm 1.03 \text{ }^\circ\text{C}$, respectively) than
223 during campaign#2 (average of $0.18 \pm 0.04 \text{ mg m}^{-3}$ and $27.20 \pm 1.01 \text{ }^\circ\text{C}$, respectively). In terms
224 of Chla concentrations, campaign#2 was characterized by a high number of observations
225 exceeding 0.5 mg m^{-3} in February, while only one observation exceeded this value during
226 campaign#1 for the same month (**Fig.3**). Differences between campaigns were also observed
227 for the onshore winds speed (Wind_U, **Fig.2e** and **Fig.2f**) that was significantly lower
228 (ANOVA, $p < 0.01$) in campaign#1 (average: $1.04 \pm 0.27 \text{ m s}^{-1}$) than in campaign#2 ($1.29 \pm$
229 0.27 m s^{-1}). The alongshore wind (Wind_V) consisted mostly of northerly wind that blew at an
230 average of $1.06 \pm 0.50 \text{ m s}^{-1}$ during campaign#1 and $0.79 \pm 0.50 \text{ m s}^{-1}$ only during campaign#2.
231 During campaign#1, up to 60% of currents flew offshore (with an average speed of $0.04 \pm$
232 0.02 m s^{-1}) and 40% were onshore (average: $0.02 \pm 0.004 \text{ m s}^{-1}$). During campaign#2, currents
233 flew entirely onshore (average: $0.03 \pm 0.03 \text{ m s}^{-1}$). During campaign#1, over 80% of alongshore
234 currents were southward (with an average speed of $0.12 \pm 0.06 \text{ m s}^{-1}$), while this percentage
235 dropped to around 67% for campaign#2 (average: $0.09 \pm 0.06 \text{ m s}^{-1}$). Only 20% of currents
236 were northward for campaign#1 (with a steady speed of 0.30 m s^{-1}) and about 33% only for
237 campaign 2 (average: $0.07 \pm 0.02 \text{ m s}^{-1}$).

238 **Spatial Characteristics of the Benthic Substrate**

239 The characteristics of the benthic substrate at GRT and ANA were different (PERMANOVA
240 test, $R^2 = 0.30$, $p < 0.01$, **Fig.4a**). SIMPER analysis revealed that this dissimilarity was mainly
241 due to macroalgae (Table. 2). Indeed, macroalgae represented 52% of the benthic substrate in
242 ANA but ~5% only at GRT (**Fig.4b**). The higher coverage by live coral (~18%) and sand
243 (~42%) at GRT against ~11% and ~19% at ANA, respectively, also contributed to the
244 dissimilarity in the characteristics of the benthic substrate between the two sites (**Fig.4b** and
245 **Table. 2**).

246 **Spatio-Temporal Variability in Presettlement Fish Assemblages**

247 A total of 165 species belonging to 42 families were collected during the two sampling
248 campaigns, with 99 species caught at GRT and 96 at ANA (Online resource 1). About 60
249 species from 20 families had an occurrence greater than 3%, with 46 species for GRT and 51
250 species for ANA. At GRT, the most frequent species were *Ostorhinchus* [BOLD:AAC2084],
251 *Dascyllus trimaculatus*, and *Apogonichthys perdis* during campaign#1 (**Fig.5a**), *A. perdis*,

252 *Pomacentrus trilineatus*, and *Petroscirtes* [BOLD:AAE6131] during campaign#2 (**Fig.5c**). At
253 ANA, the most frequent species were *Chaetodon auriga*, *D. trimaculatus*, and *Lutjanus*
254 *fulviflamma* during campaign#1 (**Fig.5b**), *C. auriga*, *A. perdix*, and *Chromis viridis* during
255 campaign#2 (**Fig.5d**). At GRT, the most abundant species were *Ostorhinchus*
256 [BOLD:AAC2084], *A. perdix* and *Nectamia* [BOLD:AAL9262] during campaign#1 (**Fig.5e**),
257 *Lethrinus mahsena*, *Ostorhinchus* [BOLD:AAC2084], and *A. perdix* during campaign#2
258 (**Fig.5g**). At ANA, the most abundant species were *Siganus sutor*, *Lethrinus*
259 [BOLD:AAB0511], and *L. fulviflamma* during campaign#1 (**Fig.5f**), *Lethrinus mahsena*, *C.*
260 *viridis*, and *C. auriga* during campaign#2 (**Fig.5h**). The difference in terms of the most frequent
261 and abundant species reflects the dissimilarity of presettlement fish assemblages between sites
262 and campaigns, a difference that was confirmed by the MDS analyses (**Fig.6**). The results of
263 PERMANOVA tests show that these differences were significant at $p < 0.01$.

264 **Relative Contribution of the Three Types of Explanatory Variables on PFAs**

265 When the two campaigns were considered together, the result of Gradient Forest analyses
266 indicates that the variations of presettlement fish assemblages (PFAs) were not explained by
267 the characteristic of the benthic substrates. Indeed, the R^2 value corresponding to habitat
268 variables was only 0.06 for ANA (**Table.3**) and explained the variations of *S. sutor* only. For
269 GRT, the R^2 value was more important, reaching 0.35, but it explained the variation of
270 *Fistularia commersonii* abundances only. Therefore, the characteristics of the benthic substrate
271 were excluded from subsequent analyses.

272 The variables measured locally during sampling (i.e. local variables) were also not relevant in
273 explaining the PFAs at GRT ($R^2_{\max.}=0.11$, $R^2_{\text{aver.}}=0.05$), and at ANA ($R^2_{\max.}=0.21$, $R^2_{\text{aver.}}=$
274 0.09 , **Table.3**).

275 When the sites were considered separately, the PFAs appeared more predictable based on $R^2_{\text{aver.}}$
276 values (**Table.4**). Compared to the characteristics of the benthic substrate and the local
277 variables, the remotely sensed oceanic conditions (RSOC) explained better the structuration of
278 the PFAs with $R^2_{\text{aver.}}=0.22$ for GRT and 0.34 for ANA during campaign#1 and $R^2_{\text{aver.}}=0.18$
279 for GRT and 0.32 for ANA during campaign#2 (**Table.4**). Interestingly, based on these “ $R^2_{\text{aver.}}$ ”
280 RSOC appeared to explain only a low part of the variations of the explained PFAs for GRT,
281 while about a moderate part for ANA.

282 **Spatio-Temporal Variability of RSOC Contribution**

283 The species for which the abundances were better explained by RSOC differed between
284 campaigns and between sites for a given campaign (**Fig.7**). The number of these species was
285 higher during campaign#2 (**Fig.7f and h**) than during campaign#1 (**Fig.7b and d**). Moreover,
286 the number of these species were also more important in ANA with nine species during
287 campaign#1 (**Fig.7d**) and 13 species during campaign#2 (**Fig.7h**) than in GRT with four and
288 seven species respectively (**Fig.7b and Fig.7f**). At GRT, *Chromis viridis* and *C. auriga* were
289 the species for which the abundances were better described in campaign#1 (**Fig.7b**) while it
290 was *L. mahsena* and *C. petersii* in campaign#2 (**Fig.7f**). At ANA, the species for which the
291 abundances were better described were different among campaigns, except for *L. fulviflamma*
292 which was always among these species (**Fig.7d and Fig.7h**).

293 The rank and importance of the explanatory variables (based on R^2 weighted importance) varied
294 between years and sites. During campaign#1, Chla and the northerly Wind_V were the most
295 important explanatory variables of presettlement fish assemblages (PFAs) at GRT and ANA
296 sites (**Fig.7a** and **c**), Current_V also being among the most important explanatory variables at
297 ANA (**Fig.7c**). For campaign#2, five explanatory variables appeared to be the most important,
298 but the rank of these explanatory variables differed at each site (**Fig.7e** and **g**).

299 **RSOC Threshold for Presettlement Fish Assemblages**

300 During campaign#1, three important RSOC variables presented thresholds for the PFAs at
301 ANA, and up to five variables at GRT. Important shifts in the PFAs structure occurred at 0.24
302 mg m^{-3} and 0.27 mg m^{-3} for Chla, and at 0.8 and 1.4 m s^{-1} for the northerly wind for the both
303 sites, and at 0 and 0.1 m s^{-1} for southward Current_V, and at 0.05 m s^{-1} for northward Current_V
304 at ANA only (**Fig.8**). During campaign#2, important shifts in the PFAs structure observed at
305 GRT at 0.5 m s^{-1} and 1.2 m s^{-1} for northerly Wind_V, at 0 and 0.065 m s^{-1} for northward
306 Current_V, at 0.03 m s^{-1} for onshore Current_U, at 1.1 and 1.5 m s^{-1} for westerly Wind_U, and
307 at 26.5°C, 27.75 and 28.25°C for SST (**Fig.8**). At ANA, important turnover in the PFAs
308 structure were detected at 26.5°C and 28.25°C for SST, at 1.5 m s^{-1} for westerly Wind_U, at
309 1.2 m s^{-1} for northerly Wind_V, at 0.035 m s^{-1} for onshore Current_U, at 0.1 m s^{-1} for southward
310 Current_V, and at 0 and 0.05 m s^{-1} for northward Current_V (**Fig.8**).

311 **RSOC Thresholds for Species**

312 During campaign#1, the northerly wind speed at 1.2 m s^{-1} induced the important shifts in the
313 abundances of all the species for which the abundances were better described (**Fig.9a** and **b**),
314 the Chla concentrations at 0.29 mg m^{-3} played a role in the abundances of *C. viridis* at GRT
315 (**Fig.9a**), and *S. sutor*, *Pomacentrus trilineatus*, and *L. fulviflamma* at ANA (**Fig.9b**). A
316 common threshold of northward currents speed at 0.12 m s^{-1} was detected during campaign#1
317 for the five species at ANA (**Fig.9b**).

318 During campaign#2, no clear common threshold was observed for the two species at GRT
319 (**Fig.9c**) or for the five species at ANA (**Fig.9d**). GF models detected thresholds for some
320 species only. For ex., the onshore wind at 1.65 m s^{-1} generated a shift in the abundances of *L.*
321 *mahsena* at GRT, and in the abundances of *Sphyræna barracuda* and *L. fulviflamma* at ANA.
322 At ANA, onshore currents at 0.06 m s^{-1} also induced shifts in the abundance of *L. notatus*.

323 **Discussion**

324 To our knowledge, this study is one of the first describing the structure of tropical presettlement
325 fish assemblages (PFAs) based on accurate species identification obtained through DNA
326 barcoding, and using a) Gradient Forest (GF), one the most robust exploratory machine learning
327 technique, and b) remotely sensed oceanic conditions (RSOC) observed before sampling (i.e.
328 past-day RSOC). The results of GF analyses highlighted the importance of past-day RSOC in
329 explaining the PFAs compared to information recorded during sampling or the characteristics
330 of the benthic substrate.

331 **RSOC: More Useful for Explaining the PFAs**

332 In the present study, the PFAs differed significantly among sites. Such differences in PFAs
333 between GRT and ANA sites had already been observed by Jaonalison et al. (2016). These
334 authors hypothesized that the spatial differences of PFAs might be explained by the
335 characteristics of the benthic substrate as such link had been clearly defined for post-settlement
336 stage by Levin et al. (1997). Although the characteristics of the benthic substrate were different
337 between the two sites, GF analyses demonstrated that they were of low importance for
338 explaining the PFAs structure. This weak relationship between the characteristics of the benthic
339 substrate and PFAs structure may suggest the water masses characteristics play an important
340 role for PFAs. Using Canonical Correspondence Analysis (CCA), Chen et al. (2018) found that
341 the larval fish assemblages collected with plankton nets in northern South China Sea can be
342 defined by water mass characteristics, mainly SST and sea surface salinity. In the current study,
343 the water masses characteristics differed among sites. In ANA, the water masses corresponded
344 to a more coastal environment under the influence of the nearby Onilahy River with salinities
345 varying from 31.5 to 35.7 (Jaonalison et al. in review). Oppositely, at GRT the water masses
346 were less variable and more similar to oceanic water conditions due to the important tidal inflow
347 that regularly renews lagoonal water masses (Chevalier et al. 2015). Although water masses
348 characteristics have been found to explain the spatial differences in abundances (N) of
349 presettlement fish (Jaonalison et al. in review), they appeared to be less important for explaining
350 PFAs structure in the current study. Indeed, GF models obtained using locally recorded
351 variables had the goodness of fit varying between 5 and 9% only.

352 The remotely sensed oceanic conditions (RSOC) appeared to be the most important variables
353 explaining the observed presettlement fish assemblages (PFAs) - reflecting the importance of
354 coastal oceanic conditions in nearshore area for shaping PFAs. Out of the 60 species for which
355 presettlement stage occurrence was more than 3%, the abundances of 23 species only were
356 explained by gradient forest based on RSOC. Among these 23 species, 15 were among the most
357 frequent species observed in the both sites but the eight others (*Abudefduf vaingiensis*,
358 *Canthigaster petersii*, *Dascyllus abudafar*, *Fowleria* [BOLD:AAD8726], *Lutjanus notatus*,
359 *Paramonacanthus pusillus*, *Pomacentrus agassizi*, and *Sphyrna barracuda*) were not among
360 the most frequent species. Thus, the occurrence of a species does not appear to be the main
361 parameter making its abundances better described by gradient forest. Moreover, the abundances
362 of species belonging to Blenniidae, Caesionidae, Fistularidae, Holocentridae, Ostracidae,
363 Plesiopidae, Pomacanthidae, Scorpaenidae, and Syngnathidae were not described by GF
364 models although some of them such as *Caesio caerulea*, *Fistularia commersonii* were
365 among the most frequent species in the samples. These species may be sensitive to
366 environmental factors that have not been considered in this study such as sea surface salinity,
367 and SST anomaly.

368 **Spatial and Interannual Variabilities of GF Model Goodness of Fit Based on RSOC**

369 The goodness of fit of the GF models based on RSOC, and the species for which the abundances
370 were better described by the GF models differed between sites and years. The spatial variability
371 could be linked to the difference in PFAs structure among sites as species may respond
372 differently to environmental conditions. Surprisingly, the number of these species and the
373 goodness of fit of models based on RSOC were always higher at ANA although this site

374 appeared to be a more variable environment due to the influence of the nearby Onilahy River
375 (Jaonalison et al. in review). This result suggests that other characteristics of coastal water
376 masses not recorded in this present study may play an important role. Some previous studies
377 have suggested the hypothesis that coastal waters could present better feeding conditions for
378 presettlement fish than open waters (Nagelkerken et al. 2001; Cocheret de la Morinière et al.
379 2002). In New Caledonia, the biomass of zooplankton has been observed to be strongly
380 influenced by terrigenous inputs from rivers due to their high concentrations of particulate
381 organic matter (Le Borgne et al. 1989). These findings led Carassou et al. (2010) to suggest that
382 terrigenous inputs may provide best feeding conditions to presettlement fish in areas influenced
383 by nearby rivers. The highest number of species at ANA could be thus linked to the best feeding
384 conditions of presettlement fish in ANA (i.e. in coastal water masses) than at GRT (i.e. in open
385 water mass). Concerning the interannual variability of the PFAs and of the species, both
386 appeared to be associated to the importance rank of Chla. Indeed, Chla was the highest
387 explanatory variables for PFAs during campaign#1 when it presented an average concentration
388 of $0.24 \pm 0.06 \text{ mg m}^{-3}$, and the lowest explanatory variables during campaign#2 with an average
389 concentration of $0.18 \pm 0.04 \text{ mg m}^{-3}$. In New Caledonia, the Chla concentration measured at the
390 moment of sampling was found to be the main factor determining the spatial structure of PFAs
391 (Carassou et al. 2008). High Chla concentrations are known to provide the main food source
392 for zooplankton (Chassot et al. 2010). Thus, food availability for fish larvae, induced by high
393 Chla, could be the main factor determining both the spatial and interannual variability of PFAs
394 in tropical areas. This seems logical as a higher food availability increases the survival of pre-
395 settlement fish (Olivar et al. 2010) and allow them to efficiently resist predation (Owen et al.
396 1989). These findings also give some insight to the tolerance of fish larvae to food availability,
397 or to the dependency of their preys to primary production. In this present study, *S. sutor* was
398 among the species during campaign#1 when Chla concentrations were high. The abundances
399 of Siganidae larvae were also linked to high Chla concentration in New Caledonia coastal
400 waters (Carassou et al. 2008). Oppositely, during campaign#2 Chla concentrations were
401 generally low, except in February, but the abundances of Sphyraenidae (*Sphyraena barracuda*),
402 Lutjanidae (*L. fulviflamma* and *L. notatus*) and Apogonidae (*A. perdix*) were highly explained.
403 This suggests these species consume prey that are not directly linked to primary production.

404 Based on criteria applied by Roubéix et al. (2017) for significant threshold selection, three
405 significant thresholds were clearly defined during campaign#1 only: 0.29 mg m^{-3} for Chla, 1.2
406 m s^{-1} for northerly alongshore wind speed, and 0.12 m s^{-1} for northward alongshore current in
407 (**Fig.9**). In contrast, no clear thresholds were detected during campaign#2 characterized by an
408 important number of observations with Chla concentration exceeding 0.5 mg m^{-3} in February
409 (**Fig.3**). This highly skewed distribution of the values of Chla during campaign#2 may explain
410 the absence of correlation between this parameter and the PFAs. The fact that high
411 concentrations of Chla occurred in February during campaign#2 suggests they were induced by
412 tropical storms that were more frequent than during campaign#1 according to Météo-France
413 (2017). Indeed, tropical storms are known to increase the Chla concentrations (Lin et al. 2003).
414 Moreover, a significant reduction of the average SST was also observed during campaign#2
415 compared to campaign#1. Strong winds associated to tropical storm have been demonstrated to
416 decrease SST (Price et al. 2008) from a minimum of 1° (Cione et al. 2000) to a maximum of 9°

417 (Lin et al. 2003). The cold surface water related to such storms can extend to hundreds of
418 kilometres (Emanuel 2001), and the return to initial conditions can take between 5 and 30 days
419 (Dare and McBride 2011). While inducing lower SST, tropical storms may weaken (Morsink
420 2018), or delay (Reynalte-Tataje et al. 2012), the reproductive activity of some fish species.
421 Moreover, tropical storms may also induce eggs and larvae drifting away as no fish larvae can
422 swim against the maximum water current $\sim 1 \text{ m s}^{-1}$ (Fisher 2005), i.e. observed during a tropical
423 storm.

424 **Research Improvements and Perspectives**

425 In conclusion, this work highlighted the complexity of predicting PFAs in the future as their
426 structure appeared to depend mainly on Chla concentrations and hydrodynamic conditions that
427 can dramatically change from year to year. Moreover, the eventual influence of tropical storms
428 on Chla concentrations and hydrodynamic conditions reinforces this complexity and explains
429 the non-consistency of gradient forest results. Two years of survey does not seem sufficient for
430 obtaining a consistent output and not useful for getting insights about interannual variability.
431 According to Brodeur et al. (2008), at least ten years of data are needed to detect a significant
432 change in PFAs. This period of time could be sufficient for addressing interannual variability
433 in PFAs structure (Koslow and Wright 2016) and to document the effects of El Nino and La
434 Nina events as well as those of tropical storms.

435 In parallel with the acquisition of longer time series of presettlement fish, forecasting Chla
436 concentrations appears to be among the important steps to avoid the difficulty in predicting
437 PFAs. Forecasting the Chla seems, however, a challenging step in a complex marine system as
438 its concentration linked to the hydrodynamic conditions. The Wavelet-transform and Artificial
439 Neural Network (WANN) appeared to be a promising method for forecasting the Chla
440 concentration 1 month ahead in South San Francisco Bay (Rajaei and Boroumand 2015) using
441 time series data from 1994 to 2013. The performance of this approach to predict the target
442 variable with high accuracy was confirmed by Alizadeh et al. (2017). However, the forecast
443 should be initially conditioned by silica, dissolved iron, nitrate, and forcing data including zonal
444 and meridional wind stress, SST and shortwave radiation (Rousseaux and Gregg 2017). Most
445 of this information remain to be acquired in Western Indian Ocean region.

446 **Acknowledgments**

447 Our acknowledgments going to the research assistants (J. J. Marcellin, D. Fiandria, R. Tsipy,
448 Tovondrainy, and Noelson) for conducting fish sampling, and to the team of the 'Institut des
449 Sciences de l'Evolution de Montpellier' (ISEM) for sequencing our fish tissues, and the Institut
450 Halieutique et des Sciences Marines for issuing research permit. We also thank the three
451 reviewers for their constructive remarks leading to the improvement our manuscript quality.

452

453 **Funding**

454 This work was co-funded by the French National Research Institute for Sustainable
455 Development (JEA-ACOM/IH.SM-IRD) in France, and the Critical Ecosystem Partnership
456 Fund (MG-66341) in United States.

457

458 **Conflict of interest**

459 The authors declare that they have no conflict of interest.

460 **Ethical approval**

461 All applicable international, national, and/or institutional guidelines for the care and use of
462 animals were followed by the authors.

463 **Sampling and field studies**

464 All necessary permits for sampling and observational field studies have been obtained by the
465 authors from the competent authorities.

466 **Data availability**

467 The datasets generated during and/or analysed during the current study are available from the
468 corresponding author on reasonable request.

469

470 **References**

- 471 Alizadeh MJ, Kavianpour MR, Kisi O, Nourani V (2017) A new approach for simulating and
472 forecasting the rainfall-runoff process within the next two months. *J Hydrol* 548:588–
473 597
- 474 Anderson MJ (2017) Permutational Multivariate Analysis of Variance (PERMANOVA). In:
475 Balakrishnan N, Colton T, Everitt B, et al. (eds) *Wiley StatsRef: Statistics Reference*
476 *Online*. John Wiley & Sons, Ltd, Chichester, UK, pp 1–15
- 477 Brodeur R, Peterson W, Auth T, Soulen H, Parnel M, Emerson A (2008) Abundance and
478 diversity of coastal fish larvae as indicators of recent changes in ocean and climate
479 conditions in the Oregon upwelling zone. *Mar Ecol Prog Ser* 366:187–202
- 480 Carassou L, Le Borgne R, Rolland E, Ponton D (2010) Spatial and temporal distribution of
481 zooplankton related to the environmental conditions in the coral reef lagoon of New
482 Caledonia, Southwest Pacific. *Mar Pollut Bull* 61:367–374
- 483 Carassou L, Ponton D, Mellin C, Galzin R (2008) Predicting the structure of larval fish
484 assemblages by a hierarchical classification of meteorological and water column
485 forcing factors. *Coral Reefs* 27:867–880
- 486 Chassot E, Bonhommeau S, Dulvy NK, Mélin F, Watson R, Gascuel D, Pape OL (2010)
487 Global marine primary production constrains fisheries catches. *Ecol Lett* 13:495–505
- 488 Chen L-C, Lan K-W, Chang Y, Chen W-Y (2018) Summer Assemblages and Biodiversity of
489 Larval Fish Associated with Hydrography in the Northern South China Sea. *Mar Coast*
490 *Fish* 10:467–480
- 491 Chevalier C, Devenon J-L, Rougier G, Blanchot J (2015) Hydrodynamics of the Toliara Reef
492 Lagoon (Madagascar): Example of a Lagoon Influenced by Waves and Tides. *J Coast*
493 *Res* 31:1403–1416
- 494 Cione JJ, Molina P, Kaplan J, Black PG (2000) SST time series directly under tropical
495 cyclones: Observations and implications. In: *American Meteorological Society*.

496 <https://ams.confex.com/ams/last2000/webprogram/Paper12807.html>. Accessed 28
497 May 2019

498 Clarke KR, Warwick RM, Laboratory PM (1994) *Change in Marine Communities: An*
499 *Approach to Statistical Analysis and Interpretation*, 2nd edn. Plymouth marine
500 laboratory, Natural environment research council

501 Cocheret de la Morinière E, Pollux BJA, Nagelkerken I, van der Velde G (2002) Post-
502 settlement Life Cycle Migration Patterns and Habitat Preference of Coral Reef Fish
503 that use Seagrass and Mangrove Habitats as Nurseries. *Estuar Coast Shelf Sci* 55:309–
504 321

505 Collet A, Durand J-D, Desmarais E, Cerqueira F, Cantinelli T, Valade P, Ponton D (2018)
506 DNA barcoding post-larvae can improve the knowledge about fish biodiversity: an
507 example from La Reunion, SW Indian Ocean. *Mitochondrial DNA Part A* 29:905–918

508 Cooke A, Ratomahenina O, Ranaivosoin E, Razafindraibe H (2000) Madagascar. In:
509 Sheppard CRC (eds) *Seas at the millennium: an environmental evaluation*.
510 Amsterdam, pp 113–131

511 Dare RA, McBride JL (2011) Sea Surface Temperature Response to Tropical Cyclones. *Mon*
512 *Weather Rev* 139:3798–3808

513 Dumas P, Bertaud A, Peignon C, Léopold M, Pelletier D (2009) A “quick and clean”
514 photographic method for the description of coral reef habitats. *J Exp Mar Biol Ecol*
515 368:161–168

516 Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests: calculating importance gradients on
517 physical predictors. *Ecology* 93:156–168

518 Emanuel K (2001) Contribution of tropical cyclones to meridional heat transport by the
519 oceans. *J Geophys Res Atmospheres* 106:14771–14781

520 Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. *J Appl*
521 *Ecol* 43:393–404

522 Fisher R (2005) Swimming speeds of larval coral reef fishes: impacts on self-recruitment and
523 dispersal. *Mar Ecol Prog Ser* 285:223–232

524 Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L, Holling CS (2004)
525 *Regime Shifts, Resilience, and Biodiversity in Ecosystem Management*. *Annu Rev*
526 *Ecol Evol Syst* 35:557–581

527 Hsieh C, Reiss C, Watson W, Allen MJ, Hunter JR, Lea RN, Rosenblatt RH, Smith PE,
528 Sugihara G (2005) A comparison of long-term trends and variability in populations of
529 larvae of exploited and unexploited fishes in the Southern California region: A
530 community approach. *Prog Oceanogr* 67:160–185

531 INSTAT M, ICF M (2016) *Enquête Démographique et de Santé de Madagascar 2008-2009*.
532 In: World Bank Microdata Library.
533 <https://microdata.worldbank.org/index.php/catalog/1435>. Accessed on 08 July 2019

534 IOTC (2009) *Madagascar National Report*. In: Food and Agricultural Organizations of
535 United Nations. <https://iotc.org/documents/madagascar-national-report>. Accessed 19
536 Mar 2019

537 Jaonalison H, Mahafina J, Ponton D (2016) Fish post-larvae assemblages at two contrasted
538 coral reef habitats in southwest Madagascar. *Reg Stud Mar Sci* 6:62–74

539 Kohler KE, Gill SM (2006) Coral Point Count with Excel extensions (CPCe): A Visual Basic
540 program for the determination of coral and substrate coverage using random point
541 count methodology. *Comput Geosci* 32:1259–1269

542 Koslow JA, Davison PC (2016) Productivity and biomass of fishes in the California Current
543 Large Marine Ecosystem: Comparison of fishery-dependent and -independent time
544 series. *Environ Dev* 17:23–32

545 Koslow JA, Wright M (2016) Ichthyoplankton sampling design to monitor marine fish
546 populations and communities. *Mar Policy* 68:55–64

547 Le Borgne R, Blanchot J, Charpy L (1989) Zooplankton of tikehau atoll (Tuamotu
548 archipelago) and its relationship to particulate matter. *Mar Biol* 102:341–353

549 Levin PS, Chiasson W, Green JM (1997) Geographic differences in recruitment and
550 population structure of a temperate reef fish. *Mar Ecol Prog Ser* 161:23–35

551 Lin I, Liu WT, Wu C-C, Wong GTF, Hu C, Chen Z, Liang W-D, Yang Y, Liu K-K (2003)
552 New evidence for enhanced ocean primary production triggered by tropical cyclone.
553 *Geophys Res Lett* 30:1718

554 Lockwood SJ, Nichols JH, Dawson WA (1981) The estimation of a mackerel (*Scomber*
555 *scombrus* L.) spawning stock size by plankton survey. *J Plankton Res* 3:217–233

556 Mahafina J (2011) Perception et comportement des pêcheurs pour une gestion durable de la
557 biodiversité et de la pêche récifale du Sud-Ouest de Madagascar. Dissertation,
558 University of La Réunion

559 Météo-France OI (2017) Bilan des activités cycloniques dans le bassin Sud-ouest de l’Océan
560 Indien. In *Cyclone Océan Indien*. [http://www.cycloneoi.com/archives-
561 blog/cyclone/2016-2017-une-nouvelle-saison-cyclonique-peu-active.html](http://www.cycloneoi.com/archives-blog/cyclone/2016-2017-une-nouvelle-saison-cyclonique-peu-active.html). Accessed
562 on 30 April 2019

563 Morsink K (2018) Hurricanes, Typhoons, and Cyclones. In: Smithsonian Institution Ocean.
564 [http://ocean.si.edu/planet-ocean/waves-storms-tsunamis/hurricanes-typhoons-and-
565 cyclones](http://ocean.si.edu/planet-ocean/waves-storms-tsunamis/hurricanes-typhoons-and-cyclones). Accessed 28 May 2019

566 Nagelkerken I, Kleijnen S, Klop T, van den Brand R, de la Morinière E, van der Velde G
567 (2001) Dependence of Caribbean reef fishes on mangroves and seagrass beds as
568 nursery habitats: a comparison of fish faunas between bays with and without
569 mangroves/seagrass beds. *Mar Ecol Prog Ser* 214:225–235

570 Nichols JH, Bennett DB, Symonds DJ, Grainger R (1987) Estimation of the stock size of
571 adult *Nephrops norvegicus* (L.) from larvae surveys in the western Irish Sea in 1982. *J*
572 *Nat Hist* 21:1433–1450

573 Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O’Hara
574 RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2019) Community
575 Ecology Package. In: Cran r. <https://cran.r-project.org/web/packages/vegan/vegan.pdf>.
576 Accessed on 29 April 2019

577 Olivar MP, Emelianov M, Villate F, Uriarte I, Maynou F, Álvarez I, Morote E (2010) The
578 role of oceanographic conditions and plankton availability in larval fish assemblages
579 off the Catalan coast (NW Mediterranean). *Fish Oceanogr* 19:209–229

580 Owen RW, Lo NCH, Butler JL, Theilacker GH, Alvarino A, Hunter JR, Watanabe Y (1989)
581 Spawning and Survival Patterns of Larval Northern Anchovy, *Engraulis mordax*, in

582 Contrasting Environments-A Site-Intensive Study. Fishery bulletin (Washington, D.C.
583 : 1971) 87:673–688

584 Price JF, Morzel J, Niiler PP (2008) Warming of SST in the cool wake of a moving hurricane.
585 J Geophys Res 113:C07010

586 R Core Team (2018) R: A Language and Environment for Statistical Computing. In: R
587 Foundation for Statistical Computing. <https://www.R-project.org/>. Accessed on 11
588 June 2018

589 Rajae T, Boroumand A (2015) Forecasting of chlorophyll-a concentrations in South San
590 Francisco Bay using five different models. Appl Ocean Res 53:208–217

591 Ratnasingham S, Hebert PDN (2013) A DNA-Based Registry for All Animal Species: The
592 Barcode Index Number (BIN) System. PLOS ONE 8:e66213

593 Reynalte-Tataje DA, Zaniboni-Filho E, Bialetzki A, Agostinho AA (2012) Temporal
594 variability of fish larvae assemblages: influence of natural and anthropogenic
595 disturbances. Neotropical Ichthyol 10:837–846

596 Roubex V, Daufresne M, Argillier C, Dublon J, Maire A, Nicolas D, Raymond J-C, Danis P-
597 A (2017) Physico-chemical thresholds in the distribution of fish species among French
598 lakes. Knowl Manag Aquat Ecosyst 418:1–14

599 Rousseaux CS, Gregg WW (2017) Forecasting Ocean Chlorophyll in the Equatorial Pacific.
600 Front Mar Sci 4:236

601 Smith JA, Miskiewicz AG, Beckley LE, Everett JD, Garcia V, Gray CA, Holliday D, Jordan
602 AR, Keane J, Lara-Lopez A, Leis JM, Matis PA, Muhling BA, Neira FJ, Richardson
603 AJ, Smith KA, Swadling KM, Syahailatua A, Taylor MD, van Ruth PD, Ward TM,
604 Suthers IM (2018) A database of marine larval fish assemblages in Australian
605 temperate and subtropical waters. Sci Data 5:180207

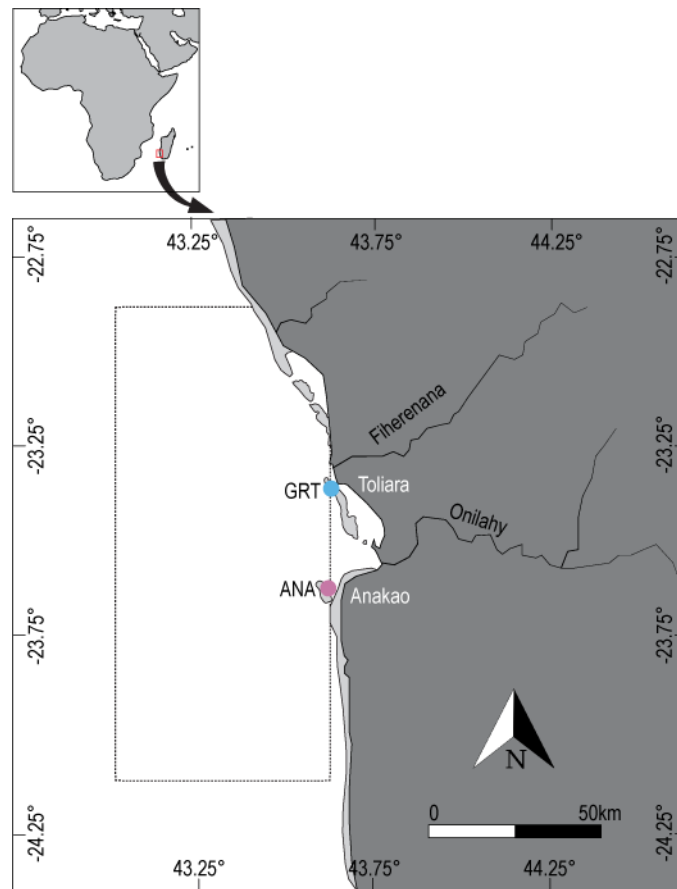
606 Takahashi M, Watanabe Y (2004) Growth rate-dependent recruitment of Japanese anchovy
607 *Engraulis japonicus* in the Kuroshio-Oyashio transitional waters. Mar Ecol Prog Ser
608 266:227–238

609

610

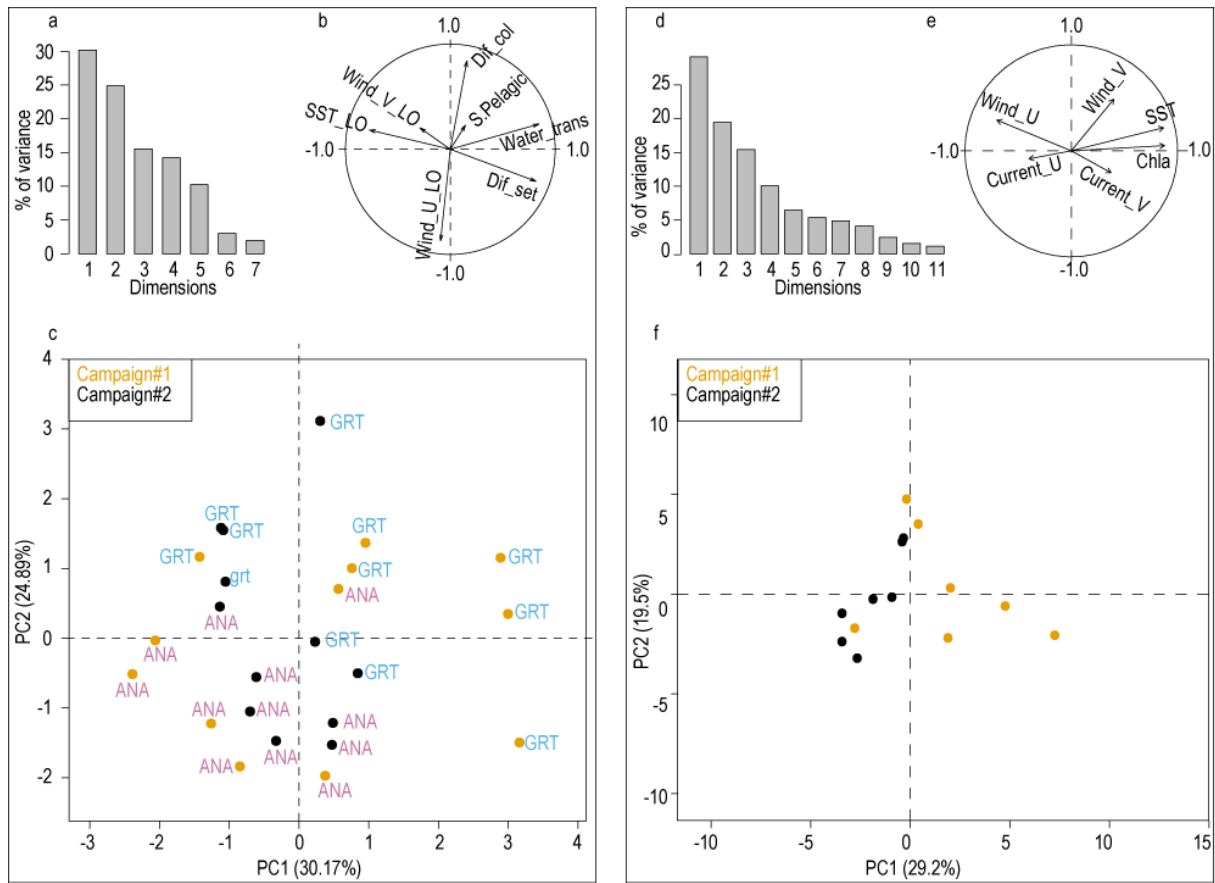
611

Figures



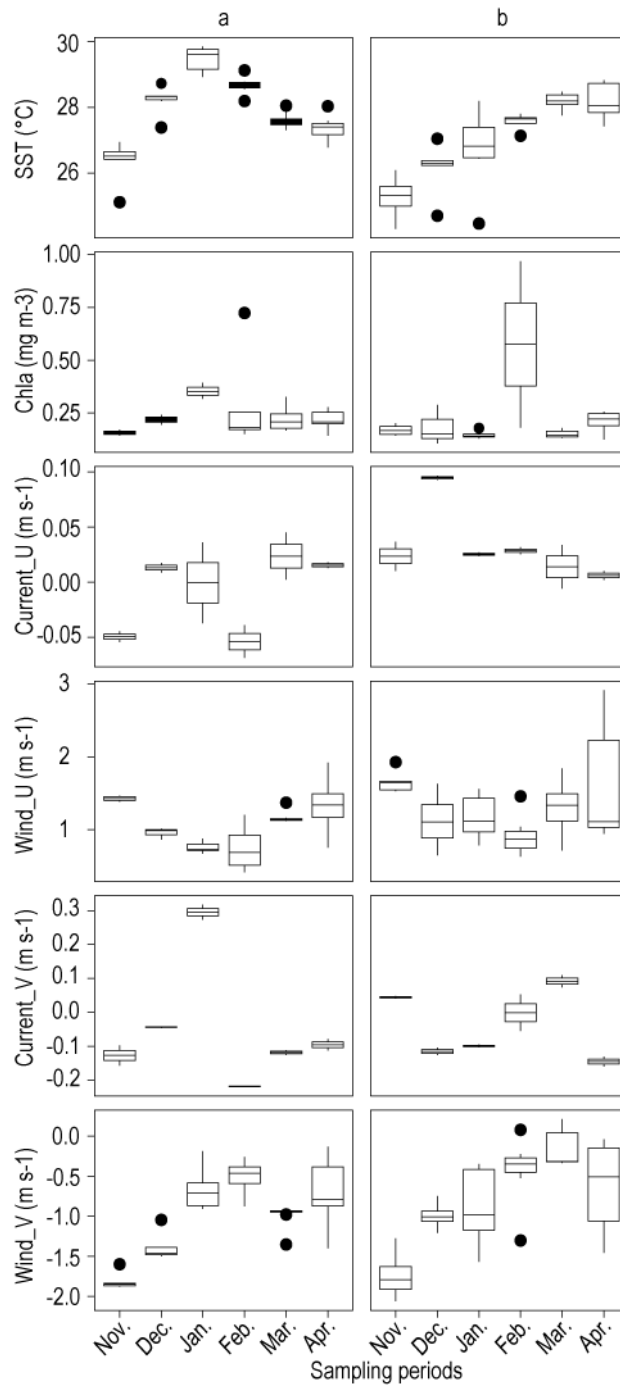
612

613 **Fig.1** Sampling sites with blue dot: Great Barrier Reef of Toliara (GRT) and pink dot: Anakao
614 Reef (ANA). The dotted rectangle corresponds to the extraction area for the remotely sensed
615 oceanic conditions



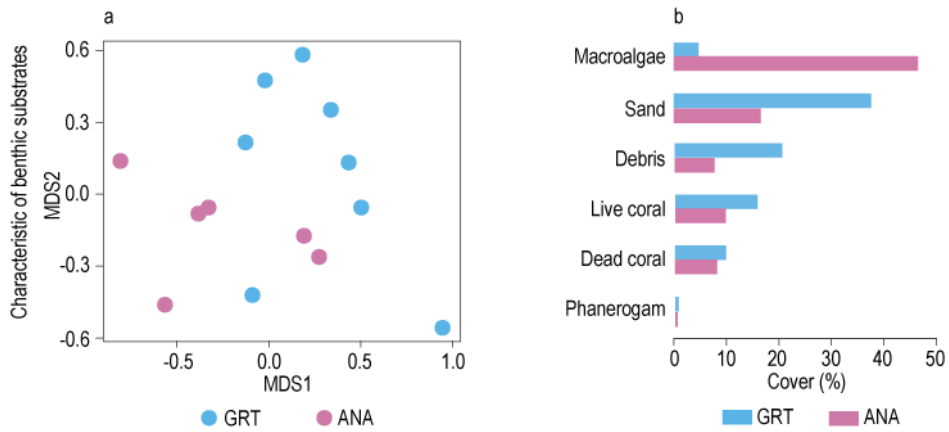
616

617 **Fig.2** PCA of local variables (in left) with a: percentage of variance for each dimension, b:
 618 factor variables, c: PC1-PC2 plane with black dots for campaign#1 and yellow for
 619 campaign#2. Each point corresponds to the environmental record per site (in blue for GRT
 620 and pink for ANA) and per month. PCA of remotely sensed oceanic conditions (in right).
 621 With d: percentage of variance for each dimension, e: factor variables, f: PC1-PC2 plane with
 622 black dots for campaign#1 and yellow dots for campaign#2



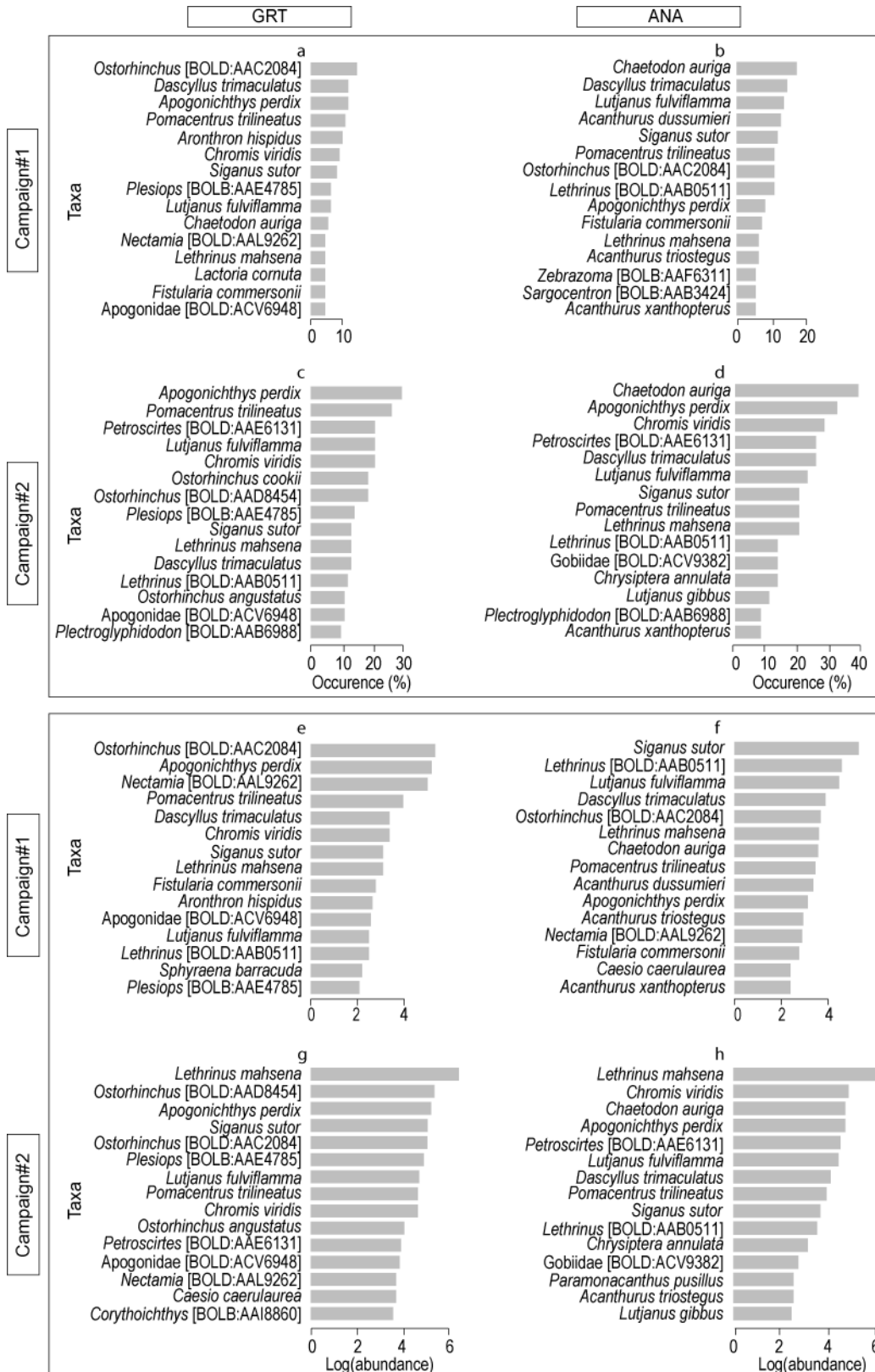
623

624 **Fig.3** Interannual variability of the remotely sensed oceanic conditions for each month of
 625 sampling, with a: campaign#1 (2016-2017) and b: campaign#2 (2017-2018). The lower and
 626 upper boundaries of the boxes correspond to the 25th and 75th percentile, respectively. The
 627 horizontal lines within the box correspond to the median values and the vertical lines show the
 628 range of values that fall within 1.5 times the interquartile range individual points correspond to
 629 values outside three times of the interquartile range



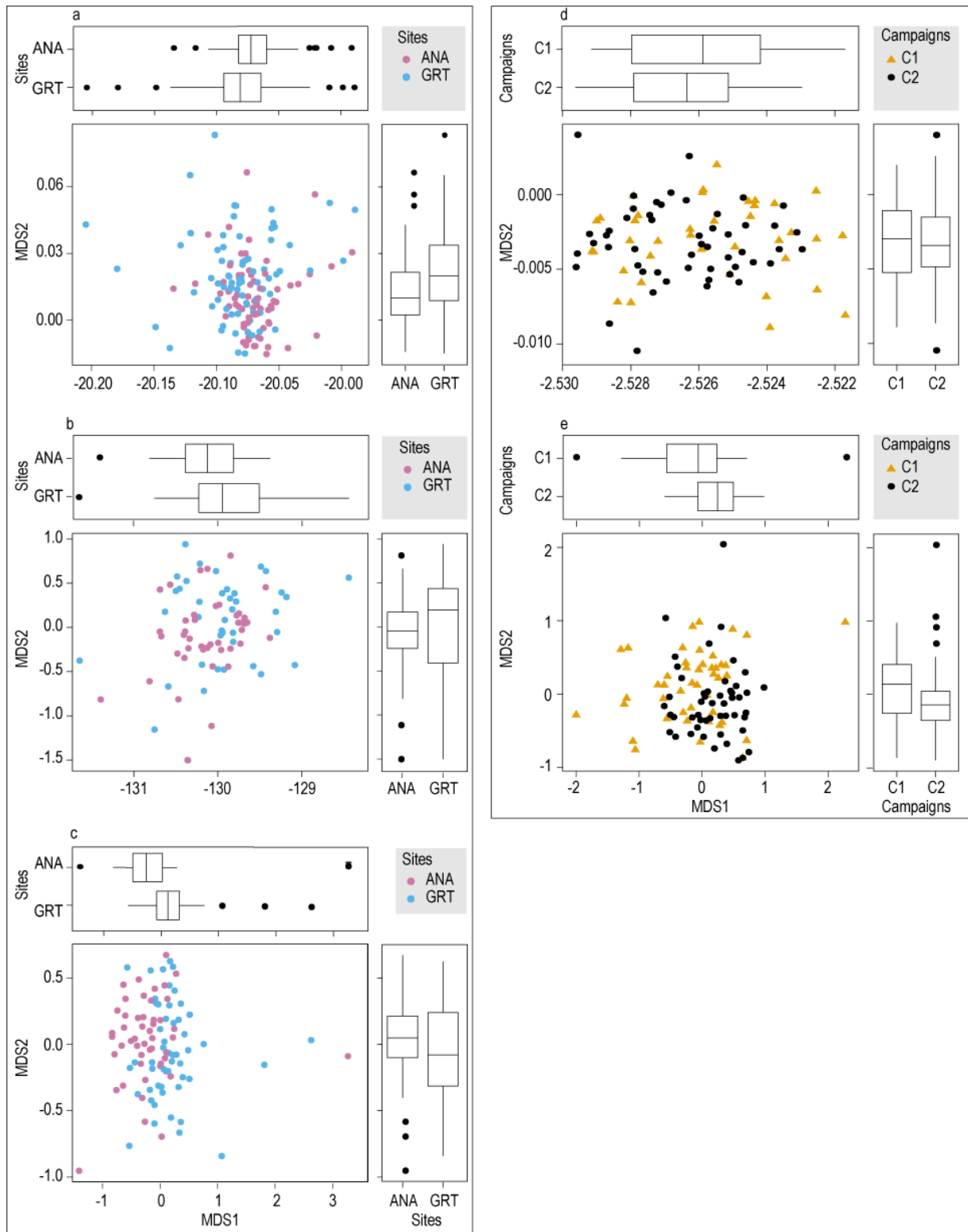
630

631 **Fig.4** Distribution of the characteristics of benthic substrate of coral reef habitat in GRT (blue
 632 bars and dots) and ANA (pink bars and dots), with a: barplot of the percentage cover of the
 633 characteristics of benthic substrate and b: ordination of the characteristics of benthic substrate
 634 with multidimensional scaling (MDS) ordination



635

636 **Fig.5** Occurrence of the first 20 most-occurring fish species observed at GRT during
 637 campaign#1 (a) and campaign#2 (c) and ANA during campaign#1 (b) and campaign#2 (d) in
 638 the top panel; and log-transformed abundances of the 20 most-abundant fish species observed
 639 at GRT for campaign#1 (e) and campaign#2 (g) and ANA for campaign#1 (f) and campaign#2
 640 (h) in the down panel

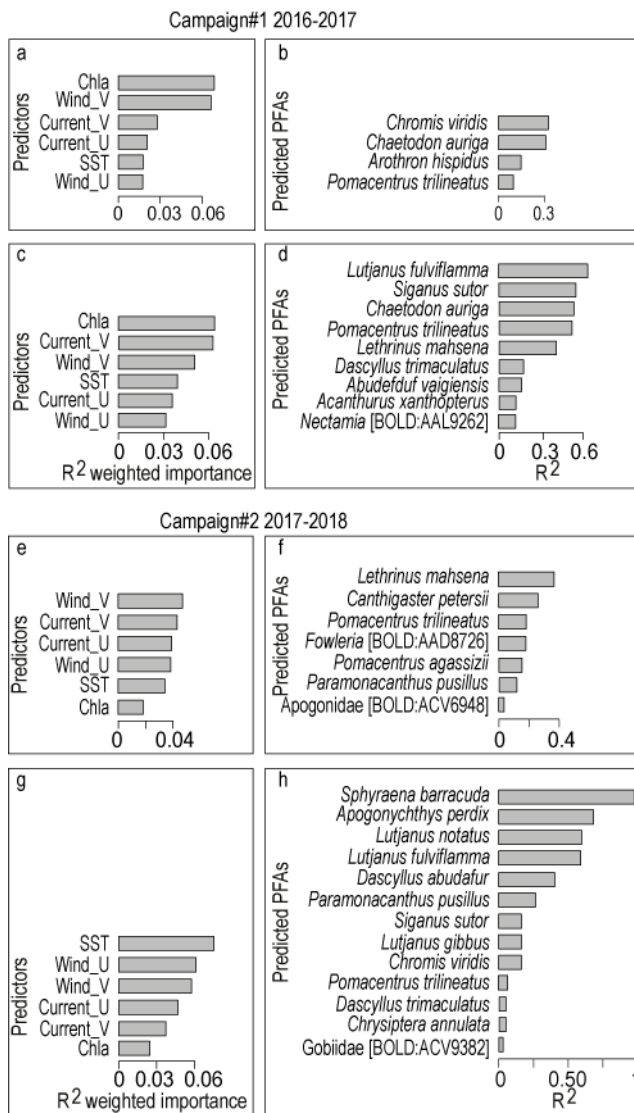


641

642 **Fig.6** Multidimensional scaling (MDS) enlightening the spatial variability of the presettlement
 643 fish assemblages (left panel), a: for both campaigns together, b: for campaign#1, and c: for
 644 campaign#2 with blue points for GRT and pink points for ANA, and the interannual variability
 645 of the presettlement fish assemblages (right panel), d: for GRT, e: for ANA with yellow
 646 triangles corresponding to campaign#1 (C1) and black circles to campaign#2 (C2). Boxplots
 647 indicate the spatial (left panel) and interannual (right panel) variability of presettlement fish

648 assemblages along each axis based on the distance between each observation and its
 649 corresponding centroid

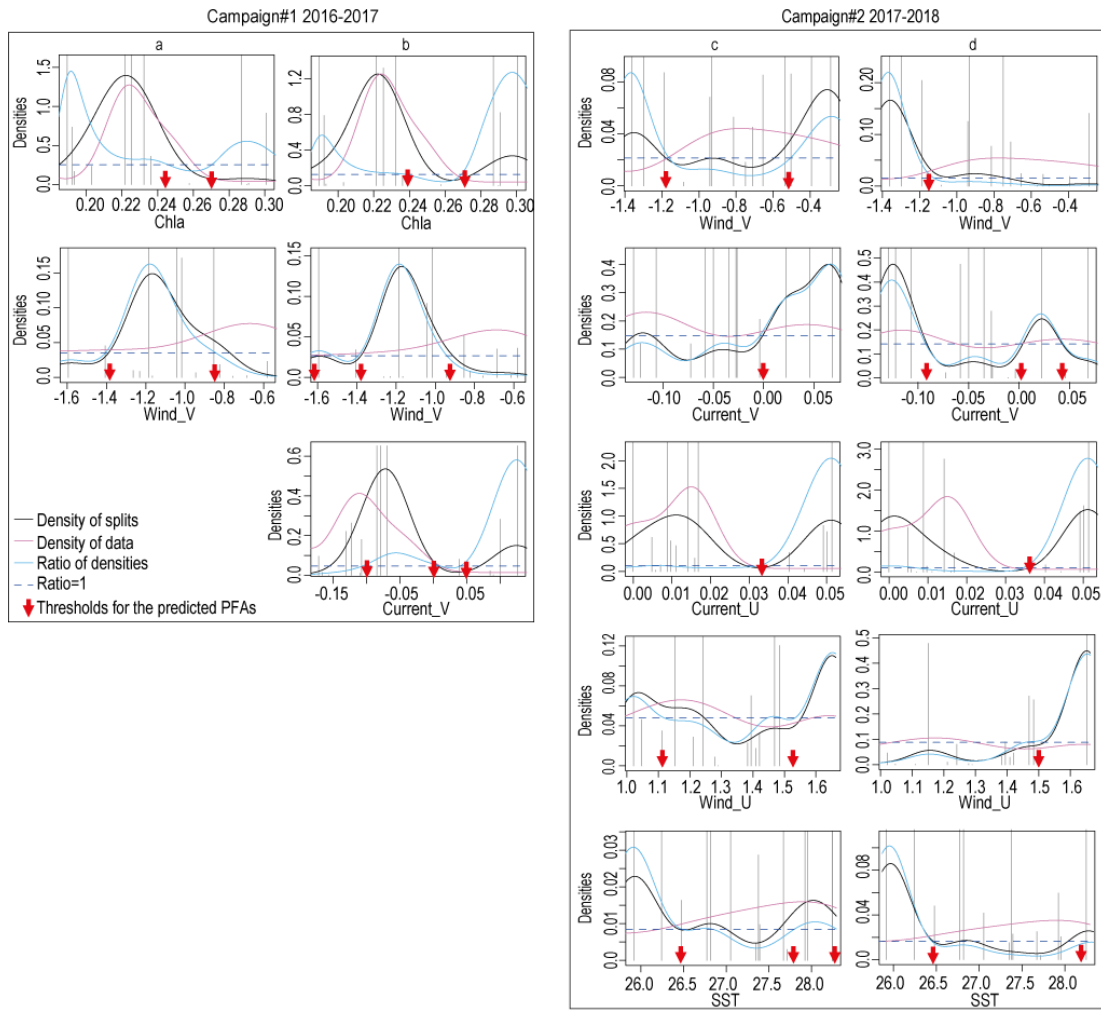
650



651

652 **Fig.7** Gradient forest results for GRT (a, b, e, and f) and ANA (c, d, g, and h) with the overall
 653 RSOC variables importance plot (a and c for campaign#1 and e and g for campaign#2), and the
 654 species performance plot showing the goodness of fit R² for overall species (b and d for
 655 campaign#1 and f and h for campaign#2). Species with R² ≤ 0 are not represented

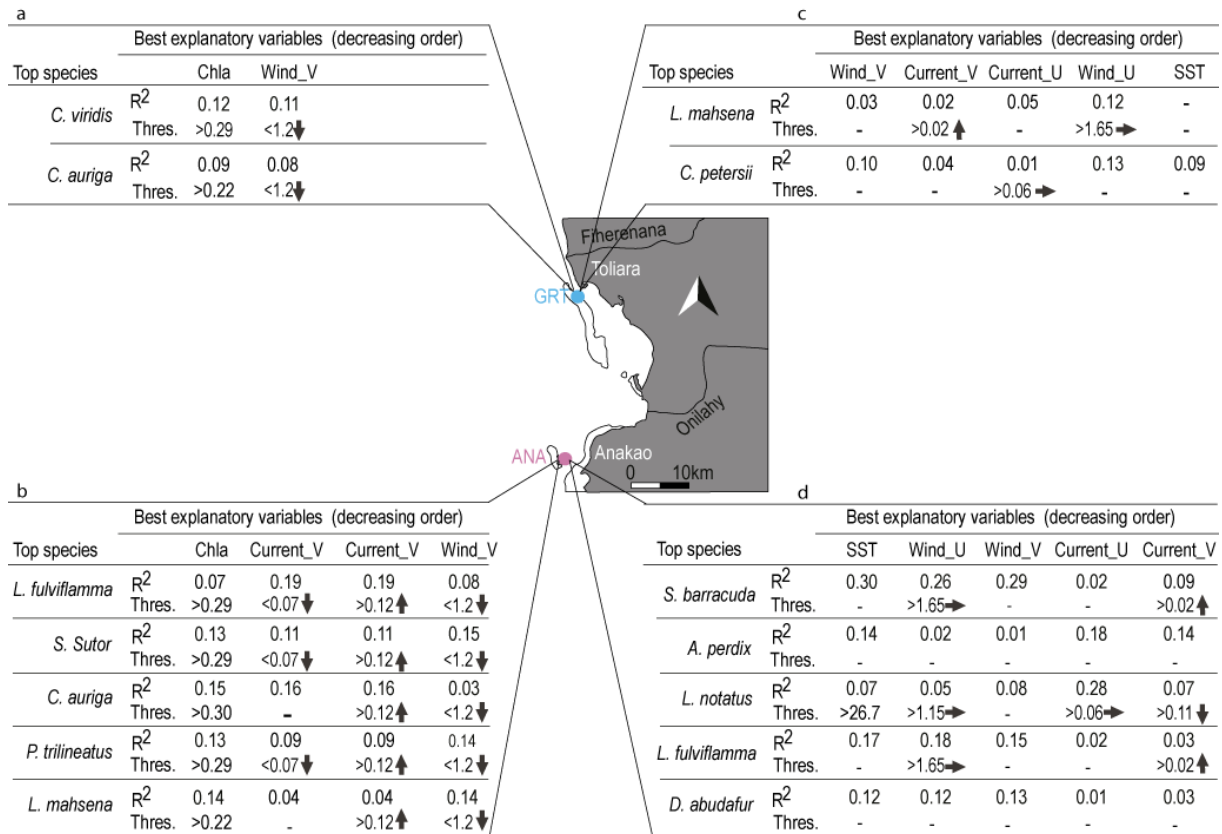
656



657

658 **Fig.8** Split density plot for campaign#1 (a, b) and campaign#2 (c, d) showing the thresholds
 659 (outlined by the red arrows) for the explained presettlement fish assemblages corresponding to
 660 the range of value of the ratio of densities (blue full line) superior to the ratio=1 (blue dashed
 661 line), the black and pink full line denote the density of splits and the density of data,
 662 respectively. With a and c for GRT, and b and d for ANA

663



664

665 **Fig.9** Interannual and spatial variability of accuracy importance (based on R² value) of the most
 666 important explanatory variables (in decreasing importance) for each of the top species for which
 667 the abundances were better described (in decreasing order) with a and b for GRT and ANA
 668 during campaign#1 and c and d for GRT and ANA during campaign#2. The thresholds (thres.)
 669 values were obtained and synthesized from the specific cumulative importance curves in **Online**
 670 **Resource3** for campaign#1, and **Online Resource4** for campaign#2. Dash (-) denotes the
 671 absence of thresholds and the arrows relate to the direction of wind and currents vectors

672

Supplementary materials

673 **Online resource1:** [https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-](https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-01068-6/MediaObjects/12526_2020_1068_MOESM1_ESM.pdf)
674 [01068-6/MediaObjects/12526_2020_1068_MOESM1_ESM.pdf](https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-01068-6/MediaObjects/12526_2020_1068_MOESM1_ESM.pdf)

675 **Online resource2:** [https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-](https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-01068-6/MediaObjects/12526_2020_1068_MOESM2_ESM.pdf)
676 [01068-6/MediaObjects/12526_2020_1068_MOESM2_ESM.pdf](https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-01068-6/MediaObjects/12526_2020_1068_MOESM2_ESM.pdf)

677 **Online resource3:** [https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-](https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-01068-6/MediaObjects/12526_2020_1068_MOESM3_ESM.pdf)
678 [01068-6/MediaObjects/12526_2020_1068_MOESM3_ESM.pdf](https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-01068-6/MediaObjects/12526_2020_1068_MOESM3_ESM.pdf)

679 **Online resource4:** [https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-](https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-01068-6/MediaObjects/12526_2020_1068_MOESM4_ESM.pdf)
680 [01068-6/MediaObjects/12526_2020_1068_MOESM4_ESM.pdf](https://static-content.springer.com/esm/art%3A10.1007%2Fs12526-020-01068-6/MediaObjects/12526_2020_1068_MOESM4_ESM.pdf)